GRIFFITH UNIVERSITY

Breast cancer progression is modulated by inherited genetic variance in multiple independent cohorts

Author

Hsieh, Szu-Min (Lintell)

Published

2010

Thesis Type

Thesis (PhD Doctorate)

School

School of Medical Science

DOI

10.25904/1912/753

Rights statement

The author owns the copyright in this thesis, unless stated otherwise.

Downloaded from

http://hdl.handle.net/10072/366638

Griffith Research Online

https://research-repository.griffith.edu.au

Breast cancer progression is modulated by inherited genetic variance in multiple independent cohorts

Szu-Min Hsieh, BBSc. (Hons)

Submitted in fulfilment of the requirements of the degree of Doctor of Philosophy

School of Medical Science
Griffith Health
Griffith University

September 2009

Abstract

Cancer metastasis remains a major health issue and it is the main cause of cancer related mortality, responsible for an estimated 90% of solid tumour related deaths. Further research into the elements that contribute to cancer metastasis is therefore important in identifying key molecular factors and pathways modulating this complex process. This would subsequently enhance the understanding of the metastatic process and facilitate development and deployment of new and improved clinical management for patients with metastatic diseases.

Genetic sequence variation is one of the fundamental elements influencing the development of cancer and cancer metastasis. Evidence suggests that inherited genetic variations are one of the essential factors regulating breast cancer metastatic susceptibility; different inbred mouse strains with inherited polymorphisms have wide ranges of transgene-induced mammary tumour metastatic potency, with inherited genetic polymorphisms believed to give rise to a diverse range of metastatic susceptibilities. Molecular analyses, bioinformatics research and literature review have identified a number of potential breast cancer metastasis susceptibility modulating genes.

This study extended the research of potential metastasis modulator genes and investigated these genes in human populations, aiming to elucidate the connection of the inherited genetic nature of these potential breast cancer metastasis modulators and breast cancer patients. The hypothesis of this research is that the inherited genetic variance of these putative breast cancer metastasis susceptibility regulating genes, which were identified in previous mouse model research, are predictive of breast cancer

patient diagnosis and prognosis. The identification of correlations between inherited genetic variation within novel breast cancer susceptibility genes, and clinical outcome would provide evidence supporting these genes as novel metastasis modulating genes.

ABI genotyping assays kits were utilized to perform SNP analysis of potential metastasis modulating gene variants in eight independent population cohorts (mostly Caucasian). This research first screened 49 candidate SNPs from 16 potential metastasis susceptibility modulating genes across two pilot cohorts that were chosen based on mouse model study results. Seven candidate SNPs from four different genes (*SIPA1*, *ARAP3*, *RRP1B*, *BRD4*) showed association with cancer survival in two pilot cohorts. These SNPs were subsequently assayed in other larger independent cohorts.

The results showed that two *ARAP3* gene SNPs, rs440279 and rs3763120, were associated with important clinical markers - response to first line chemotherapy and reduced patient survival (with and without stratification of lymph node metastasis status and oestrogen receptor status). Four *SIPA1* gene SNPs (SNP rs931127, rs746429, rs2448490 and rs3741378) were also found to be associated with patient outcome; rs746429 (missense SNP) was associated with poor outcome in metastasis-free, disease-free and overall-survival in the largest (Rotterdam, the Netherlands) population. The SNP rs2448490 (intronic SNP) from the same gene was found to be associated with better survival rate in the lymph node metastasis negative/oestrogen receptor positive patients. Since the SIPA1 protein and the RRP1B protein were shown to physically interact with each other and the *SIPA1* SNP rs2448490 and *RRP1B* SNP rs9306160 both showed association with survival in the lymph node metastasis negative/oestrogen receptor positive class of patients; a combined analysis of the *SIPA1*

SNP rs2448490 and *RRP1B* SNP rs9306160 was performed, giving the best prognosis in the lymph node metastasis negative/oestrogen receptor positive subgroup.

SNPs in the other *SIPA1* protein binding partner, *BRD4*, also showed association with progression-free survival. SNPs from the three genes, *SIPA1*, *RRP1B* and *BRD4* which are known to be the corner-stone of the Diasporin pathway (Crawford, *et al*., 2008), a novel pathway in the metastasis process, showed association with breast cancer metastasis and metastasis related survival in multiple independent populations, adding support for these genes as metastasis modulator genes. In addition, the SNP rs3741378 from the *SIPA1* gene also showed association with breast cancer incidence, indicating that the *SIPA1* gene may not only be a breast cancer metastasis modulating gene but also a breast cancer susceptibility gene. To further investigate the role of this gene in breast cancer, molecular characterisation of an identified *Sipa1* protein binding partner, Calmodulin 2, was also undertaken as part of this research. *Calm2* was first identified as a potential *Sipa1* protein binding partner from yeast two hybrid analysis, (Myriad Genetics, Salt Lake City, UT) with confirmation of binding verified by Co-immunoprecipitation assay. Stable cell lines with up-regulation of *Calm2* gene were subsequently created. The stable cell lines were used in *in vivo* analyses to investigate the possible role of the *Calm2* in breast cancer metastasis. However, deriving any firm conclusions from the *in vivo* analysis of *Calm2* can't be drawn as the control samples of this study didn't provide adequate data and further investigations of *Calm2's* possible role in metastasis are required.

These novel metastasis susceptibility modulating genes have previously shown association with breast cancer at the transcriptional and translational level. The results of this study provide evidence that these genes are engaged in human breast cancer

survival at the genetic level. The data provide evidence of association between subtle inherited genetic variation of these novel breast cancer metastasis modulators and breast cancer patient outcomes in various independent human populations. The results of this research also provide additional evidence that genetic variation is involved in human breast cancer metastasis and survival. The observation that genetic variation may be an independent indicator for lymph node metastasis, negative/oestrogen receptor positive status, suggests a novel hypothesis in relation to breast cancer biology, ie breast tumours that do and do not colonize to the lymph node system may represent two distinct subtypes of breast cancer, with distinct tumour cell biology and mechanisms of progression.

Statement of Originality

The material presented in this work has not previously been submitted for a degree or diploma in any university, and to the best of my knowledge contains no material published or written by another person except where due acknowledgement is made in the thesis itself.

Table of Contents

	Page
Abstract	i
Statement of Originality	v
Table of Contents	vi
List of Figures	xi
List of Tables	xiii
Publications Arising from Work Described in this Thesis	xvi
Acknowledgement	xviii
List of Abbreviations	xx

Chapter 1 Literature Review 1

1.1 Aims and Significance 2

1.1.1 Introduction 2

1.1.2 Aims 3

1.1.3 Significance 3

1.2 Cancer Overview 4

1.2.1 Initiation and Promotion 4

1.2.2 Metastasis 13

1.3 Breast Cancer Overview 18

1.3.1 Breast Cancer Subtypes 18

1.3.1.1 Basal-like Breast Cancer 20

1.3.1.2 Ductal Carcinoma *in situ* 23

1.3.1.3 *In situ* Lobular Lesions and Invasive Lobular Carcinoma 26

1.3.2 Hormone Receptors and Breast Cancer 30

1.3.2.1 Oestrogen Receptor 30

1.3.2.1.1 Oestrogen Receptor Subtypes 32

1.3.2.1.2 Oestrogen Receptor Interactions 33

1.3.2.1.3 Oestrogen Receptor and

Treatment Response 34

1.3.2.2 Progesterone Receptor 35

1.3.2.2.1 Progesterone Receptor Subtypes 36

1.3.2.2.2 Progesterone Receptor Interactions 37

1.3.2.3 Hormone Receptors and Breast Cancer 37

1.3.3 Breast Cancer Metastasis 39

1.3.3.1Breast cancer Metastasis Epidemiology 40

1.3.3.2 Breast Cancer Metastasis Predictors 42

1.3.3.3 Molecular and Cellular Changes in Metastatic Breast Cancer 45

1.3.3.4 Metastatic Breast Cancer Molecular Signatures 51

Chapter 2 Methods and Material 55

2.1 Study Design 56

2.1.1 Research Background 56

2.1.2 Hypothesis 63

2.1.3 Methodology Background 64

2.1.4 Genes of Interests for Functional Analyses 68

2.2 Equipment and Reagents 79

2.2.1 Equipment 79

2.2.2 Reagents 80

2.2.3 Materials 81

2.3 Study Cohorts 83

2.3.1 Southern California, USA 83

2.3.2 Greater Baltimore, USA 84

2.3.3 Minnesota, USA 86

2.3.4 Fred Hutchinson Cancer Research Centre, USA 88

2.3.5 Rotterdam, the Netherlands 89

2.3.6 Germany 96

2.3.7 Queensland, Australia 96

2.3.8 Umea, Sweden 97

	2.4	Methods	99
		2.4.1 SNP Analysis Methods	99
		2.4.2 Molecular Characterisation Analysis Methods	106
Chapter 3	Single Nucleotide Polymorphism Analyses Results		109
	3.1	Introduction	110
	3.2	*ARAP3* Gene Results	114
	3.3	*SIPA1* Gene Results	131
	3.4	*RRP1B* Gene Results	155
	3.5	*BRD4* Gene Results	168
	3.6	Other Genes Results	176
	3.7	Genotyping Results Summary	189
Chapter 4	Calmodulin 2 Molecular Study Results		191
	4.1	Introduction	192
	4.2	Methods	196
		4.2.1 *In vitro* Yeast-2-Hybrid Result Confirmation	196
		4.2.2 *In vivo* Metastasis Assay	204
	4.3	Results	208
		4.3.1 *Calm2* interacting with *Sipa1* in GT49 Cell Line	208
		4.3.2 *Calm2* Up-Regulated Clone Selection	213
		4.3.3 Calm2 Up-Regulated Clones Implantation in Mouse Mammary Fat-Pad	215
	4.4	Summary of the Novel Sipa1 Protein Binding Partner	

Identification Results 216

Chapter 5 Summary and Discussion of Results 219

5.1 Introduction 220

5.2 Summary and Discussion of the Epidemiology Studies 223

5.2.1 The *ARAP3* gene and its candidate SNPs (SNP rs440279 and rs3763120) 225

5.2.1.1 Introduction of the *ARAP3* Gene 225

5.2.1.2 Summary of the *ARAP3* Gene SNPs Genotyping Results 226

5.2.1.3 Discussion of *ARAP3* Gene SNPs Genotyping Results 231

5.2.2 The *SIPA1* gene and its candidate SNPs (SNP rs746429, rs931127, rs2448490 and rs3741378) 234

5.2.2.1 Introduction of the *SIPA1* Gene 234

5.2.2.2 Summary of the *SIPA1* Gene SNPs Genotyping Results 235

5.2.2.3 Discussion of *SIPA1* Gene SNPs Genotyping Results 242

5.2.3 The *RRP1B* gene and its candidate SNP (SNP rs9306160) 247

5.2.3.1 Introduction of the *RRP1B* Gene 247

5.2.3.2 Summary of the *RRP1B* Gene SNP Genotyping Results 248

5.2.3.3 Discussion of *RRP1B* Gene SNP Genotyping Results 252

5.2.4 The *BRD4* gene and its candidate SNPs (SNP rs4809130 and rs4808272) 253

5.2.4.1 Introduction of the *BRD4* Gene 253

5.2.4.2 Summary and Discussion of the *BRD4* Gene SNPs Genotyping Results 254

5.3 Future Directions of the Epidemiology Studies 256

5.4 Summary and Discussion of the Novel Sipa1 Protein Binding Partner Identification Results 259

5.5 Future Direction of *Calm2* and *Sipa1* Association Research 264

5.6 Discussion 266

5.6.1 The Diasporin Pathway 266

5.6.2 Inherited genetic variance in breast cancer and cancer metastasis 269

Chapter 6 References 275

List of Figures

	Page
Figure 1.1: Schematic representation of the steps of hemotagenous metastasis.	14
Figure 1.2: Comparison of the progression and the transient compartment models of metastasis efficiency.	15
Figure 1.3: Diagram of ductal and lobular anatomical region in normal breasts.	20
Figure 2.1: Effect of genotype on mammary tumour metastatic efficiency.	58
Figure 2.2: QTL mapping of metastasis efficiency on the updated AKXD RI panel.	59
Figure 2.3: eQTL mapping of the Col1a2 locus.	63
Figure 3.1: Raw output of SNP assay.	111
Figure 3.2: Linkage Disequilibrium block for *ARAP3* SNPs.	117
Figure 3.3: Locations of *ARAP3* SNPs.	120
Figure 3.4: Locations of *SIPA1* SNPs.	134
Figure 3.5: *SIPA1* SNP rs7465429 disease-free survival data.	139
Figure 3.6: *SIPA1* SNP rs7465429 metastasis-free survival data.	140
Figure 3.7: *SIPA1* SNP rs7465429 overall survival data.	141
Figure 3.8: *SIPA1* SNP rs2448490 metastasis-free survival data.	142
Figure 3.9: *SIPA1* SNP rs2448490 overall survival data.	143
Figure 3.10: *SIPA1* rs2448490 Kaplan-Meier analysis.	148
Figure 3.11: Locations of *RRP1B* SNPs.	156
Figure 3.12: *RRP1B* SNP rs9306160 metastasis-free survival data.	160
Figure 3.13: *RRP1B* SNP rs9306160 disease-free survival data.	161
Figure 3.14: *RRP1B* SNP rs9306160 overall survival data.	162
Figure 3.15: A summary of metastasis-free survival of *SIPA1* and *RRP1B* genotypes.	165
Figure 3.16: Linkage Disequilibrium block for *BRD4* SNPs.	170
Figure 3.17: Location of *BRD4* SNPs.	173
Figure 3.18: Linkage Disequilibrium block for *PI16* SNPs.	180
Figure 3.19: Linkage Disequilibrium block for *LUC7L* SNPs.	185
Figure 4.1: Yeast 2 Hybrid results.	194
Figure 4.2: Vector Map	197
Figure 4.3: Sipa1, Calm2 and Acin1 up-regulation confirmation	209
Figure 4.4: *Sipa1*, *Calm2* and *Acin1* Co-IP Assay.	212

Figure 4.5: *Calm2* up-regulation qPCR plot. 214
Figure 5.1: The Diasporin Pathway. 268
Figure 5.2; The Metastasis process. 270

List of Tables

Page

Table 2.1: SNPs and genes screened in the pilot cohorts. 77

Table 2.2: Participant characteristics by localized and regional/metastatic disease. 84

Table 2.3: Characteristics of the Greater Baltimore Breast Cancer Cohort. 85

Table 2.4: Characteristics of the Minnesota, USA Cohort. 87

Table 2.5: Characteristics of the Fred Hutchinson Cancer Research Centre, USA Cohort. 89

Table 2.6: Patient and tumour characteristics of the Rotterdam, the Netherlands Cohort. 90

Table 2.7: Characteristics of patients included in the prognostic study, n=1864. 93

Table 2.8: Characteristics of patients treated with first-line chemotherapy, n=227. 95

Table 2.9: Tumour characteristics at diagnosis, Umeå cohort. 98

Table 3.1: SNPs Genotype Frequencies. 113

Table 3.2: Southern California, USA, *ARAP3* gene SNPs, Univariate Analysis. 114

Table 3.3: Southern California, USA, *ARAP3* gene SNPs, Multivariate Survival Analysis. 118

Table 3.4: Greater Baltimore, USA, *ARAP3* gene SNPs, Univariate Analysis. 118

Table 3.5: Greater Baltimore, *ARAP3* gene SNP, Multivariate Cox Regression Survival Analysis. 119

Table 3.6: Minnesota, USA. Mammographic Density, *ARAP3* SNP rs440279. 121

Table 3.7: Umea, Sweden, *ARAP3* Candidate SNP, Univariate Analysis. 122

Table 3.8: Fred Hutchinson Cancer Research Centre, USA, *ARAP3* Candidate SNP, Univariate Analysis. 123

Table 3.9: The Netherlands, *ARAP3* SNP rs440279, Prognosis Univariate Analysis.125

Table 3.10: The Netherlands, *ARAP3* SNP rs440279, Prognosis Multivariate Analysis. 126

Table 3.11: The Netherlands, *ARAP3* SNP rs440279, Prognosis Multivariate Analysis Correct for Adjuvant Therapy. 127

Table 3.12: The Netherlands, *ARAP3* SNP rs440279, Tamoxifen Analysis. 128

Table 3.13: The Netherlands, *ARAP3* Candidate SNPs Genotype Frequency, Patients received first line chemotherapy-Univariate Analysis. 129

Table 3.14: Summary of *ARAP3* SNPs Genotyping Results. 130

Table 3.15: Fred Hutchinson Cancer Research Centre, USA, *SIPA1* Candidate SNP, Multivariate Survival Analysis. 135
Table 3.16: The Netherlands, *SIPA1* Candidate SNP Genotype Frequencies. 137
Table 3.17: The Netherlands, *SIPA1* Candidate SNP Genotype Frequency, Patients received first line chemotherapy-Univariate Analysis. 145
Table 3.18: The Netherlands, *SIPA1* Candidate SNPs, Patients received first line chemotherapy-Multivariate Survival Analysis. 147
Table 3.19: Australia, *SIPA1* Candidate SNP, Breast cancer Incidence Analysis. 153
Table 3.20: Australia, *SIPA1* Candidate SNP, Breast cancer Incidence Analysis, Monte-Carlo style Chi-square test. 153
Table 3.21: Summary of *SIPA1* SNPs Genotyping Results. 154
Table 3.22: Fred Hutchinson Cancer Research Centre, USA, *RRP1B* Candidate SNP, Univariate Analysis. 157
Table 3.23: The Netherlands, *RRP1B* Candidate SNP Genotype Frequency, Patients received first line chemotherapy-Univariate Analysis. 158
Table 3.24: The Netherlands, *RRP1B* SNP rs9306160, multivariate analysis for Metastasis-Free Survival in all patients[a]. 159
Table 3.25: The Netherlands, *RRP1B* SNP rs9306160 + *SIPA1* SNP rs2448490, multivariate analysis for MFS in ER+ lymph node-negative patients. 163
Table 3.26: The Netherlands, *RRP1B* SNP rs9306160 Genotype Frequency, Patients received first line chemotherapy-Univariate Analysis. 166
Table 3.27: Summary of *RRP1B* SNPs Genotyping Results. 167
Table 3.28: Southern California, USA, *BRD4* gene SNP, Univariate Analysis. 169
Table 3.29: Southern California, USA, *BRD4* gene SNP, Multivariate Survival Analysis. 171
Table 3.30: Greater Baltimore, USA, *BRD4* gene SNP, Univariate Analysis. 171
Table 3.31: Great Baltimore, USA, *BRD4* gene SNP, Multivariate Cox Regression Survival Analysis. 172
Table 3.32: The Netherlands, *BRD4* Candidate SNPs Genotype Frequency, Patients received first line chemotherapy-Univariate Analysis. 174
Table 3.33: The Netherlands, *BRD4* Candidate SNPs, Patients received first line chemotherapy-Multivariate Survival Analysis. 175
Table 3.34: Summary of *BRD4* SNPs Genotyping Results. 175
Table 3.35: Southern California, USA, *PI16* gene SNPs, Univariate Analysis. 178
Table 3.36: Southern California, USA, *PI16* gene SNP, Multivariate Survival

Analysis. 180
Table 3.37: Greater Baltimore, USA, *PI16* gene SNPs, Univariate Analysis. 182
Table 3.38: Southern California, USA, *LUC7L* gene SNPs, Univariate Analysis. 184
Table 3.39: Greater Baltimore, USA, Multiple Genes, Univariate Analysis 187
Table 3.40: Summary of All Other Genes SNPs Genotyping Results. 188
Table 4.1: Primary Tumour Weight of Control Clone 7. 215

Publications Arising from Work Described in This Thesis

Publications:

Hsieh S.M., Lintell N.A., Hunter K.W. (2007) Germline polymorphisms are potential metastasis risk and prognosis markers in breast cancer. Breast Disease;(2006-2007)26:157-62.

Szu-Min Hsieh[1*], Maxime P. Look[2*], Anieta M. Sieuwerts[2], John A. Foekens[2] and Kent W. Hunter[1#] Distinct inherited metastasis susceptibility exists for different breast cancer subtypes. (Submitted and Under Review) Breast Cancer Research.

*These authors contributed equally to this manuscript

Hsieh S.M[1,2], Smith R.A.[2], Lintell N.A[1], Hunter K.W[1] and Griffiths L.R.[2] Polymorphisms of the *SIPA1* gene and sporadic breast cancer susceptibility. (Submitted and Under Review) BioMed Central-Cancer.

Nicholas Lintell, **Szu-Min Hsieh,** Renard Walker and Kent W. Hunter. *Ttc9c*, a novel transcriptional regulator of the Diasporin inherited metastasis susceptibility pathway. (Submitted and Under Review) Oncogene.

Szu-Min Hsieh[1*], Maxime P. Look[2*], Argyros Ziogas, Anieta M. Sieuwerts, William Rowe, Jinghui Zhang, Hoda Anton-Culver, John A. Foekens and Kent W. Hunter. *CENTD3* predicts outcome and therapeutic response in node-positive, ER-positive breast cancer. (Final draft preparation).

*These authors contributed equally to this manuscript

Conference Presentations:

Szu-Min Hsieh, Al Ziogas, W. Rowe, Jinghui Zang, Stefen Ambs, Hoda Anton-Culver, and Kent Hunter. Germline Polymorphism in Breast Cancer: A dual cohort pilot study. *National Cancer Institute Center for Cancer Research Fellows and Young Investigators Retreat,* 16th-18th March, 2007, Ocean City. Poster

Szu-Min Hsieh, Stefan Ambs, Anton-Culver Hoda and Kent Hunter. Germline Single Nucleotide Polymorphisms in Breast Cancer: 3 Independent Patient Cohort Studies. *Laboratory of Cancer Biology and Genetics/CCR/NCI Retreat.* 17th-19th November, 2007, Flintstone. Poster

Szu-Min Hsieh, Stefan Ambs, Anton-Culver Hoda and Kent Hunter. Germline Single Nucleotide Polymorphisms in Breast Cancer: Three Independent Patient Cohort Studies. *National Cancer Institute Center for Cancer Research Fellows and Young Investigators Colloquium,* 3rd-5th March, 2008, Ocean City. Oral presentation

Szu-Min Hsieh, Maxine Look, Robert Smith, Nicholas Lintell, Anieta Sieuwerts, John Foekens, Griffiths Lyn, Kent Hunter. Polymorphisms in the SIPA1 gene are associated with sporadic breast cancer susceptibility and hallmarks of poor prognosis in two independent studies. *American Association of Cancer Research*, 19th-22nd April, 2009, Denver. Oral presentation

Acknowledgments

This research was supported by the pre-doctoral fellowship award from the Laboratory of Cancer Biology and Genetics, Centre for Cancer Research, National Cancer Institute, National Institutes of Health, United States of America.

I would like to thank my supervisors Lyn and Kent for giving me the opportunity to pursue a career in cancer research. I am especially grateful for their patience and guidance throughout the years, both inside and outside of the laboratory.

I appreciate all the collaborators, especially Dr Stefan Ambs and Dr John Foekens, for giving me valuable advice, interesting projects and the chance to thoroughly investigate my hypothesis in multiple independent cohorts from around the world.

I would like to thank all the past and present lab members; Marbin and Lutecia who taught me, from scratch, how to set up the genotyping experiments and using the 384-well plate platform. I would not be able to set up efficient experimental procedures for this research without their help. I am also grateful for their friendship outside the laboratory, which made my transition of moving half way across the world and stepping from an Honours student to a PhD research student, smoother.

I appreciate all the support given from the Hunter lab members, past and present, and their valuable scientific teaching over my projects.

I would like to express my great gratitude to Sharon from Stefan's lab and Robert from Lyn's lab, who repetitively guide me through the hateful statistic analyses with great patience, I would not be able to complete the thesis without their help.

I also thank the good friends, Tiffany, Sharon, Aaron, Anil, Scott, Dan, Daoud, Sreedevi and Chi-Ping, who helped me cope with day to day life over this very different land, as well as going through the horror of writing up the thesis, with their endless encouragements.

I would like to express my appreciation to my family and Nick's family for their continuous support and encouragement throughout the years.

Nick my fiancé is the one person I would like to thank the most. All the experiences, research and thesis would not be possible without him.

List of Abbreviations

A	Adenine
ABI	Applied Biosystems
ADH	Atypical ductal hyperplasia
AI	Aromatase inhibitors
ALH	Atypical lobular hyperplasia
ARAP3	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3
AQP2	Aquaporin 2
ARID4B	AT rich interactive domain 4B (RBP1-like)
BLBC	Basal-like breast cancer
BRD4	Bromodomain containing 4
C	Cytosine
CALM2	Calmodulin 2
cDNA	Copy DNA
CGH	Comparative genomic hybridization
CR	Complete remission
CSF1R	Colony stimulating factor 1 receptor
Ct	Threshold cycle
DCIS	Ductal carcinoma *in situ*
DFS	Disease-free survival
ECM	Extracellular matrix
EMT	Epithelia Mesenchymal Transition
eQTL	expression QTL
ER	Oestrogen Receptor
EZH2	Enhancer of zeste homolog 2
FISH	Fluorescence *in situ* hybridization
G	Guanine
ht-SNP	Haplotype-tagging SNPs
IHC	Immunohistochemical
LCIS	Lobular carcinoma *in situ*
LD	Linkage Disequilibrium
LOH	Loss of heterozygosity
LUC7L	LUC7-like
LYVE1	Lymphatic vessel endothelial hyaluronan receptor 1

MAPK14	Mitogen-activated protein kinase 14
MFS	Metastasis-free survival
MYO7A	Myosin VIIA
NC	No change
NCBI	National Centre for Biotechnology Information
OS	Overall survival
PCR	Polymerase Chain Reaction
PBS	Phosphate Buffered Saline
PFS	Progression-free survival
PI16	Peptidase inhibitor 16
PR	Progesterone Receptor
PRS	Post-relapse overall survival
Qld	Queensland
qPCR	Quantitative Real-Time PCR
QTL	Quantitative Trait Locus
RI	Recombinant inbred
SERD	Selective oestrogen receptor down-regulators
SERM	Selective oestrogen receptor modifiers
SNP	Single Nucleotide Polymorphism
STAB1	Stabilin 1
T	Thymine
TBS	Tris Buffered Saline
TDLU	Terminal duct lobular units
TTC9C	Tetratricopeptide repeat domain 9C

Chapter 1

Literature Review

1.1 Aims and Significance

1.1.1 Introduction

Cancer metastasis is the main cause of cancer related mortality, with an estimated 90% of deaths associated with solid tumours resulting from the pathophysiological impact of metastatic secondary disease. Despite advances in both basic science and applied clinical research over recent years, metastasis remains the major health concern in cancer related health issues. Further investigations into the myriad of factors associated with metastatic disease are therefore important to identify critical molecular nodes and targets in this complex process. Such studies will enhance an understanding of the metastatic process and facilitate development and deployment of new or improved clinical management for cancer patients.

One of the essential factors for breast cancer is inherited metastatic susceptibility. Recently, using a mouse model system, it was demonstrated that germline polymorphisms had significant effects on the metastatic potency of a transgene-induced mouse mammary tumour. Subsequent molecular analyses, bioinformatics research and literature review identified potential breast cancer metastasis modulator genes.

This study builds on the data obtained previously on the potential metastasis modulator genes and extends the research into human populations, aiming to elucidate the relevance of these potential novel breast cancer metastasis modulators in breast cancer patients.

1.1.2 Aims

The hypothesis of this research is that putative breast cancer metastasis modifier genes identified in the mouse model based on literature reviews and microarray data from differential metastasis potency mice, are novel breast cancer prognosis and survival indicators in human breast cancer.

1.1.3 Significance

Breast cancer is a very common disorder associated with significant mortality and morbidity. This study is aimed at identifying molecular markers involved in breast cancer development and progression so that better forms of diagnosis and treatment can be developed. The data gathered in this study provide evidence of association between novel breast cancer metastasis modulators and breast cancer patient outcomes in various independent human populations. The results support these candidate genes as novel human breast cancer metastasis modulators, as the subtle inherited variance of these genes is associated with human breast cancer progression in the clinic. The data gathered from this research adds evidence to the theory that subtle inherited variation is involved in human breast cancer metastasis and survival. The results also suggest a novel hypothesis of breast cancer biology – in that breast tumours that do and do not colonize to the lymph node system, represent two distinct sub-types of breast cancer, with distinct tumour cell biology and mechanisms of progression.

1.2 Cancer Overview

Cancer is a major public health concern in developed countries, and its effect is expected to increase, in part due to the aging society. In Australia, the death rate per 100,000 for women aged 40–44 years was 17 increasing to 66 per 100,000 in women aged 65–69 years and escalating to 182 deaths per 100,000 in women aged 85 years or over in year 2005 (AIHW, 2008). A recent review on the breast cancer disease in Australia revealed that breast cancer is the most common invasive cancer diagnosed in females in Australia. It is also the leading cause of cancer death in the female cancer patient group (AIHW, 2006). Cancer-related mortality is a serious issue in other parts of the world, for example, a recently released study indicated that cancer is the second leading cause of death in the US encompassing 23.1 % of deaths and more deaths from cancer than heart disease in people under 85 years of age (Jemal *et al.*, 2007).

1.2.1 Initiation and Promotion

Cancer is a disease involving uncontrolled cell growth and the subsequent spreading to other distant parts of the body. One of the models most often used to describe the origins of carcinogenesis is the multi-step model. In the multi-step model, cancer development can be viewed as a multi-step process, which is supported by clinical observations of pre-malignant cell lesion growths and malignant tumours (Coleman and Tsongalis, 2006). This model suggests that carcinogenesis initiates from normal cells acquiring mutations, which leads to deregulation of normal cellular growth. The multi-step model is also often described by the analogy of Darwinian evolutionary theory. Cells can have random genetic mutations that give rise to a selective advantage in

certain individuals, as viewed by Darwin's natural selection theory. In respect to carcinogenesis, mutations that give the most advantage in survival; i.e. evading normal homeostatic mechanisms such as apoptosis, are passed on to future generations of malignant cells and can be viewed as the "fittest" surviving cells (Ilyas *et al*., 1999; Bertram, 2000). This natural selection process of malignant tumour cells is supported by clinical observations, in that cancers that survive chemotherapy treatments are characterized by an increase in growth rate, metastasis and drug-resistance (Bertram, 2000). This analogy may also provide an explanation for the extremely heterogeneous nature of cancer, in general. It is known from clinical observations that cancer exhibits phenotypical variation; for instance, pancreatic cancer tends to be very aggressive, whereas prostate tumours are normally more organ-confined. Furthermore, tumours that originate from the same tissue can exhibit a range of cellular pathologies; from benign hyperplasia to invasive malignancies (Sonnenschein and Soto, 2008). Despite the heterogeneity observed, malignant tumour cells also share common traits such as proliferating independent of external growth signals and in the presence of anti-growth signals, resistance to programmed cell death-apoptosis, promotion of blood vessel growth (angiogenesis) and metastasis. These hallmarks of carcinogenesis enable the malignant tumour cells to breach cellular barriers against cell expansion and dissemination. These common characteristics may also suggest common tumour initiation mechanisms existing among cancers, regardless of the different tissues of origin (Bertram, 2000; Sonnenschein and Soto, 2008).

The hallmarks of cancer initiation are thought to be formulated through mutagenic events that up-regulate oncogenes and down-regulate tumour suppressor genes. Oncogenes were first identified in the viruses that act as carcinogenic agents, which transform normal cells into malignant tumour cells (Bishop, 1981). It was then

discovered that oncogenes are mutated versions of normal cellular genes, i.e. proto-oncogenes; that had been incorporated into the viral genome (Prada *et al.*, 1982). Tumour suppressor genes in the normal cells generally function as negative cell growth regulators (Weinberg, 1991). The mutations in oncogenes are often dominant and promote uncontrolled cell growth; on the other hand, tumour suppressor genes usually have recessive mutations that inactivate gene functions (Haber and Harlow, 1997). Therefore, inactivation of both copies of a tumour suppressor genes is usually required for tumourgenesis (Ponder, 2001). It was observed that spontaneous carcinogenesis most often developed late in life and the incidence of cancer increases dramatically with age. This indicates that mutations in a single oncogene or tumour suppressor gene are not sufficient for tumourgenesis. Furthermore, *in vitro* research utilizing cell lines and *in vivo* models of carcinogenesis support the multiple-hit hypothesis for the majority of cancers. Several major cell growth regulation mechanisms, such as cell cycle checkpoints, apoptosis, telomere length and growth factor sensitivity are often altered in carcinogenesis (Sonnenschein and Soto, 2008; Ponder, 2001). Cell cycle progression is tightly regulated in normal cells through cell cycle checkpoint proteins. These proteins integrate both external and internal cellular signals to determine the progression of the cell cycle. The tumour suppressor retinoblastoma protein (*Rb*) is a good example. The hypo-phosphorylated *Rb* binds to transcription factor *E2F*, which impedes DNA synthesis. The hyper-phosphorylated *Rb* releases *E2F* and facilitates subsequent transcription of DNA synthesis genes. The loss of *Rb* function disrupts the tight regulation of DNA synthesis in normal cells (Knudson, 1978).

Cell death through mechanisms such as apoptosis and senescence are important mechanisms to ensure normal cell growth, and at the same time prevent an accumulation of harmful mutations. A dysregulation of proliferation alone is not

sufficient for cancer formation; a suppression of apoptotic signalling is also needed (Igney and Krammer, 2002; (Reed, 2003); Gosslau and Chen, 2004). Apoptosis is a response that is involved in a whole array of normal physiologic processes, including immune defence, tissue homeostasis, and development (Gosslau and Chen, 2004). These events lead to DNA fragmentation, blebbing, the formation of apoptotic bodies and ultimately cell death (Saikumar *et al.,* 1999; Preston *et al.,* 2001). The dying cell is engulfed by phagocytes due to the exposure of phosphatidylserine and changes in surface sugars (van den Broek and Hengartner, 2000; Reed and Green, 2002; van Grup *et al*., 2003). The critical step in apoptosis is the activation of the mitochondrial-mediated death pathway, which triggers an increase in the permeability of the mitochondrial outer membrane (the other cell death pathway, necrosis, is initiated by the permeability increase in the plasma membrane, not the mitochondrial) that causes mitochondrial swelling, rupture of the outer membrane, and release of proapoptotic factors from the intermembranous space (Cavalli and Liang, 1998); Lemasters *et al.,* 1998; Kroemer *et al.,* 1998; Susin *et al.,* 1998; Skulachev, 1999; van den Broek and Hengartner, 2000; Reed and Green, 2002; van Gurp *et al.,* 2003). Cancer cells acquire resistance to apoptosis by over expression of antiapoptotic proteins such as *Bcl-2* and/or by the down regulation or mutation of proapoptotic proteins (Bax, *Apaf-1*, caspase-8, and death receptors) (Gosslau and Chen, 2004). Over expression of antiapoptotic *Bcl-2* and Bcl-xL has been estimated to occur in more than 50% of all cancers (Reed, 2003; Yang *et al.,* 2004).

The *p53* protein plays a dual but contrary role in the apoptosis of cancer cells; whilst *p53* mutations increase the resistance to chemotherapy (Lee and Beinstein, 1993) possibly due to a decrease of Bax (Yin *et al.,* 1997), the reintroduction of *p53* into *p53* mutant tumour cells can result in apoptosis by chemotherapy (Lowe, 1999). Cellular

stress responses are primarily mediated via *p53*, which can initiate cell-cycle arrest, DNA repair, apoptosis and senescence. These functions of *p53* are important in suppressing tumour formation and in cancer treatments. *p53* and its pathway are studied in depth to obtain knowledge of carcinogenesis and for developing therapy against it (Vazquez *et al*., 2008). In addition, the *PI3K*/*AKT* pathway may be hyper-activated in some tumours because *Ras*, the catalytic subunit of *PI3K*, and *Akt* are over expressed in several cancers (Gosslau and Chan, 2004). In contrast, phosphatase and tensin homolog deleted on chromosome 10, the cellular antagonist of *PI3K*, is frequently down regulated in various cancer types (Lowe, 1999; Igney and Krammer, 2002).

Another regulatory element impaired in carcinogenesis is cellular senescence, which regulates the lifespan of the cell. Telomeres; the repetitive sequences located at the end of chromosomes, are important in modulating genomic stability and cellular lifespan. The length of telomeres shortens with each DNA replication, and cells reach senescence when the telomere length reaches a critical length. It was found that malignant cells maintained a stable telomere length through up-regulating telomerase, which extends telomere DNA (Hiyama and Hiyama, 2003; Kim, 1997; Shay and Gazdar, 1997). Growth factor independence is another hallmark of cancer. Normal cells respond to environmental signals to regulate growth and proliferation. Growth factor receptors such as epidermal growth factor receptor (*EGFR*) and platelet-derived growth factor receptor (*PDGFR*); and non- receptor proteins such as *ras* and *myc* are often found to have altered regulation in malignant tumours (Colburn *et al*., 1986; Feramisco *et al*., 1985; Guroff, 1987; Kudlow, 1988; Parker and Katan, 1990).

The multi-step model for cancer initiation is a complex model; the tumour-platform model, on the other hand reduces the mechanisms into one simple condition a

malignant cell grows and proliferates while there are stimulations of proliferation and blocking of cell death at the same time. It was observed from early experiments that up-regulation or activation of proto-oncogenes often led to growth arrest or apoptosis. For instance, activation of *ras* in rat embryo fibroblasts gives rise to cell growth arrest (Evan *et al.*, 1992; Hirakawa and Ruley, 1988; Tavoloni *et al.*, 1994); cell death is promoted by activating *src* and *myc,* while *c-myc*, *e2f*, *v-jun*, *CDK*s and certain cyclins also share these dual abilities to promote both proliferation and cell death or growth arrest. Interestingly, it was also found that *myc*-induced apoptosis can be inhibited by serum or insulin-like growth factors (*IGF*), whereas in the absence of serum, *E2F* gives rise to high cell death rates in fibroblasts (Harrington *et al.*, 1994; Hueber and Evan, 1998).

These observations propose firstly that activating proliferating pathways may cause net cell loss as much as cell growth. It also proposes a close relationship between cell proliferation, arrest, death and environmental signals. It was indicated that activation of certain proliferating pathways stimulates both proliferation and death; a cell proliferates only when proper environmental cues that support growth are presented and subsequently block cell death. The cancer platform model therefore proposes that carcinogenesis initiates when a cell gains two mutations simultaneously in a single rate limiting step. This platform of uncontrolled cellular proliferation gives the cell conditions to accumulate further mutations that favour tumourgenesis through interacting with its environment (Pedraza-Farina, 2006; Sonnenschein and Soto, 2008). In this model, carcinogenesis will only initiate when a cell harbours mutations that promote proliferation and at the same time block cell death or provide survival factors; and it is this condition that provides a cancer platform (Pedraza-Farina, 2006; Sonnenschein and Soto, 2008).

Microarray analysis has been applied to generate expression signatures that classify different tumour types. The results of the analyses showed that tumours demonstrated expression signatures recognizable for both individual tumour and tumour families with shared characteristics. This supports the notion that during the cancer initiation process, tumours from different cells or tissues may preferentially deregulate specific pathways that favour the process. Therefore, it is possible that an expression signature may exist in different types of carcinogenesis sorted by both tissue and cell of origin, indicative of the oncogene and tumour suppressive lesions it has undergone. Even though mutations occur at random, once the first few lesions have been selected for and fixed in a clone, the new mutations the tumour cells acquire could be influenced by internal and external environmental selection pressures. Therefore, the initial mutations may predict the types of subsequent mutations in the later stage of a malignant cell's life and different tumours demonstrate similarities that can be ascribed to the tissue or cell-type of origin in molecular classifications (Pedraza-Farina, 2006). Clinical outcomes can be predicted by molecular signatures of tumours. For instance, histological indistinguishable breast cancer can be further sub-grouped into four categories: luminal A, luminal B, HER2/ER negative and normal basal-like breast cancer. Normal basal-like breast cancer is a predictor of poor outcome. These findings suggested that there are subsets of mutations that correlate with specific types of cancer and subsets of genes that are linked to the degree of malignancy of specific tumours. Based on the observation of the existence of these particular genetic pathways, it is implied that late-stage tumours are still dependent on the original mutations for survival. Alternatively, it can also be argued that new mutations may not be influenced by earlier mutations. Once a specific control of proliferation is deregulated, new mutations are selected independently of the original mutations (Perou *et al.*, 2000; Sorlie *et al.*, 2001).

It was hypothesized that oncogene inactivation could lead to tumour regression in malignant tumours, since tumours remain dependent on their initial transforming oncogenic mutation for growth and survival. This hypothesis was put to test via mouse models and several studies reported that *myc* oncogene inactivation led to sustained tumour regression with concomitant promotion of either differentiation or apoptosis in *myc* over-expressed lymphoid and epidermal tissues (Felsher and Bishop, 1999; Pelengaris *et al.*, 2002; Rudolph *et al.*, 2000). In other models, a fraction of malignant tumour cells were refractory to *myc* inactivation. Presumably these cells had acquired new mutations that facilitated *myc*-independent growth, and suggests that new mutations can arise independently of pre-existing ones and may often replace their functions (Beer *et al.*, 2004; Boxer *et al.*, 2004; Jonkers and Berns, 2004; Shachaf *et al.*, 2004). Dysregulation of proliferation and cell death have been the main focus of many cancer initiation and progression models, however, most malignant tumours also share another characteristic in common – poor differentiation. It is evident in the clinic that a decrease in the degree of differentiation often correlates to increased malignancy of lesions (Pedraza-Farina, 2006). Several oncogenes have demonstrated functions not only in regulation of proliferation, and cell death but also in cell differentiation. For instance, *c-myc* expression in bone marrow cells leads to a loss of cell-renewal activity in hematopoietic stem cells leading to differentiation (Wilson *et al.*, 2004); over-expression of *ras* and *src*, which are highly expressed in developing neurons leads to neurite outgrowth in PC12 cells (Ingraham *et al.*, 1989; Kremer *et al.*, 1991; Le Beau *et al.*, 1991; Sassone-Corsi *et al.*, 1989). Hence, in cases when oncogene activations do not lead to cell death and may even protect against it, promotion of terminal differentiation could be an alternative means to obstruct tumour progression. It was therefore proposed to add dysregulation of cellular differentiation as a component of

the cancer-platform model. The expanded model would include that malignant tumours arise only when oncogene-induced differentiation is effectively blocked or when the cellular environment fosters the proliferating function of the oncogene. Sufficient platform for cancer initiation and expansion would be based on promoting proliferation while simultaneously preventing differentiation. In other instances, blocking apoptosis and differentiation at the same time in conjunction with promoting proliferation may be needed to establish a cancer platform (Pedraza-Farina, 2006).

1.2.2 Metastasis

Cancer metastasis accounts for approximately 90% of cancer related deaths. Its impact on society is expected to escalate, due in part to the increasing age of populations world-wide (Weigelt *et al*, 2005). Cancer metastasis refers to the complex process in which cancer cells spread from the primary tumour site and subsequently grow at secondary sites. To colonize a secondary site a cancer cell first needs to separate from the primary tumour, and then invade adjacent tissues and basement membranes (Hsieh *et al*., 2006; Hunter, 2003). The detached cells have to survive entry into the circulation, lymphatic or the peritoneal space before they are able to settle into a distant target organ. At the secondary site the cell typically extravagates into the surrounding tissue, where it can undergo proliferation and subsequently angiogenesis (Figure 1.1) (Hsieh *et al*., 2006; Hunter, 2003). All these steps must be accomplished by the tumour cell whilst it is evading apoptotic death and/or an immunological response (Hsieh *et al*., 2006; Hunter, 2003).

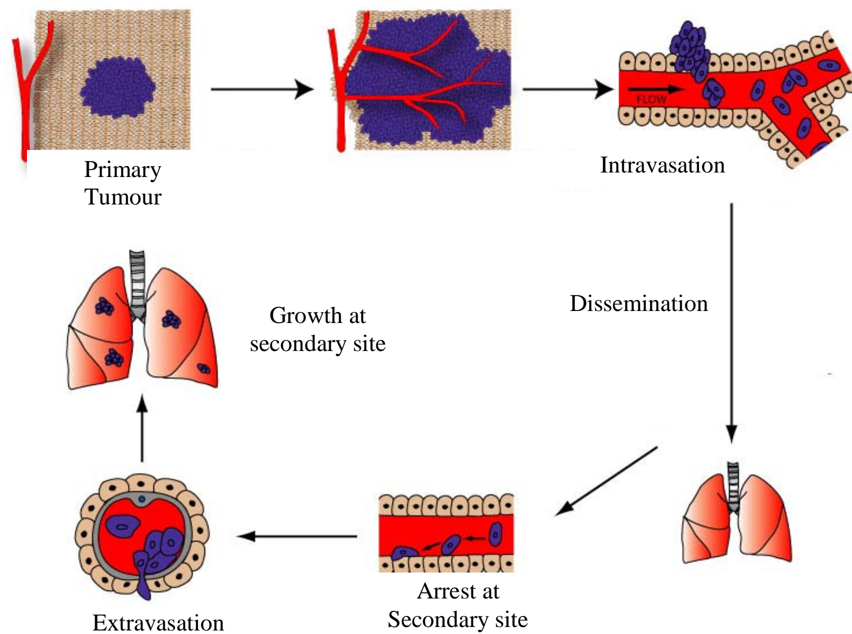
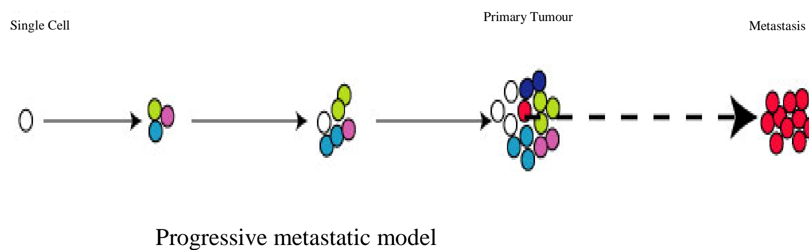


Figure 1.1: Schematic representation of the steps of hemotagenous metastasis. Tumours grow and invade surrounding tissues then intravasate into the vasculature. Tumour cells then disseminate through the body until they arrest, extravasate, then proliferate.

There are several models for metastasis, with each one of them explaining some aspects of metastasis (Figure 1.2, overleaf). However, as yet, none of these models for metastatic progression explain all of the observations associated with metastasis (Hunter, 2005). One widely acknowledged model is the progression (classical) model. It proposes that some tumour cells from the primary site undergo a series of somatic mutational events during tumour growth and eventually acquire all the necessary characteristics to form secondary tumours (Nowell, 1976; Fidler and Kripke, 1977; Hunter, 2005). A strong correlation between primary tumour size (a possible reflection of the increasing age of the tumour) with the risk of metastatic reoccurrence and long-term clinical outcome in a variety of tumour types supports this model (Poste and Fidler, 1980). The presence of mutations that silence metastatic suppressors and up-regulate metastasis promoting elements in a variety of tumours is another indicator for this model of metastasis (Kauffman *et al*, 2003; Steeg, 2003). A caveat for this model is that 5% of patients with disseminated cancers present with no clinically detectable

primary tumour or only a small well differentiated lesion, even after autopsy is performed (Riethmuller and Klein, 2001). This model is further challenged by the inconsistent increase in metastatic capacity of secondary tumours compared to primary ones (Weiss and Ward, 1990).



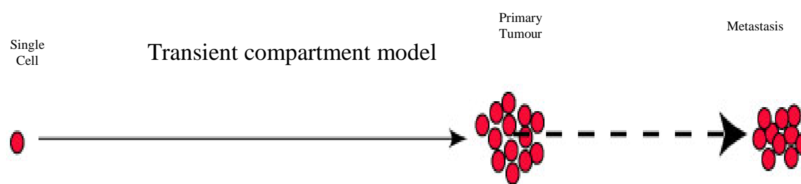


Figure 1.2: Comparison of the progression and the transient compartment models of metastasis efficiency. In the progression model a series of heritable random mutational events occurs within a tumour resulting in a small subpopulation that acquires all of the necessary alterations for metastatic competency. In the transient model, all of the tumour cells have the basic ability, but due to reversible epigenetic events and position within the tumour, not all cells maintain the ability at all times. (Adapted from Hunter, Nature Reviews Cancer, 6, pp. 1-6, 2006)

The other well established model for metastasis is the transient metastatic compartment model, which was proposed by Weiss and attempts to explain all the above observations. This model suggests that all viable cells in a primary tumour have the basic ability to metastasize, but due to reversible epigenetic events and positions within the primary tumour not all the cells are able to maintain the metastatic ability at all times (Weiss and Ward, 1990). The transient compartment model does not, however, explain the clonal nature of metastatic cancer cells. Since primary tumours are known to be heterogeneous, if every cell had metastatic ability that was modulated only by

transient epigenetic events, then it is less likely that significant proportions of secondary tumours would appear to be of clonal origin (Kripke and Fidler, 1980; Talmadge and Fidler, 1982; Fidler, 1986).

The evidence in support of the transient metastatic model is from the recently published data from van de Vijver *et al* in 2002, which was based on micorarray analysis of solid breast cancer samples (van de Vijver *et al.*, 2002). They demonstrated that a particular 70-gene expression signature identified in these samples was a predicator for metastatic progression. These findings have been shown to be reproducible (van de Vijver *et al*, 2002), with a number of different laboratories reporting similar results (Ramaswamy *et al*, 2003). These results indicate that, since the predictive gene signature profile was observed in samples taken from bulk tumour tissue, the majority of tumour cells possess the ability to metastasize. This additionally signifies that the metastatic potential of breast tumour cells may be determined at an early stage, as opposed to the subsequent somatic events that are proposed to occur in the progressive model for metastasis (Bernards and Weinberg, 2002).

Recent evidence has suggested that gene expression signature profiles can be significantly impacted by genetic background. This is theorized to include at least some of the metastatic prognosis predictive genes and therefore the data presented by van de Vijver *et al* could also be explained by genetic germ line polymorphisms. The germ line polymorphism hypothesis proposes that the allelic composition of the host genome is a major determinant in metastatic efficiency, having possibly more effect on this process than somatic mutation events. The combination of subtle changes in gene function, mediated by polymorphisms in coding sequence, splice sites, promoters and enhancers, is theorized to affect how the secondary events associated with the

metastatic process eventuate. Secondary events (which include deletions, amplifications and epigenetic modulations) are necessary for a neoplasm to complete the metastatic cascade and therefore germ line polymorphisms may play a more significant role in cancer metastasis than previously thought. Germ line polymorphisms would not only exert their affects within the tumour cells and primary tumour stroma but they would also be an important factor in the microenvironment at distant secondary sites, which has been previously identified as playing a key role in metastasis formation (Hunter, 2005).

1.3 Breast Cancer Overview

Based on the new report released by the American Cancer Society at 2007, breast cancer is responsible for 15% of cancer related deaths and is the second leading cause of cancer death after the lung and bronchus cancer in women, yet it is the most common cancer amongst females and accounts for 26% of new cases in the States (Jemal *et al*, 2007). Breast cancer is the most common invasive cancer diagnosed in females in Australia. It is also the leading cause of cancer death in females, reported by the Australian Institute of Health and Welfare and the National Breast Cancer Centre. This report projected 14,800 new cases in Australia at year 2011 (AIHW, 2006). Breast cancer is known to be the leading cause of cancer-related death in women world-wide (Polyak, 2007).

1.3.1 Breast Cancer Subtypes

Breast cancer is known to be a heterogeneous disease with numerous subtypes. Large amounts of gene expression profiling data generated by independent groups and technologies have revealed four major molecular subtypes of breast cancer: luminal A, luminal B, HER2/ER negative and normal basal-like breast cancer (Polyak, 2007). The normal basal-like subtype tends to have the worst outcome out of the four subtypes and luminal A subtype the best prognosis (Polyak, 2007). Research has indicated that these subtypes are conserved across ethnic groups and are evident as early as the ductal carcinoma *in situ* stage suggesting distinct tumourgenesis pathways for each tumour type (Polyak, 2007).

The natural history of breast tumourgenesis starts with ductal hyper-proliferation with subsequent evolution into *in situ* and invasive carcinomas (Polyak, 2007). *In situ* breast carcinoma is known as the transition phase in the evolution of invasive malignancy from normal breast tissue (Frykberg, 1999). The malignant epithelial cells are still contained within the breast membranes of their original ductal and lobular structures and the maintenance of the normal histological architectures of these elements is evident in *in situ* breast carcinoma (Frykberg, 1999). In 1975, Wellings *et al* demonstrated that most breast cancers arise from the same locality, the terminal duct lobular unit (Wellings *et al*., 1975). There are five common observed types of breast carcinomas: the columnar cell lesions (CCLs), the apocrine lesions, the basal-like breast cancer (BLBCs), ductal carcinoma *in situ* (DCIS) which primarily occupies the major lactiferous ducts, and lobular carcinoma *in situ* (LCIS) which involves the terminal duct-lobular apparatus of the breast (Frykberg, 1999). Figure 1.3 illustrates the ductal and lobular anatomical regions.

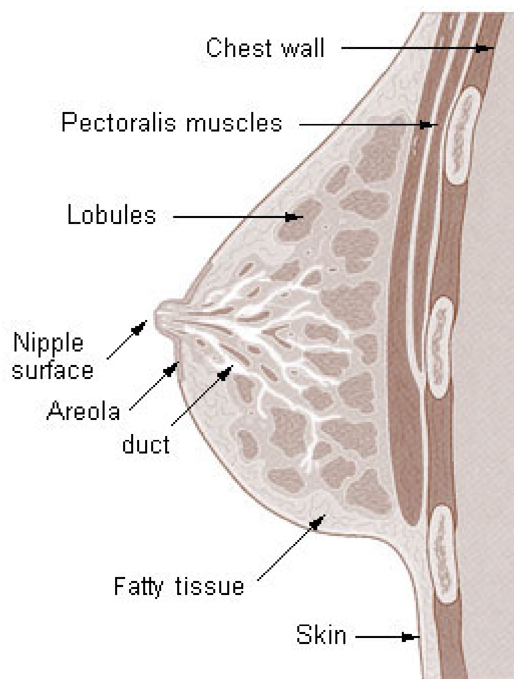


Figure 1.3: Diagram of ductal and lobular anatomical region in normal breasts. Each breast is comprised of 15 to 20 lobes arranged in a circular fashion. The subcutaneous adipose tissue covers the lobes and each lobe is comprised of many lobules, at the end of which are sacs (tiny bulb like glands), where milk is produced in response to hormonal signals. Ducts connect the lobes, lobules, and glands and deliver milk to openings in the nipple in nursing mothers. The areola is the darker-pigmented area around the nipple (http://training.seer.cancer.gov/ss_module01_breast/unit02_
sec01_anatomy. html).

1.3.1.1 Basal-Like Breast Cancer

Basal-like breast cancers were first defined by gene-expression profiling, which indicated that distinct molecular subclasses of breast cancer are present within tumours that are apparently morphologically similar (Fadare and Tavassoli, 2008). BCBLs were named because the neoplastic cells have a similar gene-expression profile to normal basal/myoepithelial cells (Gusterson *et al*., 2005; Lakhani and O'Hare, 2001; Page *et al*., 1999). BCBL expression profiles often include high-molecular-weight basal cytokeratins (*CK*; *CK5/6*; *CK14* and *CK17*), vimentin, p-cadherin, αB crystalline, fascin and caveolins 1 and 2 (Nielsen *et al*., 2004; van de Rijn *et al*., 2002; Sorlie *et al*.,

2001; Savage *et al*., 2008; Hu *et al*., 2006). BCBLs tend to be relatively heterogenous according to a multitude of clinicopathologic parameters.

Even though BCBLs were defined by gene-expression profiling, the majority of the published reports identify BCBLs via immunohistochemistry (Fadare and Tavassoli, 2008). Lacking an internationally accepted definition for BCBLs, a wide range of IHC markers that correlate closely with the basal like gene-expression profile are employed by different studies, nonetheless, most researchers have required, at minimum, positivity of markers for *CK5* or *CK5/6* to classify the tumour as basal-like breast cancer (Fadare and Tavassoli, 2008; Reis-Filho and Tutt, 2008). Under various definitions, BCBLs have been reported to constitute 17.4%-21.2% of sporadic grade III ductal carcinomas, 7.6%-24.6% of sporadic pangrade cancers and 55.7% of triple-negative (ER negative; PR negative and HER2 negative) tumours (Jones *et al*., 2004; Banerjee *et al*., 2006; Fulford *et al*., 2006, 2007). BCBLs are often reported as having more aggressive clinical behaviour (Reis-Filho and Tutt, 2008). One study has also reported different BCBL prevalence among different ethnic groups; the Carolina Breast Cancer Study reported that basal-like phenotype is over-represented in premenopausal African Americans. In this report, almost 40% of breast cancers in premenopausal African American were basal-like, compared with 14% in postmenopausal African American and 16% of cancers in non-African Americans. It was observed that among all the cases studied, patients with BCBLs were more than twice as likely to be African American and the exact mechanisms underling such observations are still unknown (Carey *et al*., 2006).

Morphologically, BCBLs are often associated with the following: high histological grade, high mitotic index, the presence of a central necrotic zone, pushing borders,

spindle cells, squamous metaplasia, high mitotic count (>40 per 10 high-power field), high nuclear to cytoplasmic ratio and conspicuous lymphocytic infiltrate (Hoadley *et al*., 2007; Reis-Filho and Tutt, 2008). Furthermore, BCBLs are more often presented with metaplastic elements and medullary/atypical medullary features in comparison to other types of breast cancer (Fadare and Tavassoli, 2008; Reis-Filho and Tutt, 2008). BCBLs are characterized with morphological features indicative of a highly proliferative state, as genes associated with cellular proliferation are highly expressed in this type of breast cancer (Fadare and Tavassoli, 2008). It was reported that BCBLs have distinct metastasis patterns, in that they disseminate to axillary nodes and bones less frequently and favour a haematogenous spread, with a peculiar proclivity to develop metastatic deposits in the brain and lungs (Fadare and Tavassoli, 2008; Reis-Filho and Tutt, 2008). The increased tendency of brain metastasis from BCBLs has been confirmed in other studies (Hicks *et al*., 2006; Fulford *et al*., 2007; Albiges *et al*., 2005). Fulford *et al* observed that BCBLs were less likely to produce bone and liver metastasis in comparison to non-basal grade III ductal carcinomas, with no difference found between basal and non-basal cancers in terms of lung and pleural metastasis (Fulford *et al*., 2007). Furthermore, published data are conflicting regarding whether BCBLs were more or less likely to show axillary lymph-node metastases, with several studies finding no significant differences between the basal-like and non-basal-like breast cancers (Fadare and Tavassoli, 2008). These contradictory findings can be explained by the lack of universal definition for this type of breast cancer and/or the use of grade III breast cancer as the control group, which might over-represent HER2 positive cases that are more likely to be lymph-node positive (Fulford *et al*., 2007).

In summary, basal-like breast cancer is a newly classified subtype of breast cancer (categorized by gene-expression profiling) with a distinct metastasis pattern. The growing knowledge in this subtype of breast cancer is going to aid in breast cancer

patient managements in the clinic and understanding of the complex phenotypes of human breast cancer.

1.3.1.2 Ductal carcinoma *in situ*

Ductal carcinoma *in situ* is characterized by the proliferation of malignant breast duct epithelial cells that are confined within the basement membrane of the ducts and have not invaded the stroma (Irvine and Fentiman, 2007). There are a few different prognostic classifications in use in the clinic, such as the Van Nuys Prognostic Index and the European Organization for Research and Treatment of Cancer grading system (well, intermediate or poorly differentiated) (Irvine and Fentiman, 2007). The European Organization for Research and Treatment of Cancer grading system grading system is based on the cyto-nuclear pattern, with DCIS exhibiting a tenfold increase in growth and a 15-fold increase in apoptosis and has been modified in screening programs to low, intermediate and high grade (Irvine and Fentiman, 2007). It is estimated that DCIS is responsible for 25% of screen-detected breast cancer. One study has reported that 70% of studied patients of DCIS developed local recurrence or progression to invasive disease (Irvine and Fentiman, 2007; Betsill *et al*., 1978). Ductal carcinoma *in situ* has been considered as a disease unlikely to progress to invasive disease, however, Collins and Colleagues reported otherwise. They reviewed the histology from 1,877 benign breast biopsies in women participating in the Nurses' Health Study in the United States and reported a 77% (10 out of 13 cases) of recurrent incidence in ductal carcinoma *in situ* (Collins *et al*., 2005). The increased risk is observed in all grades of DCIS (Irvine and Fentiman, 2007).

De Mascarel and colleagues reviewed 1,248 patients and reported two distinct types of microinvasions for DCIS (De Mascarel *et al*., 2002). Microinvasion was defined in

1997 by the American Joint Committee on Cancer as malignant tumour cells extending beyond the basement membrane into the adjacent tissue with no focus larger than 0.1 cm (*AJCC Cancer Staging Manual*, 1997). Type 1 microinvasion was described as invasion of the periductal stroma by a few isolated tumour cells; type 2 microinvasion was described as invasion of the periductal stroma with clustered tumour cells (Irvine and Fentiman, 2007). It was reported that patients with basic ductal carcinoma *in situ* or type 1 microinvasion DCIS had a very low level of lymph node metastasis after resection (1.4% and 0%, respectively); whereas patients with small area invasion of DCIS (less than 80% of the entire lesion) and type 2 microinvasion had a higher rate of lymph node metastasis (10.1 % and 27.6%, respectively) (De Mascarel *et al.*, 2002). Molecular research further described genetic instabilities in DCIS. Hwang *et al.* investigated the loss of heterozygosity in high grade DCIS and reported frequent gain of 17q in this study group (Hwang *et al.*, 2004). Smeds *et al.* observed greater levels of loss of heterozygosity (LOH) in well-differentiated DCIS compared to poorly differentiated ductal carcinoma *in situ* (Smeds *et al.*, 2005).

Other factors related to DCIS and the increased potential for invasion or increased cell growths have also been described. For example, higher levels of matrix metalloproteinases have been reported in DCIS and may participate in basement membrane penetration and metastasis (Irvine and Fentiman, 2007). Expression of plasminogen activator inhibitor (*PAI*) that may facilitate invasion through the basement membrane has been described in high grade ductal carcinoma *in situ* myoepithelial cells (Hildenbrand and Arens, 2004). Furthermore, loss of putative tumour suppressor gene *NES1* has also been reported in high grade ductal carcinoma *in situ* and serves as a predictor for the presence of invasive disease on final histology (Yunes *et al.*, 2003). An increased higher apoptotic index, such as the proliferation markers mitotic index

and *Ki67* have been observed in ER-negative and high grade ductal carcinoma *in situ* and are associated with a higher cell proliferation rate and high mortality rate (Gandhi *et al*., 1998). It was known that larger tumours at a more advanced stage show a higher level of cyclooxygenase (*COX*)-*2* expression and the level of *COX-2* expression correlates with the grade of breast tumour (Shim *et al*., 2003). *COX-2* is regarded as anti-apoptotic and higher expression is found in recurrent breast tumours; it was observed that *COX-2* expression increased as the grade increased in ductal carcinoma *in situ* (Barnes *et al*., 2003). This is supported by Boland's finding that *COX-2* expression is associated with a more aggressive DCIS phenotype including higher proliferation rates, higher nuclear grade, higher level of HER2 expression and more frequent ER negativity (Boland *et al*., 2004).

Ductal carcinoma *in situ* is also associated with hormonal status; Hawkins *et al*. reported that 77% of invasive breast cancers were ER positive compared to 44% of DCIS being ER positive in their studied group (Hawkins *et al*., 1987). Other studies have described a direct link between ductal carcinoma *in situ* differentiation and the likelihood of ER positivity (Bur *et al*., 1992; Pallis *et al*., 1992; Poller *et al*., 1993). Another important marker for breast cancer, HER2 status, has been found to be associated with DCIS. Van de Vijver *et al* reported HER2 over-expression in 14% of 189 studied cases and 42% of 45 DCIS cases (van de Vijver *et al*., 1988). Another study using fluorescence *in situ* hybridization (FISH) analyses on 100 cases of pure DCIS and 100 with an associated invasive component showed that HER2 amplification was more frequent in higher than lower grade ductal carcinoma *in situ* lesions (56% vs 19%). HER2 over-expression is more frequent in those samples with an invasive component (44%) than pure DCIS (2%) (Hoque *et al*., 2002).

In summary, ductal carcinoma *in situ* represents approximately one-quarter of the screen-detected breast cancers, and microinvasion associated with DCIS is often an indication for lymph node metastasis. In addition, patients with ER-positive ductal carcinoma *in situ* are likely benefit from adjuvant tamoxifen after breast-conserving surgery.

1.3.1.3 *In situ* Lobular Lesions and Invasive Lobular Carcinoma

Lobular lesions are typically characterized by populations of small aberrant cells with small nuclei, individual private acini and a lack of cohesion between cells. Furthermore, this type of breast cancer has increased propensity for transcoelomic metastases and dissemination to the gastrointestinal tract (Savanis *et al*., 2006; Kadivar *et al*., 2006; El Khoury *et al*., 2004; Giagounidis *et al*., 2005). Lobular carcinomas have a higher rate of incidence, in women aged≥50 years (Li *et al*., 2000). Atypical lobular hyperplasia (ALH) and LCIS are the two most common non-invasive lobular neoplasias (Hanby and Hughes, 2008). Atypical lobular hyperplasia is characterized by small discohesive cells whose cytoplasm may contain variable conspicuous private acini, occupying part of or the entire terminal duct lobule of the breast; the colonized duct lobular units are not expanded, nor are their lumina obliterated (Hanby and Hughes, 2008). Lobular carcinoma *in situ* has expanded colonized structure and losses of lumina, it is often found in close proximity with ALH lesions and sharing similar cyto-morphological characteristics (Hanby and Hughes, 2008). Lobular carcinoma *in situ* was first described in 1941 as a distinct pathologic entity and is thought to represent a transitional intra-epithelial, or *in situ*, stage in the evolution of breast cancer from hyperplastic breast epithelium (Foote and Stewart, 1941). The average age for LCIS diagnosis is about 45 years old which is 10 to 15 years younger than the average diagnosis of invasive breast cancer (Frykberg, 1999). Furthermore, LCIS has a higher

rate of ERpositivity than invasive carcinomas, which suggests that hormonal influence is important in its development and biology (Frykberg, 1999).

Lobular carcinoma *in situ* is characterized by proliferation of uniform, bland and homogenous malignant cells that have no mitoses or necrosis. The malignant cells are within the terminal duct-lobular apparatus as part of the lobular architecture, to the point of distending to the lobular lumen, with the investing basement membrane remaining intact and no evidence of invasion into the surrounding stoma (Frykberg and Bland, 1993). Cells of LCIS are typically low histological and nuclear grade, having diploid DNA morphology; are highly ERpositive; and have a low proliferative index and oncogene expression with high tumour suppressor gene expression (Frykberg, 1999). The characteristics of LCIS indicate a benign and indolent biological behaviour with low potential for invasion and metastasis. This is very different from ductal carcinoma *in situ*, which is typically more aggressive biologically with associated cytological characteristics, such as low ERand high proliferation (Frykberg, 1999). LCIS is often found with multicentricity and bilaterlity (more than 50% of patients may have LCIS in the opposite breast); furthermore, even though the risk of malignancy applies to both breasts, the risk is higher in the ipsilateral than the contralateral breast (Rosen, 1991). Lobular carcinoma *in situ* is assumed to be widely disseminated throughout all breast tissues and the foci of LCIS are often found incidentally outside the area of abnormality (Frykberg, 1999; Hanby and Hughes, 2008). The absence of any clinical or mammographic evidence of LCIS presence can be explained by the observation that LCIS does not stimulate calcium deposition or creation of a discrete mass. This may due to slow growth, absence of mitosis and extracellular secretions and its diffuse infiltrative growth pattern (Sonnenfeld *et al.*, 1991). Even though LCIS imparts a 12-fold increased risk of subsequent invasive breast carcinoma, it is

considered as a marker of risk rather than a true pre-malignant lesion based on its natural history (Frykberg, 1999).

Molecular studies on both LCIS and invasive lobular carcinoma revealed that loss or down-regulation of E-cadherin (*CDH1*), an adhesion molecule localized at the *zonula adherens* of epithelial cells that enhances cellular cohesion by homotypic interactions, is the hallmark molecular feature of this type of breast lesion (Boller *et al*., 1985; Takeichi, 1991; Gumbiner *et al*., 1988). Classic lobular neoplasms are epidermal growth factor receptor 1 and HER2-negative; and *34bE12* (cytokeratins 1, 5, 10, 14), ERand progesterone receptor positive (Wheeler *et al*., 2004; Brattauer *et al*., 2002). Cytogenetic studies revealed that classic lobular lesions have relatively lower numbers of changes compared with other types of breast carcinomas, typically presenting with loss of 16q and a gain of 1p (Roylance *et al*., 1999; Vos *et al*., 1999). It is important to note that lobular lesions closely resemble grade-one ductal carcinomas and are commonly mixed up with one another, leading to the suggestion that these lesions are closely related cytologically and biologically (Hanby and Hughes, 2008). Pleomorphic lobular carcinoma *in situ* (PLCIS) is a subset of LCIS, that has general characteristics similar to LCIS, yet tend to have more pleomorphic nuclei and are more associated with comedo-type necrosis and micro-calcification (Hanby and Hughes, 2008). It was observed that PLCIS shares cytological and immunohistological features with both invasive pleomorphic lobular carcinoma and some high-grade ductal neoplasms and therefore often has a morphology resembling high-grade DCIS (Brattauer *et al*., 2002).

Invasive lobular carcinomas (ILC) are estimated to represent up to 14% of all invasive breast cancers and are classified into six major groups; the classic subtype and the non-classic subtype, which include the solid subtype, alveolar subtype, pleomorphic

subtype, apocrine/histiocytoid subtype and the tubule-lobular subtype (Singletary *et al.*, 2005). It is known that the non-classic subtypes have worse prognosis than the classic subtype (Hanby and Hughes, 2008).

In summary, lobular carcinoma *in situ* is one of the most common non-proliferative lobular proliferations. It is clear that lobular carcinoma is a diverse group of diseases. The morphological and cytogenetic diversity suggests a spectrum of neoplasms with the common feature of loss of E-cadherin function and a close relationship to low-grade ductal carcinoma. More research is needed to elucidate the biology underlines the risk association with pleomorphic lobular carcinoma *in situ* and the genetic basis of the frequent bilaterality (Hanby and Hughes, 2008).

1.3.2 Hormone Receptors and Breast Cancer

1.3.2.1 Oestrogen Receptor

Oestrogen is a steroid hormone, a key intracellular modulator for the processes involving differentiation, homeostasis and development. Systematically, oestrogen is involved in maintenance of reproductive functions, lipid and cholesterol homeostasis in women and homeostasis of the bone tissues (Shelly *et al.*, 2008). Studies in steroid hormone knock-out mice have revealed that oestrogen, along with its receptors, promote the growth of ducts that invade the mammary fat pad emanating from the nipple (Lange, 2008). Pathologically, oestrogen is involved in breast cancer progression and its exposure is an established risk factor for breast cancer (Shelly *et al.*, 2008; Payne *et al.*, 2008; Tonini *et al.*, 2008). Oestrogen that passively diffuses through the cell membrane exerts its actions via the ER, a specific high affinity protein located in

the cell nucleus (Shelly *et al.*, 2008; Bedard *et al.*, 2008; Nicolini *et al*, 2006). The ER status, as a tissue marker and one of the most powerful individual predictive factors examined in breast cancer, has been routinely used for many years to help determine patient suitability for endocrine therapy (Payne *et al.*, 2008; Nicolini *et al*, 2006). ER is estimated to be expressed in approximately 70% of patients with breast cancer (Bedard *et al.*, 2008; Tonini et a., 2008; Lange, 2008) and it is estimated that up to two thirds of women less than 50 years of age and about 80% of women over 50 years of age (Payne *et al.*, 2008). It was reported that breast tumours that are ER+ and/or PR+, which are largely independent of other clinical tumour characteristics, have a lower mortality risk compared with ER- and/or PR- tumours (Payne *et al.*, 2008).

The oestrogen receptors belong to the steroid-retinoid receptor super-family, which includes other steroid hormones such as retinoic acid, vitamin D and thyroid-hormones (Bedard *et al.*, 2008; Cui *et al*, 2005). In the absence of oestrogen, the ERresides in a large macromolecular complex, with the receptor bound to heat-shock proteins (HSPs) (Nilsson *et al.*, 2001). The binding of 17 [β]-oestradiol (the major form of oestrogen in humans) to the ERs causes a conformational change in the ER protein structure, which leads to dissociation and release of the ERs monomers from the chaperon heat-shock proteins complex (Smith *et al.*, 1997). McDonnell and colleagues showed that a series of ER ligands from different ER-binding ligand complexes would induce different conformational changes (McDonnell *et al.*, 1995). The phosphorylated ER monomers dimerize and function as a transcription factor that ultimately initiate target gene transcription (Shelly *et al.*, 2008; Payne *et al.*, 2008; Bedard *et al.*, 2008; Nicolini *et al*, 2006). The activated ER-dimers then release co-repressors, such as the nuclear receptor co-repressor (*NCoR*), histone deacetylase 1 (*HDAC1*), and metastasis-associated antigen 1 (*MTA-1*). The ER-dimers also recruit co-activators, such as amplified in

breast cancer-1 (*AIB1*), nuclear-receptor-coactivator-1 (*NCoA-1/SRC1*), *p300* and *CBP*-associated factor (*PCAF*) (Osborne and Schiff, 2005; Normanno *et al*., 2005; Cui *et al*, 2005). The activated ER-dimer complex subsequently binds to DNA promoter regions known as oestrogen-response elements (EREs), which regulate target gene expression. The target genes are often involved in cell proliferation, inhibition of apoptosis and promotion of angiogenesis, invasion and metastasis (Osborne and Schiff, 2005; Normanno *et al*., 2005).

1.3.2.1.1 Oestrogen Receptor Subtypes

There are two known subtypes for the oestrogen receptor: ER-alpha (ER-α) and ER-beta (ER-β), which are located on separate chromosomes and are encoded by different genes, yet both are nuclear located proteins that bind to oestradiol with similar affinity (Payne *et al*., 2008; Bedard *et al*., 2008; Cui *et al*, 2005; Nicolini *et al*, 2006). Clinical determination of ERstatus can't distinguish the two isoforms (Speirs *et al*, 1999), it was observed that ER-α has a different affinity in binding to other ligands compared to ER-β (Payne *et al*., 2008; Bedard *et al*., 2008). ER-α and ER-β have different tissue distributions, with studies showing that ER-α is expressed either alone or in combination with ER-β in most breast cancer tumours, whereas ER-β predominates in normal breast tissue (Speirs *et al*, 1999; Bedard *et al*., 2008). The difference in binding characteristics and tissue distributions may explain the tissue-selective response to oestrogen (Bedard *et al*., 2008). It is known that positive ER status predicts benefit from anti-oestrogen therapy and the concentration of ERin the tumour correlates with endocrine manipulation responses (Robertson, 1996; Spataro *et al*, 1992; Harvey *et al*, 1999). It was reported that the ratio of ER-α:ER-β increases when breast cells become increasingly malignant, and therefore ER-β may act as a tumour suppressor via forming a dimer complex with ER-α to inhibit ER-α-driven transcription of genes involved in

tumour growth and metastasis (Speirs *et al*., 1997). Furthermore, there appears to be an inverse relationship between the level of ER-α and *EGFR* in breast cacner. It was speculated that the loss of ER-α and its association with anti-oestrogen therapy may induce higher sensitivity of cells to growth stimuli via switching to *EGF*, as it was observed that *EGFR* is up-regulated in cells that have become resistant to anti-oestrogen therapy (Sommer *et al*., 2003).

1.3.2.1.2 Oestrogen Receptor Interactions

Up to one-third of ERregulated genes reportedly don't contain ERE-promoter sequences. This indicates that the ERis involved in other tumourgenesis pathways, via functioning as a co-activator that binds to key proteins involved in the promoter complex, such as *fos* and *jun*, to stabilize the binding of other transcription factors to DNA at alternative response elements (e.g *AP-1*) (Osborne and Schiff, 2005; Normanno *et al*., 2005). Genes regulated by the non-classical ER co-activator complex, are usually proteins involved in cell proliferation and metastasis, such as insulin-like growth factor (*IGF-1*), *myc*, Cyclin D1, *Bcl-2* and collagenase (Osborne and Schiff, 2005; Normanno *et al*., 2005). The majority of cellular ER is located in the nucleus, with smaller amounts present in the cytoplasm and cell membrane (Jacob *et al*., 2006; Cui *et al*, 2005). Since the ERprotein does not contain hydrophobic trans-membrane domains or membrane localization sequence, cellular ERs localize to the cell membrane by associating with cytoplasm membrane anchors, such as caveolin, or by complexing with alternative trans-membrane proteins such as *GPR30*. Cellular ERs are postulated to be involved in post-translational modification, producing alternative splicing variants of target genes and playing roles in intracellular signalling pathways, which promote cell survival (Filsrdo and Thomas, 2005). It was observed that membrane-bound ER interacts with members of the type 1 tyrosine kinase growth

factor receptor family, such as *IGF-1R*, *EGFR* and HER2. The ERcan associate with Shc at the cytoplasm membrane surface and induces phosphorylation of *IGF-1R* which, leads to phosphoinositide 3-kinase (*PI3K*) activation and induction of the *Akt* (protein kinase B) pathway (Kahlert *et al*., 2000). Membrane-bound ERwas also reported to be associated with co-adaptor protein *PELP1/MNAR*, which contains *PI3K* activating domains that can directly activate ER*K/MAPK* while associating with ER. On the other hand, *PELP1/MNAR* may facilitate ER binding to *IGF-1R* and *EGFR* and induce *PI3K*-driven activation of *Akt* (Barletta *et al*., 2004; Pietras, 2006; Ellis *et al*., 2003; Levin, 2005). Cytoplasmic ERstimulates cell survival through non-genomic molecular cross talk, by stimulating those pathways, which phosphorylate pro-apoptotic proteins (such as *BAD*, caspase-9, Forkhead transcription factors and IkappaB kinase alpha) and activate *Akt* downstream effectors (such as mammalian target of rapamycin; mTOR) that increases transition from G1-S phase in the cell cycle (Johnson *et al*., 2003). Activated *Akt* also phosphorylates nuclear ER and its co-regulators, which further drive ER induced transcription activity (Font de Mora and Brown, 2000; Simoncini *et al*., 2000).

1.3.2.1.3 Oestrogen Receptor and Treatment Response

The ERplays an integral role in the efficiency of breast cancer treatment as the basis of hormone therapy is to block oestrogen-induced proliferation of breast cancer cells (Payne *et al*., 2008; Bedard *et al*., 2008). There are currently three broad classes of hormone therapy used clinically to treat hormone-responsive breast cancer: selective ERmodifiers (SERMs, such as tamoxifen), selective ERdown-regulators (SERDs, such as fulvestrant) and aromatase inhibitors (AIs, such as letrozole, anastrozole and exemestane) (Payne *et al*., 2008; Pietras, 2006; Lange, 2008). The most well characterized SERM, tamoxifen, binds directly to the oestrogen receptors and blocks its

access for other nuclear receptor cofactors and hence transcriptional activity (Shiau *et al*., 1998; Pietras, 2006). Tamoxifen has been the mainstay hormone therapy for breast cancer for the past 25 years, as it improves survival in early stage breast cancer as well as the quality of life for patients with metastasis and is effective in reducing the incidence of breast cancer in patients at risk and in women with ductal carcinoma *in situ* (Dellapasqua and Castiglion-Gertsch, 2005). Over the past decade, other therapies against ERpositive breast cancer have been introduced; SERDs function by inducing ERdegradation by directly binding to ER(Bedard *et al*., 2008; Tonini *et al*., 2008) and AIs reduce the production of oestrogen via inhibition of the aromatase enzyme in peripheral tissues, such as bone, muscle, and adipose tissue to inhibit proliferation of disseminated ERpositive breast tumour cells (Payne *et al*., 2008; Bedard *et al*., 2008; Tonini *et al*., 2008). These agents lead to a more effective inhibition of ER signalling and are an essential part of breast cancer treatment in both adjuvant and metastatic settings (Tonini *et al*., 2008).

1.3.2.2 Progesterone Receptor

Progesterone is an ovarian steroid hormone, whose actions are primarily mediated by the high affinity progesterone receptors (PRs) (Lange, 2008). Progesterone is essential for normal breast development during puberty and in preparation for lactation during pregnancy. Studies in steroid hormone receptor knock-out mice have revealed that oestrogen, with the oestrogen receptor, and progesterone, with the progesterone receptor, are required for the development of the terminal end-buds (TEBs) or acini located at the ends of breast ducts, that later become the milk producing structure in the lactating mammary gland (Feng *et al*., 2007; Mulac-Jericevic and Conneely, 2004; Haslam *et al*., 1993; Ruan, Monaco and Kleinberg, 2005). The progesterone receptors are ligand-activated transcription factors, belonging to a large family of related steroid

hormone receptors, which include the oestrogen receptor, androgen receptor, glucocorticoid receptor and mineralocorticoid receptor (Lange, 2008). Progesterone receptors localise to a diverse range of tissues such as, the brain (for reproductive behaviour), breast and reproductive organs (Lange, 2008).

1.3.2.2.1 Progesterone Receptor Subtypes

Expression of progesterone receptors can be modulated by ER or independently of ER and has three isoforms that are all products of a single gene located on chromosome 11 at q22-23 (Kastner *et al*, 1990; Cui *et al*, 2005; Hewitt and Korach, 2000; Kraus *et al.*, 1993; Payne, 2008). The two major isoforms (which can't be distinguished with common clinical assays) are PR-A (94 kDa; N-terminal truncated protein; required for uterine development and reproductive function) and PR-B (116 kDa; full length protein; essential for normal mammary gland development) (Kastner *et al*, 1990; Kraus *et al.*, 1993; Jacobsen *et al.*, 2002; Richer *et al.*, 2002; Payne, 2008), and they are often co-expressed in PR+ cells (Nicolini *et al*, 2006). The third isoform of PR is PR-C (60 kDa), which when expressed can enhance PR activity in breast cancer cells or function as a dominant inhibitor of PR-B in the uterus (Jacobsen *et al.*, 2002; Richer *et al.*, 2002). Transcriptional regulation of the three isoforms is governed by the use of distal or proximal promoter regions and different internal translational start sites present in common mRNA (Kraus *et al.*, 1993). Studies in the rodent mammary gland showed that the ratio of PR-A to PR-B is critical for proper mammary gland development (Lydon *et al.*, 1996). The progesterone receptor regulates target gene expression by binding directly or indirectly to their target DNA. The two major isoforms exhibit different transcriptional activities within the same promoter context, yet can recognize entirely different gene promoters (Jacobsen *et al.*, 2002; Richer *et al.*, 2002). Since progesterone receptor expression can be modulated by ER-α (in the presence of EGF),

its expression is therefore thought to indicate a function of the ER pathway and the assessment of PR should therefore assist in predicting response to hormone therapy, though, more observations are needed to support this theory (Cui *et al*, 2005; Payne, 2008; Ankrapp *et al*., 1998).

1.3.2.2.2 Progesterone Receptor Interactions

The progesterone receptors are like the oestrogen receptors in that they are also bound to chaperone proteins, such as HSPs. The chaperones in the PR complex facilitate PR protein trafficking as well as allowing proper PR protein folding and assembly of stable PR molecules competent for hormone binding (Pratt and Toft, 2003). Once progesterone binds to the progesterone receptors, they undergoe a conformation change, which induces dimerization of monomer PRs and HSP dissociation, (Lange, 2008). The activated PR dimers interact with co-factors, such as steroid receptor co-activators (SRCs 1-3) and the complexes remain in the nucleus. The activated progesterone receptor complex functions as a transcription factor that binds directly to specific progesterone response elements (PREs) and PRE-like sequences in the promoter region of target genes, such as *c-myc*, *fatty acid synthetise* and *MMTV* (Moore *et al*., 1997; Chalbos *et al*., 1987; Krusekopf *et al*., 1991). Recent studies indicate that cytoplasmic protein kinase activation is important for PR transcriptional function. It was observed that phosphorylation events are required for progesterone receptor gene regulatory activities and therefore, it is speculated that rapid phosphorylation events may primarily act to alter PR transcription activity to mediate promoter selectivity (Byron *et al*., 2006).

1.3.2.3 Hormone Receptors and Breast Cancer

During normal breast development, additional hormone such as *EGF* and *IGF-1* are required for normal oestrogen and progesterone induced proliferation of terminal end-buds and promotion of ductal outgrowth and side branching (Haslam *et al*., 1993; Ruan *et al*., 2005). The majority of normal proliferating breast epithelial cells are steroid hormone receptor negative, with ERpositive and progesterone receptor positive cells representing only 7%-10% of the luminal epithelial cell population. ERpositive and progesterone receptor positive cells are usually non-dividing; primarily lying adjacent to proliferating cells in the normal adult mammary gland (Robinson *et al*., 2000). Recent studies suggest that ER+/PR+ cells, even though capable of proliferating, are growth arrested by the expression of inhibitory molecules, such as *TGF-beta* or high levels of *p21* and *p27* and the endogenous inhibitors cell-cycle-dependent protein kinases (*CDK*s) (Robinson *et al*., 2000). Current observations suggest that ER+/PR+ cells may provide growth-promoting molecules to nearby cells, as communication between the breast epithelial and stromal compartments to mediate adjacent cell proliferation via secretion of locally active pro-proliferating molecules, such as *IGF-1* or stroma-derived hepatocyte growth factor (*HGF*) (Haslam and Woodward, 2003).

Approximately 70% of breast cancers express ERand more than half of these tumours also express progesterone receptor (Wenger and Clark, 1998). It was observed that steroid hormone receptor positive tumours are most often slower growing comparing to receptor-negative tumours (Wenger and Clark, 1998). Furthermore, positive status for PR, independent of ER status, is a prognostic marker for less aggressive and favourable disease-free survival. ERpositive and progesterone receptor positive tumours respond better to hormone therapy in comparison to having only one of the hormone receptors (Jacobsen *et al*., 2002; Shyamala *et al*., 1998). Some studies have shown that ER and

PR status can change over the natural history of breast cancer or during treatment. Reports revealed that ER levels are reduced slightly with intervening hormone therapy in sequential breast cancer biopsy studies, with complete loss of ERuncommon. In contrast, progesterone receptor levels decrease more dramatically during tamoxifen therapy, with up to half of tumours examined, showing complete loss of PR expression when resistance develops (Cui *et al*, 2005). Clinically, these ERpositive and progesterone receptor negative tumours have a more aggressive metastatic course after loss of PR, in comparison with tumours retaining PR, and give patients a worse overall survival rate (Gross *et al*., 1984; Balleine *et al*., 1999). Recent studies on human breast cancer reported that overabundance of PR-A is associated with resistance to tamoxifen and that a functional promoter polymorphism that results in increased production of PR-B is associated with an increased risk for development breast cancer (Hopp *et al*., 2004; De Vivo *et al*., 2003).

1.3.3 Breast Cancer Metastasis

Cancer metastasis is known to be the major cause of cancer related death. Despite significant advances in primary tumour treatment, the ability to predict the metastatic behaviour of cancer and detect and eradicate recurrence remains the major clinical challenges in oncology (Urquidi and Goodison, 2007). A study designed to assess the quality of life in patients with breast cancer metastasis and their families indicated that physical, functional and emotional wellbeing were seriously decreased. Other studies reported similar findings in that women with distant breast cancer metastases have a significant decline in social support, compared with disease free survivors (Northours *et al*., 2002; Oh *et al*., 2004; Dorval *et al*., 1998; Sorensen *et al*., 2004).

1.3.3.1 Breast Cancer Metastasis Epidemiology

Hortobagyi *et al* reported that the recurrence rate for breast cancer peaks in the first five years post surgery, even for those who received adjuvant treatment. This group studied 1,407 patients and described that disease-free survival continued to fall substantially in the period from 10 years to 15 years following diagnosis (Hortobagyi *et al*., 2004). One additional study reported similar findings, in that the risk of non-familial breast cancer patients developing distant metastasis was consistently much higher than the risk of patients developing loco-regional or contra-lateral breast cancer. This study was performed in a group of postmenopausal women with consecutively diagnosed ER+/unknown ER status breast cancer, who were treated with adjuvant endocrine therapy (86% of patients), in the UK with a medium follow up of 50 months (Mansell *et al*., 2006). Breast cancer metastases can occur after a prolonged period of successful treatment of the primary tumour (after completion of 5 years' adjuvant endocrine therapy) and even in node-negative patients (Goss *et al*., 2005; Chia *et al*., 2004). Weiss *et al* did a prospective multinational study of 905 post mastectomy patients and estimated that 18% of patients had recurrence 10 years or more after the initial treatment, despite the use of adjuvant chemotherapy. Some patients with positive hormone receptor status may even experience a recurrence 20 years following the initial diagnosis (Weiss *et al*., 2003). This tumour dormancy may be explained by the fact that pre-angiogen micro-metastases have an internal balance between active cell division and apoptosis that arrests the vascularisation. Moreover, persistence of solitary cells in secondary sites that fail to initiate cell division may account for long periods of an asymptomatic state (Chambers *et al*., 2002; Holmgren *et al*., 1995).

The current diagnosis criterion for invasive breast cancer is the disappearance of the myoepithelial cell layer as an organized entity. Local recurrence is when tumour cells

grow back over a period of time after the primary treatment; in the case of breast cancer regional recurrence refers to, the spread of tumour cells past the breast and the axillary lymph node and is considered more serious in terms of outcome than local recurrence (Rabbani and Mazar, 2007; Kryj *et al*., 1997). The most lethal and frequent type of recurrence is distant metastasis, which may occur soon after primary treatment and suggests the presence of subclinical deposits at the time of loco-regional treatment (Rabbani and Mazar, 2007; Rugo, 2008). Numerous studies have reported that patients with local recurrence are more likely to survive longer than patients with distant metastases. One study described that the 10 year survival rate was higher in the local recurrence patients (56% survival) than the distant metastases patients (9% survival), with the medium survival rate for local recurrence (12.9 years) higher than distant metastases (2.2 years) (Le *et al*., 2002). Another study showed that patients with local recurrence (41%) had a better five year disease free survival rate than patients with regional (20%) or distant metastases (13%; $p<0.0001$) (Elder *et al*., 2006). It is estimated that around 45,000 women succumb to breast cancer presenting with distant metastases each year in the United States, resulting in 40,000 deaths per year (Hanahan and Weinberg, 2000; Parker and Sukumar, 2003; Weigelt *et al*., 2003; Rugo, 2008). The risk of distant metastasis is known to increase in relation to increased primary tumour size and positivity in axillary lymph node metastasis. Vicini *et al* reported findings supporting this, analysing 1,169 patients with stage I and II breast cancer and found that loco-regional recurrence increased the risk for developing distant metastases (Vicini *et al*., 2003). Furthermore, clinical observations indicate that patients with residual initial local recurrence are at greater risk for developing distant metastases than patients with de novo, second primary neoplasms (Goldstein *et al*., 2005; Vicini *et al*., 2003). Studies also showed that the annual risk of developing distant metastases increases over a period of ten years in patients with loco-regional recurrences. Patients

with local recurrences within a shorter medium time; ipsilateral breast tumour recurrence or loco-regional failure after breast-conserving surgery have an increased risk of developing distant metastases (Kamby and Sengelov, 1997; Haylock *et al.*, 2000; Doyle *et al.*, 2001; Moran and Haffty, 2002; Veronesi *et al.*, 1995; Fowble, 1999; Komoike *et al.*, 2006). It was estimated that patients with local recurrence were at a >4.4-fold greater risk of relapsing at distant sites than patients without local recurrence (Cowen *et al.*, 1998). Other studies indicated that body mass, insulin status and obesity are significantly associated with loco-regional breast cancer and are directly correlated to the increased risk of distant metastases and death, and the mechanisms involved are largely unknown (Chlebowski *et al.*, 2002).

1.3.3.2 Breast Cancer Metastasis Predictors

Predictors for distant metastasis currently in clinical use include nodal status, proliferation factors, tumour size, tumour grade and hormonal status. The risk of distant metastases increases with larger primary tumour size, lymph node metastases, higher tumour grade/stage (stage T3 patients have an average time to distant metastasis 1.6 times shorter than T2 patients) and prior loco-regional recurrence (Kryj *et al.*, 1997). *H*ER*2* expression, up-regulation of *GRB7* and up-regulation of *H*ER*2* adaptor have been correlated with reduced distant recurrence-free survival. Higher expression of *Bcl-2* (B-cell leukaemia/lymphoma 2, involved in anti-apoptosis) and *TP53BP2* (tumour protein *p53*-binding protein 2) has also been linked to longer distant recurrence-free survival (Cobleigh *et al.*, 2005). Additional valuable predictive factors associated with risk of distant metastases, include age (<40 years), pre-menopausal hormone status, margin status, lymphovascular invasion, poorly differentiated DCIS and nuclear grade (Gonzalez-Angulo *et al.*, 2005; Kryj *et al.*, 1997).

It is known that the presence of metastasizing cells is not prognostic; the outcome is determined by whether the metastasized cells develop into a clinically relevant size and that the growth of metastasis depends largely on the interaction of the metastatic cells with different organ environments (Rabbani and Mazar, 2007). However, recent studies have indicated that the presence of circulating tumour cells with metastatic disease was associated with increased levels of tumour markers and disease progression (Rabbani and Mazar, 2007). One study reported that approximately 20% to 40% of patients with epithelial tumours have metastatic cells in the bone marrow, but did not have apparent lymph node metastasis or clinical symptoms of distant metastases (Elder *et al*., 2006). Breast cancer may metastasize to almost any part of the body and around 50% to 75% of patients relapse in a single organ (Weigelt *et al*., 2005; Chelouche Lev and Price, 2002). Lymph node metastasis is relatively common for breast cancer in that nearly one third of patients have lymph node metastasis (Bogenrieder and Herlyn, 2003). Common sites of breast cancer distant metastases include bone (60% of first recurrence), lung or thorax (15% of first recurrence), liver (13% of first recurrence) and the central nervous system (less than 5% of first recurrence) (Solomayer *et al*., 2000; Elder *et al*., 2006). It is important to note that the presence of tumour cells in the bone marrow, in both node negative and node positive patients is a strong predictor of distant metastases and indicative of worse prognosis (Diel *et al*., 1996; Schindlbeck *et al*., 2007; Vogl *et al*., 2006). One study reported that 10 year relapse free survival (62.7%) and overall survival (65.7%) were significantly better in patients without bone micro-metastases at presentation than in patients with bone micro-metastases (43.9% and 44.9% respectively) (Mansi *et al*., 1999). Furthermore, research has indicated that the site of distant metastases may influence patient's outcome; one study reported that the 5 year survival rate was 16% for bone metastases, 12% for lung metastases and zero for liver metastases (Elder *et al*., 2006).

A novel method of predicting breast cancer progression was developed by analysing the gene-expression patterns of primary tumours and comparing them to patient outcome and treatment. Currently a commercially marketed test kit, MammaPrint (Agendia, USA) is available and it is based on the 70-gene signature published by van't Veer *et al*. in Nature in 2002. The prognostic powers of these signatures come from analysing a group of genes; however, analysis on single genes has also been linked to predicting breast cancer outcome. For example, overexpression of transcription factor *FOXC1* in the basal-like breast cancer subtype consistently predicted poor overall survival (Ray *et al*., 2010). Reduced expression of *BLID,* a novel *Bcl-XL* binding partner, was found to be associated with poorer overall, and local relapse-free survival; while multivariate analysis indicated that it is an independent prognostic factor of distant metastasis-free survival (Broustas *et al*., 2010). Another gene found to be a potential prognostic marker is the C-terminal tensin-like gene, which is a member of the TENSIN gene family, and is involved in focal adhesion complexes and cell migration._It was found that increased expression of this gene is_associated with decreased breast cancer specific survival and metastasis-free survival (Albasri *et al*., 2010).

There are many more examples of single gene expression analyses serving as potential prognostic indicators of breast cancer progression. However, these finding suffer from the same critics as the multiple gene expression signature analyses in that they all lack significant reproducibility and consistency between studies, especially in sample collection and other experimental factors. Therefore, the utilization of these gene expression profiles as reliable prognostic factor requires additional research. Current breast cancer prognosis is still based mainly on clinical prognostic parameters such as tumour size, histology and axillary nodal metastasis, as well as ER and PR status.

1.3.3.3 Molecular and Cellular Changes of Metastatic Breast Cancer

Genetic alterations have long been observed in breast cancer metastasis. It was reported that patients with early relapse had a higher average number of chromosomal copy alterations, including gains on chromosomes 3q, 9p, 11p and 11q, as well as loss of 17p, compared to patients with distant metastasis free survival for more than ten years (Blegen *et al*., 2003). Furthermore, aberrations of chromosome 8 have been described in up to 40% of primary and metastatic breast cancers, reported by studies using conventional cytogenetics (Adeyinka *et al*., 2000; Popescu and Zimonjic, 2002). Other chromosomal regions such as 12q24 and 18p have been described as having prognostic value (Aubele *et al*. 2002). Recent studies, utilizing array-CGH, in ER-positive breast cancer tissues further support the significance of chromosomal regional change in breast cancer metastasis with loss of 11p15.5 to 11p15.4, 1p36.33, 11q13.1 and 11p11.2 described in a patient group with metastatic recurrence within five years (Han *et al*., 2006). Another study using the same method showed that the overall frequency of copy number alterations in chromosomal regions 1q, 8q24, 11q13, 12p13, 17q21 to 17q23, 16p13 and 20q13 correlated with tumour malignancy. These CGH data describe overlaping chromosomal regions, indicating that malignant breast adenocarcinomas are characterized by specific chromosomal copy number changes (Yao *et al*., 2006).

Breast cancer cells employ various motility mechanisms to facilitate invasion and metastasis. The major cellular mechanisms involved include actin cytoskeleton reorganization, focal adhesion regulations, epithelial mesenchymal transition (EMT) and cell membrane protein regulations via endocytosis. These processes are tightly regulated by hormonal factors, including growth factors, chemokines and cytokines.

Breast cancer cells have three major actin filament bundles or three-dimensional networks; lanellipodia, filopodia and invadopodia. Lamellipodia are a broad, sheet-like membrane protrusion that locates at the leading edge of cells during movements; filopodia are thin plasma membrane projections of the cells and are supported by actin bundles. Invadopodia are a three-dimensional actin filament network of cell extensions, which are responsible for cell invasion into the extracellular matrix (ECM). During cell migration, breast cancer cells attach to the ECM via binding of the integrins (transmemebrane protein) to the ECM. Focal complexes and focal adhesions are regions of attachment sites and composed mainly of stress fibres (large bundles of actin filaments) (Gatenby & Gillies, 2008).

The assembly, disassembly and cross linking of actin filaments are regulated by actin-binding proteins and their intracellular signalling mechanisms (Eccles & Welch, 2007; Gupta, 2006; Hiratsuka *et al*., 2006). In breast cancer cells,_the organisation of actin cytoskeleton, which effects cell mobility, adhesion and cytokinesis, is regulated by members of the Rho subfamily of small GTP-binding proteins (GTPases) (Hiratsuka *et al*., 2008). *Rac*, *Rho* and *cdc42* are the three well-studied cell mobility regulators from the *Ras* GTPase super family. *Rac* is essential for the regulation of focal complexe assembly and actin polymerization during lamellipodia formation, while *Rho* is involved in stress fibre formation. *Cdc42* regulates actin polymerization in filopodia formation. These three key molecules play important roles in regulating cell mobility and breast cancer metastasis (Hiratsuka *et al*., 2008; Christofori, 2006).

Breast cancer metastasis can also be achieved via epithelial mesenchymal transition, which is a series of cellular and molecular changes that induces the reduction of cell to cell adherence (Hugo *et al*., 2007). In *in vitro* studies, EMT is typically recognized with

down-regulation of epithelial markers and the acquisition of a fibroblast-like motile and invasive phenotype (Vega *et al*., 2004). The molecular hallmark of EMT is the lost of transmembrane protein E-cadherin. The extracellular region of E-cadherin is an integral part for cell to cell binding at the adherent junctions. The loss of E-cadherin allows cells to detach from their primary tissue and facilitate mobility (Hugo *et al*., 2007; Vega *et al*., 2004).

Endocytosis is another mechanism used for gaining cell mobility. It is a process by which cells uptake materials from the extracellular medium, recycle membrane components and down-regulate plasma membrane receptors once they are bound to their ligands (Jiang, Enomote & Takahashi, 2009). It was found that the distribution of integrins and cadherins are in continuous state of endo – exocytic flux and endocytosis-mediated degradation (Kim et al., 2006; Sanz-Moreno et al., 2008). Integrins provide continuous mobility drive for breast cancer cells once cell migration has commenced. Integrins facilitate this mobility via forming integrin clusters at focal adhesion sites and driving actin polymerization along the leading edge of the cell. This polymerization in turn continues the formation of integrin-based adhesion complexes and promotes cell migration (Guo & Giancotti, 2004). In addition, integrins mediate cell shape and polarity changes during cell migration by activating the *Rho* family of small GTP-binding proteins. Integrins have also been observed in mediating the direction of cell migration and break down of cell to cell adhesive structures (Bates, Edwards & Yates, 2000). These observations support that breast cancer cell mobility is also regulated by the endocytosis of these plasma membrane proteins. The research on the functions and characters of integrins in breast cancer and migration has been extensive and remains an important topic in breast cancer metastasis research (Bates, Edwards & Yates, 2000).

Once the cells have acquired mobility, invasive tumour cells can migrate either as single cells or multiple cells (in the forms of files, clusters, or sheets). For cells travel in clusters and sheets, cell to cell adhesion proteins and cadherins provide the intracellular adhesion within the groups of cells (Friedl & Wolf, 2003; Bates, Edwards & Yates, 2000). The single cells migrate in different rates and can travels at a slower rate via mesenchymal migration or at a faster rate via amoeboid travelling, which requires no proteolytic ECM remodelling (Friedl & Wolf, 2003; Lauffenburger & Horwitz, 1996).

Tumour cells in the invasion process need to overcome anoikis in order to successfully metastasize. Anoikis is a process when cell death is induced by inappropriate or loss of cell adhesion. The normal epithelial and endothelial cells actively initiate apoptotic responses during the period when cell to cell or cell to matrix interactions are lost, or when the adhesive substrates are inadequate (Frisch & Francis, 1994; Guo & Giancotti, 2004). Metastatic tumour cells therefore employ several mechanisms to suppress anoikis at different stages of metastasis, such as during invasion, migration in blood/lymph vessels and post-extravasation at distant sites (Liotta & Kohn, 2004; Geiger & Peeper, 2005; Zhu *et al*., 2001). The integrins are essential for a cell to detect its surroundings and adhere to the ECM (Guo & Giancotti, 2004). Specialized integrin complexes bind to their specific correspondent ECM molecules and induce intracellular signalling via focal adhesion kinases and SRC (v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)) family kinases, along with extensive cross communication between integrins and receptor tyrosine kinases (Hood & Cheresh, 2002). Therefore, integrins are critically involved in anoikis with integrins activation was found to protect cells against anoikis (Meredith, Fazeli & Schwartz, 1993; Frisch & Ruoslahti, 1997). Altered repertorie of integrin receptors or high levels of FAK (PTK2 protein tyrosine kinase 2) are often presented in malignant tumour cells with

stimulated proliferation, survival and migration (McLean et al., 2005; Janes & Watt, 2006; Guo & Giancotti, 2004). In addition, studies have suggested that integrin activation may induce EMT, and depletion of E-cadherin protects mammary cells against anoikis (Derksen *et al*., 2006; Onder *et al*., 2008; Gimond *et al*., 1999).

It is known that distant metastasis can't be achieved via cell invasion alone, it also requires a supporting network of blood/lymph vessels. If tumours grow beyond a critical size without sustained angiogenesis, which is the process of generating new vessels from existing ones, malignant tumours are forced into dormancy due to limited oxygen and nutrient supply from passive diffusion (Holash *et al*., 1999). Under normal condition, angiogenesis is tightly regulated by the balanced maintenance of several key angiogenesis activators, such as vascular endothelial growth factor A, fibroblast growth factors, platelet-derived growth factors, epidermal growth factor and angiogenic inhibitors; such as thrombospondin 1, angiostatin, endostatin and tumstatin (Bergers & Benjamin, 2003; Kalluri, 2003). Once a tumour activates vascularisation, pro-angiogenic events are induced,_which is achieved via several tumour cell intrinsic factors and stromal cells (Geiger and Peeper 116. 117). Several studies have showed that proteases are essential in angiogensis, such as during the release of pro-angiogenic factors from ECM and during the activation of angiogenic inhibitors (Bergers & Benjamin, 2003; Kalluri, 2003; Overall & Kleifeld, 2006).

The last stage of metastasis is the outgrowth of tumours at secondary sites. During the early stages of cancer research, two important distinctions about micrometastasis have been made. First, the metastatic pattern is not random: different types of tumours often have their preferred secondary organs. For example, breast cancer preferentially metastasizes to the bones, the liver, but not the spleen; whereas stomach cancer rarely

metastasizes to the bones (Fidler, 2003; Paget, 1889; Fuchs, 1882). A second observation was that the outgrowth from a single tumour cell to micro and macro metastases is an inefficient step of the metastatic cascade. These observations suggest that a secondary tumour is only established if the microenvironment of the secondary site is compatible with the properties of the disseminated cells (Fidler, 2003; Paget, 1889; Fuchs, 1882). It has been reported that the endothelia of vessels in different tissue express different adhesion molecules, and thus each vascular bed may have its specific molecular address. Tumour cells expressing the corresponding receptor can use this sytem for specific secondary site homing. For example, tumour surface protein metadherin adheres specifically to lung vessels (Ruoslahti & Rajotte, 2000; Brown & Ruoslahti, 2004). In addition, breast cancer cells that express the CXCR4 receptor can successfully form secondary lesions at tissues that express the ligand for CXCR4 receptor, CXCL12, which is expressed in the lung, liver, lymph node and bone marrow, which are all common sites for breast cancer metastasis. Additional research found that CXCR4-blocking antibodies hinder lung metastasis in SCID mouse xenograpft experiments (Muller *et al*., 2001). This further supports the specificity of disseminated cells to secondary site selection.

In summary, breast cancer metastasis requires a wide range of molecular and cellular alterations for the tumour cells to successfully survive and outgrow at a distant site. Further investigations into the unbalanced maintenance of key molecules in breast cancer such as integrins, ECM proteins and cell signalling proteins, may offer further understanding of the metastatic process and better management of the disease.

1.3.3.4 Metastatic Breast Cancer Molecular Signatures

Recent studies, utilizing microarray analysis to generate gene expression profiles, have been used to predict outcome in breast cancer patients (Rabbani and Mazar, 2007). An example of using molecular profiling by clustered analysis to identify risk of recurrence and applying it to the clinic in breast cancer patients is the recently commercially available diagnostic assay Oncotype DX™ (Genomic Health Inc., Redwood City, CA, USA). This assay can detect the likelihood of breast cancer recurrence in patients with newly diagnosed, node negative, ER positive and early stage breast cancer. The assay is based on a 21 gene expression status, including *H*ER*2* and oestrogen receptor, and the assay can also assess the magnitude of chemotherapy response (Rabbani and Mazar, 2007). Several additional studies reported microarray profiling in predicting distant metastases. Wang *et al*. reported a 76-gene signature profile that accurately predicted distant metastasis in lymph node negative patients, with a sensitivity of 93% and specificity of 48% (Wang *et al*., 2005). This 76-gene signature was further validated by a TRANSBIG multicentre test with independent populations and the 76-gene profile accurately predicted the development of distant metastases and overall survival in a low-risk prognosis group (Dal Lago *et al*., 2006). The breast cancer metastases microarray results influence not only the clinical practice but also the fundamental theories of cancer metastasis. Up until recently many studies theorized that the malignant cell acquires its metastatic potential at a later stage of the tumourgenesis progression (Urquidi and Goodison, 2007). However, recent microarray data suggests that the ability to metastasize may be an early and inherent genetic property of certain breast tumours (Urquidi and Goodison, 2007). Early reports by Ramaswamy *et al* described the derivation of a gene signature that distinguished primary from metastatic adenocarcinomas. An expression pattern of 17 genes that best distinguished primary and metastatic adenocarcinomas was generated from the comparison of expression

profiles of 64 primary adenocarcinomas, including breast carcinoma, to 12 unmatched metastases resulting from the same adenocarcinomas (Ramaswamy *et al*., 2003). It is important to note that the 17-gene metastases signature translated closely to protein translation. Subsequent research applying this 17-gene signature to 279 primary tumours, including 78 stage one breast carcinomas indicated that primary tumours carrying the 17-gene expression signature were more likely to be associated with metastasis and poor outcome (Ramaswamy *et al*., 2003). Since the metastasis signature can be generated from bulk primary tumour tissue, it is speculated that the metastatic potential of malignant cells may be an early or inherent event (Urquidi and Goodison, 2007). Other research supports this observation, for example, Weigelt *et al*. examined sets of matched human breast cancer tumours and described that gene expression profiles in the primary tumour are preserved in the associated distant metastases, with respect to both subtype profile and poor prognosis signature, even when metastases developed after a long interval (Weigelt *et al*., 2003, 2005). Microarray analysis has also been applied to understanding metastatic progression and site-specific metastases in primary breast cancer (Rabbani and Mazar, 2007). For example, a 31-gene signature was described to predict bone relapse with a specificity of 50%, and Minn *et al* reported a signature that predicted a propensity for lung metastases (Minn *et al*., 2005). Furthermore, the European project MetaBre described gene profiles for lung, liver, bone and brain metastases in breast cancer patients and the signatures can distinguish primary breast cancer tumours metastasizing to different organs (Rabbani and Mazar, 2007).

It is important to note that these prognostic signatures are methods used for patient survival prediction and that they are not a collective biological functional analysis. Even though some of the genes included in the signatures are known metastasis-related

genes, such as genes in apoptosis (BCL2-related protein A1), cell cycle (BTG family, member 2), cell proliferation (Tumour necrosis factor (ligand) super family, member 13) angiogenesis (Vascular endothelial growth factor A) etc, the presence of such genes in the signature doesn't explain its actual involvement in the pathology of the disease (Weigelt, Peterse & van't Veer, 2005; Jézéquel *et al*., 2009). In addition, the non-malignant cells present in the tumour bulk, which were included in the signature analyses, may contribute to the signatures depending on the relative amount of infiltrated cells in the samples. It has also been observed that not only oncogenic mutations but also the genetic background of the patient influences the signature. Therefore, these signatures are best viewed as prognostic tools and since these signatures were not generated on the basis of biological function, their gene ontology and functional relationships with regard to breast cancer development and progression remains largely unknown.

In summary, metastatic breast cancer remains the major health burden for breast cancer and more effective treatments are needed. Chromosomal changes and gene expression alterations in metastatic breast cancer have been studied, yet the mechanisms involved in the metastatic process are still unclear and more researche is needed to provide an in depth understanding of this process and better patient management.

Chapter 2

Methods and Materials

2.1 Study Design

2.1.1 Research Background

This project is designed to validate potential breast cancer metastasis modulators, which were previously identified in mouse model, in independent human breast cancer cohorts. This research utilized the genotype frequency versus patient clinical outcomes association study method, which is based on the common disease common variant hypothesis. This particular methodology was used to explore single nucleotide polymorphism (SNP) genotype frequencies of the candidate genes in patient cohorts, with corresponding clinical records, to translate the observations from the mouse model into the human population.

Research on quantitative genetic traits related to human complex disorders, such as cancer metastasis and cardiovascular disease, which have diverse disease phenotypes over a range of distribution in a population and are influenced by multiple genes and environmental factors, requires simplified and higher degrees of control over the genetic variation and environmental exposures (Peters *et al*., 2007; Hunter & Crawford, 2008). Mouse models that provide controlled genetic variations and environmental exposures therefore are suitable tools for studying quantitative genetic traits.

The candidate genes were selected based on microarray results gathered from the mouse model and literature reviews. Previous experiments from the Hunter lab showed that breast cancer metastasis potency is correlated with genetic background. These experiments utilised the AKXD recombinant inbred (RI) strains, which have previously been generated by interbreeding AKR/J (high metastatic genotype) mice with DBA/2J mice (low metastatic genotype) (Bedigian *et al*., 1983; Yang *et al*., 2007). Each of the resulting inbred strains is

composed of unique chimeric compositions. The hypothesis that germline genetic background influences cancer metastasis capacity was examined by breeding the highly metastatic Polyma middle-T (PyMT) transgenic FVB/NJ mouse (which develops mammary tumours and extensive (>90%) pulmonary metastases by 100 days) to 18 of the 25 AKXD sub-lines to generate 18 different strains. Analysing complex traits such as metastasis efficiency in this manner has several advantages. Mice within each substrain in an RI panel are inbred (ie. Not all mice in the panel are identical) and thus genetically identical (within each substrain). Once the initial genotyping has been performed, traits can be mapped cumulatively by phenotyping across the panel and comparing the trait segregation to the known genetic map. Furthermore, since the genotypes across the panel can be assayed multiple times, the background noise in assays determining traits, such as quantitative trait locus (QTL) and expression QTL (eQTL), can be minimized (Park *et al*., 2003; Hunter & Crawford, 2006; Hunter, 2006; Hsieh *et al*., 2006). The F1 progenies exhibited various propensities for developing lung metastasis, with some F1 strains developing about 3 times more metastasis compared to the original FVB/NJ inbred strain, whilst other F1 strains had 10 times less metastasis (Figure 2.1, overleaf). Since all tumours that developed were induced by the PyMT antigen, the difference in metastasis efficiency is best explained by germline variation between each inbred strain. It is important to note that the strains listed in Figure 2.1 were not from the AKXD panel. Therefore the total numbers of strains analysed didn't match the 18 strains mentioned in the earlier text. The strains analysed in Figure 2.1 were well studied inbred strains that were randomly selected to include in the study and additional strains were included in the analyses in order to provide extra support for the findings.

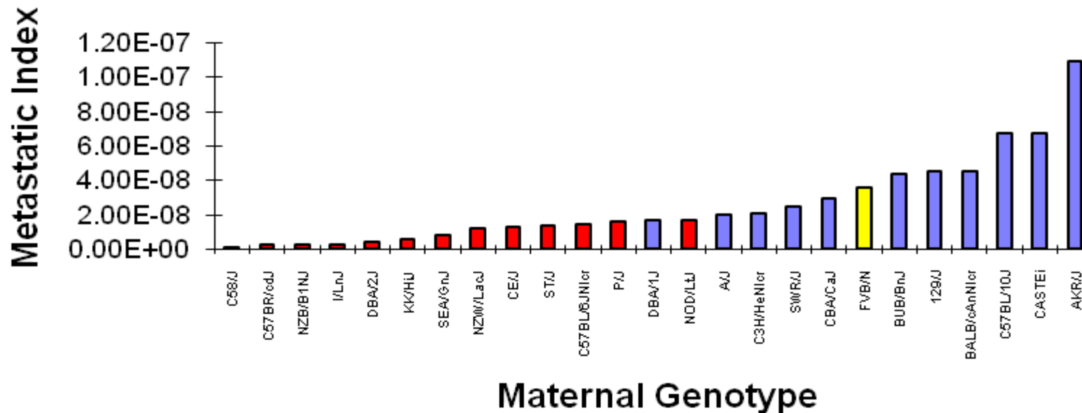


Figure 2.1: Effect of genotype on mammary tumour metastatic efficiency. The maternal inbred-strain of the F1 outcrosses is indicated on the x-axis. The yellow bar depicts the original homozygous FVB/NJ genotype. The y-axis represents the metastatic efficiency, as measured by the number of metastases per square micron of lung tissue. Red bars indicate strains that had statistically significantly different metastatic efficiencies compared to the FVB/NJ homozygous background (Hunter, 2003). The FVB/N-TgN(MMTV-PyMT)^{634Mul} mice are transgenic mice carrying the (MMTV-PyVT) transgene with albino appearance. This strain of mice is homozygous for the retinal degeneration allele $Pde6b^{rd1}$. The female mice can develop palpable mammary tumours as early as 5 weeks of age and pulmonary metastases are observed in 80-94% of tumour-bearing female mice. These mice have an intact immune system. Further information can be obtained from the Jackson Laboratory website http://jaxmice.jax.org/strain/002374.html.

To identify potential metastasis modulators, primary mammary tumours from the F1 progenies of the PyMT x AKXD crosses were harvested and microarray analysis on these samples was subsequently performed. The gene expression data was normalized and averaged for each RI substrain and loaded into the publicly accessible internet based QTL analysis package, WebQTL. Correlation analysis was performed, using the Pearson correlation option, to identify transcripts with high correlation to the metastatic efficiency of each AKXD RI subline. Suggestive peaks with subsequent analyses revealed important loci

on chromosomes 19 (*SIPA1* locus; *Ttc9c* locus) and 13 (*Arid4b* locus) (Figure 2.2), with the genes located on the regions earmarked for functional analysis (unpublished data) (Yang *et al*., 2005).

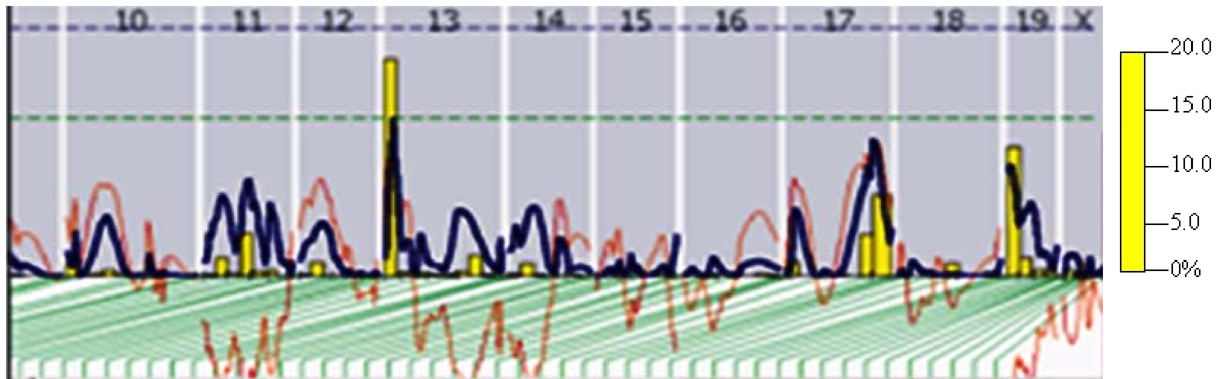


Figure 2.2: QTL mapping of metastasis efficiency on the updated AKXD RI panel. The chromosomes are depicted lying head-to tail along the X-axis. The boundaries of each chromosome are represented by the gray boxes. The blue line depicts the LRS (likelihood ratio statistics) score of the statistical likelihood of the presence of a QTL at any given point in the genome. The *p*-value is 0.05. The yellow bar indicates the most likely location of a QTL under each peak. The red line represents the additive effect of the QTL. This figure illustrates the significant mapping result of chromosomes 13 and 19.

In addition to the direct correlation study using the AKXD RI panel, an additional experiment was conducted utilizing both the mouse microarray data obtained previously and human breast cancer gene expression data. First, based on previous microarray analysis on the normal tissues (non-malignant breast tissue from the same mice) from the high (AKR/J) and low (DBA/2J) metastatic genotype strains obtained from the PyMT F1 progeny, it was found that extracellular matrix (ECM) genes were differentially expressed. This indicates a correlation between genetic background and ECM expression in metastatic cancer cells. Secondly, it was reported that ECM gene expression levels have been linked to metastasis predictive expression profiles in human breast tumour tissue in several studies (Bergamaschi *et al*., 2008; Sorlie *et al*., 2006; Grigoriadis *et al*., 2006). For example, Bergamaschi *et al.* found that over expression of certain set of serpin family protease inhibitors is correlated with

favourable clinical outcome, while over expression of integrins and metallopepdases (MMPs) in conjunction with low expression of several laminin chains had a poorer prognosis (Bergamaschi *et al*., 2008). Furthermore, recent evidence demonstrated that wound healing signatures (partially composed of ECM and collagen genes) increases the prognostic power of certain metastasis predictive gene expression profile (Chang *et al*., 2005). Taken together, the differential expression of ECM genes is considered to be either a marker for metastatic risk or a causative factor (Hunter & Crawford, 2006; Hunter, 2003; 2005; 2006).

The ECM is composed of a complex mixture of proteins, including fibronectin proteoglycans and adhesive glycoproteins such as collegens, laminins and other macromlecules (Bergamaschi *et al*., 2008; Iseri *et al*., 2009). The ECM provides mechanical support to cells and tissues and is essential for cell division, survival, morphology and mobility regulations. Since the microenvironment is crucial in tumour cell survival, migration and invasion, it was proposed that tumour cells remodel its microenvironment via reciprocal interactions with the ECM and create a microenvironment favourable for proliferation and migration (Bergamaschi *et al*., 2008; Iseri *et al*., 2009). Histology observations support the proposal that the stroma of breast cancer is known to be altered compared to the non-malignant breast tissues. Different breast lesions also posses different ECM histology, for example, certain lesions have loose stroma, with high hyaluronic acid content, where as other lesions may have dense stroma, which is rich in collagen and laminin, resembling the mammary gland stroma (Tamimi & Ahmed, 1987).

It was proposed that differential expression of the ECM genes and MMP genes may contribute to the ECM remodelling and therefore tumour invasion and metastasis. Initial reports suggested that tumours with invasive characteristics had increased levels of ECM

components, such as fibronectin, laminin, collagen type IV and increased MMP-2 activity (Mitsumoto *et al.*, 1998). Alterations in ECM components effect not only the invasiveness but also the survival of tumour cells; it was reported that the sensitivity of tumour cells to growth factor induced apoptosis is dependent on cell to matrix and cell to cell interactions via ECM to integrin signalling (Morin, 2003).

Based on the ECM literature review, the Hunter Lab hypothesized that the germline polymorphisms modulating ECM expressing are the same ones that modulate metastasis efficiency. The hypothesis was addressed by performing eQTL analysis on the same microarray data obtained earlier (the PyMT x AKXD F1 progenies array data since it gives low background noise in QTL or eQTL analysis). In this analysis, the program treated the ECM gene expression patterns as complex genetic traits and subsequently generated peaks portraying transcripts and loci that are involved in modulating metastasis-predictive ECM expression. The three loci that exhibited the highest correlation with metastasis efficiency were on chromosomes 7, 17 and 18 (Figure 2.3, overleaf). Since mouse chromosomes 7, 17 and 18 gave reproducible ECM eQTLs, correlation analysis was performed to identify potential candidate genes that modulate ECM gene expression. Utilizing the Trait Correlation function of the WebQTL (a publicly available internet-based analytical program), genes that displayed a high degree of correlation with a low p value were selected for further analysis (Lancaster *et al.*, 2005). In depth analysis of these loci identified six candidate genes (in human homolog); Ribosomal RNA processing 1 homolog B (*RRP1B*), Centaurin, delta 3 (*ARAP3*), Colony stimulating factor 1 receptor, formerly McDonough feline sarcoma viral (*v-fms*) oncogene homolog (*CSF1R*), Peptidase inhibitor 16 (*PI16*), LUC7-like (*LUC7L)* and Bromodomain containing 4 (*BRD4*), that were earmarked for further molecular functional analysis (Hunter, 2005). Other genes within these loci with lower correlation, yet with higher

p-values were: Enhancer of zeste homolog 2 (*EZH2*), Stabilin 1 (*STAB1*), Myosin VIIA (*MYO7A*), lymphatic vessel endothelial hyaluronan receptor 1 (*LYVE1*) and Mitogen-activated protein kinase 14 (*MAPK14*). These five genes were selected to be examined by genotyping only; with SNP analysis performed using two independent pilot cohorts to determine possible involvement of these genes in human breast cancer and survival. Only if significant associations were observed in relation to human breast carcinogenesis would these five particular transcripts be marked for further functional analysis.

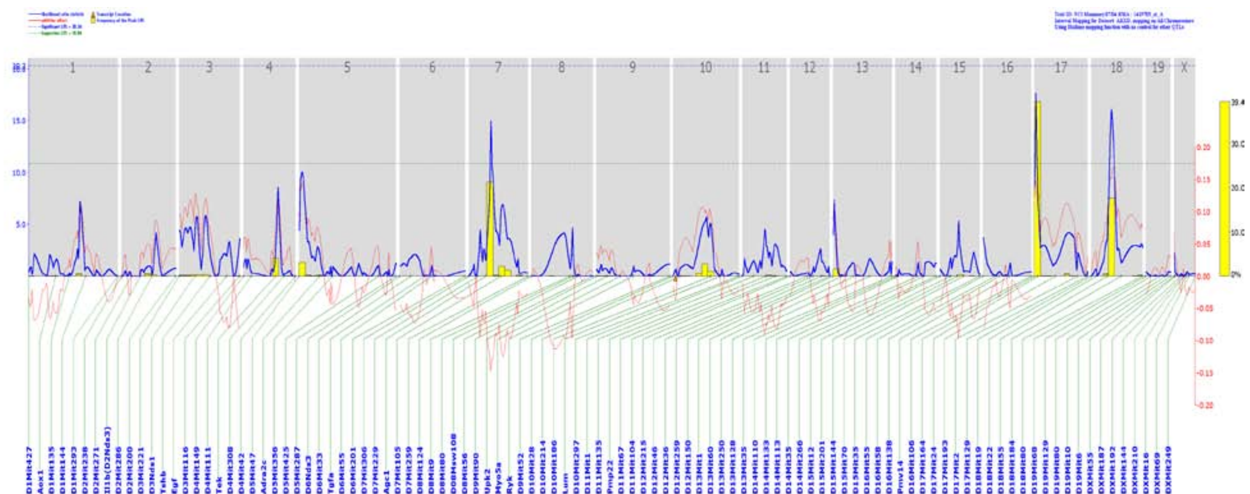


Figure 2.3: eQTL mapping of the Col1a2 locus. The peaks observed in chromosome 7, 17 and 18 were reproducibly observed in many of the ECM genes examined. The blue line depicts the LRS score of the statistical likelihood of the presence of an eQTL at any given point in the genome. The *p*-value is 0.05. The yellow bar indicates the most likely location of an eQTL under each peak (Yang *et al*., 2005).

Two additional genes; calmodulin 2 (*CALM2*) and aquaporin 2 (*AQP2*), were selected for genotyping in the same pilot cohorts due to the fact that these two transcripts directly interact with the *SIPA1* protein, which was the first validated and characterized metastasis efficiency modulator gene identified by the Hunter lab (Park *et al*., 2005).

2.1.2 Hypothesis

The hypothesis of this research is that putative breast cancer metastasis modifier genes identified in the mouse model based on literature reviews and microarrays data from differential metastasis potency mice, are breast cancer prognosis and survival indicators in human breast cancer.

2.1.3 Methodology Background

Ample evidence indicates that germline genetic variations correlate to disease development in common complex trait diseases such as cancer. For example, a SNP (SNP309) of the Murine Double Minute p53 Binding Protein homolog (mouse) oncogene (*MDM2*) was found to be associated with poor prognostic breast cancer features such as high grade tumours and greater lymph node involvement in Scottish cohort (Paulin *et al.*, 2008). In addition, a promoter region SNP (rs16949649) from the non-metastatic cell 1 (*NME1*) was found to be associated with higher breast cancer-specific mortality and the minor allele of SNP rs2302254 (from the same promoter region) altered NME1's nuclear proteins binding capacity resulting in reduced promoter activity (Qu *et al.*, 2008).

It has been established that carcinogenesis is influenced by multiple genes, with each gene and the inherited genetic variations in them having subtle effects on cancer development (Bernig & Chanock, 2006). Among all the different types of germline genetic variations in the human genome, SNPs are the most frequent variation. Single nucleotide polymorphisms are a single-base, stable substitution, having a minor allele frequency equal to or greater than 1% in at least one studied population (Bernig & Chanock, 2006; Savage & Chanock, 2004).

The particular SNP's in each of the candidate genes that were to be assayed were chosen based on the common disease common variant hypothesis. Traditional linkage analyses are effective in researching traits that display a simple Mendelian inheritance pattern, but not complex traits, such as sporadic cancer, diabetes, heart diseases and mental illness. Complex traits are most likely shaped by multiple genes that exert weak allelic effects, with differential distributions of allele frequencies and effect sizes across different populations. The common disease common variant hypothesis therefore hypothesizes that common traits are most likely due to common variants, with small to modest effects on complex traits that may have escaped stringent selection pressure. Hence, association studies, which this research employed, may have greater power in detecting common variants with small effects and subsequently give insights into the complex trait in question: breast cancer metastasis (Chen *et al.*, 2007; Cargill & Daley, 2000).

There is mounting evidence that suggests common genetic polymorphisms play a role in complex diseases. The majority of SNPs have no measurable effect on protein function and are known as synonymous SNPs. Synonymous SNPs are "silent" as they do not alter the amino acid residue sequence of a protein, yet these SNPs can still have important impact on the stability, splicing variation and expression of the gene (Chanock, 2001; Savage & Chanock, 2004; Conne *et al.*, 2000; Duan *et al.*, 2003; Hoogendoorn *et al.*, 2003; Pagani & Baralle, 2004). For example, a recent study reported that a synonymous SNP from the Multidrug Resistance 1 (*MDR1*) gene, which was not expected to change the function of the protein, gave rise to a gene product (P-pg protein) with altered drug and inhibitor interactions. Similar mRNA and protein levels of the gene were observed in this study yet altered conformations of the polymorphic P-pg were found. The authors hypothesize that the presence of the rare codon, which is marked by the synonymous polymorphism, may affect

the timing of cotranslational folding and insertion of P-pg into the membrane, hence altering the structure of substrate and inhibitor interaction sites (Kimchi-Sarfaty *et al*., 2007). Furthermore, SNPs in the promoter region have also been linked to altered gene expression, as illustrated in the promoter SNP rs16949646 from the *NME1* promoter region, which altered this transcripts nuclear protein binding capacity, as well as reducing its promoter activity (Qu *et al*., 2008). SNPs in other regions of genes, such as the $3^{'}$ untranslated region ($3^{'}$ UTR), have also been reported in regulating gene expression. SNP rs15705 in the $3^{'}$ UTR of the bone morphogenetic protein (*BMP*) 2 gene, which is genetically associated with osteoporosis and osteoarthritis, was found to disrupt a putative posttranscriptional regulatory motif an AU-rich element within the *BMP2* $3^{'}$ UTR. The affinity of specific proteins for this SNP differs between the wild-type and variant allele, with the effect of subsequently altering the *in vitro* decay rate of the RNA. Such change in protein:mRNA interactions was suggested to influence the posttranscriptional regulation of *BMP2* gene expression and influence the development or physiology of bone (Fritz *et al*., 2006).

Nonsynonymous SNPs, which change an amino acid residue, are located in the coding regions of a gene and have been the focus for the large majority of genotyping-based studies (Savage & Chanock, 2004; Chanock, 2001). For example, the two common non-synonymous SNPs ($Ile^{105}Val$ and $Ala^{114}Val$) alterations in the Glutathion S-transferase P1 (*GSTP1*) gene, which is heavily involved in chemotherapy drug metabolizing and cell cycle regulation, are known to be associated with variations in cancer risk and clinical response to anti-neoplastic drug therapy. The Val^{105} substitution for instance, was found to result in steric restriction of the H-site due to shifts in the side chain of several amino acids. This shift in side-chain arrangement affects the *GSTP1* protein to accommodate less bulky substrates than the Ile^{105}

allele and displays substrate specificities that differ from the wild-type allozyme (Moyer *et al.*, 2008).

It has been observed that SNPs in the human genome are not inherited randomly and the density of SNPs varies between chromosomes, as well as between regions on chromosomes (Taillon-Miller *et al.*, 2000; Eaves *et al.*, 2000). This observation may be due to several factors, such as different evolutionary pressure on biologically important functions and different rates of mutation and recombination (Chanock, 2001; Erichsen & Chanock, 2004; Bernig & Chanock, 2006). Since SNPs are inherited as a combination of alleles in the same chromosome, they form haplotype blocks. Haplotype blocks can be defined by the presence of strong linkage disequilibrium (LD), which is a trait characterized by the non-random association of alleles in the same chromosome (Bernig & Chanock, 2006; Daly *et al*, 2001; Gabriel *et al*, 2002; Stephens *et al*, 2001). Since SNPs are inherited in LD with limited haplotype diversity, it is possible to select subsets of SNPs in a haplotype block as markers, to represent the majority of common diversity of that haplotype block. Haplotype-tagging SNPs (ht-SNPs) are widely employed to investigate LD and haplotype blocks of a gene or locus (Low *et al.*, 2006; Johnson *et al*., 2001; Carlson *et al*., 2004; Stram *et al*., 2003). Utilizing ht-SNPs to investigate germline polymorphisms can be applied to one or multiple genes of a biological pathway and therefore association can be derived with complex disease such as cancer and its clinical outcome (Bernig & Chanock, 2006). It is important to acknowledge the diversity of LD and haplotype block patterns between populations while choosing ht-SNPs. The diversity of LD patterns across different populations and the existence of population private SNPs may be due to many factors, such as genetic drift (the significant change of allele frequencies in a population due to various reasons), population growth, bottlenecks (a significant percentage of a population prevented from breeding, resulting in a

serious decrease in population size), admixture (inter-racial breeding) and natural selection (Hawks *et al.*, 2000; Ardlie *et al.*, 2002; Carlson *et al.*, 2003). Although different populations share some LD patterns, certain ht-SNPs may not represent the polymorphism at the same genetic region in different populations (Pritchard & Przeworski, 2001).

2.1.4 Genes and SNPs of Interests for Statistical Analyses

The description of the molecular function of the genes of interests will be brief. Detailed descriptions of their molecular characteristics and their possible involvement in cancer and cancer metastasis will be given in the results chapters for those genes that showed significant associations with the multiple human cohorts tested.

Signal-induced proliferation-associated 1 (SIPA1; Chromosome locus 11q13*)*

Two alternatively spliced variants of this gene encoding the same isoform have been described (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=6494&ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008). *SIPA1* is a mitogen-inducible GTPase activating (GAP) protein and is specific for members of the Ras-related proteins: Rap1 and Rap2 (Kurachi *et al.*, 1997). *In vitro* studies revealed that abnormally and/or prematurely expressing the *SIPA1* protein severely impedes the mitogen-induced cell cycle progression (Hattori *et al.*, 1995). *SIPA1* was also found to play a role in cell adhesion via its Rap1GTP domain; with *SIPA1* suppression (via RNAi) inducing increased cell adhesion and over-expression reducing cell attachment (Park *et al.*, 2005). A pilot epidemiology study indicated that three SNPs

(rs931127, rs3741378 and rs746429) in this gene are associated with breast cancer survival in a non-Hispanic Caucasian population from Southern California (Crawford *et al*., 2006).

Tetratricopeptide repeat domain 9C (TTC9C; Chromosome locus 11q12.3*)*

The full length cDNA clone of this gene has only been recently established and no known function has been reported. Its only known domain is a tetratricopeptide repeat domain (TPR), which has been identified in various other proteins functioning as a protein-protein interactor (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=283237&ordinalpos=3&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008).

AT rich interactive domain 4B (RBP1-like) (ARID4B; chromosome locus 1q42.1-1q43*)*

Alternate transcriptional splice variants encoding different isoforms of this gene have been characterized(http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=51742&ordinalpos=2&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008). This gene is also known as breast cancer associated antigen 1 gene (BRCAA1) and encodes a 1,214 amino acid (93.857 kb) protein composed of 18 exons and 17 introns. The *ARID4B* protein is known to have 10 glycosylate sites and two known antigen epitopes (KASIFLK and IKPSLGSKK). The *ARID4B* protein shares the same antigen epitope, IKPSLGSKK, with the retinoblastoma-binding (RB) binding protein 1 and is speculated to play a role in cell proliferation and regulation. A novel *ARID4B* antigen epitope, SSKKQKRSHK, was found in the cytoplasm of breast cells and the expression of this particular antigen epitope is closely associated with ER and PR status in the breast cancer tissues examined (Cui *et al*., 2004).

ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 (ARAP3; chromosome locus 5q31.3*)*

An alternatively spliced transcript has been found for this gene, but its biological validity has not been determined as reported by molecular research. (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=64411&ordinalpos=4&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008). This gene is a phosphoinositide binding protein that contains ARF-GAP, RHO-GAP, RAS-associating, and pleckstrin homology domains. This GTPase activating protein is reported to be part of the PtdIns(3,4,5)P3/PtdIns(3,4)P2 (PI 3-kinase) signalling pathway and mediates rearrangements in the cell cytoskeleton and cell shape. *ARAP3* is phosphorylated by tyrosine kinase, which modulates cell adhesion to the extracellular matrix (ECM) substrate fibronectin (Krugmann *et al*., 2004; I *et al*., 2004). It was reported that up-regulating of *ARAP3* expression in epithelial cells increases cell rounding and adhesion, membrane process formation and cell clustering on ECM substrates, yet inhibited cell spreading. Based on these results *ARAP3* is considered to play a role in the integrin-mediated tyrosine kinase signalling pathway, which controls Rho GTPases and cell spreading (Krugmann *et al*., 2004; I *et al*., 2004). A recent study postulated a role for *ARAP3* in human breast cancer progression by identifying a signature derived from increased *ARAP3* expression in the mouse *in vitro* model system and applying it to publicly available human breast cancer datasets. Clustering the patient samples into groups that appeared to have up-regulated *ARAP3* versus those that did not, revealed a survival benefit in the former class (Crawford *et al*., 2008). This study also indicated *ARAP3* as part of the diasporin pathway which is a tumour progression related transcriptional network that predicts breast cancer survival (Crawford *et al*., 2008).

Bromodomain containing 4 (BRD4; chromosome locus 19p13.1*)*

Two alternatively spliced transcript variants have been reported by molecular studies (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=23476&ordinalpos=2&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008). The *BRD4* protein, from the bromodomains and extraterminal (BET) family, is involved in multiple cell activities, including transcription, replication (*BRD4* protein is homologous to the murine mitotic chromosome-associated protein (*MCAP*), which is associated with chromosomes during mitosis), signal transduction pathways and cell cycle progression. The *BRD4* protein, with two bromodomains, conducts its functions mostly via protein association with acetylated chromatin, with the tandem bromodomains binding to diacetylated H4-AcK5/K12 and H3-AcK9/K14 peptides to induce transcriptional activation (Liu *et al*., 2008).

Peptidase inhibitor 16 (PI16; chromosome locus 6p21.2*)*

The *PI16* protein has been identified in several organs, such as the prostate, testis, ovary and intestine. The study that identified these localisations also found that *PI16* protein's levels were significantly lower (P=0.0014) in the serum of a prostate cancer patient population (n=65), compared with a control population (n=70) (Reeves *et al.*, 2005).

LUC7-like (LUC7L; chromosome locus 16p13.3*)*

The *LUC7L* gene is widely transcribed and may represent a mammalian heterochromatic gene. A study revealed that *LUC7L* encodes a putative RNA-binding protein similar to the yeast *Luc7p* subunit of the U1 snRNP splicing complex, which is normally required for 5-prime splice site selection (Tufarelli *et al*., 2001).

Mitogen-activated protein kinase 14 (MAPK14; chromosome 6p21.3-p21.2*)*

There are four alternatively spliced transcript variants of *MAPK14* encoding distinct isoforms that have been observed in molecular studies (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=1432&ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008). The *MAPK14* protein belongs to the *MAP* kinase family, which functions as an integration point for various cellular signal pathways and plays integral role in a wide range of cellular processes, including proliferation, differentiation, transcription regulation and development. *MAPK14* is reported to be activated by various environmental stresses and proinflammatory cytokines via phosphorylation by *MAP* kinase kinases (*MKK*s), or autophosphorylated by interacting with the *MAP3K7IP1/TAB1* protein (LaMarca *et al*., 2008). The known substrates of *MAPK14* include the transcription regulators *ATF2*, *MEF2C*, and *MAX*; the cell cycle regulator *CDC25B* and tumour suppressor *p53.* These interactions suggest that *MAPK14* plays a part in stress related transcription responses, cell cycle regulation, and genotoxic stress response (LaMarca *et al*., 2008).

Colony stimulating factor 1 receptor, formerly McDonough feline sarcoma viral (v-fms) oncogene homolog (CSF1R; chromosome 5q33-q35*)*

The *CSF1R* protein is a trans-membrane receptor that is specific for cytokine colony stimulating factor 1 (CSF1, also known as macrophage colony-stimulating factor (*M-CSF*)). It is a multifunctional growth factor that regulates monocytic lineage/macrophage cell proliferation, differentiation and survival. The *CSF1R* gene is the receptor for *CSF1* and the *CSF1R* gene product forms homodimeric tyrosin kinase to mediate actions of *CSF1*. *CSF1R* belongs to the growth factor family subclass III, which are activated via oligomerization and transphosphorylation and are regulated in monocytic cells by cytokines, such as interleukin 4

(Zhu *et al*., 2002). Recent studies indicated that *CSF1R* expression is a significant prognostic factor for predicting metastasis in prostate cancer (Richardsen *et al*., 2008).

Stabilin 1 (STAB1; chromosome 3p21.1*)*

STAB1 encodes a large, homologous transmembrane receptor protein composed of the following domains: seven fasciclin, sixteen epidermal growth factor (EGF)-like, two laminin-type EGF-like domains, and one C-type lectin-like hyaluronan-binding Link module. The STAB1 protein is reported to be involved in cell-cell and cell-matrix interactions in vascular, inflammatory and receptor scavenging process and is primarily expressed in the sinusoidal endothelial cells of the liver, spleen and lymph node (Politz *et al*., 2002). STAB1 is also involved in trafficking between early/sorting endosomes and the trans Golgi network in human macrophages, in addition to its endocytosis/recycling functions (Kzhyshkowska *et al*., 2007).

Enhancer of zeste homolog 2 (EZH2; chromosome 7q35-q36*)*

EZH2 has been characterized as part of the Polycomb-group (PcG) family, which form multimeric protein complexes that regulate long-term gene silencing in multiple organisms, and subsequently maintain the transcriptional repressive state of genes over successive cell generations. EZH2 is especially important among the PcG proteins as it has histone methyltransferase (HMTase) activity that catalyzes the trimethylation of histone H3 at Lys 27 (H3-K27) (Tiwari *et al*., 2008). As part of the PcG complex, EZH2 is involved in embryonic ectoderm development and self-renewal and pluripotency of embryonic stem cells. EZH2 is known to be associated with the VAV1 oncoprotein and the X-linked nuclear protein, with differential expression of EZH2 reported in various types of carcinomas, including hepatocellular carcinoma, prostate, colon, gastric, breast, bladder carcinomas and uveal

melanoma (Sasaki *et al*., 2008; Hoffmann *et al*., 2007; Fujii *et al*., 2008; Hinz *et al*., 2008; Holling *et al*., 2007; Raaphorst *et al*., 2003).

Myosin VIIA (MYO7A; chromosome locus 11q13.5*)*

The *MYO7A* gene was first identified as playing an integral role in Usher syndrome. More than 90 different *MYO7A* gene mutations have been observed in the Usher syndrome, which is characteriszed by deafness, reduced vestibular function and, in humans, retinal degeneration. The MYO7A protein is a mechanochemical protein (actin-based molecular motor) and a member of the myosin gene family. The MYO7A protein comprises a motor domain, an actin-binding domain, a neck domain that interacts with other proteins, and a tail domain that serves as an anchor, the tail domain is unconventional very short for MYO7A. Multiple transcript variants derived from alternative splicing have been reported in the literature (Reiners *et al*., 2006).

Lymphatic vessel endothelial hyaluronan receptor 1 (LYVE1; chromosome locus 11p15*)*

The LYVE1 protein primary function is thought to be involved in regulation in cell trafficking within lymphatic vessels and lymph nodes. The LYVE1 protein, which comprises a 322-residue (60-70 kDa), integral membrane glycoprotein with a single Link module located at the N terminus of the extracellular domain. The extracellular domain is continued with a membrane-proximal domain, which is modified by O-glycosylation. The LYVE1 protein is a receptor for the extracellular matrix mucopolysaccharide hyaluronan and has been an important marker for studies on embryonic and tumour-induced lymphagiogenesis. It is also used for lymphatic endothelial cells detection and isolation. LYVE1 protein is a type I integral membrane glycoprotein, which belongs to the Link protein family, with the

hyaluronan receptor directly involved in leukocyte migration and tumour metastasis (Jackson, 2004).

Ribosomal RNA processing 1 homolog B (RRP1B; chromosome locus 21q22.3*)*

A recent report suggested the *RRP1B* gene and its gene product as a novel breast cancer metastasis modulator. Functional assays indicated that *RRP1B* protein is a binding partner of a previously identified breast cancer metastasis gene *SIPA1*. *In vitro* expression of *RRP1B* was found to alter extracellular matrix expression, which is a previously known marker for breast cancer metastasis. Furthermore, *in vitro* expression of *RRP1B* was found to reduce primary breast tumour growth and lung metastasis capacity in the mouse model. This study also reported a gene expression signature, which is generated by indicative ectopic expression of *RRP1B*, predicts breast cancer patient survival (Crawford *et al*., 2007).

Aquaporin 2 (collecting duct) (AQP2; chromosome locus 12q12-q13*)*

The AQP2 protein is a water channel protein located in the subapical storage vesicles in the kidney collecting tubule. The AQP2 functions to transport water across the apical membrane of the collecting duct, with the trafficking strictly governed by vasopressin (Noda and Sasaki, 2008; Balasubramanian *et al*., 2008). The AQP2 gene belongs to the MIP/aquaporin family and mutations in this gene have been linked to autosomal dominant and recessive forms of nephrogenic diabetes insipidus (Marr *et al*., 2002).

Calmodulin 2 (phosphorylase kinase, delta) (CALM2; chromosome locus 2p21*)*

CALM2 belongs to the Calmodulin family of major intracellular Ca^{2+}-binding proteins. Calmodulins are found mainly in the central nervous system and mediate many physiological

functions in response to changes in the intracellular Ca^{2+} concentration (Shirasaki *et al*., 2006). A recent study reported that p53 mutant 175H can stimulate the gene expression of calmodulin 2 (*CaM 2*) specifically, but not *CaM 1* or *3* (Knaup and Roemer, 2004).

In summary, the following genes and their SNPs were chosen for epidemiological analyses for associations with breast cancer progression, as summarized in Table 2.1. *SIPA1* was chosen because in vivo analyses of this gene showed association with breast cancer metastasis. *ARAP3*, *PI16*, *BRD4*, *LUC7L*, *CSF1R*, *EZH2*, *STAB1*, *MYO7A*, *LYVE1* and *MAPK14* were chosen because the eQTL analysis showed that these genes were located in loci that were involved in modulating metastasis-predicting ECM expression. *TTC9C* and *ARID4B* were chosen because the expression of these two genes were strongly associated with breast cancer metastasis potency in the microarray analyses. *CALM2*, *RRP1B* and *AQP2* were chosen because they are the protein binding partners of the SIPA1 protein.

Table 2.1: SNPs and genes screened in the pilot cohorts.

Gene	Gene Selection Rational	SNP
SIPA1	In vivo analyses of this gene showed association with breast cancer metastasis.	rs931127[e], rs3741378[e], rs746429[e], rs2448490[e]
RRP1B	Binding partner of SIPA1 protein.	rs9306160[e]
ARAP3	Correlated with metastasis-predicting ECM expression.	rs440279[e], rs1031904[b], rs3763120[e], rs6891143[e], rs2306339[e], rs4912610[e], rs6895094[e], rs17208495[e]
PI16	Correlated with metastasis-predicting ECM expression.	rs707542[e], rs707998[e], rs708006[a], rs734905[e], rs1405069[b], rs4147297[a], rs6901560[e], rs7755143[d]
ARID4B	Correlated with metastasis potency.	rs946027[e], rs4660114[d], rs10925436[e], rs12117563[e]
TTC9C[f]	Correlated with metastasis potency.	rs584845[c], rs6675[c]
BRD4	Correlated with metastasis-predicting ECM expression.	rs4808272[e], rs4809130[e], rs8104223[e], rs11880801[e]
LUC7L	Correlated with metastasis-predicting ECM expression.	rs1203981[e], rs1211375[e], rs3918352[e], rs8044711[e], rs11642609[e]
CSF1R[f]	Correlated with metastasis-predicting ECM expression.	rs10079250[b]
EZH2[f]	Correlated with metastasis-predicting ECM expression.	rs230242[d]
STAB1[f]	Correlated with metastasis-predicting ECM expression.	rs4434138[b], rs12636502[b]
MYO7A[f]	Correlated with metastasis-predicting ECM expression.	rs1052030[b], rs2276288[b]
LYVE1[f]	Correlated with metastasis-predicting ECM expression.	rs16907980[b]
MAPK14[f]	Correlated with metastasis-predicting ECM expression.	rs1050579[e]
CALM2[f]	Binding partner of SIPA1 protein.	rs1027478[e], rs1723493[a]
AQP2[f]	Binding partner of SIPA1 protein.	rs467323[e], rs2878771[e], rs3741559[e], rs3759129[e]

Table 2.1: SNPs selected for screening in the two pilot cohorts.

[a]3' Near gene SNP.--[b]non-synonymous coding SNP.--[c]Synonymous coding SNP.--[d]intergene SNP.--[e]intronic SNP-- [f]SNPs from the *TTC9C* gene and all the genes listed after the *LUC7L* genes were not genotyped on the Southern California, USA cohort.

2.2 Equipment and Reagents

2.2.1 Equipment

The following specific equipment was used for research described in this thesis:

OSP-105 Power box (Owl Scientific Plastics, Inc., USA); Western Transfer Units (Bio-Rad Laboratories, USA); Novex Mini-Cell XCell SureLock Electrophoresis Cell (Invitrogen, USA); Kodak X-Omatic Cassette (Kodak, USA); X-OMAT 2000A Processor (Kodak, USA); Platform Vari Mix Shaker (Thermolyne, USA); Isotemp 210 water bath (Fisher Scientific, USA); Hera Cell Incubator (Kendro laboratory Products, USA); Micro One Bench Top Centrifuge (Tomy, USA); Eppendorf Centrifuge 5417R (Eppendorf, USA); Savant DNA120 Centrifuge (Thermo Electron Corporation, USA); Savall Legend RT Centrifuge (Savall, USA); Hydra96 (Robbins Scientific, USA); OHAUS GT2100 Balance (Ohaus, USA); Stirrer/Hot Plate (Corning Inc., USA); Pipetman Pipettes, 2 µl, 10 µl, 20 µl, 10-100 µl, 200 µl, 1000 µl (Rainin Gilson, USA); Multi-Channel Automatic Pipettes, 12.5 µl, 250 µl, 1250 µl (Matrix, USA); Hood Mate 300 Automatic Pipette (Drumnond, USA); Daigger Vortex Genie 2 (A Daigger & Co., Inc., USA); NALGENE Cryo 1 ºC Freezing Container (Nalgene, USA); Mastercycler ep384, Eppendorf (Eppendorf, USA); 7900HT Fast Real-Time PCR System (Applied Biosystems, USA); GeneAmp PCR System 9700 (Applied Biosystems, USA); Desiccater, 5317-0120 (Nalgene, USA); Spectrometer ND-1000 (NanoDrop, USA); ELWD 0.3/OD75 Microscope (Nikon, Japan); Cellometer Auto T4 (Noxcelom Bioscience, USA); OPTImax Tunable Microplate Reader (MTX Lab Systems Inc., USA); Dell Optiplex Desktop Computer (Dell, USA); Surgical Scissors E25-500 (Khosla, Germany); Surgical Tweezers 6-107 (Miltex, Germany).

2.2.2 Reagents

The following specific reagents were used for research described in this thesis:

Dµlbecco's Modified Eagle Medium High Glucose 1x (DMEM) (Gibco/ Invitrogen, USA); BenchMark Fetal Bovine Serum (FBS) (Gemini Bio-Products, USA); Phosphate Buffer Saline 1x (PBS) (Gibco/ Invitrogen, USA); Trypsin 0.05%-EDTA 0.1% (Quality Biological Inc., USA); Puromycine Ready Made Solution (Sigma, USA); Penicillin/ Streptomycin (Cellgro, USA); Cryoprotective Medium (Lonza, MD, USA); Ethyl Alcohol, U.S.P. (The Warner-Graham Company, USA); M-PER Extraction Reagents (Pierce Biotechnology, USA); Halt Protease Inhibitor Cocktails (Pierce Biotechnology, USA); Anti-V5 mouse monoclonal IgG_{2a} antibody (Invitrogen, USA); HA. 11 polyclonal Antibody (Covance, USA); Posi-Tag Epitope Tag Control Protein (Covance, USA); ECL Plus Western Blotting Detection System (GE Healthcare Amersham, UK); 10x Tris/Glycine Buffer (Bio-Rad Laboratories, USA); 10x TRIS Buffered Saline pH 7.4 (Quality biological Inc., USA); Tween 20/Polysorrbate 20; Polyoxyethylene sorbitan monolaurate (American Bioanalytical, Natick, USA); NuPAGE MOPS SDS Running Buffer (20x) (Invitrogen, USA); BCA Protein Assay Kit (Pierce Biotechnology, USA); A454-4 Methanol (Fisher Scientific, USA); Sample Reducing Buffer (Invitrogen, USA); NuPAGE Antioxidant Buffer (Invitrogen, USA); NuPAGE LDS Sample Buffer (4x) (Invitrogen, USA); Instant Nonfat Dry Milk (Safeway, USA); Rainbow Protein Ladder (GE Healthcare, USA); Page Ruler Prestained Protein Ladder (Fermentas, USA); GenomiPhi DNA Amplification kit (GE Healthcare, UAS); TaqMan Universal Genotyping Master Mix (Applied Biosystems/Roche, USA); TaqMan, SNP genotyping Assays, Pre-Designed, Small-Scale (Applied Biosystems, USA); QuantiTech SYBR Green PCR Kit (Qiagen, USA);PCR Primers (Integrated DNA Technology, USA); DEPC Treated Water 100ml (Quality Biological, USA); LB Agar (Sigma, USA); SuperFect Transfection Reagent (Qiagen, USA); EndoFree Plasmid Maxi Kit (Qiagen, USA); 2-

Propanol, ≥99.5%, A.C.S. reagent (Sigma-Aldrich, USA); ProFound HA Tag IP/Co-IP Kit and Application Set (Pierce, IL, USA); iScript Select cDNA Synthesis Kit (Bio-Rad Laboratories, USA).

2.2.3 Materials

The following specific materials were used for research described in this thesis:

Eppendorf tubes, 1.5 ml, 2.0 ml (Eppendorf, USA); 8-strip of caps for 0.2 ml thin wall PCR (Thomas Scientific, USA); 8-strip of tubes for 0.2 ml thin wall PCR (Thomas Scientific, USA); Polypylene round-bottom tubes, 5 ml, 14 ml, 50 ml (Falcon Dickinson Labware, USA Dickinson Labware, USA); CryoTube vials, 1.8 ml (Nuncbrand, USA); Stripette, individually wrapped, 5 ml, 10 ml, 25 ml, 50 ml (Corning Inc., USA); Woodpecker reload 10 µl pipette tips (Denville, USA); Filter tips, 12.5 µl, 200 µl, 1250 µl (Matrix, USA); SpaceSaver tip rack refills, 250 µl, 1000 µl (Rainin, USA); Fisherbrand nitrile gloves, small (Fisher Scientific, USA); Multiwell tissue culture plate, 6 well, 12 well, 96 well (Falcon Dickinson Labware, USA, USA); Nonpyrogenic 100 mm x 20 mm, 250 mm x 20 mm dish (Corning Inc., USA); MicroAmp pptical adhesion film (Applied Biosystems, USA); MicroAmp optical reaction plates, 96 well, 384 well (Applied Biosystems, USA); NuPAGE 4-12% Bis-Tris gel 1.0mmx12well (Invitrogen, USA);Anhydrous calcium sulfate (W.A. Hammond Drierite, USA);Polyvinylidene fluoride (PVDF) microporous membrane; the Millipore immobilon-P transfer membrane (Millipore, USA); Fiber pads, mini trans-blot cell (Bio-Rad Laboratories, USA); Amersham hyperfilm ECL; high performance chemiluminescence film (GE Healthcare, UK); Corning Inc. Erlenmeyer flask, 1000 ml (Corning Inc. Inc., USA); 1 cc insulin syringe $28_G 1/2$ (Becton Dickinson, USA).

2.3 Study Cohorts

2.3.1 Southern California, USA

Patients were breast cancer probands diagnosed between 1 March 1994 and 28 February 1995, who were identified through the population-based cancer registry of the Cancer Surveillance Program of Orange County. The research was approved by the institutional review boards of the US National Institutes of Health and the University of California (Irvine, CA, USA). This population-based cohort includes 152 patients with regional or metastatic breast cancer and 148 randomly selected localized disease patients (Table 2.2). Regional disease indicates direct extension of the primary tumour to the skin, muscle, chest wall, lymph nodes, or a combination of the above; metastatic disease indicates all other forms of involvement beyond regional disease. Pathologic diagnoses were obtained at the time of initial presentation in all cases and the methods used for data collection were standardized. Data on tumour characteristics and stage at diagnosis were collected from the cancer registry. These data were based on pathology reports and medical records of the cancer patients. Blood samples were collected from all cancer patients, DNA was extracted from each sample and aliquots were sent to the US National Cancer Institute (IRB #HS-2004-3832) (Crawford *et al*., 2006).

Table 2.2: Participant characteristics by localized and regional/metastatic disease.

Characteristic		Localized (n = 148)	Regional/metastatic (n = 152)	P
Age at diagnosis	<50 years	44 (30.1%)	57 (37.0%)	0.0042[a]
	≥50 years	102 (69.9%)	97 (63.0%)	
	Mean age (years; mean ± SD)	59.7 ± 14.3	55.2 ± 12.6	
	Range (years)	33–91	25–87	
ER status	Positive	109 (74.7%)	105 (68.2%)	0.1741[b]
	Negative	22 (15.0%)	33 (21.4%)	
	Unknown	15 (10.3%)	16 (10.4%)	
PR status	Positive	96 (65.7%)	87 (56.5%)	0.0489[b]
	Negative	33 (22.7%)	451 (33.1%)	
	Unknown	17 (11.6%)	16 (10.4%)	
Tumour size	< 2 cm	102 (69.9%)	55 (35.7%)	<0.0001[c]
	≥ 2 cm	34 (27.4%)	84 (54.6%)	
	Unknown	4 (2.7%)	15 (9.7%)	
	Mean size (cm; mean ± SD)	1.6 ± 1.0	3.1 ± 2.7	
	Range (cm)	0.1–6.5	0.4–18.0	
Grade	Well to moderate	86 (58.9%)	46 (29.9%)	<0.0001[b]
	Poor	26 (17.8%)	79 (51.3%)	
	Unknown	34 (23.3%)	29 (18.8%)	

Values are expressed as *n* (%), unless otherwise stated.--[a]Based on a t-test.--[b]Based on a Fisher's exact test excluding the missing values.--[c]Based on a t-test on the logarithmic scale. ER, estrogen receptor; PR, progesterone receptor; SD, standard deviation (Crawford *et al.*, 2006).

2.3.2 Greater Baltimore, USA

This cohort included blood and survival information collected from 248 patients with pathologically confirmed breast cancer from African-American or Caucasian descent. Surgical breast cancer cases were recruited at the University Of Maryland Medical Centre, the Baltimore Veterans Affairs Medical Centre, Union Memorial Hospital, Mercy Medical Centre, and the Sinai Hospital in Baltimore, Maryland between February 15, 1993 and August 27, 2003. These patients had pathologically confirmed breast cancer, were of African-American or Caucasian descent by self-report and were diagnosed with breast cancer within the last six months prior to recruitment with no previous history of the disease (by self-report)

(Table 2.3). Patients were excluded if they were HIV, hepatitis B virus, or hepatitis C virus carriers, intravenous drug users, institutionalized, or were physically or mentally unable to sign consent and complete the questionnaire. The ER-α status, disease stage, treatment, and survival records were obtained from medical records and pathology reports, the Social Security Death Index, and the National Death Index. Disease staging was performed according to the TNM system of the American Joint Committee on Cancer/ the Union Internationale Centre le Cancer (AJCC/UICC). The Institutional Review Boards at the participating institutions approved the study (Crawford *et al.*, 2007).

Table 2.3: Characteristics of the Greater Baltimore Breast Cancer Cohort.

Patient Characteristic		*N*	%
Race/Ethnicity	African-American	143	58
	Caucasian	105	42
Survival	Alive	189	76
	Deceased	59	24
Tumour histology	Ductal	189	76
	Lobular	34	14
	Others	25	10
Node status	Negative	144	63
	Positive	84	37
Stage at diagnosis (TNM)[a]	≤ Stage I	66	29
	Stage II	118	52
	≥ Stage III	44	19
Chemotherapy	No	99	43
	Yes	132	57
Estrogen receptor-α	Negative	102	41
	Positive	145	59
Her2/neu	Low to weak	154	62
	Moderate to strong	93	38
Tumour Size	<2 cm	58	27
	2-5 cm	154	73

Values are expressed as *n* (%), unless otherwise stated.--[a]TNM: Tumour, Node, Metastasis system, the American Joint Committee on Cancer/ the Union Internationale Centre le Cancer (AJCC/UICC) (Crawford *et al.*, 2007).

2.3.3 Minnesota, USA

There were 843 patients included in this cohort, with 843 samples included for the breast cancer risk analysis and 625 samples for the mammary density analysis (samples may overlap) (Table 2.4). The patients were selected from the Mayo Clinic mammography screening practice in Rochester, Minnesota. Patients were diagnosed with primary invasive cancer or ductal carcinoma *in situ* and those with bilateral mastectomies or breast implants before diagnosis were excluded. Weight, height, and hormone replacement therapy were abstracted from the Mayo Clinic medical record for all serial mammogram taken dates. Weight and height, which were used to construct body mass index (BMI) in kilograms per meters squared, were routinely measured and recorded at medical exams associated with the mammogram. Hormone replacement therapy was abstracted from patient reported medication use in the medical record over the entire mammogram period (Celine *et al*., 2007). This study was approved by the Mayo Clinic Institutional Review Board.

Table 2.4: Characteristics of the Minnesota, USA Cohort.

Characteristic		Controls n=843 (risk analysis)		Controls n=625 (density analysis)	
	Level	N	(%)	N	(%)
Age, years	20-39	48	(5.7)	19	(3.0)
	40-49	166	(19.7)	129	(20.6)
	50-59	274	(32.5)	206	(33.0)
	60-69	207	(24.6)	159	(25.4)
	70+	148	(17.6)	112	(17.9)
BMI[a], kg/m^2	Median (25th-75th percentile)	25.8	(23.0-29.9)	27.0	(23.0-29.8)
Family history, 1st or 2nd degree relative with breast or ovarian cancer	Yes	345	(42.9)	244	(40.8)
	No	460	(57.1)	354	(59.2)
	Missing	38	(4.5)	27	(4.3)
Postmenopausal status	Yes	579	(72.5)	439	(71.8)
	No	220	(27.5)	172	(28.2)
	Missing	44	(5.2)	14	(2.2)
Age at menarche, years	< 12	122	(16.2)	89	(15.9)
	12	184	(24.4)	138	(24.6)
	13	238	(31.6)	178	(31.7)
	≥ 14	209	(27.8)	156	(27.8)
	Missing	90	(10.7)	64	(10.2)
Oral contraceptive use	0 Months	243	(31.3)	189	(32.8)
	1-48 Months	189	(24.3)	142	(24.6)
	48+ Months	345	(44.4)	246	(42.6)
	Missing	66	(7.8)	48	(7.7)
Postmenopausal hormone use	0 Months	366	(49.3)	252	(45.9)
	1-60 Months	160	(21.6)	124	(22.6)
	60+ Months	216	(29.1)	173	(31.5)
	Missing	101	(12.0)	76	(12.2)

Values are expressed as *n* (%), unless otherwise stated.--[a]BMI: Body Mass Index.

2.3.4 Fred Hutchinson Cancer Research Centre, USA

There were 903 patients included in this cohort and the details of the clinical characteristics are listed in Table 2.5. Patients were diagnosed with primary invasive breast cancer (diagnosed between 1 January 1993 and 30 June 2001) and continuously enrolled at the

Group Health Cooperative for at least one year subsequent to their diagnosis. Patients were identified through the Seattle-Puget Sound Surveillance, Epidemiology, and End Results (SEER) cancer registry (Velicer *et al.*, 2005). This study was approved by the Group Health Cooperative Human Subjects Review Committee.

Table 2.5: Characteristics of the Fred Hutchinson Cancer Research Centre, USA Cohort.

Clinical Characteristics			N	%
Oestrogen Receptor Status	Positive		663	73.42
	Negative		169	18.72
	Missing		71	7.86
Progesterone Receptor Status	Positive		596	66.00
	Negative		235	26.02
	Missing		72	7.98
Tumour Size	<= 2 cm		594	65.78
	> 2 cm		250	27.69
	Missing		59	6.53
Nodal Status	Positive		600	66.45
	Negative		261	28.90
	Missing		42	4.65
Stage	Local		623	68.99
	Regional/Distant Metastasis		280	31.01
	Missing		0	0
Grade	Well Differentiated		135	14.95
	Moderate Differentiated		287	31.78
	Poor/Undifferentiated		342	37.87
	Missing		139	15.4
Age	<50 years old	Alive	284	31.45
		Dead	48	5.32
	50+ years old	Alive	446	49.40
		Dead	124	13.73
	Missing		1	0.1
Survival	Alive		730	80.85
	Dead		172	19.05
	Missing		1	0.1

Values are expressed as *n* (%), unless otherwise stated.

2.3.5 Rotterdam, the Netherlands

This cohort used coded 1,922 primary tumour tissue samples from female Dutch patients with primary operable breast cancer who entered the clinic in Rotterdam between 1978 and 2002. The usage of these samples is in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (http:/www.fmvv.nl). All 1,922 tumours were included, to analyse the potential association of the variant allele frequencies of the various SNPs with patient and tumour characteristics. Adjuvant radiotherapy was given to patients undergoing lumpectomy (43%; 842 patients) or modified mastectomy (97%, 1080 patients) (Table 2.6). Tumour ER and PR levels were determined in cytosolic extracts by routine ligand binding assay or by enzyme immunoassay (Foekens *et al.*, 1989). The cut point to classify primary breast tumours as ER and/or PR positive was 10 fmol/mg of cytosolic protein and none of the patients had received neo-adjuvant therapy.

Table 2.6: Patient and tumour characteristics of the Rotterdam, the Netherlands Cohort.

Characteristics		N	%
Age	≤ 40	246	12.80
	41-55	750	39.02
	56-70	621	32.31
	>70	305	15.87
Menopausal status	Premenopausal	840	43.70
	Postmenopausal	1082	56.30
Lymph nodes involved	0	1101	57.28
	1-3	353	18.37
	>3	451	23.47
	Missing	17	0.88
Tumour Size	pT1 (≤ 2 cm)	695	36.16
	pT2 (2-5 cm)	1009	52.50
	pT3 (> 5 cm)	218	11.34
Grade	Poor	1044	54.32
	Poor/moderate	285	14.83
	Unknown	593	30.85
Oestrogen Receptor Status	Positive	1347	70.08
	Negative	481	25.03
	Missing	94	4.89
Progesterone Receptor Status	Positive	1120	58.27
	Negative	600	31.22
	Missing	202	10.51

Values are expressed as *n* (%), unless otherwise stated.

In the analysis of the association of the SNPs with prognosis, 1,864 patients were studied (Table 2.7). Inclusion criteria were that none of the patients showed distant metastasis within the first two months after primary surgery, and that none of the 1,096 lymph node-negative (LNN) patients had received adjuvant systemic therapy, thus allowing the analysis of the pure prognostic value. Five hundred forty-one of the 768 lymph node-positive (LNP) patients were treated with adjuvant systemic therapy, of whom 224 (29%) received hormonal therapy, 312 (41%) received chemotherapy, and 5 (1%) received combination therapy. The median follow-up time of patients alive was 90 months (range 4 to 231 months), with 1,015 events in the analysis of disease-free survival (DFS, no recurrent of breast cancer after the initial administration), 816 events in the analysis of distant metastasis-free survival (MFS, no distant metastasis detected after the initial administration), and 724 events in the analysis of overall survival (OS). Patients who showed a relapse and were subsequently treated with first-line tamoxifen monotherapy or polychemotherapy were evaluated according to the standard International Union Against Cancer (Geneva, Switzerland) classification criteria (Hayward, 1977) for complete remission (CR), partial remission (PR), and no change of disease (NC). NC was classified as NC > 6 months and NC ≤ 6 months for patients who received tamoxifen therapy. For patients who received chemotherapy, NC was classified as NC > 4 months and NC ≤ 4 months based on the evaluation after 3 and 6 cycles of treatment.

Three hundred and fifteen ER-positive patients (76 premenopausal, 239 postmenopausal) were included in the analysis for the relationships of the various SNPs with the efficacy of first-line tamoxifen therapy in recurrent disease. None of these patients had received prior adjuvant hormonal therapy, though 65 patients had received prior adjuvant chemotherapy (21 anthracycline-based, 44 non-anthracycline-based). These 315 patients included 276 patients

who showed a relapse in the prognostic analysis described above, and 39 additional patients who were not eligible for the prognostic study (relapse already at diagnosis, no axillary dissection, or information not complete enough to allow analysis of DFS or MFS). The median age of the patients at the start of tamoxifen therapy for recurrent disease was 60 years (range 29 to 87 years) and the median follow-up time of patients alive was 43 months (range 4 to 104 months). Two hundred and fifteen patients experienced a clinical benefit from first-line tamoxifen therapy, of which 10 patients (5%) had a CR, 41 patients (19%) had PR, and 164 patients (76%) showed NC > 6 months. During follow-up, 292 patients showed tumour progression (PD or NC $\leq$ 6 months) and were counted as events in the analysis for progression-free survival (PFS). Two hundred and twenty eight patients died and were considered events in the analysis of post-relapse overall survival (PRS).

Table 2.7: Characteristics of patients included in the prognostic study, n=1864.

Patient characteristics: prognostic study		Number of patients	%[a]
All patients[b]		1,864	100
Age at surgery (years)	≤ 40	239	13
	41 - 55	736	39
	56 - 70	597	32
	> 70	292	16
Primary diagnosis	M0	1,864	100
Surgery	lumpectomy	821	44
	mastectomy	1,043	56
Axillary clearance	yes	1,864	100
Adjuvant therapy[c]	none	1,323	71
	CMF	222	12
	FEC/FAC	95	5
	hormonal (mainly tamoxifen)	224	12
Nodal status	0	1,096	59
	1 - 3	350	19
	> 3	418	22
Tumour size	pT1:≤ 2 cm	686	37
	pT2:2 - 5 cm	978	52
	pT3/4:> 5cm	200	11
ER[d] status	negative	487	26
	positive	1,368	73
PR[e] status	negative	601	32
	positive	1,150	62
Grade[f]	good	27	1
	moderate	255	14
	poor	1,008	54
	unknown	574	31
Secondary breast cancer	no	1,734	93
	yes	130	7

[a]Due to rounding and missing values the numbers do not always add up to 100%.--[b]Median follow-up time was 90 months (range, 4 – 231 months), with 1015 events in the analysis for disease-free survival, 816 events in the analysis for distant metastasis-free survival, and 724 events in the analysis for overall survival.--[c]Adjuvant systemic therapy was only given to lymph-node positive patients.--[d]ER : Oestrogen receptor.--[e]PR : Progesterone receptor.--[f]Grade was assessed by regional pathologists and reflects the current practice during the years the tumours were collected.

Two hundred and twenty-seven (100 ER positive, 125 ER negative, 2 ER unknown) patients were included in the analysis for the relationships of the various SNPs with the efficacy of first-line chemotherapy (150 anthracycline-based, 77 non-anthracycline-based) in recurrent disease (Table 2.8). Fifty patients had received prior adjuvant chemotherapy (12 anthracycline-based, 38 non-anthracycline-based), and 60 patients had received adjuvant hormonal therapy. These 227 patients included 208 patients who showed a relapse in the

analysis for DFS and 19 additional patients who were not eligible for the prognostic study. The median age of the patients at the start of chemotherapy for recurrent disease was 51 years (range 25 to 82 years) and the median follow-up time of patients alive was 22 months (range 3 to 183). A different sub-group of patients (370 patients) include 149 patients who experienced a clinical benefit from first-line chemotherapy, of which 14 patients (9%) had a CR, 77 patients (52%) showed a PR, and 58 patients (39%) showed NC > 4 months. Two hundred and twenty-one patients showed tumour progression (PD or NC ≤ 4 months) and were counted as events in the analysis for PFS. Two hundred and six patients died and were considered events in the analysis of PRS.

Table 2.8: Characteristics of patients treated with first-line chemotherapy, n=227.

Patient characteristics: chemotherapy		Number of patients	%
All patients[a]		227	100
Primary diagnosis	M0	212	93
	M1	15	7
Age at start therapy (years)[b]	≤ 40	39	17
	41 - 55	112	49
	56 - 70	39	30
	> 70	7	3
Menopausal status at start therapy	pre menopausal	112	49
	post menopausal	115	51
Dominant site of relapse	local regional relapse	31	14
	bone	35	15
	other	161	71
Disease free survival	0 - 12 months	83	37
	12 - 36 months	104	46
	> 36 months	40	18
Adjuvant therapy	None	101	44
	chemo	50	22
	hormonal	60	26
	not applicable	16	7
Type of response[c]	CR	14	6
	PR	77	34
	NC > 4 months	58	26
	NC ≤ 4 months	14	6
	PD	64	28
ER[d] status	negative	100	55
	positive	125	44
	unknown	2	1
Type of first-line chemotherapy[4]	CMF[e]	77	34
	FEC/FAC/anthracycline	150	66

[a]During follow-up, 221 patients showed tumor progression and were counted as events in the analysis for progression-free survival. Two hundred and six patients died and were considered events in the analysis of post-relapse overall survival.--[b]The median age of the patients at start of tamoxifen therapy was 60 years (range, 29 – 87 years).--[c]Patients were evaluated according to the standard International Union Against Cancer (Geneva, Switzerland) classification criteria for complete remission (CR), partial remission (PR), no change (NC) classified as NC > 4 months and NC≤ 4 months (based on the evaluation after 3 and 6 cycles), and progressive disease (PD). Clinical benefit was defined as CR + PR + NC > 4 months.--[d]ER: oestrogen receptor.--[e]CMF: cyclophosphamide, methotrexate, 5-fluorouracil; FEC/FAC: 5-fluorouracil, epirubicin or adriamycin, cyclophosphamide.

2.3.6 Germany

This cohort includes 200 DNA samples from normal bone marrow tissue, segregated into 100 patients with cytokeratin (CK) staining positive (bone metastasis positive) bone marrow cells and 100 samples without CK-positive bone marrow cells (bone metastasis negative). Patients with primary breast cancer (stage M_0) and advanced (stage M_1) were recruited for this study (from Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany). Fine needle bone marrow aspirates were taken from one site of the upper iliac crest at the time of surgery, with the time of bone marrow aspiration varied between patients. Volunteering non-tumour patients (healthy bone marrow donors) were selected as a negative control group (Woelfle *et al*., 2005). All Patients and healthy volunteers (controls) gave their informed consent and this study was approved by the local Ethics Review Board.

2.3.7 Queensland, Australia

The population screened in this cohort comprised 200 female individuals previously diagnosed with breast cancer and 200 female that had no cancer history of any type. The affected population was predominately recruited from 416 breast cancer affected individuals in collaboration with the Pathology Department of the Gold Coast Hospital, Southport. In this population, originally 416 breast cancer affected individuals were recruited; however through out the years, some samples were used up by other experiments and some patients lacked the clinical records needed in this study and thus were excluded from the sample population. Therefore only 200 samples were available for the analysis in this study. Further affected individuals were obtained through the Genomics Research Centre, Griffith University, Southport, Queensland, Australia. Each affected individual was without a known family

history of breast cancer, although unidentified familial cases would not be expected to exceed 5%. The control population in whole was recruited through the Genomics Research Centre over a 4-year period. At the time of blood donation, each individual completed a questionnaire outlining their personal and family history of several disorders and only those individuals with no family history of cancer or precancerous conditions were selected for involvement in this study. All volunteers were informed of the objectives of the study and consent was received for their participation. A blood sample was collected from each volunteer and DNA extracted using standard procedures (Smith *et al*., 2001). Control samples were age (±5years), sex and ethnicity (all Caucasian to grandparents) matched to cases.

2.3.8 Umea, Sweden

The population-based Medical Biobank at Umeå University contains blood samples and questionnaires collected in the Northern Sweden Health and Disease Study (NSHDS), which consists of the Västerbotten Intervention program (VIP), the Northern Sweden MONICA project and the Västerbotten Mammary Screening Program. The biobank, which started specimen collection in 1986, contains 170,000 blood samples from 95,000 subjects. All samples studied in this research were drawn after informed consent. The breast cancer cases (N=801) (Table 2.9) were identified by linkage to the Regional Cancer Registry by means of personal civic numbers. All blood samples from breast cancer cases in this study were collected prior to the diagnosis of breast cancer.

Table 2.9: Tumour characteristics at diagnosis, Umeå cohort.

Characteristic		N	%
Age at diagnosis	<= 50 years	176	22.0
	> 50 years	625	88.0
Estrogen receptor	Positive	415	51.8
	Negative	245	30.6
	missing	141	17.6
Progesterone receptor	Positive	290	36.2
	Negative	283	35.3
	missing	228	28.5
Tumour size	<= 2 cm	523	65.3
	2-5 cm	222	27.7
	> 5 cm	14	1.7
	missing	40	5.0
Histologic grade	1	159	19.9
	2	365	45.6
	3	231	28.8
	missing	46	5.7
Distant metastases	Negative	773	96.5
	Positive	14	1.7
	Missing	14	1.7
Regional lymph node metastases	Negative	479	59.8
	Positive	224	28.0
	Missing	98	12.2
Stage	0	2	0.2
	1	413	51.6
	2A	230	28.7
	2B	102	12.7
	3A	14	1.7
	3B	12	1.5
	4	16	2.0
	missing	12	1.5

Values are expressed as *n* (%), unless otherwise stated.

2.4 Methods

2.4.1 SNP Analysis Methods

Whole Genome Amplification

To increase the amount of genetic DNA, which is extracted from patient's lymphocytes, DNA was whole genome amplified using the commercially available kit, GenomiPhi DNA Amplification kit (GE Healthcare, UAS).

The amplification was performed according to the short protocol from the company. Briefly, 1 µl of template DNA, which needed to be at least 10ng/ µl was mixed with 9 µl of sample buffer. The mixture was heated to 95 °C for 3 min. then cooled to 4 °C. To prepare the amplification reaction, 9 µl of reaction buffer was mixed with 1 µl of enzyme on ice, and then the cooled sample was added. The mixture was incubated at 30 °C for 16-18 hours. To inactivate the reaction, the mixture was heated to 65 °C for 10 minutes and cooled to 4 °C. The amplification process was then complete.

Single Nucleotide Polymorphism Selection

The epidemiology study is based on the common-disease, common-variation model; therefore, only common SNPs are of interest to this study. The SNPs were chosen based on the following criteria: First, the genomic region that encodes the longest transcript (if there are multiple isoforms) +/- 2 kb is the region of interest for choosing candidate SNPs. dbSNP build 126 was used as the reference SNP database, along with NCBI's build 36.1 (Mar. 2006) as the reference genome assembly. Secondly, all SNPs with functional importance (non-synonymous, splice variant) were chosen regardless of their allele frequency. Those that were

found to be invalid SNPs in previous genotype assays were eliminated based on the Population Diversity information provided in the dbSNP.

The LD bin structure analysis was the third criterion and was implemented using the HapMap and Perlegen genotype database. SNPs that have been assayed by the HapMap project in Caucasian and Yuroban African with ≥ 10% allele frequency were identified and LD bins were constructed using SNPs (with ≥ 10% allele frequency) by running the program LD Select. LD bins were built first with Caucasian and Yuroban separately and then secondly for a combined population. TagSNPs in LD bins were then selected, with, TagSNPs that appeared in both populations given a higher priority. TagSNPs from this list with a profile threshold $r2 \geq 0.80$ were then selected as candidate SNPs in this study.

Two SNPs of the *ARAP3* gene were selected based on haplotype-tagging. These two SNPs are intronic: NCBI SNP rs440279 and NCBI SNP rs3763120. rs440279 (C/G) is located in intron 6 and rs3763120 (C/T) is located in intron 15.

Genotyping

The SNP analysis was carried out using TaqMan Universal PCR Master Mix (Applied Biosystems, USA) and TaqMan SNP Genotyping Assays (Applied Biosystems, USA).

The TaqMan SNP Genotyping Assays are composed of primer and probes that are pre-designed and validated by the company. The probes are conjugated with VIC-MGB or FAM-MGB dyes, one to each allele. Once the PCR is complete the 7900HT Sequencing Detection System (Applied Biosystem, USA) distributes the data points according to the signals generated depending on the allele composition of each patient. The genotype determination

was made by a signal distribution comparison to several control samples (derived from CEPH Family DNA) for which the genotypes are known.

The assays were carried out following the instructions from the company. For a 5 µl PCR reaction, the DNA needs to be dried (by placing it in desiccators) before adding the other reagents. The amount of volume required for each 5 µl reaction was as follows: 2.5 of TaqMan Universal PCR Master Mix, 0.25 µl TaqMan SNP Genotyping Assay and 2.25 µl of DEPC Treated Water (Quality Biological, INC, USA). The reaction conditions were as follows: 95 °C for 10 minutes, and 50 cycles of 92 °C for 15 seconds and 60 °C for 1 minute. After the PCR was finished, the plates were then read by the 7900HT Sequencing Detection System (Applied Biosystems, USA).

Statistical Analyses

The statistical methods used to analyse each component of the genotyping data will be outlined below, with each collaborators using different analytical methods for their own population. Details of the statistical analyses results and discussion of these results will be outlined in Chapter 3.

Most populations used univariate and multivariate analyses to assess the possible associations between the SNP and different clinical parameters. Univariate analysis analysed the association between one SNP and a single variable, while the multivariate analysed multiple variabl phenotypes such as ER status, Pr status, age etc. Univaiate analysis was used in analysing the result to determine if each variable, such as age, hormonal receptor status etc was independently associated with the SNP from other variables. The multivariate analysis

was then carried out to determine the possible association between the SNPs and the combined effect of different variables.

A number of populations used the dominant and recessive models in the analyses. The dominant model assesses the association of a dominant allele in a SNP. Typically this model compares the group of the homozygous dominant allele to the heterozygotes and homozygous variant allele. The recessive model examines the opposite of the dominant model. This model compares the group of the homozygous recessive allele to the heterozygotes and homozygous dominant allele. These models were used to determine which allele is associated with the variants.

This study analysed multiple SNPs in each population and it is possible to suggest that the p-value should be adjusted to account for such multiple testing issues. Multiple testing corrections adjust p-values to correct for occurrences of false positive results. This calculation is commonly used for analyses where multiple variables, such as multiple genes and SNPs, are analysed as a single event, such as the microarray analysis and genome wide association study. This study analysed individual genes that had already been selected by other techniques that indicated that they may play a part in breast cancer progression. As such, each gene analysed is analysed as a single event, with each SNP of the gene analysed in a similar manner. This is because each SNP is haplotype tagged and represent succinct regions of particular genetic loci. Each SNP and each gene were carefully selected for the specific criteria they represent and they were analysed as separate factors that could potentially modulate breast cancer progression. In the case of the *SIPA1* and *RRP1B* SNP interaction study, the SNPs were still treated as individual events. The premise of the study

was to identify if the polymorphism status of each gene could be combined to provide a more through sub-population stratification, which indeed it did.

Southern California and Greater Baltimore, USA Cohorts

The Southern California, USA population and The Greater Baltimore, USA population was analysed with the Student's *t* test and the Wilcoxon's sum rank test to compare medians. For linkage disequlibrium determination, which was only performed in the Southern California, USA population, Lewontin's D prime (D') and correlation coefficient (r^2) were calculated as two measures of linkage disequlibrium. The frequencies of the most common haplotypes (with a frequency >1%) in this population were estimated using the E-M algorithm and the expected haplotypes for each subject were imputed. For genotyping data analyses, χ^2 test or Fisher's exact tests were used to analyse differences between categorical variables and the Hardy-Weinberg equilibrium test. In addition, those variables that were not normally distributed, such as tumour size, were log transformed. For multivariate analyses, unconditional logistic regression adjusting for multivariate covariates, such as age at diagnosis, was used; likelihood ratio tests were used to calculate *p*-values comparing a model with haplotypes versus a model without. Cox regression models were used to perform multivariate survival analysis. All *p*-values presented in this study were for two tailed tests and were considered to be statistically significant if they were below 0.05 (Crawford *et al.*, 2006). For all multivariate survival analyses performed in the Greater Baltimore population, such as Table 3.6, only the association between the SNP and survival rate was determined. This analysis didn’t examine the genotype or allele responsible for the association and only aimed to determine whether the SNP as a single event was related to survival rate or not. Therefore, no genotype/allele association can be shown for this particular analysis.

Minnesota, USA Cohort

The distributions of demographic and other clinical variables of the Minnesota, USA samples were summarised and presented as means and standard deviations or counts and percentages. The baseline risk factors considered in this study were BMI (body mass index), menopausal status, family history of breast cancer, age at first birth, number of births, and HRT (hormone replace therapy) use. For the genotyping data analyses, conditional logistic regression was used to calculate odds ratios (OR) with adjustment for potential confounding factors (Celine *et al*., 2007).

Fred Hutchinson Cancer Research Centre, USA Cohort

The data collected from the Fred Hutchinson Cancer Research Centre, USA cohort were analysed with age-adjusted unconditional logistic regression to test for associations between single nucleotide polymorphisms and breast cancer risk factors and tumour characteristics. All analyses were performed using Intercooled Stata Version 8.2 (Stata Corp, College Station, TX) and a *p*-value less than 0.05 was used to determine statistical significance (Velicer *et al*., 2005).

Rotterdam, Netherlands Cohort

In the Rotterdam, Netherlands cohort, the relationship of the variant SNP alleles with patient and tumour characteristics was analysed with Pearson's chi-squared statistic. The hazard ratios (HRs) for SNPs were determined with Cox proportional hazards models for both uni- (disease free survival, metastasis free survival, and overall survival) and multivariate regression analyses. Metastasis free survival was considered the major endpoint for the prognostic study. The endpoint for disease free survival was defined as any recurrence of the disease including secondary breast cancer in the contralateral breast. Metastasis was defined

as any distant recurrence not including secondary breast cancer or local or regional recurrences. Death from any cause was considered an event for overall survival. Hazard Ratios were represented with their 95% confidence intervals (95% CI). Survival curves were generated using the Kaplan-Meier method, a log-rank test was used to test for differences between the survival curves or when appropriate the logrank test for trend. Computations were performed with the STATA statistical package, release 10.0 (STATA Corp, College Station, TX). All p-values were two-sided, and $p < 0.05$ was considered statistically significant.

Germany Cohort

The data collected from the bone marrow samples from Germany were analysed with a two-sided exact Wilcoxon signed rank test in the software package R (Woelfle *et al*., 2005).

Queensland, Australia Cohort

The case-control study of the Australian cohort was performed with the chi-square method (STATA statistical package), allele and genotype frequencies were compared to determine whether any significant differences in polymorphism frequencies occurred between the case and control populations. In addition, the Monte Carlo style CLUMP analysis program was used when necessary to analyse the data, since the minor allele frequency is low (n <5) and the standard Chi-square method's assumptions are violated (it needs counts of at least 5 in all categories) in this particular cohort.

Umea, Sweden Cohort

The data collected from the Umea, Sweden cohort were analysed with SPSS and P value < 0.05 was regarded as statistically significant. The data were then analysed by the same method used in the Southern California, USA cohort as described previously.

2.4.2 Molecular Characteristics Analysis Methods

Several experiments were previously performed by the Hunter lab aiming to characterise the novel metastasis modulating gene *Sipa1*. One of the experiments performed was yeast-two-hybrid (Y2H) analysis on the *Sipa1* protein, in an attempt to identify potential *Sipa1* protein binding partners. Several proteins were identified in the Y2H experiment as putative *Sipa1* protein binding partners. They include *p21Arc* (actin related protein 2/3 complex, subunit 3), *Acin1* (apoptotic chromatin condensation inducer 1), *Npc1* (Niemann Pick type C1), *Gart* (phosphoribosylglycinamide formyltransferase), *Gtf2h2* (general transcription factor II H, polypeptide 2) and *Calm2* (calmodulin 2).

The aim of the molecular characteristics analysis was to confirm the Y2H results with Co-immunoprecipitation (Co-IP) experiments. Once a putative *Sipa1* protein interacting partner was confirmed with the Co-IP analysis, *in vivo* analysis of this putative Sipa1 protein interacting partner was performed to elucidate the potential role of this gene in breast cancer metastasis.

The methods used to address the aim of the molecular characteristics analysis will be outlined below and described in detail in Chapter 4, along with the results and discussion of this section of research.

In vitro Yeast-2-Hybrid Result Confirmation

Vector: the plasmid vectors containing the putative *Sipa1* protein interacting partner gene were harvested from *E coli* using the EndoFree Plasmid Maxi extracting kit form Qiagen, USA. The exact procedure performed on vector purification will be described in Chapter 4.

Transfection: the *Sipa1* protein interacting partner genes were subsequently transiently co-transfected with *Sipa1* in the GT49 cells, which is a Chinese hamster cell line used previously in the lab and known to be highly transfectable. The specific methods used for tissue culturing, transfection, protein extraction, protein concentration estimation, SDS-PAGE and Western Blot analysis will be described in detail in Chapter 4.

Co-IP: Proteins collected from the cells were first analysed for up-regulation of both *Sipa1* and *Sipa1* protein interacting partner genes (via Western blot analyses). And then co-immunoprecipitation assays were performed using the ProFound HA Tag IP/Co-IP Kit and Application Set from Pierce, IL, USA, to determine whether those two proteins physically interact. Detailed description on the Co-IP analysis is listed in Chapter 4.

In vivo Metastasis Assay

Stable Cell Line Establishment: The sole identified Sipa1-interacting protein, Calmodulin 2, was stably transfected into the Met-1 mouse mammary carcinoma cell line. The increased expression of the *Calm2* gene was confirmed by real-time PCR. The methods used for RNA extraction, reverse-transcriptase reaction and quantitative real-time PCR analysis will be described in detail in Chapter4.

In vivo Metastasis Analysis: to determine whether the up-regulation of *Calm2* increases or decreases mouse lung metastasis, stable *Calm2* up-regulated cells were orthotopically transplanted into the anterior mouse mammary fat pad. The details of the protocol followed

ethical approval, with the methods for animal handling and orthotopic implantation of the mouse mammary fat-pad described in Chapter 4.

Chapter 3

Single Nucleotide Polymorphism Results

3.1 Introduction

Previously the Hunter lab demonstrated that inherited variation influences mammary tumour metastasis potency in the Polyoma middle T mouse model (Hunter, 2005). Subsequent molecular and genetic research identified regions that were differentially expressed between high metastatic and low metastatic mouse strains, with fine mapping tools to identify candidate genes. A cross reference between this list of genes and genes that modulated extracellular matrix (ECM) gene expression, which are frequently differentially expressed in metastasized tumours and are regarded as an indicator of breast cancer metastasis, was subsequently performed to select candidate genes for the epidemiology studies. This is based on the hypothesis that genes that modulate ECM gene expression are potentially metastasis potency modulators. This project is designed to validate potential breast cancer metastasis modulators, which were previously identified in the mouse model, in independent human breast cancer cohorts. This study explored SNP genotype frequencies of 16 candidate genes in independent patient cohorts, with corresponding clinical records, aiming to validate and translate the observations obtained in the mouse model into the human population.

The single nucleotide polymorphisms in each candidate gene were selected by Dr Jinghui Zhang (Laboratory of Population Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America). These selected SNPs were screened in two pilot cohorts, with SNPs showing significant association with patient survival then subsequently screened in other independent larger cohorts. For large scale SNP analysis, the experimental procedure was as follows: the sample DNA (provided by each collaborator in low amounts) was whole genome amplified to

gain larger quantities of material for large amounts of SNP screening. Amplified sample DNAs were distributed in 384-well plates (Applied Biosystems, USA) and allowed to dry in each well. Subsequent genotyping assays were carried out utilizing primers and probes pre-designed by Applied Biosystems (Applied Biosystems, USA) and the raw data (an example is provided in Figure 3.1) were gathered and translated into corresponding genotypes.

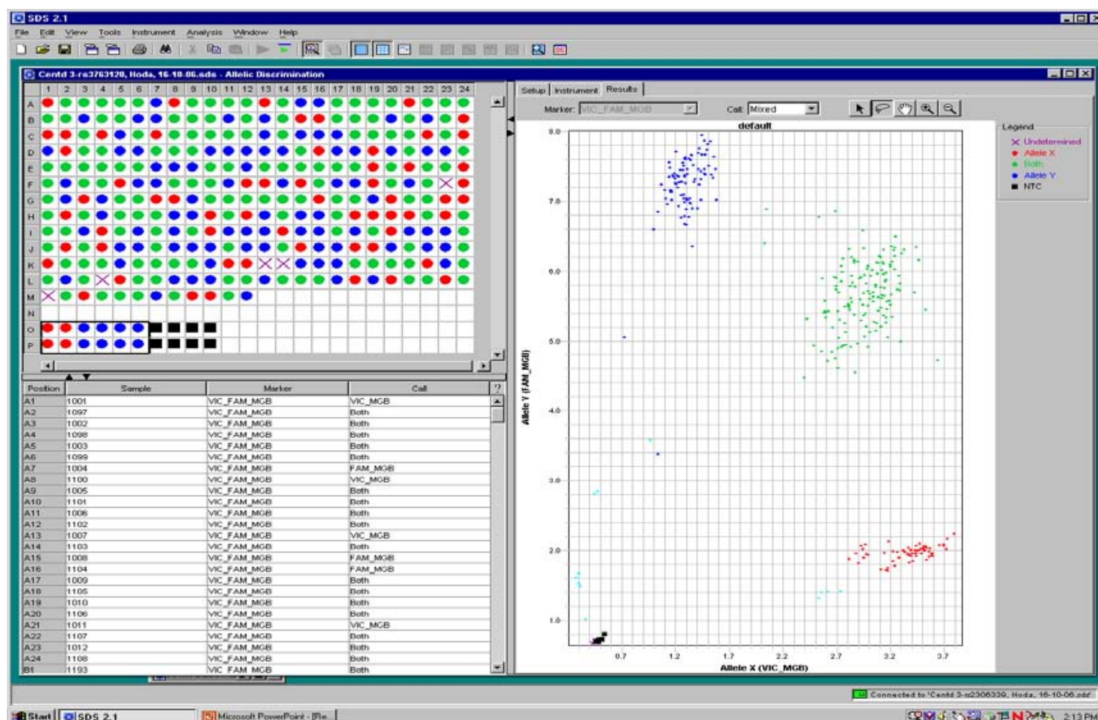


Figure 3.1: Raw output of SNP assay. This is an example of the raw output of SNP rs3763120 (*ARAP3* gene) genotyping in the Southern California, USA cohort. The blue and red dots are homozygous genotypes samples (blue: the FAM-dye reporter, red: VIC-dye reporter), green colour dots are heterozygous genotype. The actual genotype of each sample was assigned according to the manufacture's Genotyping Kit information sheet (Applied Biosystems, USA). The lighter blue dots are positive control samples with known genotypes of this SNP.

The genotyping data for each tested SNP was then forwarded to each collaborator for statistical analysis. The statistical analysis on data gathered from the Southern California, USA cohort was performed by Dr Argyrios Ziogas (from Dr Hoda Anton-Culver's lab, Epidemiology Division, Department of Medicine, University of

California Irvine, Irvine, California, United States of America). The data from the greater Baltimore, USA cohort was analysed by Dr Stefan Ambes (Laboratory of Human Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America). Data obtained from the Umeå, Sweden cohort samples were analysed by Dr Håkan Jonsson (Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden), the Fred Hutchinson Cancer Research Centre, USA cohort by Mr David Doody (from Dr Kathi Malone's lab, Cancer Epidemiology Research Coop, Fred Hutchinson Cancer Research Centre, Seattle, Washington, United States of America) and the Hamburg, Germany cohort by Prof. Dr Burkhard Brandt (Institute for Tumour Biology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany). The Minnesota, USA cohort data was analysed by Mr Christopher Scott (from Dr Celine Vachon's lab, Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, USA). The largest population, the Rotterdam, Netherlands cohort samples were prepared by Miss Anieta Sieuwerts and data analysed by Miss Maxime Look (from Dr John Foekens's lab, Erasmus Medical Centre, Rotterdam, The Netherlands). The statistical analyses results will be presented in the subsequent sections of this chapter.

A summary of the HWE calculation over some SNPs is included in Table 3.1. The genotype frequencies of these SNP were examined in the Southern California, USA cohort (n=300), with Hardy-Weinberg equilibrium (HWE) determined (Information from Gemone Variation Server, Seattle, USA). None of the SNPs deviated from HWE in this population. Only those HWE data that were available to me were listed in this table. The HWE of each SNP has been determined using the chi-square analysis in excel with automatic generated genotype frequencies for the predicted population.

Table 3.1: SNP Genotype Frequencies.

Gene	refSNP ID[a]	N[b]	*ARAP3* gene SNP Genotype Frequencies						Std[c]	p-HW[d]	Alleles[e]
ARAP3	rs2306339	278	GG	60.80%	GA	33.80%	AA	5.40%	1.1	0.7287	A/G
	rs6895094	278	TT	34.90%	GT	47.10%	GG	18.00%	1.5	0.6235	G/T
	rs6891143	278	GG	42.10%	GA	45.70%	AA	12.20%	1.4	1.0000	A/G
	rs17208495	280	CC	75.40%	CT	21.40%	TT	3.20%	0.9	0.0818	T/C
	rs3763120	280	GG	30.40%	GA	48.60%	AA	21.10%	1.5	0.7166	A/G
	rs440279	278	CC	71.10%	GC	26.40%	GG	2.50%	0.8	1.0000	G/C
	rs4912610	262	TT	76.30%	CT	22.10%	CC	1.50%	0.7	1.0000	C/T
	rs1031904	253	CC	83.40%	GA	16.20%	GG	0.40%	0.4	1.0000	G/C
BRD4	rs4808272	275	AA	29.50%	GA	47.60%	GG	22.90%	1.5	0.4643	G/A
	rs4808272	275	AA	29.50%	GA	47.60%	GG	22.90%	1.5	0.4643	G/A
	rs11880801	253	GG	66.40%	GT	27.70%	TT	5.90%	1.1	0.0487	T/G
	rs8104223	276	AA	46.00%	GA	44.20%	GG	9.80%	1.3	0.8827	G/A
	rs4809130	276	CC	79.00%	CT	20.30%	TT	0.70%	0.5	0.7622	T/C
PI16	rs1405069	279	AA	32.30%	AC	47.00%	CC	20.80%	1.5	0.4568	C/A
	rs707542	277	GG	33.90%	GA	47.70%	AA	18.40%	1.5	0.7121	A/G
	rs707998	279	GG	64.90%	GA	32.30%	AA	2.90%	0.9	0.5573	A/G
	rs734905	279	CC	64.90%	AC	33.00%	AA	2.20%	0.8	0.1730	A/C
	rs6901560	281	GG	58.00%	GC	37.70%	CC	4.30%	1.0	0.3978	C/G
	rs7755143	281	CC	65.80%	CT	32.00%	TT	2.10%	0.8	0.2355	T/C
	rs708006	280	CC	71.10%	AC	27.10%	AA	1.80%	0.7	0.6410	A/C
	rs4147297	270	TT	32.20%	TA	49.30%	AA	18.50%	1.5	1.0000	A/T
LUC7L	rs1211375	278	CC	37.10%	AC	46.80%	AA	16.20%	2.1	0.7097	A/C
	rs3918352	279	GG	34.10%	GA	45.90%	AA	20.10%	2.2	0.2676	A/G
	rs1164260	278	CC	23.00%	CT	47.50%	TT	29.50%	2.2	0.4638	T/C
	rs1203981	243	CC	17.70%	CT	44.90%	TT	37.40%	2.3	0.2883	T/C
	rs8044711	277	GG	79.10%	GA	20.20%	AA	0.70%	1.3	0.7576	A/G

[a]refSNP ID: Reference SNP Cluster Identification Number.--[b]Total number of samples.--[c]Standard Deviation.--[d]*p*-value of Hardy-Weinberg equilibrium--[e]The first alleles are the variant alleles.

3.2 *ARAP3* Gene Results

ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 (*ARAP3*) was previously identified as a candidate breast cancer progression and metastasis modifier gene in a mouse model by its high mRNA transcript correlation with extra-cellular matrix (ECM) genes, which have been frequently found to be associated with multiple independent metastasis predictive gene expression signatures (Crawford *et al*., 2008). Molecular research reveals that *ARAP3* (Chromosome 5q31.3) is a phosphoinositide binding protein containing ARF-GAP, RHO-GAP, RAS-associating, and pleckstrin homology domains. The function of the ARF-GAP and RHO-GAP domains is to co-operate in mediating rearrangements in the cell cytoskeleton and cell shape. Though the nature of *ARAP3* is not yet known, it has been identified as a specific PtdIns(3,4,5)P3/PtdIns(3,4)P2-stimulated Arf6-GAP protein. Research has identified an alternatively spliced transcript for *ARAP3* (Raaijmakers *et al*., 2007). Evidence supporting a role for *ARAP3* in human breast cancer progression was developed by identifying a signature of *ARAP3* expression in the mouse *in vitro* model system and applying it to publicly available human breast cancer datasets. Clustering the patient samples into groups that appeared to have up-regulated *ARAP3* versus those that did not revealed a survival benefit in the former class, consistent with the mouse model data (Crawford *et al*., 2008).

The pilot study cohorts included the Southern California, USA cohort (n=269) and the Greater Baltimore, USA cohort (n=248 surgical breast cancer patients; 58% African-American, 42% Caucasian).

Southern California, USA

Three of the *ARAP3* SNPs (rs440279, rs1031904 and rs17208495) gave significant association with major clinical characteristics, such as oestrogen receptor status, breast cancer stage and lymph node metastasis (univariate analysis). The TT genotype of rs17208495 was more prevalent in lymph node metastasis positive patients (87.5%) and was significantly associated with lymph node metastasis status (p=0.0320) (see Table 3.2). The rest of the SNPs screened (rs3763120, rs6891143, rs2306339, rs4912610 and rs6895094) did not show association with clinical markers. The clinical information was not available to the author and logistic regression analysis was not performed on these data, therefore the strength of the association observed could not be further determined.

Table 3.2: Southern California, USA, *ARAP3* gene SNPs, univariate analysis.

Clinical Markers and [c]refSNP ID	Genotype	Numbers and Frequency			
Stage rs440279		Non-localized		Localized	
		N	%	N	%
	GG	58	41	59	43
	GC	60	43	67	49
	CC	23	16	11	8
	Total	141	100	137	100
	[b]Dominant model: GG + GC v.s. CC, p=0.8082				
	[b]Recessive model: GG v.s GC + CC, p=0.0436[a]				
Oestrogen Receptor Status rs1031904		Positive		Negative	
		N	%	N	%
	CC	39	75	154	87.5
	CG	14	25	21	12
	GG	0	0	1	0.5
	Total	52	100	176	100
	[b]Dominant model: GG + GC v.s. CC, p=0.0171[a]				
	[b]Recessive model: GG v.s GC + CC, p=1.000				
Lymph Nodal Metastasis Status rs17208495		Positive Node		Negative Node	
		N	%	N	%
	CC	93	75	102	77.2
	CT	25	20	29	22
	TT	7	5	1	0.8
	Total	125	100	132	100
	[b]Dominant model: TT + TC v.s CC, p=0.6622				
	[b]Recessive model: TT v.s TC + CC, p=0.0320[a]				

[a]p< 0.05.--[b]Fisher's Exact Chi test.--[c]refSNP ID: Reference SNP Cluster Identification Number.

Three of the *ARAP3* SNPs, which are in the same LD block in this population (Figure 3.2) screened in the Southern California, USA cohort were found to be associated with patient survival in the multivariate survival analysis.

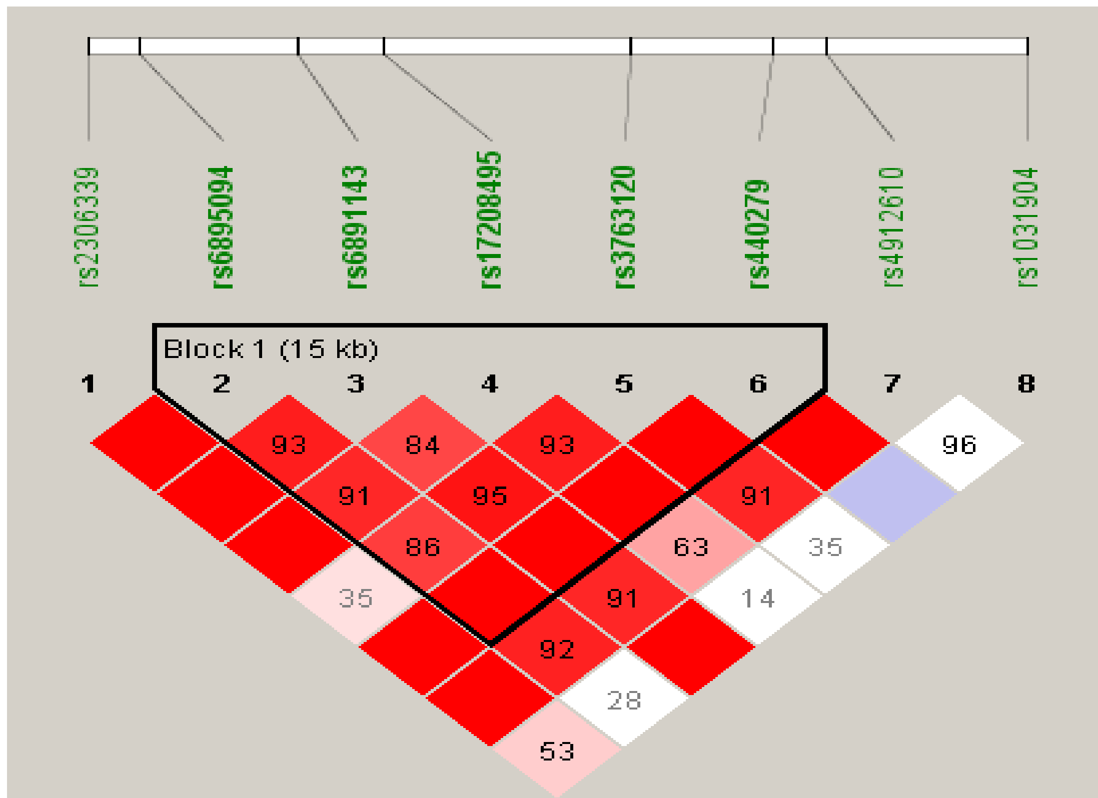


Figure 3.2: Linkage Disequilibrium block for *ARAP3* SNPs. This is a figure of the Linkage Disequilibrium block for the eight candidate single nucleotide polymorphisms screened in the Southern Californian, USA cohort. The three SNPs found to be associated with patients survival (rs440279, rs3763120 and rs6891174) are in the same LD block.

The survival analyses revealed that the GG genotype from *ARAP3* gene SNP rs3763120 was associated with poor survival in both the five and ten year analyses (5 years: HR=2.52, 95% CI= 1.33-4.77, p=0.0045; 10 years: HR=1.98, 95% CI= 1.15-3.41, p=0.0142) (see Table 3.3). Single nucleotide polymorphism rs440279 was found to be associated with better survival rate in the 10 year analysis and rs6891143 was found to be associated with poor outcome in the same analysis (Table 3.3).

Table 3.3: Southern California, USA, *ARAP3* gene SNPs, multivariate survival analysis.

refSNP ID[a]	Model	5 Years[h]			10 Years[i]		
		HR[f]	95% CI[g]	p	HR[f]	95% CI[g]	p
rs3763120	Dom[c]. GG + GA v.s AA	1.37	0.67-2.80	0.3941	1.92	1.00-3.71	0.0507
	Add[d]. GG v.s. GA v.s. AA	1.65	1.06-2.58	[b]0.0268	1.67	1.15-2.42	[b]0.0065
	Rec[e]. GG v.s. GA + AA	2.52	1.33-4.77	[b]0.0045	1.98	1.15-3.41	[b]0.0142
rs440279	Dom[c]. GG + GC v.s. CC	0.56	0.30-1.04	0.0671	0.55	0.33-0.92	[b]0.0225
	Add[d]. GG v.s GC v.s CC	0.62	0.38-1.01	0.0562	0.60	0.40-0.90	[b]0.0139
	Rec[e]. GG v.s GC + CC	0.49	0.15-1.60	0.2374	0.44	0.16-1.20	0.1085
rs6891143	Dom[c]. GG + GA v.s AA	1.08	0.54-2.15	0.8213	1.22	0.71-2.12	0.4734
	Add[d]. GG v.s. GA v.s. AA	1.15	0.61-2.18	0.6693	1.34	0.80-2.23	0.2622
	Rec[e]. GG v.s. GA + AA	2.81	0.37-21.02	0.3152	5.24	1.21-22.71	[b]0.0270

[a]refSNP ID: Reference SNP Cluster Identification Number.--[b]$p < 0.05$.--[c]Dom.: Dominant model.--[d]Add.: Additive model.--[e]Rec.: Recessive model.--[f]HR: Hazard Ratio.--[g]95% CI: 95% Confidence Interval.--[h]Analysis is adjusted with breast cancer stage and PR status in the 5 year survival analysis.--[i]Analysis is adjusted with breast cancer stage, PR status, tumour size and age at diagnosis in the 10 year survival analysis.

Greater Baltimore, USA

The univariate analysis on the Greater Baltimore, USA cohort (n=248) revealed that the AA genotype from *ARAP3* gene SNP rs3763120 was associated with larger tumour size at diagnosis (p=0.011) (Table 3.4). The rest of the 7 SNPs screened showed no association with clinical markers in this population.

Table 3.4: Greater Baltimore, USA, *ARAP3* gene SNPs, univariate analysis.

Clinical Markers and refSNP ID[a]	Genotype	Numbers and Frequency				Pearson chi2 (2)	Pr
		<2 cm		2-5 cm			
		N	%	n	%		
Tumour Size rs3763120	GG	15	26	58	38	9.0832	0.011
	AG	26	45	77	50		
	AA	17	29	19	12		
	Total	58	100	154	100		

[a]refSNP ID: Reference SNP Cluster Identification Number.

The multivariate Cox regression survival analysis revealed that three SNPs; rs440279, rs3763120 and rs4912610, from the *ARAP3* gene were associated with breast cancer survival (Table 3.5). The association found with rs4912610 and poor survival was not

observed in the Southern California, USA cohort and thus this SNP was ranked as a lower priority SNP for further epidemiology analyses. The associations found with SNP rs440279, which showed better survival rate in both pilot cohorts, and rs3763120, which was associated with poor outcome in both pilot cohorts, were ranked as high priority for further epidemiology analyses in multiple lager independent cohorts in order to confirm these observations.

Table 3.5: Greater Baltimore, *ARAP3* gene SNP, multivariate Cox regression survival analysis.

refSNP ID[a]	Hazard Ratio	Standard Error	z	p> z	95% Confidence Interval
rs440279	0.395	0.166	-2.20	0.027	0.173-0.902
rs3763120	2.155	0.687	2.41	0.016	1.154-4.026
rs4912610	1.966	0.633	2.10	0.036	1.046-3.70

[a]refSNP ID: Reference SNP Cluster Identification Number.

All the survival analysis are adjusted with age at diagnosis, ER Status, TNM stage and receipt of chemotherapy and are the Caucasian subpopulation of this cohort.

In summary of the pilot studies, three of the *ARAP3* SNPs showed association with survival. However, two out of the three (rs440279 and rs 3763120) *ARAP3* SNPs screened showed significant association with breast cancer survival in both pilot cohorts and and were selected for further epidemiology analyses in larger independent cohorts. A few selected *ARAP3* gene SNPs were associated with various clinical features, including hormonal status and lymph node metastasis status in one of the pilot cohorts.

The locations of *ARAP3* candidate SNPs rs440279 and rs3763120 is included to illustrate their relevant genomic location to each other and the host gene.

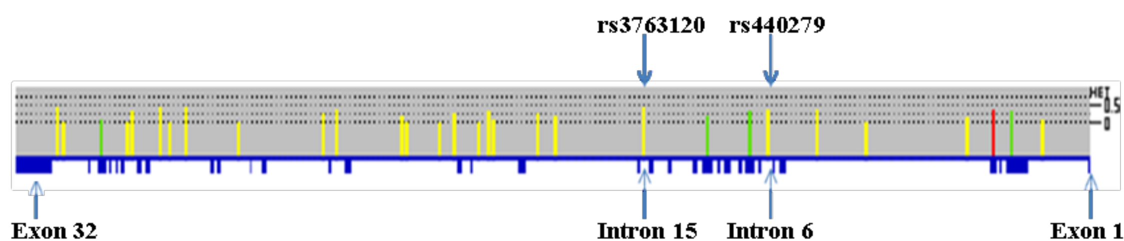


Figure 3.3: Locations of *ARAP3* SNPs. Single nucleotide polymorphism rs440279 is located in intron 6 and rs3763120 is in intron 15 in the *ARAP3* gene.

Minnesota, USA Cohort (n=618)

The two SNPs rs440279 and rs3763120 from *ARAP3* were screened in the Minnesota, USA cohort (n=618), which primarily focuses on potential association with mammographic density, a major clinical marker for breast cancer. Candidate SNP rs440279 was found to be associated with mammographic density: patients with the GG genotype tend to have a higher dense area (dense area: 28.1, 95% CI: 24.3-31.8, p=0.020) (Table 3.6) in the multivariate analysis, which was adjusted for age, BMI, menopausal status, postmenopausal hormone use, age at first birth and alcohol intake. However, further permutation and other analyses (data not shown) did not reproduce this observation and therefore this finding is regarded as a preliminary indication of association between SNP rs440279 and mammographic density. No association was found in candidate SNP rs3763120.

Table 3.6: Minnesota, USA. Mammographic Density, *ARAP3* SNP rs440279.

Genotype	N	%	Dense Area	95% Confidence Interval	Ordinal p trend
CC	272	44.00	23.5	21.5-25.5	0.020
CG	269	43.53	25.9	23.9-27.9	
GG	77	12.47	28.1	24.3-31.8	

Umea, Sweden Cohort (n=801)

Candidate SNPs rs440279 and rs3763120 from the *ARAP3* gene were screened in the Umea, Sweden cohort (n=801). Various clinical characteristics including tumour size and hormonal status showed association with these SNPs, yet no breast cancer survival association was observed in this cohort. Candidate SNP rs440279 was found to be associated with tumour size and hormonal receptor status (oestrogen receptor and progesterone receptor); patients with larger tumours (>20mm) were more likely to have the GG genotype (Odds Ratio=1.99, 95% CI=1.19-3.32), the variant GG genotype was also found to be more prevalent in patients with PR negative tumours (Odds Ratio=1.94, 95% CI=1.09-3.43) and ER negative tumours (Odds Ratio=2.76, 95% CI=1.6-4.78) (Table 3.7).

Table 3.7: Umea, Sweden, *ARAP3* Candidate SNP, univariate analysis.

Clinical Markers	refSNP ID[a]	Genotype	Numbers and Frequency				Odds Ratio	95% Confidence Interval
Tumour Size (mm)			>20		<=20			
			N	%	N	%		
	rs440279	CC	94	41	257	49	1	reference
		CG	106	46	219	42	1.32	0.95-1.84
		GG	32	13	44	9	1.99	1.19-3.32
		Total	232	100	520	100		
Progesterone Receptor Status			Negative		Positive			
			N	%	N	%		
	rs440279	CC	122	44	143	50	1	reference
		CG	120	43	122	42	1.15	0.81-1.63
		GG	38	13	23	8	1.94	1.09-3.43
		Total	280	100	288	100		
Oestrogen Receptor Status			Negative		Positive			
			N	%	N	%		
	rs440279	CC	98	40	205	50	1	reference
		CG	109	45	176	43	1.3	0.92-1.82
		GG	37	15	28	7	2.76	1.6-4.78
		Total	244	100	409	100		

[a] refSNP ID: Reference SNP Cluster Identification Number.

Fred Hutchinson Cancer Research Centre, USA Cohort (n=903)

Both *ARAP3* candidate SNPs screened in this larger population showed significant association with breast cancer, in that SNP rs3763120 was found to be associated with different clinical characteristics (hormonal receptor status and tumour grade) but not breast cancer survival, SNP rs440279 was found to be associated only with breast cancer survival and with none of the other clinical markers. The G allele from candidate SNP rs3763120 was more frequently observed in patients with ER positive tumours (Odds Ratio=1.5, 95% CI= 1.0-2.2), PR negative tumours (Odds Ratio=1.5, 95% CI=1.0-2.1) and higher tumour grade (poor/undifferentiated tumours) (Odds Ratio=1.4, 95% CI= 1.0-2.0) (Table 3.8).

The univariate survival analysis is included in Table 3.22, with SNP rs440279 found to be associated with breast cancer survival, in that the G allele was associated with poor outcome (CC vs CG/GG: Odds Ratio=1.4, 95% CI= 1.0-1.9).

Table 3.8: Fred Hutchinson Cancer Research Centre, USA, *ARAP3* Candidate SNP, univariate analysis.

Clinical Markers	[a]refSNP ID	Genotype	Numbers and Frequency				Odds Ratio	95% Confidence Interval
			Positive		Negative			
			N	%	N	%		
Oestrogen Receptor Status	rs3763120	GG	199	30	37	22	1.0	reference
		AG	315	48	95	57	1.6	1.1-2.5
		AA	147	22	34	21	1.2	0.7-2.1
		Total	661	100	166	100		
		GG	199	84	37	16	1.0	reference
		AG/AA	462	89	129	11	1.5	1.0-2.2
Progesterone Receptor Status	rs3763120	GG	182	31	54	23	1.0	ref
		GA	283	48	127	55	1.5	1.0-2.2
		AA	128	21	52	22	1.4	0.9-2.1
		Total	593	100	233	100		
		GG	182	77	54	23	1.0	reference
		AG/AA	411	70	179	30	1.5	1.0-2.1
Tumour Grade			well/mod. Diff[c]		poor/undiff[d]			
			N	%	N	%		
	rs3763120	GG	130	31	83	24	1.0	reference
		AG	201	48	185	54	1.5	1.0-2.1
		AA	90	21	72	22	1.3	0.8-2.0
		Total	421	100	340	100		
		GG	130	61	83	39	1.0	reference
		AG/AA	291	53	257	47	1.4	1.0-2.0
Survival[b]	rs440279		Alive		Dead			
			N	%	N	%		
		CC	331	45	72	42	1.0	reference
		CG	314	43	85	49	1.4	1.0-1.9
		GG	83	12	15	9	0.8	0.5-1.5
		Total	728	100	172	100		
		CC	331	82	72	18	1.00	reference
		CG/GG	397	80	100	20	1.2	0.9-1.7

[a] refSNP ID: Reference SNP Cluster Identification Number.--[b]Survival is adjusted for age of diagnose and receipt of hormonal therapy and chemotherapy.--[c]well/mod. Diff.: well to moderate differentiated tumours.--[d]poor/undiff: poor to undifferentiated tumours.

The Netherlands Cohort (n=1922)

The *ARAP3* gene candidate SNPs rs440279 and rs3763120 were genotyped in all 1,922 samples in this population. The distributions of the genotype of each SNP in all 1,922 patients were obtained and compared to the expected distribution, to determine

whether the distribution of the genotypes of the candidate SNPs deviated from the expected HWE (data not shown); the distributions of the genotypes were found to not deviate from HWE. Pearson's chi-squared statistic was used to study the relationship of the variant SNP alleles with patient and tumour characteristics, with no association between the candidate SNPs and the clinical markers identified. Candidate SNP allele frequencies were subsequently examined for their associations with prognosis, using univariate and multivariate survival analysis performed on all 1,864 patients who were eligible for the analysis of prognosis.

The univariate survival analysis indicated that the grouping of genotypes CG+GG of *ARAP3* candidate SNP rs440279, implying at least one copy of the G allele, was associated with favourable disease-free survival (Hazard Ratio=0.76, 95% CI=0.63-0.91, p=0.003), metastasis-free survival (Hazard Ratio=0.74, 95% CI=0.60-0.90, p=0.004) and overall survival (Hazard Ratio=0.75, 95% CI=0.60-0.93, p=0.009) in all patients eligible for the prognosis analysis (Table 3.9).

Table 3.9: The Netherlands, *ARAP3* SNP rs440279, Prognosis univariate analysis.

Test	Genotype	All N[a]	All Patients			LN + / ER +[b]		
			HR[d]	chi 2	95% CI[e]	HR[d]	chi 2	95% CI[e]
Disease Free Survival	CC	748	1.00	-	reference	1.00	-	reference
	CG	809	1.00	0.999	0.88-1.14	0.89	0.285	0.72-1.10
	GG	285	0.75	0.005[c]	0.62-0.92	0.57	0.001[c]	0.41-0.81
	CC+CG v.s GG	1557 v.s 285	0.76	0.003[c]	0.63-0.91	0.61	0.003[c]	0.44-0.84
	CC v.s CG+GG	748 v.s 1094	0.93	0.267	0.82-1.06	0.81	0.036[c]	0.66-0.99
Metastasis Free Survival	CC	748	1.00	-	reference	1.00	-	reference
	CG	809	0.98	0.757	0.84-1.13	0.89	0.314	0.71-1.12
	GG	285	0.73	0.004[c]	0.58-0.91	0.61	0.008[c]	0.43-0.88
	CC+CG v.s GG	1557 v.s 285	0.74	0.004[c]	0.60-0.90	0.65	0.015[c]	0.46-0.92
	CC v.s CG+GG	748 v.s 1094	0.91	0.172	0.79-1.04	0.82	0.067	0.66-1.02
Overall Survival	CC	748	1.00	-	reference	1.00	-	reference
	CG	809	0.99	0.922	0.85-1.16	0.93	0.578	0.73-1.19
	GG	285	0.75	0.013[c]	0.59-0.94	0.68	0.058	0.46-1.01
	CC+CG v.s GG	1557 v.s 285	0.75	0.009[c]	0.60-0.93	0.71	0.070	0.49-1.03
	CC v.s CG+GG	748 v.s 1094	0.92	0.294	0.80-1.07	0.87	0.245	0.69-1.10

[a]All N: All patients eligible for prognosis analysis, however, due to missing cases, the total number may not always add up to 1864.--[b] LN + / ER +: Lymph node metastasis positive and oestrogen receptor positive.--[c] Significant result as $p<0.05$.--[d]HR: Hazard ratio.--[e]95% CI: 95% confidence interval.

When this patient population was stratified by ER status and lymph node metastasis status, the variant G allele was found to be associated with positive ER and positive lymph node metastasis in all three types of survival classification (disease-free survival, metastasis-free survival and overall survival).

The analysis stratified for ER status because the assumption of proportionality was violated for ER. We didn't correct all patients for adjuvant therapy as only a small

subset of patients received treatment and we were most interested in the natural progression of breast cancer progression.

The association of the variant G allele of SNP rs440279 with better survival was reproduced in the multivariate analysis; this allele was associated with favourable survival in all three end-points (disease-free survival (Hazard Ratio=0.77, 95% CI=0.64-0.93, p=0.007), metastasis-free survival (Hazard Ratio=0.77, 95% CI=0.62-0.95, p=0.016) and overall survival (Hazard Ratio=0.77, 95% CI=0.62-0.97, p=0.026)) in all patients (Table 3.10).

Table 3.10: The Netherlands, *ARAP3* SNP rs440279, Prognosis multivariate analysis.

Test	Genotype	All N[a]	All Patients			LN + / ER +[b]		
			HR	chi 2	95% C I	HR	chi 2	95% C I
Disease Free Survival	CC	748	1.00	-	reference	1.00	-	reference
	CG	809	0.99	0.840	0.86-1.13	0.93	0.492	0.74-1.15
	GG	285	0.76	0.009[c]	0.63-0.93	0.59	0.003[c]	0.42-0.83
	CC+CG v.s GG	1557 v.s 285	0.77	0.007[c]	0.64-0.93	0.61	0.004[c]	0.44-0.85
	CC v.s CG+GG	748 v.s 1094	0.93	0.234	0.81-1.05	0.83	0.080	0.67-1.02
Metastasis Free Survival	CC	748	1.00	-	reference	1.00	-	reference
	CG	809	0.98	0.761	0.84-1.14	0.92	0.476	0.72-1.16
	GG	285	0.76	0.018[c]	0.61-0.95	0.63	0.013[c]	0.44-0.91
	CC+CG v.s GG	1557 v.s 285	0.77	0.016[c]	0.62-0.95	0.66	0.018[c]	0.46-0.93
	CC v.s CG+GG	748 v.s 1094	0.92	0.243	0.80-1.06	0.84	0.119	0.67-1.05
Overall Survival	CC	748	1.00	-	reference	1.00	-	reference
	CG	809	1.00	0.967	0.85-1.19	0.93	0.592	0.72-1.21
	GG	285	0.78	0.037[c]	0.61-0.99	0.74	0.145	0.50-1.11
	CC+CG v.s GG	1557 v.s 285	0.77	0.026[c]	0.62-0.97	0.77	0.175	0.53-1.12
	CC v.s CG+GG	748 v.s 1094	0.94	0.432	0.81-1.10	0.88	0.321	0.69-1.13

[a]All N: All patients eligible for prognosis analysis, however, due to missing cases, the total number may not always add up to 1864.--[b] LN + / ER +: Lymph node metastasis positive and oestrogen receptor positive.--[c] Significant result as $p<0.05$.--[d]HR: Hazard ratio.--[e]95% CI: 95% confidence interval.

The univariate results were reproduced in the multivariate analysis; the G allele of SNP rs440279 was associated with favourable disease-free survival and metastasis-free survival, yet not in the overall survival, in the positive lymph node metastasis subgroup and in analysis of positive lymph node metastasis patients with positive oestrogen receptor (Table 3.10) The association of the G allele and favourable survival rate remained true after multivariate analysis adjusted for adjuvant therapy (Table 3.11, 3.12).

Table 3.11: The Netherlands, *ARAP3* SNP rs440279, Prognosis multivariate analysis Correct for Adjuvant Therapy.

Test	Genotype	LN + Corr. Therapy[b]			LN + / ER + Corr. Therapy[c]		
		HR	chi 2	95% C I	HR	chi 2	95% C I
Disease Free Survival	CC	1.00	-	reference	1.00	-	reference
	CG	0.99	0.918	0.82-1.19	0.94	0.583	0.75-1.17
	GG	0.69	0.010[d]	0.51-0.92	0.61	0.005[d]	0.43-0.86
	CC+CG v.s GG	0.69	0.007[d]	0.52-0.91	0.63	0.006[d]	0.45-0.87
	CC v.s CG+GG	0.91	0.279	0.76-1.08	0.85	0.117	0.69-1.04
Metastasis Free Survival	CC	1.00	-	reference	1.00	-	reference
	CG	0.95	0.618	0.78-1.16	0.93	0.525	0.73-1.18
	GG	0.74	0.049[d]	0.54-1.00	0.66	0.025[d]	0.46-0.95
	CC+CG v.s GG	0.76	0.057	0.57-1.01	0.68	0.032[d]	0.48-0.97
	CC v.s CG+GG	0.89	0.246	0.74-1.08	0.85	0.160	0.68-1.07
Overall Survival	CC	1.00	-	reference	1.00	-	reference
	CG	0.94	0.615	0.76-1.17	0.95	0.676	0.73-1.23
	GG	0.74	0.074	0.54-1.03	0.77	0.193	0.51-1.14
	CC+CG v.s GG	0.76	0.086	0.56-1.04	0.79	0.218	0.54-1.15
	CC v.s CG+GG	0.89	0.274	0.73-1.09	0.90	0.402	0.70-1.10

[a]All N: All patients eligible for prognosis analysis, however, due to missing cases, the total number may not always add up to 1864.--[b]LN + Corr. Therapy:Lymph node metastasis positive correct for adjuvant therapy.--[c]LN + / ER + Corr. Therapy: Lymph node metastasis positive and oestrogen receptor positive correct for adjuvant therapy.--[d]Significant result as $p<0.05$.

In relation to therapy response and survival, computations were performed with the STATA statistical package, release 10.0 (STATA Corp, College Station, TX). All *p*-value s were two-sided, with $P < 0.05$ considered statistically significant. The total number of patients who participated in the Tamoxifen analysis was not available; therefore such data was not included in this result section. The variant G allele of SNP rs440279 was found to be associated with better survival in both the univariate (Hazard Ratio=0.37, 95% CI=0.18-0.74, p=0.005) and multivariate analysis (Hazard Ratio=0.34, 95% CI=0.16-0.74, p=0.006) (Table 3.12). In this sub-group, progression free survival was the major endpoint for the study of efficacy of Tamoxifen therapy in recurrent patients; *ARAP3* candidate SNP rs440279 variant G allele was found to be

associated with poor progression free survival in the multivariate analysis (Hazard Ratio=1.44, 95% CI=1.00-2.08, p=0.047) (Table 3.12).

Table 3.12: The Netherlands, *ARAP3* SNP rs440279, Tamoxifen Analysis.

Test	Genotype	Univariate Analysis			Multivariate Analysis		
		HR	chi 2	95% C I	HR	chi 2	95% C I
Response to Tamoxifen	CC	1.00	-	reference	1.00	-	reference
	CG	1.31	0.303	0.78-2.21	1.20	0.514	0.69-2.10
	GG	0.42	0.020[a]	0.20-0.87	0.38	0.017[a]	0.17-0.84
	CC+CG v.s GG	0.37	0.005[a]	0.18-0.74	0.34	0.006[a]	0.16-0.74
	CC v.s CG+GG	1.00	0.994	0.62-1.61	0.92	0.747	0.55-1.53
Progression Free Survival	CC	1.00	-	reference	1.00	-	reference
	CG	0.86	0.217	0.67-1.10	0.88	0.315	0.69-1.13
	GG	1.32	0.148	0.91-1.92	1.36	0.120	0.92-1.99
	CC+CG v.s GG	1.43	0.053	1.00-2.04	1.44	0.047[a]	1.00-2.08
	CC v.s CG+GG	0.93	0.526	0.74-1.17	0.95	0.688	0.75-1.21

[a]Significant result as p<0.05.

The distributions of SNPs rs440279 and rs3763120 in patients who received first line chemotherapy of this cohort were obtained and compared to the expected distribution, to determine if there is an association between genotype and various clinical markers. The detail of the percentages of each genotype for this group of patients and the Pearson chi-sqaure *p*-values can be seen in Table 3.13. None of the *p*-values are under the 0.05 indicating that there is no association.

Table 3.13: The Netherlands, *ARAP3* Candidate SNPs Genotype Frequency, Patients received first line chemotherapy-univariate analysis.

[a]refSNP ID			rs440279				rs3763120			
Clinical Markers		Expected (%)	Test (%)			p	Test (%)			p
			CC	CG	GG		GG	AG	AA	
Response to first line therapy	No Response	45.1	50.5	43.3	33.3	0.23	39.5	44.3	56.8	0.18
	Response	54.9	49.5	56.7	66.7		60.5	55.7	43.2	
Age in classes (years)	≤40	17.3	13.1	17.5	30.0	0.32	22.4	17.0	9.1	0.36
	41-55	49.1	50.5	49.5	43.3		50.0	47.2	52.3	
	>55	33.6	36.4	33.0	26.7		27.6	35.8	38.6	
Menopausal status	pre	49.1	46.5	49.5	56.7	0.62	53.3	46.2	45.5	0.42
	post	50.9	53.5	50.5	43.3		44.7	53.8	54.5	
Dominant site of relapse	LRR[b]	13.7	10.1	16.5	16.7	0.73	13.2	16.0	9.1	0.64
	Bone	15.5	16.2	15.5	13.3		18.4	12.3	18.2	
	Other	7.08	73.7	68.0	70.0		68.4	71.7	72.7	
Disease free survival	0-12 months	36.7	38.4	38.1	26.7	0.15	36.8	36.8	36.4	0.42
	12-36 months	46.0	42.4	50.5	43.3		46.1	48.1	40.9	
	>36 months	17.3	19.2	11.3	30.0		17.1	15.1	22.7	
Oestrogen receptor status	Negative	55.8	55.6	55.7	53.3	0.96	61.8	51.9	52.3	0.42
	Positive	44.2	43.4	43.3	46.7		38.2	47.2	45.5	
Total number of samples included		226	99	97	30		76	106	44	

[a] refSNP ID: Reference SNP Cluster Identification Number.--[b]LRR: local regional relapse.

Univariate analysis showed significant association of the variant G allele of SNP rs440279 and poor first line chemotherapy response in ER positive patients (Hazard ratio=2.94, 95% CI=1.27-6.96, p=0.014), yet this result wasn't reproduced in the multivariate analysis (full table of the negative results of the univariate and multivariate analyses not included).

In summary, the association study of the pilot cohorts showed that two SNPs rs440279 and rs3763120 from *ARAP3* were consistently associated with breast cancer survival. These two SNPs were screened in six larger independent populations to verify the findings obtained from the pilot cohort. As outlined below, the SNPs from *ARAP3* showed association with mammographic density in the Minnesota, USA cohort (n=618), various clinical markers and breast cancer survival in the Umea,

Sweden (n=801), Fred Hutchinson Cancer Research Centre, USA (n=903) and the Netherlands (n=1922) population. No significant association was identified in the Germany cohort (bone marrow metastasis) and the Australian population (breast cancer incidence). A table summarize the results is included.

Table 3.14: Summary of *ARAP3* SNPs Genotyping Results.

Gene	SNP	Genotype	Trait	p-value	population
ARAP3	rs440279	C allele	Increased non-localized tumour	0.0436	Southern California
		G allele	Increased 5 yr survival	0.0562	Southern California
		G allele	Increased 10 yr survival	0.0225	Southern California
		-	Increased survival	0.027	Greater Baltimore
		GG	Higher mammographic dense area	0.02	Minnesota
		GG	Larger tumour size	-	Sweden
		GG	Higher in ER negative tumours	-	Sweden
		GG	Higher in PR negative tumours	-	Sweden
		G allele	Decreased survival	-	Fred Hutchinson
		G allele	Increased DFS in all patients	0.003	Netherlands
		G allele	Increased MFS in all patients	0.004	Netherlands
		G allele	Increased OS in all patients	0.009	Netherlands
		G allele	Increased DFS in ER+/LN+	0.007	Netherlands
		G allele	Increased MFS in ER+/LN+	0.016	Netherlands
		G allele	Increased OS in all ER+/LN+	0.026	Netherlands
		G allele	Decreased PFS in Tamoxifen treated	0.047	Netherlands
	rs3763120	G allele	Decreased 5 yr survival	0.0045	Southern California
		G allele	Decreased 10 yr survival	0.0142	Southern California
		AA	Increased tumour size	0.011	Greater Baltimore
		-	Decreased survival	0.016	Greater Baltimore
		G allele	Higher in ER positive tumours	-	Fred Hutchinson
		G allele	Higher in PR negative tumours	-	Fred Hutchinson
		G allele	Higher grade	-	Fred Hutchinson
	rs1031904	G allele	Higher in ER negative tumours	0.0171	Southern California
	rs17208495	T allele	Higher in node positive patients	0.032	Southern California
	rs6891143	G allele	Decreased 10 yr survival	0.027	Southern California
	rs4912610	-	Decreased survival	0.036	Greater Baltimore

3.3 *SIPA1* Gene Results

A tumour suppressor function of *SIPA1* in leukaemia was first proposed based on the observation that restoring *SIPA1* expression in the *SIPA1*-deficient leukemic blast cell line resulted in dissolution of Rap1GTP accumulation and concomitant loss of leukemogencity *in vivo* (Ishida *et al*., 2003). The product of this gene is a mitogen-inducible GTPase activating protein (GAP) for members of the Ras-related proteins; RAP1 and RAP2, but not for RAS, RHO, CDC42, RAC and RAN, with comparable specific activity to the RAP1GAP encoded protein (Kurachi *et al*., 1997). The SIPA1 protein can severely hamper the mitogen-induced cell cycle progression when abnormally and/or prematurely expressed (Hattori *et al*., 1995); it participates in cell adhesion via interacting with Rap1GTP activities. Up-regulation of *SIPA1* suppressed Rap1 activity and reduced HeLa cell adhesion on a fibronectin coated surface; this suggests that *SIPA1* negatively regulates Rap1, which is required for cell adhesion induced by both extracellular matrix and soluble factors (Ishida *et al*., 2003).

Two *SIPA1* protein interacting partners were identified; Aquaporin-2 (*AQP2*) protein and bromodomain containing 4 (*BRD4*) proteins. Signal-induced proliferation-associated 1 protein interacts with the *AQP2* protein via its PDZ domain, *AQP2*'s positioning is vital in regulating body water homeostasis; trafficking of *AQP1* to the apical membrane is critical for the re-absorption of water in renal collecting ducts. The GAP activity of *SIPA1* is required for APQ2 trafficking to the apical membrane (Noda *et al*., 2005) and a nonsynonymous amino acid polymorphism in the PDZ domain of the mouse *SIPA1* gene was reported to affect the *SIPA1* Rap-GAP function and influences the binding efficiency to *AQP2* (Park *et al*., 2005).

Signal-induced proliferation-associated 1 binding protein *BRD4* binds to chromatin and regulates cell cycle progression at multiple stages. The partnership of the two proteins was identified via mass spectrometry and immune-purification. *SIPA1* and *BRD4* were found interacting with each other in the nucleus of living cells via Bifluorescence complementation. *SIPA1* was found localized in the cytoplasm and to a lesser degree in the nucleus while *BRD4* localized solely in the nucleus, while *BRD4* enhances the Rap GAP activity of *SIPA1* both *in vitro* and *in vivo*. Up-regulating *SIPA1* and *BRD4* indicated redirected sub-cellular localization of the partner protein and disrupted normal cell cycle progression. The re-localization can be reversed by injecting a proper amount of *SIPA1* or *BRD4* and together this suggested that a proper balance between *SIPA1* and *BRD4* expression in the G2 phase is required for proper cell division (Farina *et al*., 2004).

Mouse *Sipa1* was established as a candidate for underlying the breast cancer metastasis efficiency modifier locus *Mtes1*, which was recognized as a genetic region that substantially influences the metastatic efficiency of mammary tumours of the mouse (Park *et al*., 2005). The mouse *Mtes1* locus is orthologous to human chromosome 11q12-11q13, which is known to harbour the well-characterised metastasis suppressor gene *BRMS1*. Utilizing the Multiple Cross Mapping strategy, mouse *Sipa1* was identified as a potential candidate for the *Mtes 1* locus and molecular research into *Sipa1* revealed that cellular *SIPA1* level was correlated with cellular metastatic capacity. Since previous studies demonstrated that *SIPA1* over-expression reduces cell attachment, the Hunter lab demonstrated that *SIPA1* suppression (via RNAi) induces increased cell adhesion (Park *et al*., 2005). Moreover, *SIPA1* is reported to modulate the expression of two well-characterized metastasis-

associated genes *Kai1* (metastasis suppressor) and *Tfpi* (reduction of metastasis in *in vivo* experiments). An inverse correlation between the level of *SIPA1* and of both *Kai1* and *Tfpi* suggests *SIPA1* may be an upstream regulator of these two metastasis inhibitor genes (Park *et al*., 2005). *SIPA1* is further recognized as a metastasis modulator gene as an ectopically expressing *SIPA1* cell line developed about two fold more surface pulmonary metastases compared to the control cell line (p=0.0004) (Park *et al*., 2005).

The *SIPA1* SNPs were only genotyped in one of the pilot cohorts and showed associations with major clinical parameters. *Sipa1* is one of the few breast cancer metastasis promoters that have been identified. Increasing the protein expression of Sipa1 in a mouse mammary carcinoma model dramatically increased the rate of lung metastasis. Conversely, knocking down expression of Sipa1 by shRNA, ablated the ability of highly metastatic mouse mammary carcinoma cell line to metastasize to the lung. Given the high correlation of the expression of *SIPA1* to breast cancer metastasis potency and the association of *SIPA1* SNPs in clinical outcome in the pilot cohort study, SNPs in the *SIPA1* gene were selected for additional epidemiology studies, aiming to translate the metastasis modulating role of *Sipa1* found in the mouse model into the human population and further support *SIPA1* as novel breast cancer metastasis modulating gene.

The three *SIPA1* SNPs (rs931127, rs3741378 and rs746429) are in the same pairwise linkage disequlibrium block and were selected based on their genomic location (from the NCBI SNP database), which are within the regulatory or coding regions of *SIPA1* to characterise this gene; SNP rs931127 is located in 5'-untranslated region of *SIPA1*

and is considered to be within the promoter region of the gene, SNP rs3741378 and rs746429 are located within coding regions, one within exon 1 (545C>T [F182S]; rs3741378) and one within exon 12 (2760G>A [A920A]; rs746429). Univariate analysis in the Southern California, USA pilot cohort revealed that rs931127 and rs746429 were associated with lymph node metastasis (p=0.0139, p=0.0062 respectively) and rs3741378 was associated with hormonal status (ER status, p=0.006; PR status, p=0.035) (Crawford *et al.*, 2006).

The location of the candidate SNPs in the *SIPA1* gene that was analysed for association and their relevant position in the gene and in relation to each other are included for reference (Figure 3.4).

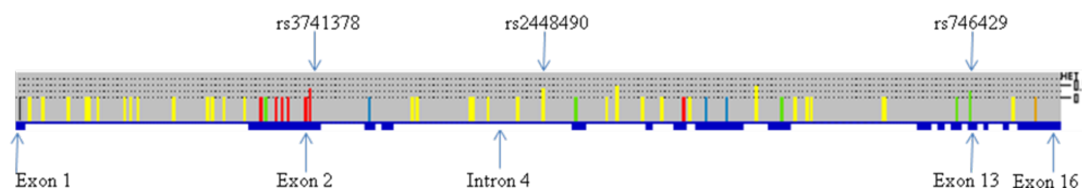


Figure 3.4: Locations of *SIPA1* SNPs. Two of the *SIPA1* gene candidate SNPs are exonic, one intronic and one 5' near gene, which is not shown in this figure.

Fred Hutchinson Cancer Research Centre, USA Cohort (n=903)

Signal-induced proliferation-associated 1 candidate SNP rs746429 was found to be associated with breast cancer survival in the Fred Hutchinson Cancer Research Centre, USA population; the AA genotype was associated with better outcome in this cohort (95% odds ratio=1.7, CI=1.0-2.7) (Table 3.15). No other candidate SNPs showed association with either clinical markers or patient's survival in this studied population.

Table 3.15: Fred Hutchinson Cancer Research Centre, USA, *SIPA1* Candidate SNP, multivariate survival analysis.

Clinical Markers	[a]refSNP ID	Genotype	Numbers and Frequency				Odds Ratio	95% Conf. Interval
			Alive		Dead			
			N	%	N	%		
Survival[b]	rs746429	GG	314	43	69	40	1.0	reference
		AG	335	46	77	45	1.0	0.7-1.5
		AA	76	11	25	15	1.7	1.0-2.7
		Total	725	100	171	100		
		GG	314	82	69	18	1.0	reference
		AG/AA	411	80	102	20	1.2	0.8-1.6

[a] refSNP ID: Reference SNP Cluster Identification Number.--[b]Survival analysis is adjusted for age at diagnosis, hormonal therapy and chemotherapy receipt.

The Netherlands Cohort (n=1922)

The variant allele frequencies for the *SIPA1* candidate SNP in all 1,922 breast cancer cases included in the study (Table 3.16) were determined; 47.5% for the rs971127 G allele, 13% for rs3741378 T allele, 38% for rs2448490 A allele and 33.5% for rs746429 A allele. None of the SNP, except rs3741378 in the PR status (p=0.005), showed association with the clinical markers. Association of the candidate SNPs with patient menopausal status and hormonal status were observed. The TC genotype from SNP rs3741378 was more prevalent in patients with progesterone receptor negative breast tumours (Table 3.16).

A significant association of the *SIPA1* SNP rs746429 with menopausal status was observed (*P* =0.049), with the variant alleles over-represented in premenopausal patients (Table 3.16). Candidate SNP rs2448490, which was only genotyped in this population, showed association with hormonal status; the variant A allele was under-represented in patients with negative oestrogen and/or progesterone receptor breast tumours (Table 3.16)

Table 3.16: The Netherlands, *SIPA1* Candidate SNP Genotype Frequencies.

Characteristic		n	rs931127						rs3741378						rs2448490						rs746429					
			AA		GA		GG		CC		TC		TT		GG		AG		AA		GG		GA		AA	
			N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
	Total	1922[d]	565	(30)	886	(47)	454	(24)	1428	(76)	415	(22)	45	(2)	771	(40)	836	(44)	299	(16)	831	(43)	868	(45)	213	(11)
Age (years)	≤40	246	72	(29)	107	(44)	66	(27)	182	(75)	59	(24)	2	(1)	100	(41)	101	(41)	45	(18)	107	(44)	119	(49)	19	(8)
	41-55	750	224	(30)	358	(48)	162	(22)	566	(77)	152	(21)	17	(2)	310	(42)	328	(44)	103	(14)	311	(42)	347	(47)	86	(12)
	56-70	621	180	(29)	280	(46)	152	(25)	459	(76)	134	(22)	14	(2)	237	(39)	279	(45)	99	(16)	268	(43)	271	(44)	79	(13)
	>70	305	89	(29)	141	(46)	74	(24)	221	(73)	70	(23)	12	(4)	124	(41)	128	(42)	52	(17)	145	(48)	131	(43)	29	(10)
			P = 0.730						*P = 0.278*						*P = 0.551*						*P = 0.236*					
Menopausal status	Pre-menopausal	840	251	(30)	398	(48)	188	(22)	630	(76)	180	(22)	17	(2)	344	(41)	370	(44)	118	(14)	346	(41)	406	(49)	85	(10)
	Post-menopausal	1082	314	(29)	488	(46)	266	(25)	798	(75)	235	(22)	28	(3)	427	(40)	466	(43)	181	(17)	485	(45)	462	(43)	128	(12)
			P = 0.454						*P = 0.687*						*P = 0.280*						*P = 0.049*[a]					
Lymph nodes involved	0	1101	321	(30)	501	(46)	262	(24)	811	(76)	237	(22)	24	(2)	441	(41)	462	(42)	185	(17)	481	(44)	477	(44)	133	(12)
	1-3	353	98	(28)	172	(49)	83	(24)	266	(76)	76	(22)	10	(3)	135	(38)	166	(47)	51	(14)	148	(42)	175	(50)	30	(9)
	>3	451	142	(31)	202	(45)	107	(24)	335	(75)	101	(23)	11	(2)	188	(42)	199	(44)	62	(14)	198	(44)	206	(46)	47	(10)
			P = 0.797						*P = 0.972*						*P = 0.343*						*P = 0.211*					
Tumour size	pT1	695	209	(31)	314	(46)	158	(23)	519	(77)	139	(21)	19	(3)	282	(41)	296	(43)	109	(16)	298	(43)	309	(45)	82	(12)
	pT2	1009	295	(29)	469	(47)	242	(24)	746	(75)	232	(23)	20	(2)	402	(40)	450	(45)	152	(15)	442	(44)	453	(45)	110	(11)
	pT3/4	218	61	(28)	103	(47)	54	(25)	163	(77)	44	(21)	6	(3)	87	(40)	90	(42)	38	(18)	91	(42)	106	(49)	21	(10)
			P = 0.944						*P = 0.551*						*P = 0.853*						*P = 0.805*					
Grade	Poor	1044	309	(30)	473	(46)	252	(24)	777	(76)	226	(22)	25	(2)	417	(40)	455	(44)	166	(16)	460	(44)	468	(45)	110	(11)
	Good/moderate	285	89	(32)	121	(43)	71	(25)	216	(78)	54	(19)	7	(3)	114	(41)	122	(44)	44	(16)	123	(43)	122	(43)	39	(14)
			P = 0.438						*P = 0.823*						*P = 0.994*						*P = 0.514*					
ER[b] status	Positive	1347	381	(29)	629	(47)	325	(24)	1014	(77)	274	(21)	32	(2)	517	(39)	601	(45)	216	(16)	582	(43)	606	(45)	152	(11)
	Negative	481	155	(33)	213	(45)	108	(23)	341	(72)	123	(26)	10	(2)	216	(45)	189	(39)	74	(15)	210	(44)	220	(46)	48	(10)
			P = 0.253						*P = 0.064*						*P = 0.046*[a]						*P = 0.736*					
PR[c] status	Positive	1120	309	(28)	533	(48)	267	(24)	852	(78)	220	(20)	27	(2)	426	(38)	511	(46)	171	(15)	484	(43)	510	(46)	121	(11)
	Negative	600	198	(33)	260	(44)	136	(23)	418	(71)	158	(27)	12	(2)	269	(45)	230	(39)	98	(16)	257	(43)	270	(45)	68	(11)
			P = 0.059						*P = 0.005*[a]						*P = 0.008*[a]						*P = 0.936*					

[a]p<0.05.--[b]ER: oestrogen receptor.--[c]PR: progesterone receptor.--[d]Due to missing cases, the total number under each category may not always added up to 1922.

SNP allele frequencies were subsequently examined for their association with prognosis. Univariate Cox regression analysis of all 1,864 patients who were eligible for the analysis of prognosis, as well as for the clinically relevant subgroups defined by lymph node and ER status, was performed. Since these genes had been previously associated with risk of metastatic progression, MFS was considered the primary endpoint of interest. Disease-free survival and overall survival were also included to examine the association of these genes with all forms of disease relapse, as well as disease survival. Preliminary univariate analysis of *SIPA1* candidate SNPs rs931127 and rs3741378 showed no association with DFS, MFS or OS in the total patient population or in the patients stratified by lymph node and ER status and these two SNPs were therefore excluded from any further analysis. Significant associations were observed for the candidate SNPs rs746429 and rs2448490.

SIPA1 candidate SNP rs746429 was found to be associated with disease-free survival, with the AA genotype associated with poor disease-free survival in positive lymph node metastasis patients with oestrogen receptor negative breast tumours (Hazard ratio=2.13, 95% CI=1.20-3.77) (Figure 3.5)

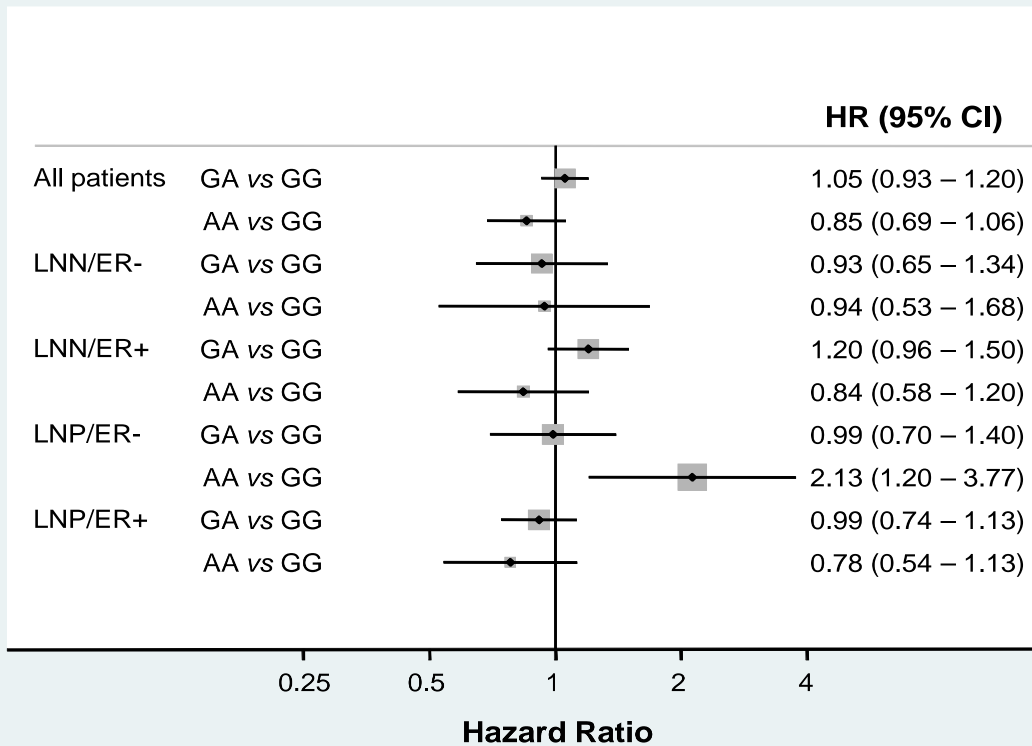


Figure 3.5: *SIPA1* SNP rs7465429 disease-free survival data. The AA genotype from *SIPA1* rs746429 is associated with poor disease-free survival in a subset of patients. LNN: Lymph node negative; LNP: Lymph node positive; ER+: oestrogen receptor positive; ER-: oestrogen receptor negative.

Most important of all, SNP rs746429 was found to be associated with metastasis-free survival; the AA genotype was also found associated with poor metastasis-free survival in a different sub-group of patients, those that have lymph node metastasis with negative oestrogen receptor breast tumours (Hazard ratio=1.96, 95% CI=1.07-3.62). The GA genotype was associated with poor metastasis-free survival in lymph node metastasis negative patients, with positive oestrogen receptor breast tumours (Hazard ratio=1.32, 95% CI=1.01-1.73) (Figure 3.6).

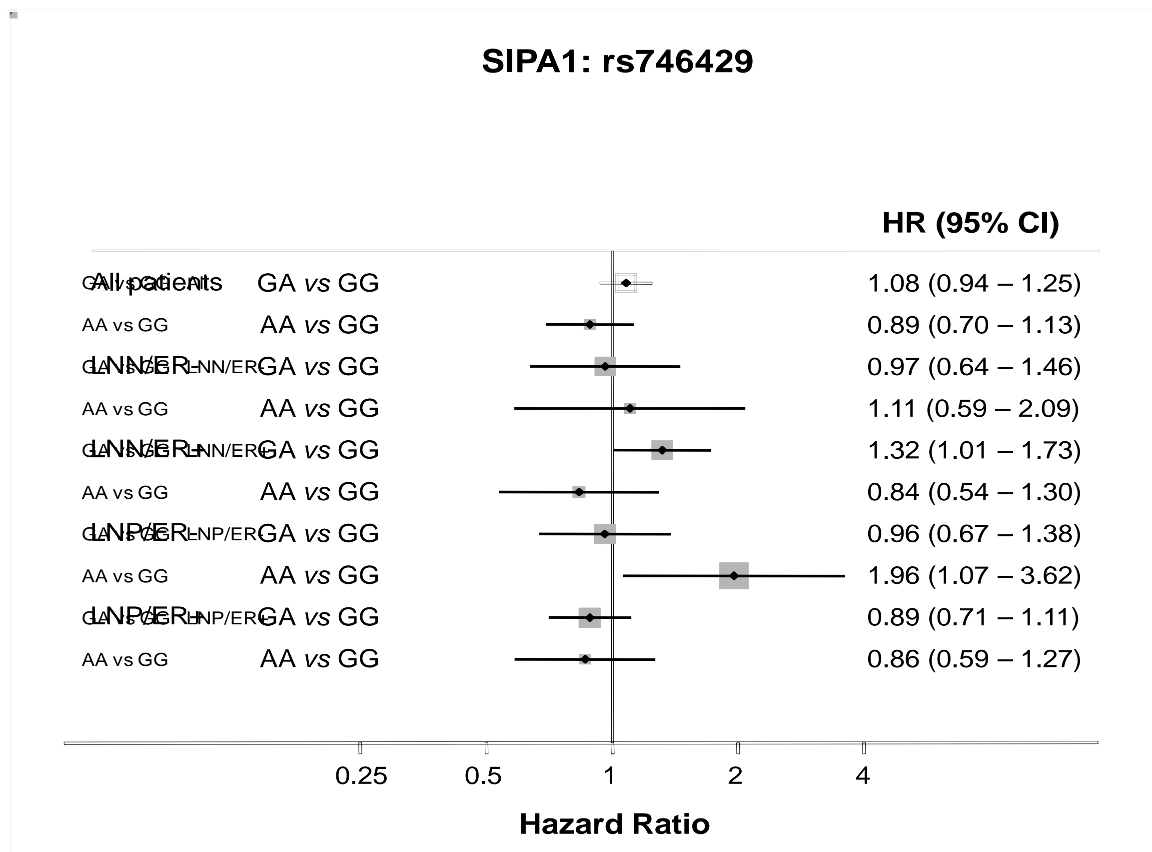


Figure 3.6: *SIPA1* SNP rs7465429 metastasis-free survival data. Candidate SNP rs746429 was found to be associated with poor metastasis-free survival in two different patient stratifications. Patients homozygous for the variant A allele had a approximately 2-fold higher risk of developing distant metastases compared with the GG genotypes. LNN: Lymph node negative; LNP: Lymph node positive; ER+: oestrogen receptor positive; ER-: oestrogen receptor negative.

SIPA 1 candidate SNP rs746429 was not only associated with poor metastasis-free survival and disease-free survival, it was also found to be associated with poor overall survival. The AA genotype from this SNP was associated with poor overall survival in patients with lymph node positive/oestrogen receptor negative breast tumours (LNP+/ER-) (Hazard ratio=2.08, 95% CI=1.13-3.84) (Figure 3.7).

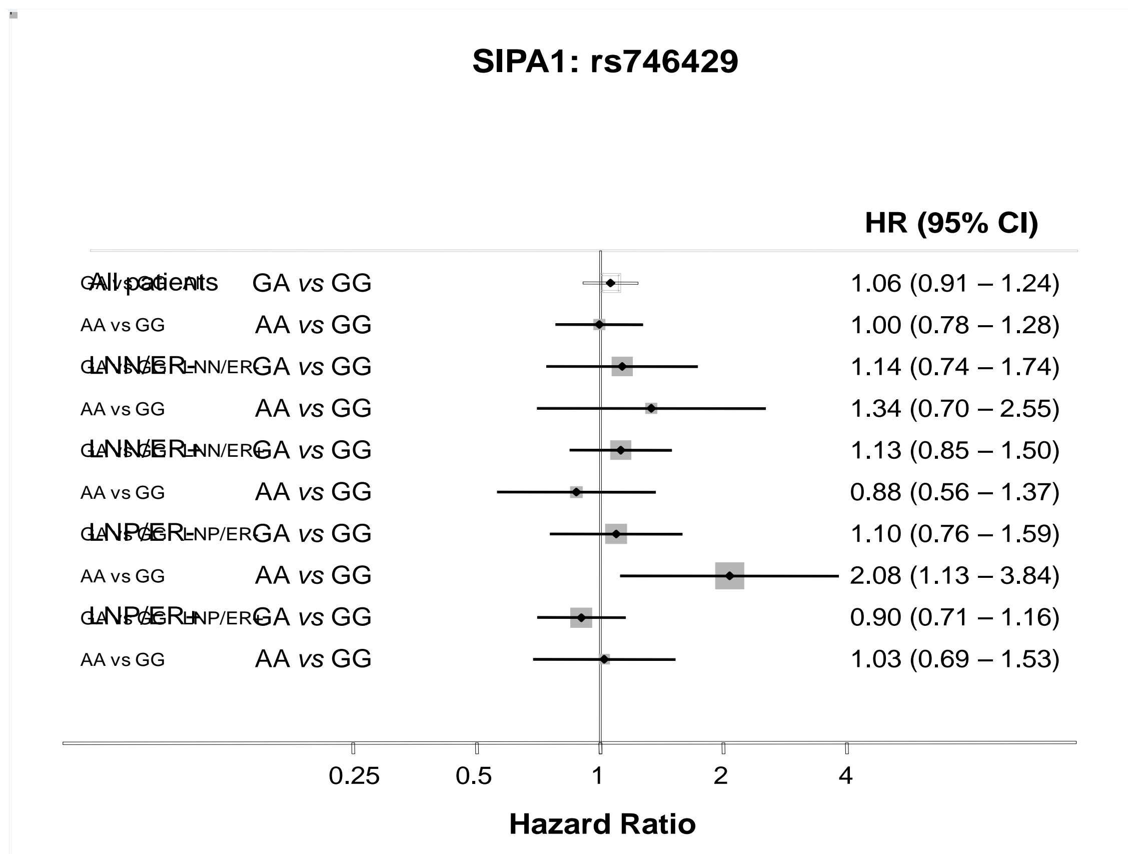


Figure 3.7: *SIPA1* SNP rs7465429 overall survival data. Poor overall survival was identified in lymph node metastasis positive patients with negative oestrogen receptor detection in the homozygous variant A allele of *SIPA1* SNP rs746429. LNN: Lymph node negative; LNP: Lymph node positive; ER+: oestrogen receptor positive; ER-: oestrogen receptor negative.

Another *SIPA1* gene candidate SNP found to be associated with outcome was SNP rs2448490, which earlier was identified as having association with hormonal status. No association was found with disease-free survival, yet significant association with metastasis-free survival was identified. The AA genotype from SNP rs2448490 was associated with better metastasis-free survival in lymph node metastasis negative patients with positive oestrogen receptor detection on tumours (Hazard ratio=0.65, 95% CI=0.43-0.97) (Figure 3.8).

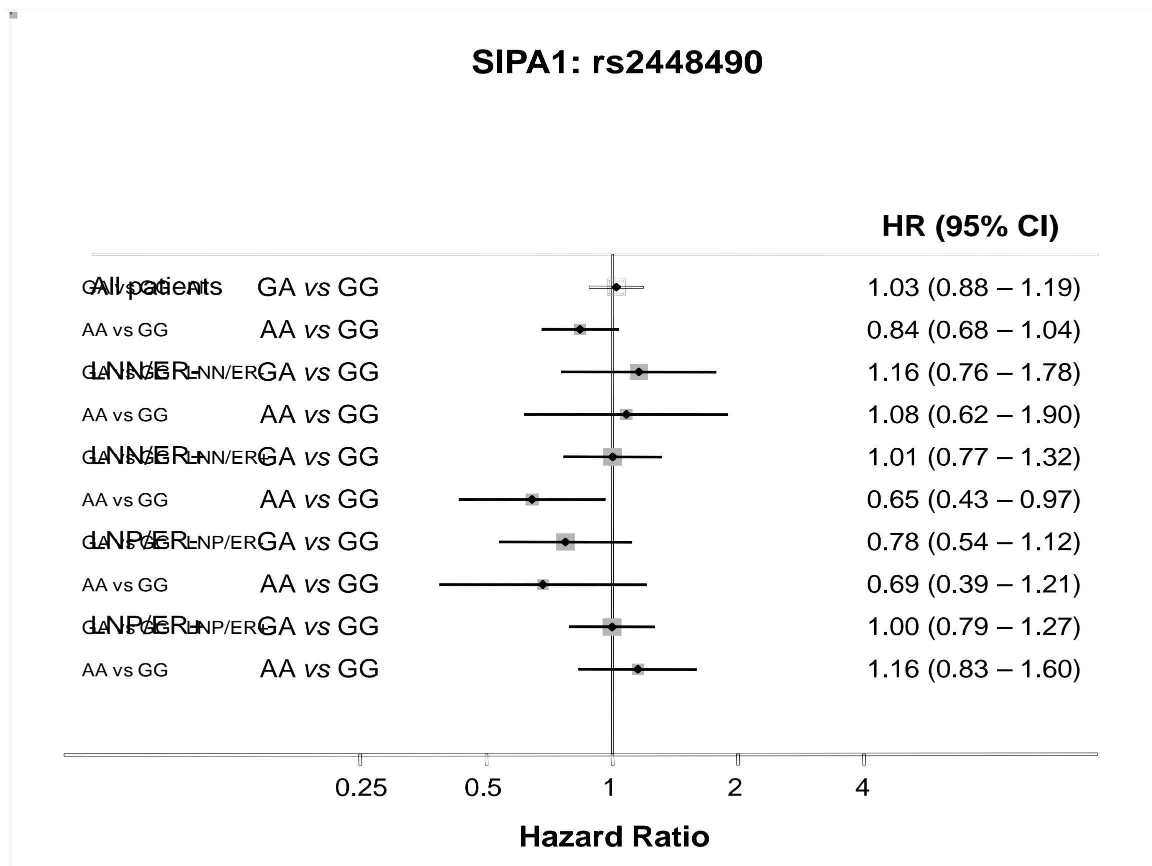


Figure 3.8: *SIPA1* SNP rs2448490 metastasis-free survival data. The AA genotype of *SIPA1* SNP rs2448490 is associated with better metastasis-free survival in the lymph node metastasis negative and oestrogen receptor positive patients. LNN: Lymph node negative; LNP: Lymph node positive; ER+: oestrogen receptor positive; ER-: oestrogen receptor negative.

A similar finding was observed in the overall survival study; patients homozygous for the A allele had a favourable overall survival in the total patient population (HR=0.79, 95% CI=0.63-0.99) and in the lymph node metastasis negative with oestrogen receptor positive patient group (HR=0.64, 95% CI=0.41-0.98) (Figure 3.9). The same trend, although not statistically significant, was observed in the analysis for disease-free survival (data not shown).

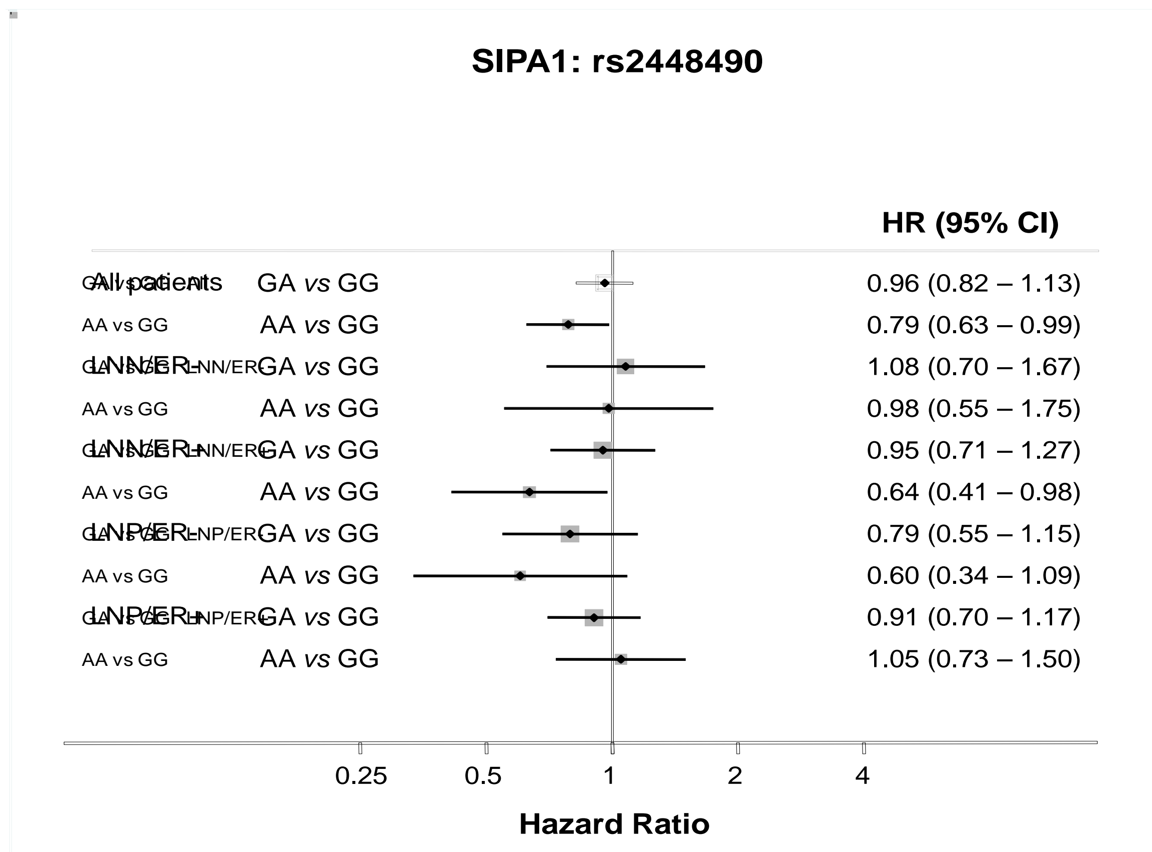


Figure 3.9: *SIPA1* SNP rs2448490 overall survival data. The variant A allele of the *SIPA1* SNP rs2448490 was found to be associated with better overall survival. LNN: Lymph node negative; LNP: Lymph node positive; ER+: oestrogen receptor positive; ER-: oestrogen receptor negative.

The variant A allele of the *SIPA1* SNP rs2448490 was found to be associated with better outcome; stratification of the patients by ER and lymph node status demonstrated that the association of the AA class of rs2448490 was significant with a favourable prognosis only in the LNN/ER+ subgroup for both metastasis-free survival (Figure 3.8) and overall survival.

The efficacy of first-line tamoxifen therapy was explored and no significant associations of any of the *SIPA1* candidate SNPs with the efficacy of first-line tamoxifen therapy in 315

oestrogen receptor positive recurrent breast cancer patients was observed, irrespective of the endpoint studied.

Patient survival after receipt of first line chemotherapy treatment was determined and association with candidate SNPs identified. The genotype frequency of each SNP in patients who received first line chemotherapy was determined. None of the *p*-values were under the 0.05 value indicating that the distribution of the genotypes in this sub-cohort is normal (Table 3.17).

Table 3.17: The Netherlands, *SIPA1* Candidate SNP Genotype Frequency, Patients received first line chemotherapy-univariate analysis.

[a]refSNP ID			rs931127				rs746429				rs2448490				rs3741378			
Clinical Markers		Expected (%)	Test (%)			p	Test (%)			p	Test (%)			p	Test (%)			p
			AA	GA	GG		GG	AG	AA		AA	GA	GG		CC	CT	TT	
Response to first line therapy	No Response	45.1	52.7	44.5	33.3	0.13	40.6	48.6	47.8	0.50	51.52	43.00	25.93	NA	43.29	44.44	85.71	NA
	Response	54.9	47.3	55.5	66.7		59.4	51.4	52.2		48.48	57.00	74.07		56.71	55.56	14.29	
Age in classes (years)	≤40	17.3	17.6	16.4	19.0	0.85	15.6	19.6	13.0	0.86	43.59	38.46	17.95	0.7	74.36	25.64	0.00	0.6
	41-55	49.1	47.3	52.7	42.9		49.0	49.5	47.8		43.75	46.43	9.82		73.64	23.64	2.73	
	>55	33.6	35.1	30.9	38.1		35.4	30.8	39.1		44.00	44.00	12.00		71.05	23.68	5.26	
Menopausal status	pre	49.1	44.6	50.9	52.4	0.63	51.0	51.4	30.4	0.17	42.90	42.90	14.30	NA	73.90	24.30	1.80	NA
	post	50.9	55.4	49.1	47.6		49.0	48.6	69.6		44.70	45.60	9.70		71.90	23.70	4.40	
Dominant site of relapse	LRR[b]	13.7	14.9	13.6	11.9	0.96	13.5	13.1	17.4	0.58	48.39	38.71	12.90	0.95	74.19	22.58	3.23	NA
	Bone	15.5	17.6	14.5	14.3		19.8	13.1	8.7		47.06	41.18	11.76		61.76	29.41	8.82	
	Other	7.08	67.6	71.8	73.8		66.7	73.8	73.9		42.24	45.96	11.80		75.00	23.13	1.88	
Disease free survival	0-12 months	36.7	43.2	32.7	35.7	0.25	35.4	32.7	60.9	0.15	51.22	37.80	10.98	0.36	71.08	25.30	3.61	0.36
	12-36 months	46.0	41.9	51.8	38.1		45.8	49.5	30.4		41.35	48.08	10.58		70.59	27.45	1.96	
	>36 months	17.3	14.9	15.5	26.2		18.8	17.8	8.7		35.00	47.50	17.50		82.50	12.50	5.00	
Oestrogen receptor status	Negative	55.8	60.8	50.0	59.5	0.29	54.2	57.0	52.2	0.86	57.58	53.06	59.26	NA	54.94	61.11	28.57	NA
	Positive	44.2	39.2	49.1	38.1		44.8	42.1	47.8		42.42	46.94	40.74		45.06	38.89	71.43	
Total number of samples included			74	110	42		96	107	23		136	167	57		164	54	7	

[a] refSNP ID: Reference SNP Cluster Identification Number.--[b]LRR: local regional relapse.

All four candidate SNPs were found to be associated with progression-free survival in the first line chemotherapy receipt patient subgroup. SIPA1 candidate SNPs rs746429, rs2448490 and rs3741378 were found to be associated with progression-free survival in all first line chemotherapy receipt patients; the AA genotype from SNP rs746429 was associated with poor rate of progression-free survival (odds ratio=1.40, 95% CI=1.19-2.14, p=0.002); a similar finding was found with the TT genotype of SNP rs3741378 (odds ratio=2.34, 95% CI=1.06-5.15, p=0.035). On the other hand, the GG genotype of SNP rs2448490 was found to be associated with favourable progression-free survival (odds ratio=0.62, 95% CI=0.39-0.98, p=0.039) (Table 3.18). When stratifying this sub-population by oestrogen receptor status, the TT genotype from SNP rs3741378 was shown to be associated with better progression-free survival in oestrogen receptor positive patients of this subgroup, contradicting the finding observed in the whole first line chemotherapy receipt patient subgroup; the variant A allele from SNP rs746429 was found to be associated with poor progression-free survival in oestrogen receptor negative patients (Table 3.18). The variant G allele from candidate SNP rs931127, which is a 5' gene SNP, was found to be associated with poor outcome in the oestrogen receptor negative subgroup (Table 3.18). Candidate SNP rs2448490 was found to be associated with better progression-free survival in the oestrogen receptor negative patients as well (Table 3.18).

Table 3.18: The Netherlands, *SIPA1* Candidate SNPs, Patients received first line chemotherapy-multivariate survival analysis.

Test	Group	[a]refSNP ID	Genotype	N	Odds Ratio	*p*-value	95% Confidence Interval
Progression Free Survival	All 1st line Chemotherapy Patient	rs746429	GG	96	1.0	-	reference
			AG	107	1.60	0.089	0.90-1.78
			AA	23	1.40	0.002	1.19-2.14
		rs2448490	GG	136	1.0	-	reference
			AG	167	0.80	0.115	0.60-1.06
			AA	57	0.62	0.039	0.39-0.98
		rs3741378	CC	273	1.0	-	reference
			CT	80	0.81	0.198	0.58-1.12
			TT	6	2.34	0.035	1.06-5.15
	Oestrogen Receptor Positive	rs3741378	CC	84	1.0	-	reference
			CT	24	0.80	0.329	0.52-1.25
			TT	5	0.55	0.037	0.31-0.96
	Oestrogen Receptor Negative	rs931127	AA	44	1.0	-	reference
			AG	46	2.16	0.001	1.39-3.37
			GG	21	2.06	0.047	1.01-4.20
		rs746429	GG	43	1.0	-	reference
			AG	57	2.16	0.001	1.39-3.37
			AA	11	2.06	0.047	1.01-4.20
		rs2448490	GG	54	1.0	-	reference
			AG	52	0.82	0.349	0.53-1.25
			AA	15	0.49	0.027	0.26-0.92
	1st line Anthracycline Patient	rs746429	GG	59	1.0	-	reference
			AG	77	1.50	0.028	1.04-2.16
			AA	14	1.94	0.043	1.02-3.67

[a] refSNP ID: Reference SNP Cluster Identification Number.

Survival curves were generated using the Kaplan-Meier method; a log-rank test was used to test for differences between the survival curves or when appropriate the log-rank test for trend. In the analysis for progression-free survival (PFS), in ER-negative patients, the AA allele of rs2448490 predicted a significantly (*P*=0.044) delayed disease progression, with a median time to progression of 9.0 months compared with 5.1 months for the GA genotype and 4.5 months for the GG genotype (Figure 3.10). In multivariate logistic regression analysis in ER negative patients, the AA variant allele of *SIPA1* SNP rs2448490 was the only significant factor after correcting for

age, dominant site of relapse and disease-free interval. Irrespective of the chemotherapy used, no further significant associations of any of the SNPs with the type of response to chemotherapy, PFS or post-relapse overall survival (PRS), were observed.

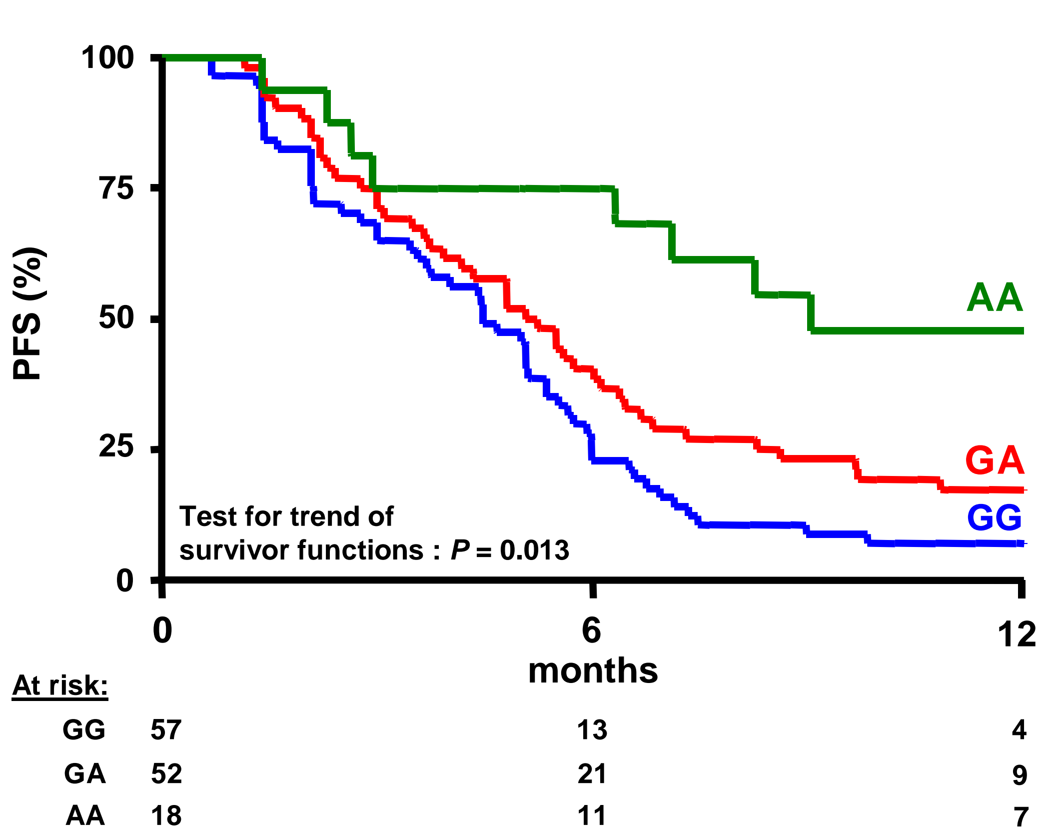


Figure 3.10: *SIPA1* rs2448490 Kaplan-Meier analysis. Kaplan-Meier analysis of progression-free survival for *SIPA1* rs2448490 in ER-negative patients treated with first-line chemotherapy. The AA genotype has better survival rate compared to the GA and GG genotypes.

Queensland, Australia Cohort (Case n=200; Control n=200)

The *SIPA1* protein has been related to increased breast cancer metastasis in the mouse model. The polymorphisms of the *SIPA1* gene have significant effects on the protein function in the mouse model and the SNPs in the human *SIPA1* gene are associated with major clinical markers. As a previous study showed that *SIPA1* RNAi down-regulation increased the primary tumour burden in the mouse model, and the *SIPA1* gene and protein are associated with poor-prognosis markers and metastasis, the possibility that the *SIPA1* gene is correlated to breast carcinoma incidence as well as prognosis was explored using a case-control incidence cohort from Australia.

The frequencies of the genotypes of SNP rs3741378 are listed in Table 3.19. Significant frequency difference was observed between the breast cancer group and the control group samples in this SNP. The TT genotype of rs3741378 was more frequently observed in the breast cancer group compared to the disease-free control group and the frequency of the TT genotype in the population was considerably much lower than the other genotypes; the TC genotype is more common in the control patients than in cancer patients, which would indicate that having both alleles is protective in some manner. Due to the extreme low counts for the TT allele, Hardy-Weinberg equilibrium analysis was performed to exclude the possibility of experimental artefacts misleading the results.

Table 3.19: Australia, *SIPA1* Candidate SNP, Breast Cancer Incidence Analysis.

[a]refSNP ID			Allele Frequency		Genotype Frequency			
	Group	Frequency	T	C	TT	TC	CC	Total
rs3741378	Breast Cancer	N	53	303	6	41	131	162
		%	15	85	3.37	23.03	73.60	-
	Control	N	57	267	1	55	106	178
		%	18	82	0.62	33.95	65.43	-

[a] refSNP ID: Reference SNP Cluster Identification Number.--Chi-Squared stat = 7.51, with Degrees of freedom = 2 and *p*-value = 0.023. Hardy Weinberg Equilibrium probability for Breast Cancer is 0.7 and Hardy Weinberg Equilibrium probability for the Control group is 0.184.

Additionally, because of the extremely low count for the TT genotype, the standard Chi-square method's assumptions are violated (it needs counts of at least 5 in all categories). Thus, a further analysis using CLUMP, a piece of software , that uses a Monte-Carlo style Chi-square test (Sham and Curtis, 1995), which depends on random numbers rather than a set probability formula was performed. The results of that analysis confirmed the original Chi-square results (Table 3.20).

Table 3.20: Australia, *SIPA1* Candidate SNP, Breast cancer Incidence Analysis, Monte-Carlo style Chi-square test.

rs3741378	T1 analysis (mimics standard Chi-Square)	T4 analysis (collapses low-count categories into the next lowest category)
Chi-Squared stat	7.514	4.988
Degrees of freedom	2	1
Probability	0.023	0.050

No other association with either clinical markers or survival were observed in other cohorts, including the German cohort (bone marrow metastasis analysis), the Minnesota, USA cohort (Mammography density analysis) and the Umea, Sweden population.

In summary, *SIPA1* SNPs showed association with breast cancer progression in multiple independent cohorts and suggest that these observations are not due to artefacts. A table summarize the significant results in survival analyses (other than Minnesota, USA cohort and Germany cohort) is included (Table).

Table 3.21: Summary of *SIPA1* SNPs Genotyping Results.

Gene	SNP	Genotype	Trait	p-value	population
SIPA1	rs931127	G allele	Higher in node positive patients	0.0139	Southern California
	rs746429	G allele	Higher in node positive patients	0.0062	Southern California
	rs3741378	C allele	Higher in ER negative tumours	0.0006	Southern California
	rs746429	C allele	Higher in PR negative tumours	0.035	Southern California
	rs746429	AA	Increased survival	-	Fred Hutchinson
	rs2448490	A allele	Decreased in ER negative tumours	0.046	Netherlands
	rs2448490	A allele	Decreased in PR negative tumours	0.008	Netherlands
	rs2448490	AA	Increased OS	-	Netherlands
	rs2448490	AA	Increased OS in LN-/ER+	-	Netherlands
	rs2448490	AA	Increased MFS in LN-/ER+	-	Netherlands
	rs2448490	GG	Increased PFS in 1st line Chemotherapy treated	0.039	Netherlands
	rs746429	AA	Decreased PFS in 1st line Chemotherapy treated	0.002	Netherlands
	rs3741378	TT	Decreased PFS in 1st line Chemotherapy treated	0.035	Netherlands
	rs3741378	TT	Increased PFS in 1st line Chemotherapy treated/ER+	0.037	Netherlands
	rs2448490	AA	Increased PFS in ER-	0.027	Netherlands
	rs746429	A allele	Decreased PFS in ER-	0.047	Netherlands
	rs746429	A allele	Decreased PFS in 1st line Anthracycline treated	0.043	Netherlands
	rs931127	G allele	Decreased PFS in ER-	0.047	Netherlands
	rs3741378	TT	Increased incidence	0.023	Australia

3.4 *RRP1B* Genes Results

Ribosomal RNA processing 1 homolog B was identified as a putative metastasis modulator in the mouse model, in that *Rrp1b* was differentially expressed in high metastatic and low metastatic potency mouse and its expression is highly correlated with extracellular matrix expression, which is a previously known marker for breast cancer metastasis. Functional assays indicated that *RRP1B* protein is a binding partner of the previously identified breast cancer metastasis gene *SIPA1. In vitro* expression of *RRP1B* was confirmed to alter extracellular matrix expression; these results taken together suggested that the *RRP1B* gene and its gene product may be a novel breast cancer metastasis modulator. Additional research revealed that *in vitro* expression of *RRP1B* reduces primary breast tumour growth and lung metastasis capacity in the mouse model; a gene expression signature was subsequently generated by indicative ectopic expression of *Rrp1b* and the signature predicts breast cancer patient survival (Crawford *et al*., 2007).

One SNP from the *RRP1B* gene (rs9306160) showed associations with survival (poor outcome) in both pilot populations (Crawford *et al*., 2007) and was selected as one of the high importance SNPs for further epidemiology analysis. The *RRP1B* gene SNP rs9306160 was analysed in both pilot cohorts and this SNP showed association with a few clinical characteristics including breast cancer stage (localized v.s non-localized, p=0.006), hormonal status (ER status (p=0.001), PR status (p=0.001)), primary tumour grade (p=0.001) and lymph node metastasis (p=0.033) in the Southern California, USA pilot cohort. In the Greater Baltimore, USA cohort, the candidate SNP from the *RRP1B* gene was found to be associated with clinical characteristics

such as progesterone receptor status (p=0.015) and lymph node positive metastasis (p=0.03). Furthermore, Kaplan-Meier survival analysis revealed that the GG genotype was associated with poor outcome (p=0.025) (Crawford *et al*., 2007).

The molecular function of RRP1B may play a critical role in breast cancer metastasis as, similar to BRD4, it is a protein binding partner of the novel breast cancer metastasis promoter SIPA1 and its expression has been shown to influence mouse lung metastasis potency in an orthotopic implantation mouse mammary carcinoma model.

The location of the candidate SNP rs9306160 in the *RRP1B* gene tested in this study is included for reference (Figure 3.11).

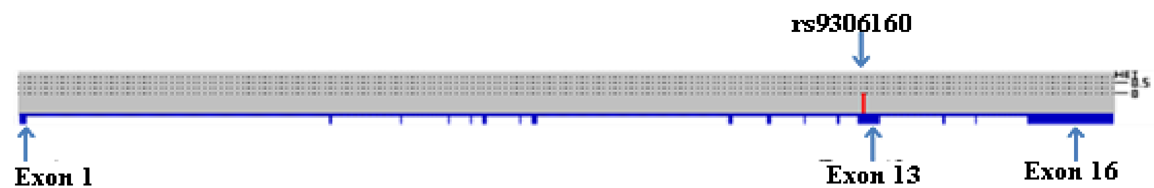


Figure 3.11: Locations of *RRP1B* SNPs. Ribosomal RNA processing 1 homolog B candidate SNP rs9306160's location in the *RRP1B* gene.

Fred Hutchinson Cancer Research Centre, USA Cohort (n=903)

Ribosomal RNA processing 1 homolog B candidate SNP rs9306160 was found to be associated with tumour size in the Fred Hutchinson Cancer Research Centre, USA (n=903) cohort. The variant A allele was associated with a smaller tumour size (<= 2 cm) (odds ratio=1.4, 955 CI=1.0-2.0) (Table 3.22).

Table 3.22: Fred Hutchinson Cancer Research Centre, USA, *RRP1B* Candidate SNP, univariate analysis.

Clinical Markers	[a]refSNP ID	Genotype	Numbers and Frequency				Odds Ratio	95% Confidence Interval
			<=2 cm		>2 cm			
			N	%	N	%		
Tumour Size	rs9306160	GG	229	39	79	31	1.00	reference
		AG	262	44	121	49	1.4	1.0-2.0
		AA	100	17	49	20	1.5	1.0-2.3
		Total	591	100	249	100		
		GG	229	74	79	26	1.00	reference
		AG/AA	362	68	170	32	1.4	1.0-2.0

[a] refSNP ID: Reference SNP Cluster Identification Number.

The Netherlands Cohort (n=1922)

The variant allele frequency for the *RRP1B* gene candidate SNP rs9306160 in all 1,922 breast cancer cases included in the study were determined (Table 3.39); the variant A allele was 42.5% and did not deviate from Hardy-Weinberg equilibrium. No *p*-value was less than 0.05 suggesting no significant or strong relationships of any of the SNP genotypes with patient's age or menopausal status, tumour size, lymph node status or hormone receptor status (Table 3.23). Tumours of the AA class of *RRP1B* SNP rs9306160 showed a non-significant trend towards smaller tumours in this cohort (P = 0.054).

Table 3.23: The Netherlands, *RRP1B* Candidate SNP Genotype Frequency. Patients received first line chemotherapy-univariate analysis.

Characteristic		Patients	*RRP1B* rs9306160						p
			GG		GA		AA		
			N	(%)	N	(%)	N	(%)	
Total		1922[a]	636	(33)	938	(49)	337	(18)	
Age (years)	≤40	246	85	(35)	127	(52)	34	(14)	0.392
	41-55	750	237	(32)	371	(50)	139	(19)	
	56-70	621	201	(33)	300	(49)	115	(19)	
	>70	305	113	(37)	140	(46)	49	(16)	
Menopausal status	Premenopausal	840	273	(33)	431	(51)	135	(16)	0.145
	Postmenopausal	1082	363	(34)	507	(47)	202	(19)	
Lymph nodes involved	0	1101	361	(33)	537	(49)	192	(18)	0.490
	1-3	353	107	(30)	183	(52)	63	(18)	
	>3	451	163	(36)	208	(46)	80	(18)	
Tumour size	pT1	695	213	(31)	345	(50)	131	(19)	0.054
	pT2	1009	343	(34)	480	(48)	182	(18)	
	pT3/4	218	80	(37)	113	(52)	24	(11)	
Grade	Poor	1044	339	(33)	529	(51)	170	(16)	0.225
	Good/moderate	285	102	(36)	123	(44)	57	(20)	
ER status	Positive	1347	446	(33)	653	(49)	240	(18)	0.812
	Negative	481	166	(35)	231	(48)	81	(17)	
PR status	Positive	1120	366	(33)	546	(49)	202	(18)	0.651
	Negative	600	206	(35)	290	(49)	99	(17)	

[a]Due to missing cases, the total number may not always add up to 1922.

The result for metastasis-free survival was however, significant in multivariate recessive model (GG v.s GA + AA) analysis, independent of the traditional prognostic factors (hazard ratio=0.83, 95% CI=0.72-0.96, p=0.013; Table 3.24), while the associations of the A allele with disease-free survival and overall survival was not significant in the multivariate analysis. Addition of adjuvant chemotherapy to the multivariate model did not affect the estimated coefficients of the *RRP1B* SNP rs9306160 genotype (data not shown).

Table 3.24: The Netherlands, *RRP1B* SNP rs9306160, multivariate analysis for Metastasis-Free Survival in all patients[a].

Characteristic		HR (95% CI)	p
Age (years)	41-50 vs. ≤40	0.88 (0.71 – 1.10)	0.257
	51-70 vs. ≤40	0.69 (0.50 – 0.97)	0.031
	>70 vs. ≤40	0.62 (0.43 – 0.89)	0.010
Menopausal status	Post vs. premenopausal	1.31 (1.02 – 1.69)	0.034
Lymph nodes involved	1-3 vs. 0	1.81 (1.50 – 2.08)	<0.001
	>3 vs. 0	2.72 (2.29 – 3.24)	<0.001
Tumor size	pT2 vs. pT1	1.56 (1.32 – 1.86)	<0.001
	pT3/4 vs. pT1	1.87 (1.46 – 2.39)	<0.001
Grade	Unknown vs. poor	0.91 (0.77 – 1.06)	0.212
	Good/moderate vs. poor	0.57 (0.44 – 0.73)	<0.001
Oestrogen Receptor	Positive vs. negative	0.79 (0.66 – 0.96)	0.019
Progesterone Receptor	Positive vs. negative	0.82 (0.69 – 0.98)	0.031
RRP1B rs9306160	GA+AA vs. GG	0.83 (0.72 – 0.96)	0.013

[a]The final model included 1740 patients

Further survival analysis revealed that patients carrying the GA (=CT) genotype of *RRP1B* SNP rs9306160 showed a significant favorable prognosis in all patients for metastasis-free survival (Hazard ratio=0.77, 95% CI=0.67-0.90) (Figure 3.12) and a borderline statistical association with better disease-free survival (Hazard ratio=0.87, 95% CI=0.76-1.00) (Figure 3.13) and overall survival (Hazard ratio=0.86, 95% CI=0.73-1.01) (Figure 3.14).

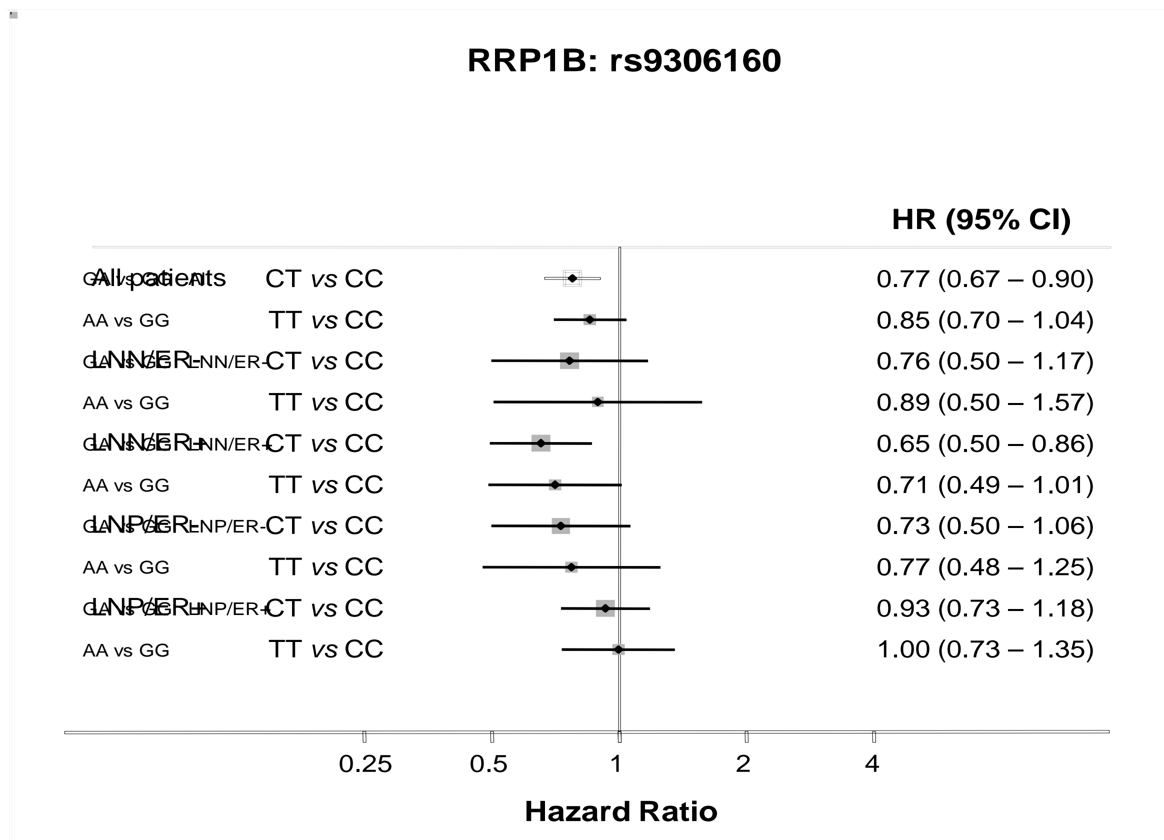


Figure 3.12: *RRP1B* SNP rs9306160 metastasis-free survival data. The CT (=GA) genotype of *RRP1B* rs9306160 is associated with better metastasis-free survival in all 1922 patients and lymph node metastasis negative with oestrogen receptor positive patients.

When stratifying the population by lymph node metastasis status and oestrogen receptor status to evaluate association in best (lymph node negative, ER-positive; LNN/ER+) and worst classes (lymph node positive, ER-; LNP/ER-), as determined by standard clinical parameters, SNP rs9306160 showed association with survival in not only metastasis-free survival but also disease-free survival and in trend for better overall survival. Association of variant allele A of *RRP1B* SNP rs9306160 with a favourable prognosis was significant only in the subgroup of LNN/ER+ patients, but not in the other patient subgroups (LNN/ER-, LNP/ER+, LNP/ER-). Lymph node metastasis negative, positive oestrogen receptor patients with the GA genotype of SNP rs9306160 had a better metastasis-free survival rate compared with the GG

RRP1B: rs9306160

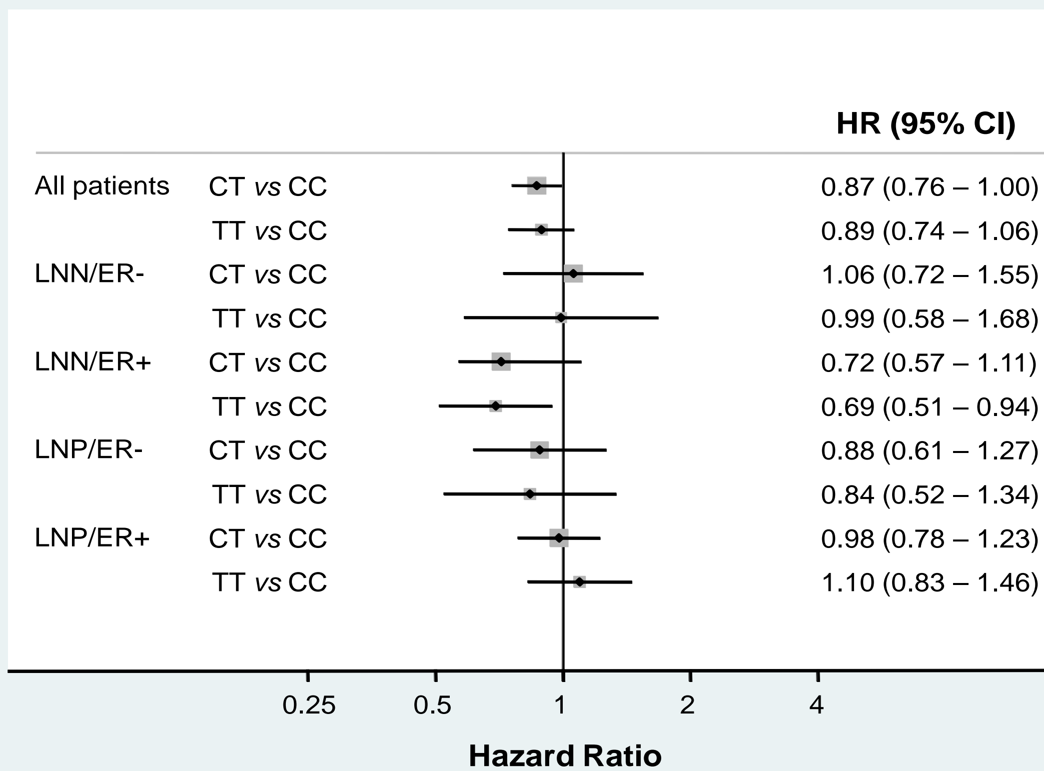


Figure 3.13: *RRP1B* SNP rs9306160 disease-free survival data. The AA (=TT) genotype showed association with better disease-free survival in patients having lymph node metastasis negative with oestrogen receptor positive patients.

Candidate SNP rs9306160 showed a trend of better survival rate in the overall survival analysis in all patients and patients stratified by lymph node status and oestrogen receptor status (Figure 3.14).

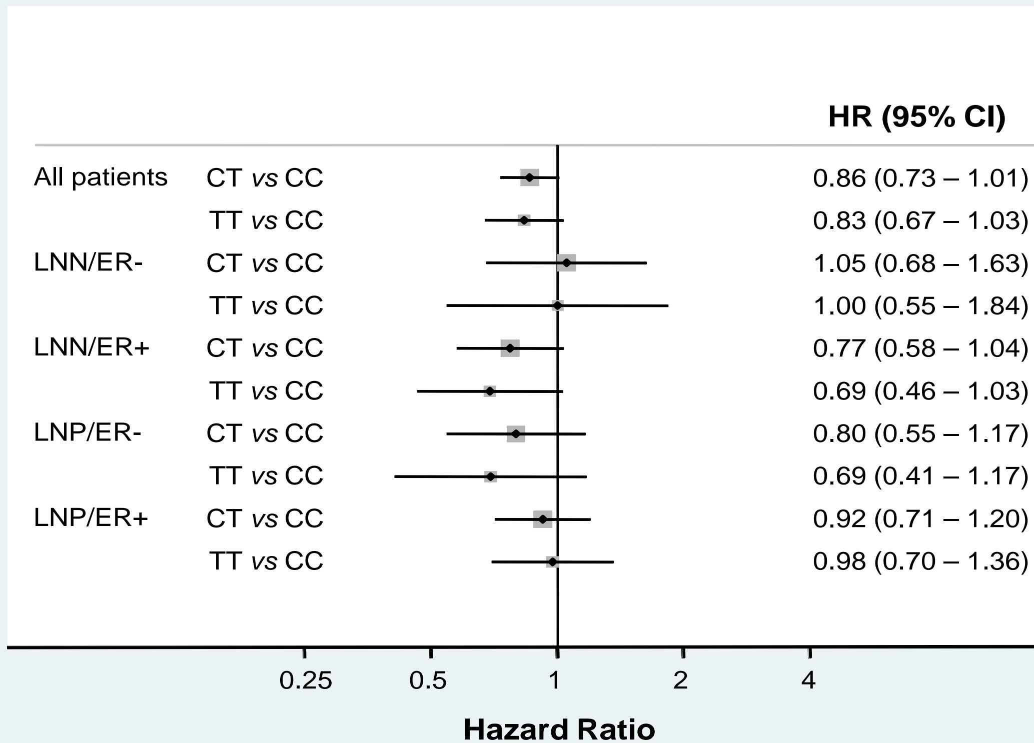


Figure 3.14: *RRP1B* SNP rs9306160 overall survival data. The CT genotype of *RRP1B* rs9306160 is in tending of better overall survival in all 1922 patients and lymph node metastasis negative with oestrogen receptor positive patients studied in the Rotterdam, the Netherlands population.

The protective association in LNN/ER+ patients remained statistically significant for all three end points after correction for the traditional prognostic factors in multivariate analysis [Disease-free survival (Hazard ratio=0.71, 95% CI=0.57–0.88, p=0.002), Metastasis-free survival (Hazard ratio=0.67, 95% CI=0.52–0.86, p=0.002) (Figure 3.15A), and Overall survival (Hazard ratio=0.75, 95% CI=0.57–0.99, p=0.040)]. Previous studies demonstrated that the *SIPA1* and *RRP1B* gene products physically interact. Since the *SIPA1* SNP rs2448490 (Figure 3.15B) and *RRP1B* SNP rs9306160 were both associated with a favorable prognosis in LNN/ER+ patients, we explored the possibility that the combination of both SNP genotypes might show

additional prognostic power. In Cox multivariate analysis, *SIPA1* SNP rs2448490 and *RRP1B* SNP rs9306160 remained independent factors predicting a prolonged MFS with Hazard ratios of 0.61 and 0.69, respectively (Table 3.25).

Table 3.25: The Netherlands, *RRP1B* SNP rs9306160 + *SIPA1* SNP rs2448490, multivariate analysis for MFS in ER+ lymph node-negative patients[a].

Characteristic		HR (95% CI)	*P*
Age (years)	41-50 vs. ≤40	0.73 (0.49 – 1.06)	0.097
	51-70 vs. ≤40	0.43 (0.28 – 0.65)	<0.001
	>70 vs. ≤40	0.43 (0.27 – 0.69)	<0.001
Tumour size	pT2 vs. pT1	1.41 (1.08 – 1.84)	0.011
	pT3/4 vs. pT1	2.00 (1.07 – 3.75)	0.030
Grade	Unknown vs. poor	1.06 (0.80 – 1.40)	0.691
	Good/moderate vs. poor	0.55 (0.37 – 0.82)	0.003
SIPA1 rs2448490	AA vs. AG+GG	0.61 (0.42 – 0.90)	0.006
RRP1B rs9306160	CT+TT vs. CC	0.69 (0.54 – 0.90)	0.013

[a]The final model included 761 patients.

Kaplan-Meier analysis for metastasis-free survival, as a function of the combined genotypes in the LNN/ER+ patients, demonstrated that the combination of the homozygous AA genotype of *SIPA1* SNP rs2448490 and at least one T allele (AA and CT+TT) of *RRP1B* SNP rs9306160 was associated with the best prognosis (Figure 3.15D). The risk for developing distant metastasis was more than 2.5-fold lower compared with that observed for the GG+GA/AA genotype (Hazard ratio=0.38, 95% CI=0.20 – 0.73, p=0.004).

SIPA1 and *RRP1B* have both been shown to play some part in breast cancer progression in several published studies (Park *et al.*, 2005; Crawford *et al.*, 2007). They have also been shown to physically interact in normal and cancerous breast cells. The SNPs chosen for analysis of these two genes represent LD blocks that encompass many different SNPs and without categorically identifying the causative

SNP that is modifying the clinical characteristics being investigated, the study could not be sure why certain SNPs may have clinical correlation. If the causative SNP/s were identified than x-ray crystallography on SIPA1 and RRP1B and how their physical interaction or mode of action could be modified by the causative SNP could be undertaken to shed light on how the SNPs investigated were modulating breast cancer progression. As neither the causative SNPs nor protein structures for SIPA1 and RRP1B have been identified, the study took the view of investigating whether the polymorphic status of either gene in combination may either be independent and/or syngeneic with regard to modifying clinical correlation, or dependent on particular alleles. Thus, the study was purely investigating polymorphic combinations from the viewpoint of modifying factors rather than truly determining whether the SNPs investigated are playing a direct role in breast cancer progression by modulating the efficacy of the SIPA1 and RRP1B complex.

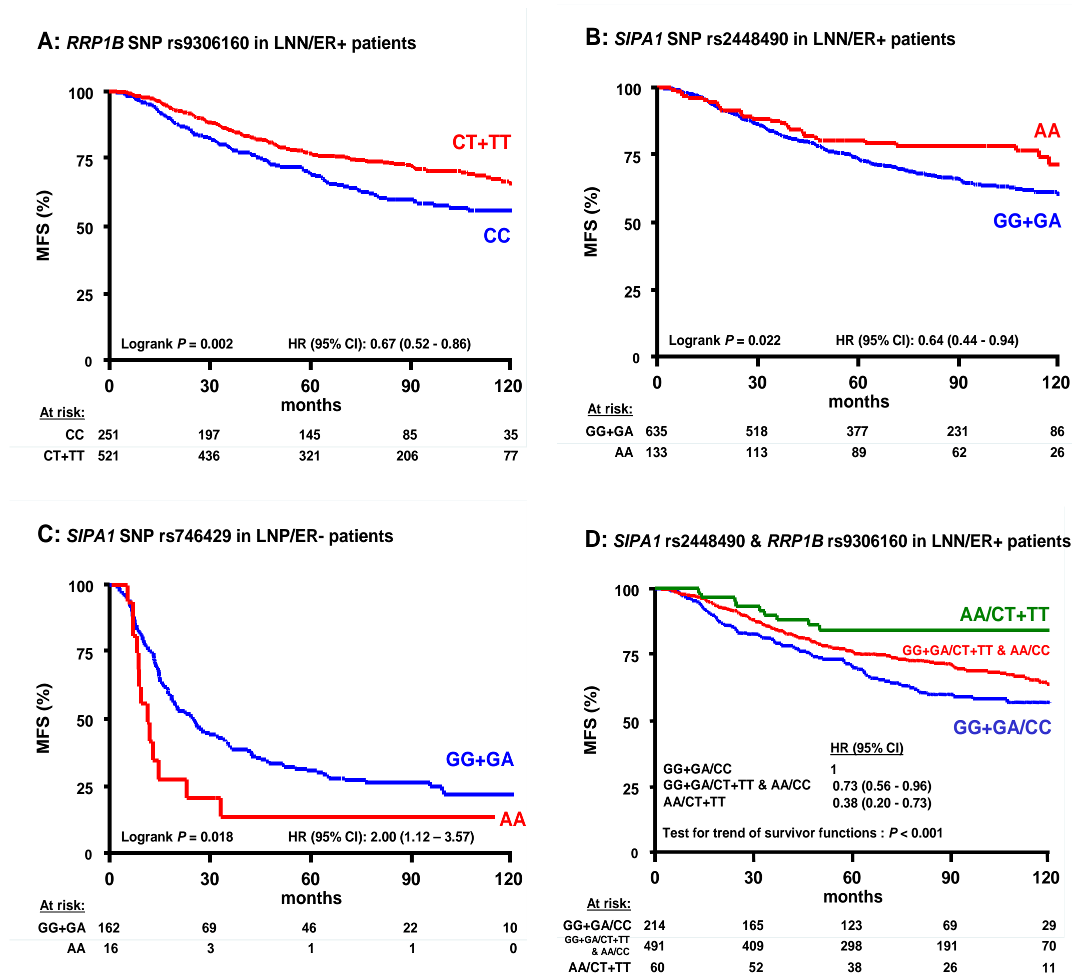


Figure 3.15: A summary of metastasis-free survival of *SIPA1* and *RRP1B* genotypes. The CT+TT genotype of the *RRP1B* SNP showed better survival in LNN/ER+ patients (panel A). The AA genotype of *SIPA1* SNP rs2448490 showed better survival in the same patient group (panel B) and the GG+GA genotype from SNP rs746429 showed better survival in a different patient group (LNP/ER-) (panel C). Combining the results in the LNN/ER+ patients group, the homozygous AA variant genotype of *SIPA1* SNP rs2448490 and the T variant allele (CT+TT) of *RRP1B* SNP rs9306160 were associated with the best prognosis (panel D).

The potential relationship between SNP rs9306160 and treatment response were subsequently analysed; no significant association of any of the SNPs with the efficacy of first-line tamoxifen therapy in 315 oestrogen receptor positive recurrent breast cancer patients was observed, irrespective of the endpoint studied.

Response to first line chemotherapy was assessed and the genotype frequency of this SNP in the Rotterdam, Netherlands population was determined. Since none of the *p*-values were under 0.05 this indicated that the distributions of the genotypes in this sub-cohort did not deviate from HWE (Table 3.26). No association of SNP rs9306160 and progression-free survival in patient receipt of first line chemotherapy was identified.

Table 3.26: Rotterdam, the Netherlands, *RRP1B* SNP rs9306160 Genotype Frequency, Patients received first line chemotherapy-univariate analysis.

Clinical Markers		Expected (%)	Test (%)			p
			CC	CT	TT	
Response to first line therapy	No Response	45.1	51.2	41.9	40.5	0.37
	Response	54.9	48.8	58.1	59.5	
Age in classes (years)	<=40	17.3	15.5	21.0	10.8	0.21
	41-55	49.1	42.9	51.4	56.8	
	>55	33.6	41.7	27.6	32.4	
Menopausal status	pre	49.1	47.6	52.4	43.2	0.60
	post	50.9	52.4	47.6	56.8	
Dominant site of relapse	LRR[a]	13.7	7.1	16.2	21.6	0.09
	Bone	15.5	13.1	15.2	21.6	
	Other	7.08	79.8	68.6	56.8	
Disease free survival	0-12 months	36.7	36.9	36.2	37.8	0.17
	12-36 months	46.0	51.2	40.0	51.4	
	>36 months	17.3	11.9	23.8	10.8	
Oestrogen receptor status	Negative	55.8	56.0	50.5	67.6	0.24
	Positive	44.2	44.0	47.6	32.4	
Total number of samples included			84	105	37	

[a]LRR: local regional relapse.

No other association with either clinical markers or survival was observed in other cohorts, including the German cohort (bone marrow metastasis analysis), the Minnesota, USA cohort (Mammography density analysis), the Queensland, Australia cohort (breast cancer incidence analysis) and the Umea, Sweden population.

In summary, *RRP1B* SNP showed associations with breast cancer survival in the large Netherland population but no other additional populations. The combined effect of *SIPA1* and *RRP1B* SNPs is especially interesting and further studies are required to provide additional support over these findings. A table summarizing the significant results is included in Table 3.27.

Table 3.27: Summary of *RRP1B* SNPs Genotyping Results.

Gene	SNP	Genotype	Trait	p-value	population
RRP1B	rs9306160	A allele	Increased localized tumour	0.006	Southern California
		A allele	Higher in ER positive tumours	0.001	Southern California
		A allele	Higher in PR positive tumours	0.001	Southern California
		A allele	Decreased grade	0.001	Southern California
		A allele	Increases survival	0.025	Greater Baltimore
		A allele	Decreased tumour size	-	Fred Hutchinson
		GG	MFS	0.013	Netherlands
		A allele	Increased DFS in LN-/ER+	0.002	Netherlands
		A allele	Increased MFS in LN-/ER+	0.002	Netherlands
		A allele	OS in LN-/ER+	0.04	Netherlands

3.5 *BRD4* Gene results

Expression QTL (eQTL) mapping in recombinant inbred mice to characterize genetic loci modulating metastasis-predictive ECM gene expression revealed bromodomain containing 4 as a potential metastasis modulating gene. Further molecular studies showed that the *BRD4* protein is physically a binding partner of the previously identified metastasis modulator *SIPA1*, supporting *BRD4* as a metastasis modulator (Crawford, *et al.*, 2008). The *BRD4* protein, from the bromodomains and extraterminal (BET) family, is involved in multiple cell activities, including transcription, replication (*BRD4* protein is homologous to the murine mitotic chromosome-associated protein (MCAP), which is associated with chromosomes during mitosis), signal transduction pathways and cell cycle progression. The *BRD4* protein, with two bromodomains, conducts its functions mostly via protein association with acetylated chromatin, with the tandem bromodomains binding to diacetylated H4-AcK5/K12 and H3-AcK9/K14 peptides to induce transcriptional activation (Liu *et al.*, 2008). *In vitro* analysis of *Brd4* ectopically expressed in a highly metastatic mouse mammary tumour cell line demonstrated significant reduction of invasiveness without altering intrinsic growth rate, yet a dramatic reduction of tumour growth and pulmonary metastasis was observed after implantation into mice, implying that activation of *Brd4* may alter response to tumour microenvironment *in vivo*. Additional *in vitro* analysis showed that *Brd4* modulates extracellular matrix gene expression, a class of genes frequently present in metastasis-predictive gene signatures. Microarray analysis of the mammary tumour cell lines identified a Brd4 activation signature that robustly predicted progression and/or survival in multiple human breast cancer datasets analysed on different microarray platforms. In addition, the *Brd4* signature closely matched a molecular classifier of low-grade tumours. Taken together, these

data suggest that dysregulation of *Brd4*-associated pathways may play an important role in breast cancer progression and underlies multiple common prognostic signatures (Crawford *et al*., 2008).

The pilot study cohorts included the Southern California, USA cohort (n=269) and the Greater Baltimore, USA cohort (n=248 surgical breast cancer patients; 58% African-American, 42% Caucasian).

Southern California, USA

The candidate SNP rs11880801 from the *BRD4* gene showed association with progesterone receptor status in the univariate analyses in the Southern California, USA pilot study; the TT genotype was more frequently observed in patients with PR negative tumours than the PR positive tumours (p=0.0017) (Table 3.28). The other three candidate SNPs (SNP rs4808272, rs8104223 and rs4809130) showed no association with clinical markers in this pilot cohort.

Table 3.28: Southern California, USA, *BRD4* gene SNP, univariate analysis.

Clinical Markers and refSNP ID[c]	Genotype	Numbers and Frequency			
		Negative		Positive	
		N	%	N	%
Progesterone Receptor Status rs11880801	GG	41	59	110	71
	GT	19	27	40	26
	TT	10	14	4	3
	Total	70	100	154	100
	Dominant[b]: GG + GT v.s TT, p=0.0657				
	Recessive[b]: GG v.s GT + TT, p=0.0017[a]				

[a]p<0.05.--[b]Fisher's Exact Chi test.--[c]refSNP ID: Reference SNP Cluster Identification Number.

Two of the four candidate SNPs are from the same LD block (Figure 3.16), with the candidate SNP rs4809130 showing association with patient survival located in different LD blocks and isn't in the same block with other candidate SNPs.

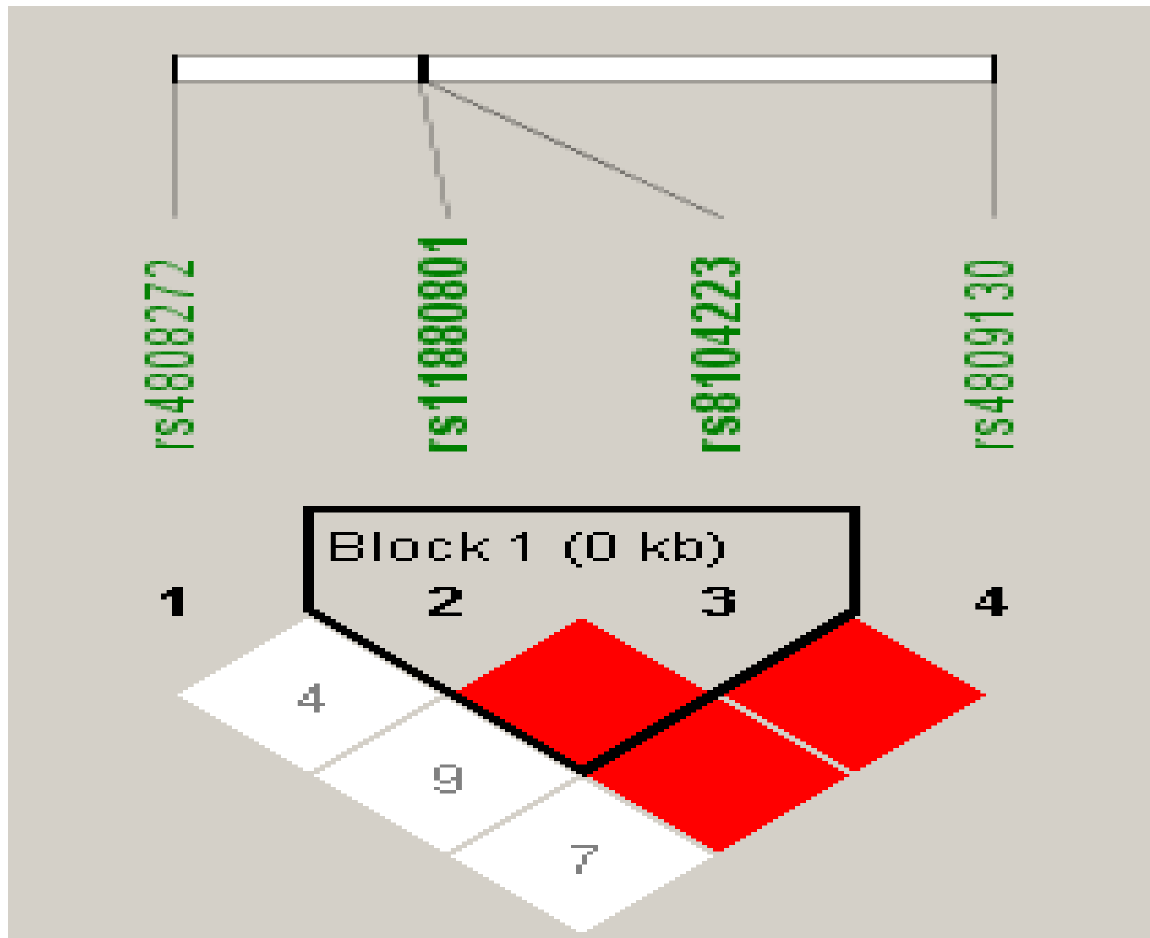


Figure 3.16: Linkage Disequilibrium block for *BRD4* SNPs. Candidate SNPs rs11880801 and rs8104223 from the *BRD4* gene are from the same LD block.

The variant T allele from SNP rs4809130 in the *BRD4* gene was associated with poor survival in the 5 year analysis (HR=20.80, 95% CI=2.37-182.36, p=0.0061) in the multivariate analysis in the Southern California, USA pilot study (Table 3.29).

Table 3.29: Southern California, USA, *BRD4* gene SNP, multivariate survival analysis.

refSNP ID[a]	Model	5 Years[h]			10 Years[i]		
		HR[f]	95% CI[g]	p	HR[f]	95% CI[g]	p
rs4809130	Dom[c]. CC + CT v.s TT	0.92	0.41-2.10	0.8520	1.35	0.75-2.42	0.3185
	Add[d]. CC v.s CT v.s TT	1.08	0.50-2.33	0.8456	1.43	0.82-2.51	0.2109
	Rec[e]. CC v.s CT + TT	20.80	2.37-182.36	0.0061[b]	9.56	1.00-91.68	0.0503

[a]refSNP ID: Reference SNP Cluster Identification Number.--[b]p < 0.05.--[c]Dom.: Dominant model.--[d]Add.: Additive model.--[e]Rec.: Recessive model.--[f]HR: Hazard Ratio.--[g]95% CI: 95% Confidence Interval.--[h]Analysis is adjusted with breast cancer stage and PR status in the 5 year survival analysis.--[i]Analysis is adjusted with breast cancer stage, PR status, tumour size and age at diagnosis in the 10 year survival analysis.

Greater Baltimore, USA

Candidate SNP rs8104223 in the *BRD4* gene showed association with tumour stage and ER status in the Greater Baltimore, USA pilot study; patients with higher stage (II and above) were more likely to carry the variant G allele (p=0.015) and all six patients have the GG genotype (Table 3.30). Since these findings were not reproduced in the other pilot population, this candidate SNP was ranked as a low priority SNP for further analysis.

Table 3.30: Greater Baltimore, USA, *BRD4* gene SNP, univariate analysis.

Clinical Markers and [a]refSNP ID	Genotype	Numbers and Frequency				Pearson chi2	Pr
Breast Cancer Stage rs8104223		I		II and above			
		N	%	N	%		
	AA	75	80	85	64	6.4772	0.039
	GA	17	18	40	30		
	GG	2	2	7	6		
	Total	94	100	132	100		
Oestrogen Receptor Status rs8104223		Negative		Positive			
		N	%	N	%		
	AA	64	65	78	79	8.3754	0.015
	GA	28	29	21	21		
	GG	6	6	0	0		
	Total	98	100	99	100		

[a] refSNP ID: Reference SNP Cluster Identification Number.

Multivariate survival analysis on data gathered from *BRD4* gene candidate SNPs revealed that SNP rs4808272 was associated with better breast cancer survival (p= 0.026) (Table 3.31), yet this association was not reproduced in the other pilot cohort, which contained less Caucasian patients compared to the Southern California, USA cohort and hence, this SNP was ranked as lower priority for further epidemiology analyses. The SNP from the Southern California, USA cohort (n=300) (SNP rs4809130) was ranked as a higher priority and screened in multiple larger independent cohorts.

Table 3.31: Great Baltimore, USA, *BRD4* gene SNP, multivariate Cox regression survival analysis.

[a]refSNP ID	Hazard Ratio	Standard Error	z	p> z	95% Conf. Interval
rs4808272	0.639	0.129	-2.22	0.026	0.430-0.948

[a] refSNP ID: Reference SNP Cluster Identification Number.

All the survival analysis were adjusted with age at diagnosis, ER Status, TNM stage and receipt of chemotherapy and are the Caucasian subpopulation of this cohort.

Two *BRD4* candidate SNPs showed association with breast cancer survival, yet this result was not reproducible between the pilot cohorts; SNP rs4809130 was found to be associated with survival in the Southern California, USA cohort, whilst SNP rs4808272 was linked with survival in the Greater Baltimore, USA cohort. Candidate SNP rs4809130 was chosen for subsequent examination in larger cohorts since its association with patient survival was found in the Southern California, USA population, which has higher amounts of Caucasian patients and matches the condition of the subsequent larger cohort tested (the additional cohorts are Caucasian population only). Several other candidate SNPs showed association with numerous clinical features in one of the pilot cohorts, yet these findings were not reproducible

between the pilot cohorts and the candidate SNPs ranked as lower priority SNPs for further analyses. However, since the functional analyses revealed that BRD4 may be an important molecule in the Diasporin pathway, the *BRD4* SNPs were then included in further analyses in the larger populations.

The location of the candidate SNPs of the *BRD4* gene was analysed further in this study are included below for reference (Figure 3.17).

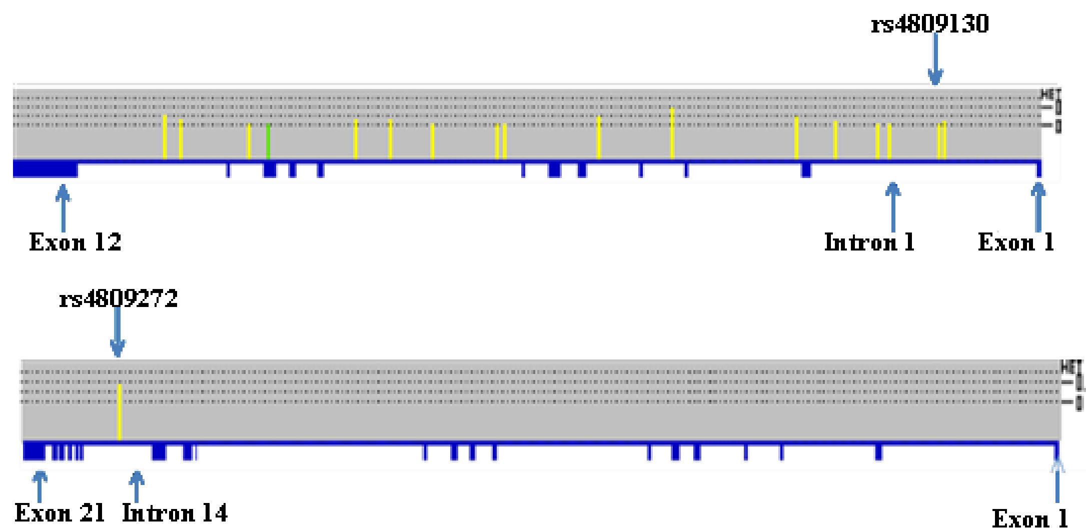


Figure 3.17: Locations of *BRD4* SNPs. Bromodomain containing 4 candidate SNPs rs4808272 and rs4809130 locations in the *BRD4* gene.

The Netherlands Cohort (n=1,922)

The analysis of the two SNPs, rs4809130 and rs4808272, from the *BRD4* gene in this cohort was performed in the first line chemotherapy receipt subgroup, but not all 1,922 samples. The genotype frequency of each SNP in the first line chemotherapy receipt subgroup were obtained and compared to the expected distribution to determine whether the distribution in this sub-population deviated from the expected HWE. The *p*-value was greater than 0.05 and thus the genotypes did not deviate from Hardy-Weinberg equilibrium (Table 3.32).

Table 3.32: The Netherlands, Various clinical markers frequency in *BRD4* Candidate SNPs, Patients received first line chemotherapy-univariate analysis.

[a]refSNP ID			rs4809130				rs4808272			
Clinical Markers		Expected (%)	Test (%)			p	Test (%)			p
			CC	CT	TT		AA	AG	GG	
Response to first line therapy	No Resp.	45.1	44.1	54.3	20.0	0.28	45.3	45.6	43.8	0.98
	Resp.	54.9	55.9	45.7	80.0		54.7	54.4	56.3	
Age in classes (years)	<=40	17.3	17.7	17.1	0.0	0.71	13.3	19.4	18.8	0.76
	41-55	49.1	48.4	48.6	80.0		52.0	49.5	43.8	
	>55	33.6	33.9	34.3	20.0		34.7	31.1	37.5	
Menopausal status	pre	49.1	47.3	54.3	80.0	0.28	50.7	49.5	45.8	0.87
	post	50.9	52.7	45.7	20.0		49.3	50.5	54.2	
Dominant site of relapse	LRR[b]	13.7	12.9	17.1	20.0	0.83	12.0	16.5	10.4	0.58
	Bone	15.5	16.1	14.3	0.0		20.0	12.6	14.6	
	Other	7.08	71.0	68.6	80.0		68.0	70.9	75.0	
Disease free survival	0-12 months	36.7	31.7	34.3	40.0	0.61	40.0	31.1	43.8	0.51
	12-36 months	46.0	45.7	51.4	20.0		44.0	48.5	43.8	
	>36 months	17.3	17.2	14.3	40.0		16.0	20.4	12.5	
Oestrogen receptor status	Negative	55.8	54.3	62.9	40.0	0.53	56.0	54.4	56.3	0.92
	Positive	44.2	44.6	37.1	60.0		42.7	45.6	41.7	
Total number of samples included		226	186	35	5		75	103	48	

[a]refSNP ID: Reference SNP Cluster Identification Number.--[b]LRR: local regional relapse.

Both candidate SNPs showed association with progression-free survival when further stratifying this subgroup; the AG genotype of SNP rs4808272 was associated with poor progression-free survival in the 1st line Anthracycline receipt patients and the CT genotype of SNP rs4809130 was associated with better progression-free outcome in patients with negative oestrogen receptor detection (Table 3.33).

Table 3.33: The Netherlands, *BRD4* Candidate SNPs, Patients received first line chemotherapy-multivariate survival analysis.

Test	Group	[a]refSNP ID	Genotype	N	Odds Ratio	chi 2	95% Confidnce Interval
Progression Free Survival	1st line Anthracycline Patients	rs4808272	AA	51	1.00	-	reference
			AG	71	0.65	0.031	0.44-0.96
			GG	28	0.65	0.077	0.40-1.05
	Oestrogen Receptor Negative	rs4809130	CC	NA	1.00	-	reference
			CT	NA	1.97	0.022	1.10-3.52
			TT	NA	0.47	0.465	0.06-3.53

[a]refSNP ID: Reference SNP Cluster Identification Number.

No other associations with either clinical markers or survival were observed in other cohorts, including the German cohort (bone marrow metastasis analysis), the Minnesota, USA cohort (Mammography density analysis), the Queensland, Australia cohort (breast cancer incidence analysis), the Fred Hutchinson Cancer Research Centre, USA cohort and the Umea, Sweden population.

In summary, *BRD4* SNPs showed associations with breast cancer progression in the Netherland cohort, more analyses maybe needed before conclusions can be drawn. A table summarizing the results is included.

Table 3.34: Summary of *BRD4* SNPs Genotyping Results.

Gene	SNP	Genotype	Trait	p-value	population
BRD4	rs11880801	TT	Increased in PR+ tumours	0.0017	Southern California
	rs4809130	T allele	Increased 5 yr survival	0.0061	Southern California
	rs8104223	G allele	Higher stage	0.039	Greater Baltimore
	rs8104223	G allele	Increased in ER+ tumours	0.015	Greater Baltimore
	rs4808272	G allele	Increased survival	0.026	Netherlands
	rs4808272	AG	Decreased PFS in 1st line Anthracycline treated	-	Netherlands
	rs4809130	CT	Increased PFS in ER-	-	Netherlands

3.6 Other Gene Results

This section collectively presents the results of SNPs that were not selected for further investigations in the larger populations. The SNPs included in this section are SNPs that only showed associations with clinical parameters and the significant findings were not reproduced between the two pilot populations.

The pilot study cohorts included the Southern California, USA cohort (n=269) and the Greater Baltimore, USA cohort (n=248 surgical breast cancer patients; 58% African-American, 42% Caucasian).

***PI16* Gene**

One SNP (rs708006) from the *PI16* gene was found to be associated with breast cancer survival in only one of the pilot populations; it was then ranked as a lower priority candidate SNP but was further examined in two additional larger cohorts. Several other SNPs were found to be associated with various clinical features in the pilot cohorts, yet these markers were not closely related to breast cancer metastasis and these SNPs were ranked as lower priority for future analyses.

Southern California, USA

Several candidate SNPs from the *PI16* gene showed association with clinical markers in the univariate analyses. Four SNPs rs707542, rs707998, rs1405069 and rs4147297 of the eight selected *PI16* gene SNPs showed significant association with the stage of breast cancer in this cohort; the rare allele (A) from SNP rs707542, rs1405069 and rs4147297 were associated with non-localised breast cancer (Table 3.35). The lymph

node metastasis status is of particular interest in this study and SNPs rs707542 and rs4147297 were found to be associated with lymph node status in this pilot cohort. Since lymph node metastasis status is one of the criteria in determining stage of breast cancer, it is not surprising that SNPs showing association with stage may also show association with lymph node metastasis. Two of the SNPs showing association with breast cancer stage, also showed association with lymph node metastasis in this pilot cohort. The AA genotype in SNP rs707542 was more frequently observed in patients with non-localised breast cancer (73% v.s localized 27%) and patients with positive lymph node metastasis (65% v.s lymph node metastasis negative 35%); a similar finding was observed with the TT genotype in SNP rs4147297 (Table 3.35).

In addition, all genotypes from candidate SNP rs708006 from the *PI16* gene were more commonly observed in patients with ductal tumours compared to the non-ductal tumours, this maybe a reflection of the larger number of patients with the ductal tumours compared to the non-ductal tumours (Table 3.35). The other candidate SNPs (rs737905, rs6901560 and rs7755143) were not found to be associated with clinical markers in the Southern California, USA population. The clinical information of this pilot population was not available to the author and logistic regression analysis was not performed on data gathered for *PI16* candidate SNPs and hence the strength of the association observed could not be further determined.

Table 3.35: Southern California, USA, *PI16* gene SNPs, univariate analysis.

Clinical Markers	refSNP ID[c]	Genotype	Numbers and Frequency			
Stage			Non-Localized		Localized	
			N	%	N	%
	rs707542	GG	40	28	54	40
		GA	65	46	67	50
		AA	37	26	14	10
		Total	142	100	135	100
		Dominant[b]: GG + GA v.s. AA, p=0.0426				
		Recessive[b]: GG v.s GA + AA, p=0.0010[a]				
	rs707998	GG	101	71	80	58
		GA	39	27	51	37
		AA	2	2	6	5
		Total	142	100	137	100
		Dominant[b]: GG + GA v.s. AA, p=0.0328[a]				
		Recessive[b]: GG v.s GA + AA, p=0.1667				
	rs1405069	AA	54	38	36	26
		AC	63	44	68	50
		CC	25	18	33	24
		Total	142	100	137	100
		Dominant[b]: AA + AC v.s CC, p=0.0407[a]				
		Recessive[b]: AA v.s AC + CC, p=0.1882				
	rs4147297	TT	56	41	31	23
		TA	59	44	74	55
		AA	20	15	30	22
		Total	135	100	135	100
		Dominant[b]: TT + TA v.s AA, p=0.0017[a]				
		Recessive[b]: TT v.s TA + AA, p=0.1581				
Nodal Status			Positive		Negative	
			N	%	N	%
	rs707542	GG	37	30	51	39
		GA	57	46	63	48
		AA	30	24	16	13
		Total	124	100	130	100
		Dominant[b]: GG + GA v.s. AA, p=0.1466				
		Recessive[b]: GG v.s GA + AA, p=0.0150[a]				
	rs4147297	TT	45	38	32	25
		TA	56	47	69	53
		AA	18	15	28	22
		Total	119	100	129	100
		Dominant[b]: TT + TA v.s AA, p=0.0289[a]				
		Recessive[b]: TT v.s TA + AA, p=0.1949				
Ductal Histology			Ductal		Non-Ductal	
			N	%	N	%
	rs708006	CC	131	67	68	81
		AC	60	31	16	19
		AA	5	2	0	0

		Total	196	100	84	100
		Dominant[b]: CC + CA v.s AA, p=0.0210[a]				
		Recessive[b]: CC v.s CA + AA, p=0.3267				
Lobular Histology			Lobular		Non-Lobular	
			N	%	N	%
	rs708006	CC	44	83	155	68
		AC	9	17	67	30
		AA	0	0	5	2
		Total	53	100	227	100
		Dominant[b]: CC + CA v.s AA, p=0.0426[a]				
		Recessive[b]: CC v.s CA + AA, p=0.5873				

[a]p<0.05.--[b]Fisher's Exact Chi test.--[c]refSNP ID: Reference SNP Cluster Identification Number.

The two SNPs that showed association with both stage of breast cancer and lymph node metastasis (SNP rs707542 and rs4147297) belong to two different LD blocks (Figure 3.18). Block 2 contains not only SNP rs4147297 that is associated with lymph node status, but also SNP rs708006, which is associated with breast cancer histology and survival (Table 3.36).

Figure 3.18: Linkage Disequilibrium block for *PI16* SNPs. Both blocks contain important SNPs that are associated with lymph node metastasis status and patient survival.

SNP rs708006 is the only SNP that showed association with survival in this pilot cohort. The variant A allele showed association with poor survival in both 5 and 10 year analyses (Table 3.36).

Table 3.36: Southern California, USA, *PI16* gene SNP, multivariate survival analysis.

refSNP ID[a]	Model	5 Years[h]			10 Years[i]		
		HR[f]	95% CI[g]	p	HR[f]	95% CI[g]	p
rs708006	Dom[c]. CC + CA v.s AA	1.74	0.91-3.36	0.0960	1.10	0.63-1.90	0.7446
	Add[d]. CC v.s CA v.s AA	1.9	1.06-3.42	0.0309	1.28	0.77-2.12	0.3346
	Rec[e]. CC v.s CA + AA	6.68	1.55-28.76	0.0108[b]	6.00	1.75-20.54	0.0043[b]

[a]refSNP ID: Reference SNP Cluster Identification Number.--[b]$p < 0.05$.--[c]Dom.: Dominant model.--[d]Add.: Additive model.--[e]Rec.: Recessive model.--[f]HR: Hazard Ratio.--[g]95% CI: 95% Confidence Interval.--[h]Analysis is adjusted with breast cancer stage and PR status in the 5 year survival analysis.--[i]Analysis is adjusted with breast cancer stage, PR status, tumour size and age at diagnosis in the 10 year survival analysis.

Greater Baltimore, USA

Three candidate SNPs from the *PI16* gene showed associations with tumour grade in the Greater Baltimore, USA cohort (univariate analysis). The CC genotype from SNP rs708006, CC genotype from SNP rs6901560 and AA genotype from SNP rs1405069 were more frequently found in patients with lower grade of breast tumour (Table 3.37). The TT genotype from SNP rs7755143, which didn't show association with clinical markers, was found to be more common in patients with oestrogen receptor positive tumours (p=0.034). In addition, SNP rs4147297 was found to be associated with breast cancer stage, which reproduced the finding obtained in the Southern California, USA pilot study. The AA genotype is the most prevalent in the higher staged (II and above) patient (p=0.045) group (Table 3.37).

Table 3.37: Greater Baltimore, USA, *PI16* gene SNPs, univariate analysis.

Clinical Markers	[a]refSNP ID	Genotype	Numbers and Frequency				Pearson chi2	Pr
Tumour Grade			1		2&3			
			N	%	N	%		
	rs708006	CC	54	54	35	36	6.6883	0.035
		AC	30	30	41	42		
		AA	16	16	22	22		
		Total	100	100	98	100		
	rs6901560	GG	66	68	75	75	7.5768	0.023
		GC	24	25	26	25		
		CC	7	7	0	0		
		Total	97	100	101	100		
	rs1405069	CC	36	36	36	36	7.3852	0.025
		AC	37	37	51	51		
		AA	26	26	12	12		
		Total	99	100	99	100		
Oestrogen Receptor Status			Negative		Positive			
			N	%	N	%		
	rs7755143	CC	80	85	98	73	6.7391	0.034
		CT	14	15	32	24		
		TT	0	0	5	3		
		Total	94	100	135	100		
Breast Cancer Stage			I		II and above			
			N	%	N	%		
	rs4147297	TT	80	47	17	44	6.2021	0.045
		TA	65	38	21	54		
		AA	27	15	1	2		
		Total	172	100	39	100		

[a]refSNP ID: Reference SNP Cluster Identification Number.

None of the candidate SNPs from the *PI16* gene showed association with breast cancer survival in the multivariate Cox regression survival analysis in the Greater Baltimore, USA pilot population. Since SNP rs708006 showed association with breast cancer patient survival in the Southern California, USA cohort and the result did not reproduce in the Greater Baltimore, USA cohort and thus, this SNP was marked as a lower priority SNP and has only been screened in two additional larger cohorts (Umea, Sweden and the Netherlands).

With regard to the genes of interest and their related SNPs, both the *PI16* and *BRD4* genes showed associations with survival in only one of the pilot population. However, due to the molecular characteristics of the *BRD4* gene and its possible significant import on breast cancer progression, only the *BRD4* candidate SNP was selected for further analyses. The *Brd4* protein is a physical binding partner of the novel metastasis modifier gene *Sipa1* and in addition to the *Rrp1B* gene, is the central node of the novel metastasis modulating Diasporin pathway (Crawford *et al*., 2008).

No more candidate SNPs from any of the other genes (*ARID4B*, *TTC9C*, *LUC7L*, CSF1R, *EZH2*, *STAB1*, *MYO7A*, *XLKD1*, *MAPK14*, *CALM2* and *AQP2*) showed association with patient survival in either of the pilot cohorts and since association between the candidate SNPs and major clinical markers were identified, these SNPs were ranked as lower priority for future analyses. Statistical analysis results from the Southern California, USA cohort are available for the *ARID4B* gene and *LUC7L* gene candidate SNPs. The candidate SNPs from the *ARID4B* gene showed no association with clinical markers in this cohort. Candidate SNPs from the *LUC7L* gene showed association with clinical characteristics.

The univariate analysis revealed that the CC genotype from SNP rs11642609 was more prevalent in patients with progesterone receptor positive breast tumours (p=0.0177), and patients with lobular histology breast cancer tend to have the CC genotype from SNP rs1203981 (p=0.0201) (Table 3.38).

Table 3.38: Southern California, USA, *LUC7L* gene SNPs, univariate analysis.

Clinical Markers and [a]refSNP ID	Genotype	Numbers and Frequency			
Progesterone Receptor Status rs11642609		Negative		Positive	
		N	%	N	%
	CC	12	15	48	29
	CT	45	56	71	43
	TT	24	29	48	28
	Total	81	100	167	100
	Dominant[b]: CC + CT v.s TT, p=0.0177[a]				
	Recessive[b]: CC v.s CT + TT, p=0.8825				
Lobular Histology rs1203981		Lobular		Non-Lobular	
		N	%	N	%
	CC	14	30	29	15
	CT	16	34	93	47
	TT	17	36	74	38
	Total	47	100	196	100
	Dominant[b]: CC + CT v.s TT, p=0.0201[a]				
	Recessive[b]: CC v.s CT + TT, p=0.8685				

[a]p<0.05.--[b]Fisher's Exact Chi test.--[c]refSNP ID: Reference SNP Cluster Identification Number.

Two of the *LUC7L* candidate SNPs are in the same LD block in the Southern California, USA pilot population and the two SNPs that showed association with clinical markers are not in the same LD block (Figure 3.19).

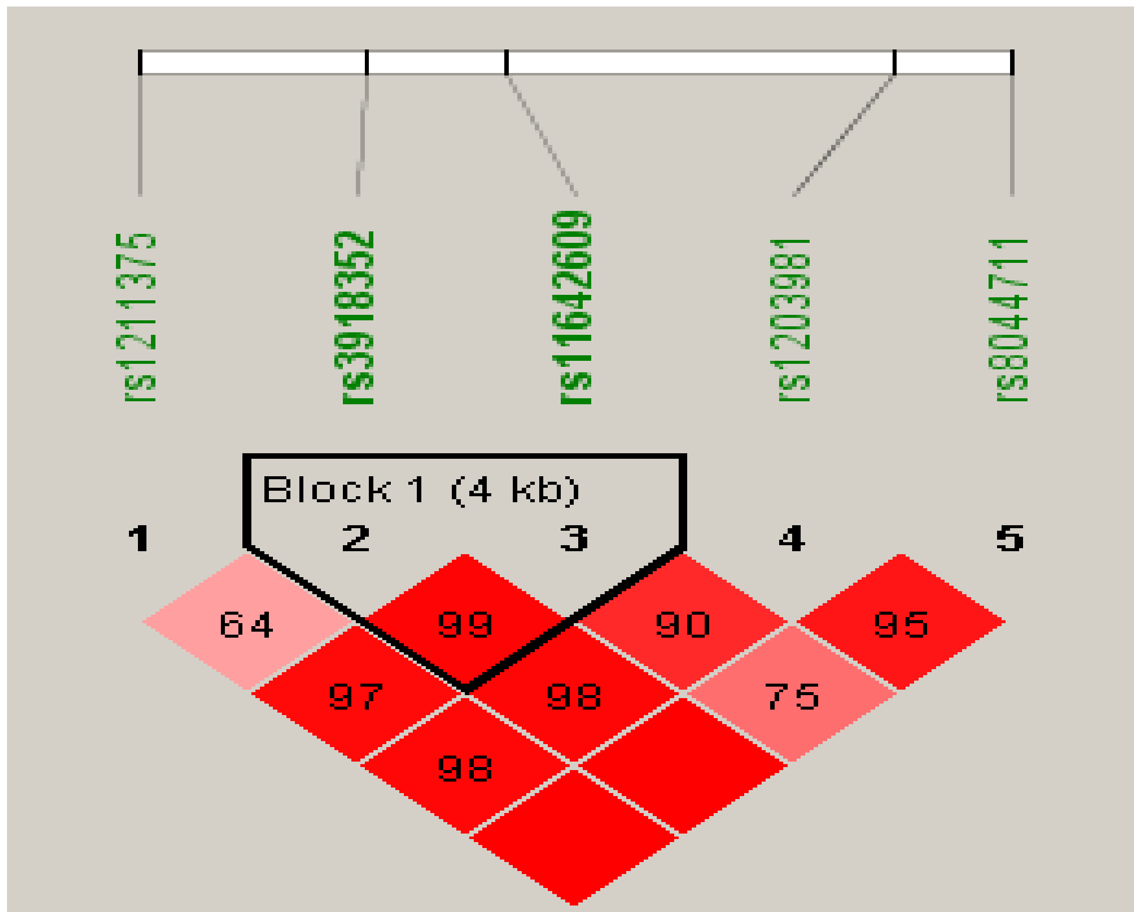


Figure 3.19: Linkage Disequilibrium block for *LUC7L* SNPs. Two of the five *LUC7L* gene candidate SNPs are in the same LD block, however the two SNPs that showed association with clinical markers are not in the same LD block.

No significant association was identified in the Greater Baltimore, USA cohort for the SNPs in the *LUC7L* gene.

Numerous candidate SNPs across different genes showed significant association with different clinical characteristics in the Greater Baltimore, USA pilot population. All the associations identified were described in the Caucasian sub-population of this cohort. Three SNPs across three different genes (*MYO7A*, SNP rs2276288; *ARID4B*, SNP rs946927; *CALM2*, SNP rs1027478) were found to be associated with oestrogen receptor status in the univariate analysis (Table 3.39). In addition, the Calmodulin 2 SNP rs1027478 was not only associated with ER status but also PR status, indicating

that this SNP and gene may play a role in hormonal receptor mediated carcinogenesis. The candidate SNP rs65916720 from the *TTC9C* gene was found to be associated with both TNM stage and the lymph node status, which are two clinical characteristics tightly correlated with each other and relevant to breast cancer metastasis. The common C allele was more frequently observed among patients with negative lymph node metastasis (p=0.008) and patients with lower TNM stage (stage I) (p=0.014) (Table 3.39). The candidate SNP rs946027 from the *ARID4B* gene was found to be associated with not only the oestrogen receptor status but also tumour grade (p=0.006), in that patients with higher tumour grades (2 & 3) of breast cancer tended to have the CC genotype of this SNP (Table 3.39).

Table 3.39: Greater Baltimore, USA, Multiple Genes, univariate analysis

Clinical Markers	Gene and [a]refSNP ID		Numbers and Frequency				Pearson chi2 (2)	Pr
oestrogen Receptor Status			Negative		Positive			
			N	%	N	%		
	MYO7A rs2276288	TT	34	37	55	41	7.7410	0.021
		TA	48	52	48	36		
		AA	10	11	30	23		
		Total	92	100	133	100		
	ARID4B rs946027	TT	74	79	83	61	8.8332	0.012
		TC	15	16	45	33		
		CC	5	5	8	6		
		Total	94	100	136	100		
	CALM2 rs1027478	CC	54	58	56	42	7.7201	0.021
		CT	33	35	54	41		
		TT	6	7	22	17		
		Total	93	100	132	100		
Progesterone Receptor Status			Negative		Positive			
			N	%	N	%		
	CALM2 rs1027478	CC	56	57	32	39	9.8157	0.007
		CT	37	37	35	42		
		TT	6	6	16	19		
		Total	99	100	83	100		
Breast Cancer Stage			I		II and above			
			N	%	N	%		
	TTC9C rs65916720	CC	125	74	23	56	8.5556	0.014
		CT	41	24	15	37		
		TT	2	2	3	7		
		Total	168	100	41	100		
Lymph Node Status			Negative		Positive			
			N	%	N	%		
	TTC9C rs65916720	CC	100	76	52	64	9.6423	0.008
		CT	32	24	24	30		
		TT	0	0	5	6		
		Total	132	100	81	100		
Tumour Grade			1		2&3			
			N	%	N	%		
	ARID4B rs946027	TT	63	63	75	75	10.0916	0.006
		TC	35	35	18	18		
		CC	2	2	8	8		
		Total	100	100	101	100		

[a] refSNP ID: Reference SNP Cluster Identification Number.

In summary, numerous genes and SNPs were examined in the pilot population, most of the SNPs showed associations with clinical parameters but not survival, except rs708006 from *PI16*, which is association with decreased survival in the Southern California cohort. However, this association was not reproduced in other populations. A table summarizing the results is included.

Table 3.40: Summary of All Other Genes SNPs Genotyping Results.

Gene	SNP	Genotype	Trait	p-value	population
PI16	rs707542	G allele	Increased localized tumour	0.001	Southern California
	rs707998	G allele	Increased non-localized tumour	0.0328	Southern California
	rs1405069	A allele	Increased non-localized tumour	0.0407	Southern California
	rs4147297	T allele	Increased non-localized tumour	0.0017	Southern California
	rs707542	G allele	Higher in ER negative tumour	0.015	Southern California
	rs4147297	T allele	Higher in ER positive tumour	0.0289	Southern California
	rs708006	C allele	Higher in non-ductal carcinoma	0.021	Southern California
	rs708006	C allele	Higher in lobular carcinoma	0.0426	Southern California
	rs708006	A allele	Decreased 5 yr survival	0.0108	Southern California
	rs708006	A allele	Decreased 10 yr survival	0.0043	Southern California
	rs708006	CC	Higher in lower grade tumours	0.035	Greater Baltimore
	rs6901560	CC	Higher in lower grade tumours	0.023	Greater Baltimore
	rs1405069	AA	Higher in lower grade tumours	0.025	Greater Baltimore
	rs7755143	TT	Higher in ER+ tumours	0.034	Greater Baltimore
	rs4147297	AA	Higher in higher stage tumours	0.045	Greater Baltimore
LUC7L	rs11642609	C allele	Higher in PR+ tumours	0.0177	Southern California
	rs1203981	C allele	Higher in lobular carcinoma	0.0201	Southern California
MYO7A	rs2276288	AA	Higher in ER+ tumours	0.021	Greater Baltimore
ARID4B	rs946027	CC	Higher in ER+ tumours	0.012	Greater Baltimore
	rs946027	CC	Higher in higher grade tumours	0.006	Greater Baltimore
CALM2	rs1027478	TT	Higher in ER+ tumours	0.021	Greater Baltimore
	rs1027478	TT	Higher in PR+ tumours	0.007	Greater Baltimore
TTC9C	rs65916720	TT	Higher in higher stage tumours	0.014	Greater Baltimore
	rs65916720	TT	Higher in node+ patients	0.008	Greater Baltimore

3.7 Genotyping Results Summary

This epidemiology study started with screening 49 candidate SNPs from 16 candidate metastasis modulator genes across two pilot cohorts. Five candidate SNPs from three different genes showed association with survival in the pilot studies and three of these were subsequently genotyped in an additional six larger independent cohorts. An additional four SNPs from two genes identified in previous studies were also examined. The four genes further investigated were *ARAP3* (rs440279, rs3763120), *SIPA1* (rs931127, rs746429, rs3741378 and rs2448490), *BRD4* (rs4809130 and rs4808272) and *RRP1B* (rs9306160).

The two candidate SNPs rs440279 and rs3763120 from the *ARAP3* gene were found to be associated with better patient survival, with and without stratification of lymph node metastasis status and oestrogen receptor status, but also with other important clinical markers, such as mammographic density. This gene also showed association with response to first line chemotherapy. There were initially three candidate SNPs for the *SIPA1* gene (SNP rs931127, rs746429 and rs3741378), with one additional SNP rs2448490 assayed in an attempt to cover greater areas of the gene. Interesting findings were observed for these SNPs; the variant allele of SNP rs746429 (missense SNP) was found to be associated with poor outcome in metastasis-free, disease-free and overall survival in the largest Rotterdam, Netherlands population. Furthermore, SNP rs2448490, an intronic SNP, was found to be associated with better rate of outcome in the best class of breast cancer patients: lymph node metastasis negative and oestrogen receptor positive class. A combined analysis of the *SIPA1* SNP rs2448490 and *RRP1B* SNP 9306160 (since the two proteins show physical

interactions with one another) in this same class of patients gave the best prognosis. The other *SIPA1* protein binding partner, *BRD4*, showed association with progression-free survival with its candidate SNPs as well. The candidate SNPs of the three genes, *SIPA1*, *RRP1B* and *BRD4* genes that are known to be the corner stone of the Diasporin pathway (Crawford, *et al*., 2008), showed association with breast cancer metastasis and metastasis related survival in multiple independent populations, adding support to the theory that these genes are metastasis modulator genes. In addition, the candidate SNP rs3741378 from the *SIPA1* gene also showed association with breast cancer incidence, indicating that the *SIPA1* gene may not only be a breast cancer metastasis modulating gene, but also a breast cancer susceptibility gene.

Chapter 4

Novel *Sipa1* Protein Binding Partner Identification

4.1 Introduction

Sipa1 (signal-induced proliferation-associated 1) was reported as a novel breast cancer metastasis modulating gene, with *in vivo* analysis revealing that increased *Sipa1* expression enhances mouse lung metastasis potency (Park *et al.*, 2005). This rare property of promoting metastasis facilitated strong interest in investigating the molecular characteristics of *Sipa1* in breast cancer metastasis. However, the molecular characteristics of *Sipa1* are largely unknown and therefore, *in vitro* analyses were designed in an attempt to further elucidate the molecular properties of this novel metastasis modifier gene. One of the experiments designed aimed to discover the protein binding partners of the *Sipa1* protein, which would shed light in finding out the molecular network and functions of *Sipa1* in breast cancer metastasis. In an attempt to address this aim a yeast-two-hybrid assay on the *Sipa1* protein was performed and putative *Sipa1* protein binding partners were identified.

Previous mouse studies have demonstrated that a polymorphism in the PDZ protein–protein interacting domain of the *Sipa1* gene is associated with metastasis. It was found that the polymorphism in the PDZ domain changed the protein binding efficiency between *Sipa1* and *Aqp2* (Aquaporin 2). In addition to the change in *Aqp2* binding efficiency, this polymorphism also impacted upon the Rap-GAP activity of *Sipa1* as well as impacting upon the metastasis potency of the highly metastatic mouse mammary carcinoma cell line, Mvt-1 (also known as Pei 1). Since the PDZ domain bears importance in *Sipa1* protein binding function in association with breast cancer metastasis and is involved in Rap-GAP activity, subsequent yeast two-hybrid screening of the *Sipa1* PDZ-domain was therefore performed to identify additional genes potentially involved in metastasis via interacting with this domain. Yeast two-hybrid screens using PDZ regions of the human *Sipa1* protein (Entrez Gene ID No:

6494) as bait were performed by ProNet technology (Myriad Genetics, Salt Lake City, UT) and the results were summarised in Figure 4.1. the yeast-two hybrid system allows for the identification of putative protein-protein interactions by screening expression libraries. The system utilises two plasmid-borne gene fusions that are co-transformed into a host yeast strain expressing inducible reporter genes. This entails encoding the protein of interest (or "bait") as a gene fusion to a DNA binding domain from either the GAL4 or LexA protein. A second protein, or a library of proteins, is fused to a transcription activation domain. if the two proteins interact, than the localisation of the transcription activation domain to the DNA of the host strain will activate transcription of the adjacent receptor gene (i.e adjacent to GAL4 or LexA) and generate an observable phenotypic signal (James *et al*., 1996). Following sequence alignment, many clones were found to bind to the PDZ domain; the clones selected for analysis in this part of research are indicated in the red brace in Figure 4.1, they are actin related protein 2/3 complex, subunit 3 (*p21Arc*), Niemann Pick type C1 (*Npc1*), general transcription factor II H, polypeptide 2 (*Gtf2H2*), phosphoribosylglycinamide formyltransferase (*Gart*), Calmodulin 2 (*Calm2*) and apoptotic chromatin condensation inducer 1 (*Acin1*).

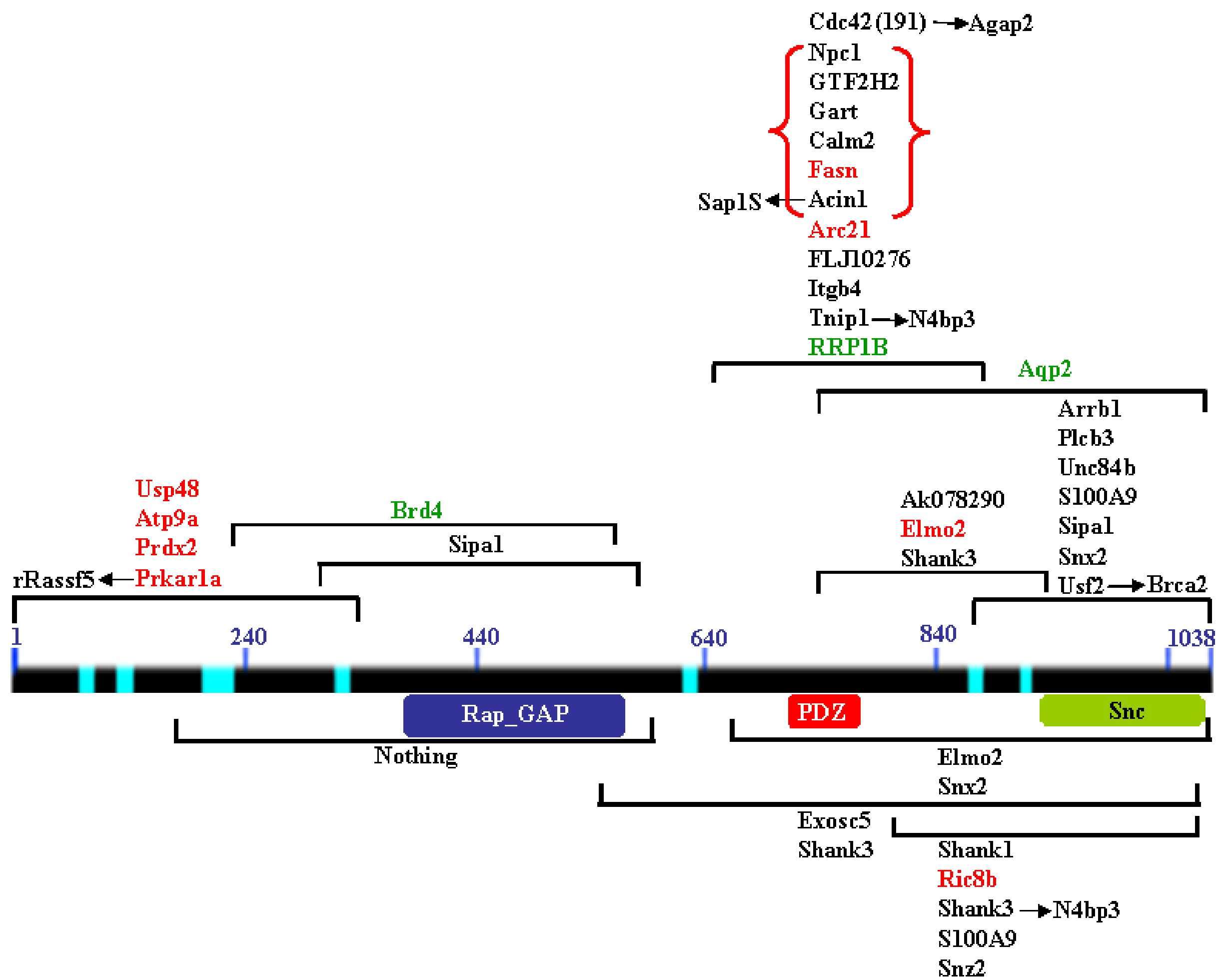


Figure 4.1: Yeast 2 Hybrid results. This figure summary the genes found to be interacting with *Sipa1* in the Y2H assay. Those in red were observed once; black represents multiple times and green indicates published interactions. The genes in the red brace were investigated in this part of research.

The six clones identified in the Y2H assay were further investigated in this part of the research. The first aim of this section of the research was to confirm the yeast-2-hybrid assay results via Co-IP assays. Four (*p21Arc*, *Npc1*, *Gtf2H2* and *Gart*) of the clones failed to have successful dual-transfection of both the gene of interest and *Sipa1* (which ensured optimal amount of proteins for IP assays) and were not selected for subsequent Co-IP assays. Co-IP assays on *Calm2* + *Sipa1* and *Acin1* + *Sipa1* dual-transfected cell lysate found that the *Calm2* protein and the *Sipa1* protein physically interact, whereas *Acin1* didn't bind to *Sipa1* in this assay. To further elaborate the nature of *Sipa1* and *Calm2* associations in breast cancer metastasis, the second aim of this section of the research investigated the *Calm2* gene in an *in*

vivo metastasis assay. The purpose to investigate *Calm2 in vivo* (or any other *Sipa1* binding partners) is to find out whether the metastasis modulating effect of Sipa1 is due to a *Sipa1* independent effect or a syngeneic affects between *Sipa1* and its binding partners.

4.2 Methods

4.2.1 *In vitro* Yeast-2-Hybrid Result Confirmation

To confirm the putative *Sipa1*- binding proteins identified in the yeast-2-hybrid assay, epitope tagged *Sipa1* and epitope tagged proteins of interest (eg. *Acin1* and *Calm2*) were transiently over-expressed together in tissue culture cells. This would theoretically ensure that there would be more protein of interest and *Sipa1* produced due to the dual up-regulation and thus if the two proteins physically interact, it would be more likely to be detected by the Co-immunoprecipitation (Co-IP) assay.

The vectors containing the genes of interest were first isolated from bacterial cells. A figure of the vector is shown in Figure 4.2. Vector pcDNA3.1/V5-His/*lac*Z: 8549 nucleotides was used for *Sipa1* and a HA tagged vector (based on the same vector used for *Sipa1*) created by the NCI Protein Expression Laboratory (PEL, Frederick, USA) was used for the rest of the genes. The cloning work for the candidate genes was performed by the same group. The detail of the experimental procedure is described below.

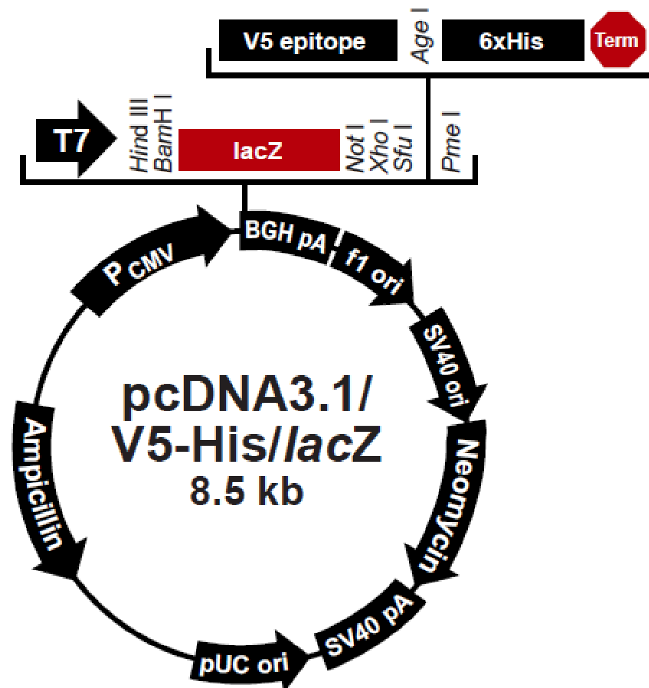


Figure 4.2: Vector map. This is amp of the vetor used for the V5 tag for Sipa1.

Vector Purification

Vectors were constructed by cloning and ligation at the Protein Expression Laboratory at the NCI Frederick Facility. Vectors were transformed into GE10 competent bacteria cells, with glycerol stock subsequently created for each vector. The cultures were grown in 37°C and shaking over night. The bacterial colonies were grown on agar plates and only one colony was picked to obtain a pure vector containing culture. Vector-containing bacterial cultures were grown for eight hours in 15 ml liquid agar culture, with 100 µl subsequently transferred to fresh 100 ml agar culture at the end of the incubation period. The 100 ml culture was incubated overnight and then ready for vector purification. Vector purification was carried out utilizing the EndoFree Plasmid Purification kit from Qiagen, USA, following the handbook's summary. Briefly, bacterial cultures were pelleted at 6000xg for 15 minutes at 4 ºC and resuspended in 10 ml of Buffer P1. Ten ml of Buffer P2 was added and mixed

thoroughly by vigorously inverting 4-6 times, followed by adding 10 ml chilled Buffer P3 and mixed by inverting vigorously 4-6 times. The lysate was poured into the barrel of the QIAfilter Cartridge and incubated at room temperature for 10 minutes. The plunger was inserted and lysate was filtered into a 50 ml tube. Two and half ml of Buffer ER was added to the filtered lysate and mixed by inverting, with the mixture incubated on ice for 30 minutes. A QIAGEN-tip 500-column was equilibrated by adding 10 ml Buffer QBT and emptied by gravity flow. The filtered lysate from the previous step was poured into the equilibrated column and allowed to enter the resin by gravity flow. The QIAGEN-tip column was washed with 2 x 30 ml Buffer QC. DNA was then eluted with 15 ml Buffer QN and precipitated by adding 10.5 ml room-temperature isopropanol to the eluted DNA, which was mixed and centrifuged at≥ 15,000 x g for 30 minutes at 4 ºC. The supernatant was removed and the DNA pellet was washed with 5 ml endotoxin-free room-temperature 70% ethanol and centrifuged at≥ 15,000 x g for 30 minutes at 4 ºC. The supernatant was removed and the pellet air-dried for 5-10 minutes, with the DNA redissolved in a suitable volume of endotoxin-free Buffer TE.

The purified vectors were then transfected into tissue culture cells, CHTG49, which is a Chinese hamster cell line that have a higher propensity for successful transfection, the handling of tissue culture cells and transfection procedures are described in detail below.

Tissue Culture

- Passaging cells

Cells were washed with 15 ml of PBS before being treated with 2 ml of Trypsin for 5 minutes. Thirteen ml of medium was then added to neutralise the Trypsin, with 500 µl of this mixture then transferred to a new T-75 flask with 14.5 ml of fresh medium. The cell passages

were all performed in T75 flasks. The cell culture was generally split 1 into 15. The cells were passaged at 80% confluency and the media used was Dμlbecco's Modified Eagle Medium High Glucose 1x (DMEM) (Gibco/ Invitrogen, USA) in 500 ml and combined with 50ml of BenchMark Fetal Bovine Serum (FBS) (Gemini Bio-Products, USA) and 5ml of Penicillin/ Streptomycin (Cellgro, USA).

- Cell freeze down

Cells in the T-75 flask were washed with PBS once and Trypsinized for 15 minutes with 2 ml of Trypsin. Trypsinized cells were neutralized with 8 ml of media and subsequently centrifuged at 100 g (~800 rpm) for 10 minutes at 4 ºC. The medium was removed and cells were resuspended in 1 ml of Cryoprotective Medium. Resuspended cells were immediately transferred into a Cryoprotective tube and placed in the NALGENE Cryo 1 ºC Freezing Container. The cells stayed in the container in a -80 freezer for overnight before transferred into a liquid nitrogen tank. The cryoprotective media is a ready to use solution Cryoprotective Medium (Lonza, MD, USA) from Lonza.

- Re-seed cells

Cryoprotective tubes containing the cell pellet and cryo-protective medium were transferred from the liquid nitrogen onto dry ice. Cells were them placed in a 37 ºC water bath for approximately two minutes and once defrosted, transferred immediately to a T-75 flask with fresh media.

Transfection

At day one, 7.5 x 10^5 of CHTG49 cells were seeded in a 10 mm dish. At day two, the transfection mixture was prepared as follows; 20 μg of vector was added to 180 μl of SuperFect Transfection Reagent and 600 μl of pure DMEM, with the solution mixed by vortexing. The transfection mixture was incubated at room temperature for 10 minuets. The

PBS coving the cells was removed and 6 ml of normal media (DMEM + FBS + Penicillin and Streptomycin) was added to the transfection solution and mixed by pipetting, which was then immediately transferred to the 10 mm dish containing the cells. The cells were incubated with this transfection mixture at 37 ºC and 5% CO_2 for 3 hours. After incubation, the transfection mixture was removed and the cells washed with PBS three times. Fresh medium was added to the cells and incubated for two days, before adding the selection antibiotics for establishing stable cell lines. For selection of Met-1 cells the puromycin selection concentrations had already been established and were 4 µg/ml.

The transient transfected CHTG49 cells were then harvested for confirmation of up-regulation of both the *Sipa1* protein and its putative protein binding partner, once the up regulation of both genes is confirmed, the same set of samples were subsequently prepared for Co-IP assays. The procedures taken up to confirmation of up-regulation of genes of interest and the subsequent Co-IP assays are described in detail below.

Protein Extraction

To extract protein out of cultured CHTG49 cells, cells were first treated with a suitable amount of Trypsin. Trypsinized cells were then neutralized with normal medium and centrifuged at 3500 x g for 10 minutes at 4 ºC. The supernatant was removed and the cells were washed with PBS and centrifuged at 3500 x g for 10 minutes at 4 ºC. The supernatant was removed and lysing buffer (M-PER : 100 x Halt Protease Inhibitor = 1 ml : 10 µl) was added and mixed, with the solution then put on a shaker at 90 RPM at 37 ºC for 5 minutes. The lysed cells were then centrifuged at 1500 x g, 4 ºC for 10 minutes and the supernatant transferred into a clean autoclaved tube for further protein analysis.

Protein Concentration Estimation

200 µl of BCA working reagent (BCA Reagent A: BCA Reagent B = 24: 1) was dispensed in each required well in a UV- transparent flat bottom 96-well plate. 25 µl of nine sequential standards Diluted Albumin (BSA) were then added with the concentrations as follows: 2000 µg/ml, 1500 µg/ml, 1000 µg/ml, 750 µg/ml, 500 µg/ml, 250 µg/ml, 125 µg/ml, 25 µg/ml and 0 µg/ml. 25 µl of samples were added to the BCA working reagent and the plate was covered with foil and incubated at 37 ºC for 30 minutes before reading at 562 nm.

SDS-PAGE and Western Blot Analysis

The concentrations of each sample were obtained after protein extraction and then the samples were diluted with the protein extraction solution (M-PER Extraction Reagents (Pierce Biotechnology, USA)) to establish equal concentration (1ug/ul) across all samples. Therefore, the same volume 16.25ul can be applied to each sample and gave the same amount of protein used for the assay. To prepare protein samples for loading on a 9-well 4-12 % Tris-Bis SDS-PAGE gel, 6.25 µl of 4 x LDS, 2.5 µl of Reducing Agent and 16.25 µl of protein sample solution were mixed, then heated at 70 ºC for 10 minutes. The SDS-PAGE electrophoresis running unit was assembled and running buffer (760 ml H_2O + 40 ml MOPS) was added to the outer chamber first (to check for leaks) and then the inner chamber, before loading the denatured protein samples to each well. The SDS-PAGE gel was run at 110 V until the loading dye reached the bottom of the gel.

To prepare the samples for Western blotting, a polyvinylidene fluoride (PVDF) blotting membrane was washed with methanol and along with 4 pieces of blotting paper and 2 transfer pads was submerged in the Western blotting transfer buffer (1400 ml H_2O + 400 ml Methanol + 200 ml 10 x Tris/Glycine Buffer). The SDS-PAGE gel was transferred onto the

PVDF membrane and the transfer sandwich assembled. The transfer sandwich was placed in the transferring tank, filled with transferring buffer and ran at 20 V overnight in the cold room at 4 ºC. The next morning the transfer unit was run at 100 V for 30 minutes at 4 ºC before transferring to a plastic membrane holder for blocking in 5% milk (5 g fat-free dry milk powder in 100 ml 1 x TBS) for seven hours at 4 ºC. The milk was rinsed off and the membrane incubated with primary antibody in fresh 5 % milk-TBS on the shaker in the cold room overnight. The primary antibody solution was removed and membrane washed five times with TBS-Tween (50 ml 10 x TBS + 450 ml H_2O + 500 µl Tween 20) the next morning. The membrane was incubated with the secondary antibody at room temperature on the shaker for one hour. The membrane was washed five times with TBS-Tween before being treated with ECL reagent (1 ml Reagent 1 + 1 ml Reagent 2) for signal development. The membrane was incubated with the ECL reagent mixture for 1 minute and placed on top of the high performance chemiluminescence film in the Kodak X-Omatic cassette for the appropriate time (between 5 minutes and then 10 minutes depending on quality of antibody and target protein concentration). The film was then developed in the Kodak film developer.

Co-Immunoprecipitation Analysis

The Co-Immunoprecipitation analysis was carried out utilizing the ProFound HA Tag IP/Co-IP Kit and Application Set. Following the instructions from the company, an appropriate amount of cell lysate was added to the Handee Mini-Spin Column (with bottom plugs on). Fifty µl of HA-tagged Positive Control diluted in 150 µl TBS was added in a separate column as a positive control. Six µl of anti-HA agarose slurry was dispensed in each column and the cap screwed on. The mixture was incubated with gentle end-over-end mixing, overnight at 4 ºC. The cap was loosened and the bottom plug removed with a collection tube placed under the column before pulse Centrifugation for 10 seconds. The column was washed by adding

0.5 ml of TBS-Tween to each column, gently inverted with the collection tube and pulse centrifuged for 10 seconds. The column was washed three times and elution of the HA-Tagged protein was performed by placing the column in a new collection tube and 25 µl 2 X Non-Reducing Sample Buffer added to the anti-HA agarose and mixed by gently tapping the tube. The column with the collection tube was warmed to 100 ºC on a heating block for 5 minutes before pulse centrifuged for 10 seconds. The eluted protein was then ready for further analysis, such as Western Blot detection.

4.2.2 *In vivo* Metastasis Assay

The Calm2 protein was identified as a protein binding partner of the Sipa1 protein in the Co-IP assay; subsequent experiments were designed to investigate the role of *Calm2* in breast cancer metastasis. This involved an *in vivo* metastasis assay on the *Calm2* up-regulated cells. The stable *Calm2* up-regulated cell line was established following the transfection methods with the same vector used in the previous section experiments. The confirmation of stable up-regulating the *Calm2* gene was assayed with quantitative real-time PCR. The procedures involved in preparations leading up to quantitative real-time PCR and the orthotopic implantation of the mammary fat-pad, are described in detail below.

RNA Extraction

RNA extraction was carried out utilizing the RNeasy Mini Kit (Qiagen, USA). The tissue cultured cells were first trypsinized and collected as a cell pellet, which was then washed with PBS. Following the RNeasy Mini Kit handbook's summary, cells were disrupted by adding Buffer RLT (350 µl) and then pipetted to mix. The cell lysate was then dispensed into a

QIAshredder (Qiagen, USA) spin column (placed in a 2 ml collection tube) and centrifuged for 2 min at full speed (~16,000 g for a bench top centrifuge). 350 µl of 70% ethanol was added to the homogenized lysate and mixed via pipetting. Up to 700 µl of the sample mixture was transferred to an RNeasy spin column (placed in a 2 ml collection tube) and centrifuged for 15 sec at ≥8000 x *g*, with the flow-through discarded. Three hundred and fifty µl of Buffer RW1 was added to the RNeasy column and spun for 15 sec at ≥8000 x *g*(discard the flow-through). DNase I was then added to Buffer RDD (10 µl DNase I Stock + 70 µl RDD Buffer per sample) and the tube was inverted gently to mix (do not vortex). 80 µl of Buffer RDD with DNase I was then added directly to the RNeasy Spin column membrane and incubated at room temperature for 15 min. 350 µl of Buffer RW1 was added to the spin column and centrifuged for 15 sec at ≥8000 x *g*(discard the flow-through), with washing of the column conducted by adding 700 µl of Buffer RW1 and centrifuging for 15 sec at ≥8000 x *g*(discard the flow-through). 500 µl of Buffer RPE with ethanol was added to the spin column and centrifuged for 15 sec at ≥8000 x *g*(discard the flow-through). To elute the RNA, the RNeasy spin column was placed in a new 1.5 ml collection tube and 30-50 µl RNase-free water was added directly to the spin column membrane and centrifuged for 1 min at ≥8000 x *g*.

Reverse-Transcriptase Reaction

The reverse-transcriptase reaction was carried out utilizing the iScript Select cDNA Synthesis Kit (Bio-Rad Laboratories, USA) following the iScript Select cDNA Synthesis Kit protocol with minor changes. To set up the reactions all components were thawed, except the iScript Reverse Transcriptase. All components were then placed on ice until use. The composition of each 20 µl reaction is as follows: 4 µl of 5x iScript Select reaction mix, 1.5 µl of Oligo(dT)$_{20}$ primer and 1.5 µl of random primer, 1 µl of iScript Reverse Transcriptase and 1 µg total

RNA from the sample. An appropriate amount of Nuclease-free water was then added to make the total to 20 µl per reaction. A master mix was first prepared and then added to each well in a 96-well plate. RNA samples were then added to each well. The entire 96-well plate with samples and reaction components was subsequently incubated in a thermo-cycler, with conditions of 5 min at 25°C, followed by 30 min at 42°C and 5 min at 85°C (to inactivate the reaction). The product cDNA was then stored at -20°C for further experiments.

Quantitative Real-Time PCR Analysis

The quantitative real-time polymerase chain reaction was carried out utilizing the QuantiTect SYBR Green PCR kit (Qiagen, USA) following the kit protocol. To set up the reaction, 2x QuantiTect SYBR green PCR Master Mix, cDNA and primers were thawed on ice till use. Each quantitative real-time reaction had 10 µl of 2x QuantiTect SYBR Green PCR Master Mix, 2 µl of cDNA, equalling to 4 µg of total cDNA) 0.3 µl of Forward and Reverse (150 nM) primers and 7.4 µl of RNase-free water to make up the reaction volume to 20 µl. The reaction solution was mixed without the cDNA thoroughly and 18 µl dispensed into each well of the 96-well PCR plates. Two µl of cDNA was subsequently added to each well and the plate was sealed and centrifuged briefly to collect the solution. The plate was then placed in the Applied Biosystems 7500 machine under the set real-time cycler conditions: 15 min of 95°C for the PCR initial activation (due to SYBR Green master mix containing Hot-Start polymerase), followed by 50 cycles of 20 sec of 95 °C for denaturation, 20 sec of 60 °C for annealing and 40 sec of 72 °C for extension. Melt curve analysis of the PCR products was performed and the results analysed by normalisation to known cDNA concentration of the house-keeper gene CypB gene to generate the absolute quantity amounts of the gene of interest by standard curve.

Orthotopic Implantation of the Mammary Fat-Pad

This procedure has been approved by the NCI-Bethesda Animal are and Use Committee from August 2, 2006 to August 1, 2009. The surgical procedure of mammary fat-pad tumour cell injection is as follows; the mice were anesthetized using a freshly prepared ketamine/xylazine dose at 0.1ml/10gm of body weight delivered intraperitonially with a sterile syringe and a 25 gauge needle. The hair around the site of injection had previously been clipped with scissors and cleaned with Betadine. A small incision (nine mm to one cm) was made using sterile scalpels, between the nipple and midline of the body. The mammary fat pad was exposed by gently everting the tissue with sterile forceps and 100,000 tumour cells in less than 100ul volume were injected into the mammary fat pad using a 27 gauge needle. The incision was closed with a wound clip (Thomas Sci Autoclip) and cleaned with Betadine. Buprenorphine was administered IP at 0.5 mg/kg with a 27 gauge needle for analgesia. The mice recovered in a warm environment and monitored until they were fully recovered from anesthesia. 1-2 cc of warm NaCl was administered sub-qutaneous to help prevent postoperative complications. The mice were observed 4-12 hours post-operation and again at 24 hours for signs of pain, infection or dehiscence. The surgery and dates were noted on the cage card along with the dates and times of each observation. Wound clips were removed 10 days post-surgery.

4.3 Results

The results of this section of research identified a novel *Sipa1* binding partner. Subsequent *in vivo* metastasis assay on this novel protein binding partner (*Calm2*) showed potential involvement in breast cancer metastasis; however, additional assays are needed for the later part of result to be conclusive. In order to increase the relevance of the result presentation, the negative results were not included in the presentation. This showed consistency in the principle of data presentation across different chapters and assays.

4.3.1 *Calm2* Interacting with *Sipa1* in CHTG49 Cell Line

To confirm the result of the yeast-two-hybrid assay, the Sipa 1 vector (with V5 tag) and vectors containing each gene (with HA tag) were first constructed by NCI, Fredrick Core Facility. GT49 cells were then co-transfected with epitope-tagged mouse candidate genes (*p21Arc*, *Npc1*, *Gtf2H2*, *Gart*, *Acin1* and *Calm2*) and *Sipa1* to ensure adequate levels of both proteins for optimal detection. Successful up-regulation on both *Sipa1* gene and its potential protein binding partner was identified in the *Calm2* + *Sipa1* dual transfection and *Acin1* + *Sipa1* dual transfection. None of the other pairs (*Sipa1* + *p21Arc*; *Sipa1* + *Gtf2H2* and *Sipa1* + *Gart*) showed up-regulation of both proteins and therefore excluded for further assays.

The Western blot assay identified *Sipa1* + *Calm2* and *Sipa1* + *Acin1* up-regulation and is presented in Figure 4.3; one half of the blot was probed with anti-V5 tag primary antibody, aiming to identify the V5 tagged *Sipa1* proteins, while the other half was probed with anti-HA tag primary antibody, aiming to identify the HA tagged *Sipa1* protein binding partners

(*Calm2*, *Acin1* and *p21Arc*). The anti-V5 blot indicated that all three pairs (*Sipa1* + *Calm2*; *Sipa1* + *Acin1* and *Sipa1* + *p21Arc*) had *Sipa1* (MW: 112 kDa) up-regulation (Figure 4.2, left hand side blot), however, the HA blot indicated that *Calm2* (MW: 16.7 kDa) and *Acin1* (MW: 67 kDa) were over-expressed in the dual transfection while *p21Arc* (MW: 20.4 kDa) was not up-regulated (Figure 4.3, right hand side blot). Significant additional binding of the HA antibody was observed in the HA blot. This could be due to the strong sensitivity of the HA antibody and is a potential downfall of this approach. Due to limited laboratory budget and the limited availability of antibodies against indigenous proteins, each of the proteins was not probed using specific antibodies.

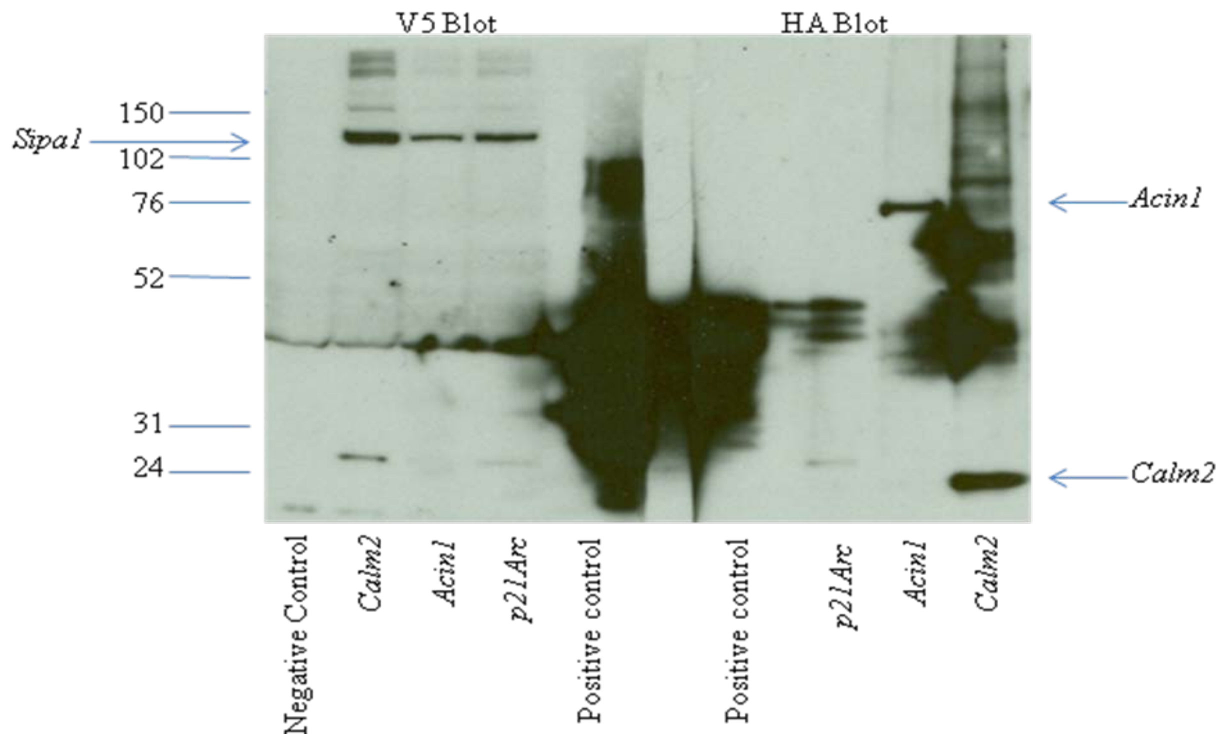


Figure 4.3: *Sipa1*, *Calm2* and *Acin1* up-regulation confirmation. The left side of blot was probed for the V5 tag, the bands between the 102 kDa and 150 kDa indicated the presence of V5 tagged *Sipa1* proteins (MW: 112 kDa) in all three dual over-expression pairs (*Sipa1* + *Calm2*; *Sipa1* + *Acin1* and *Sipa1* + *p21Arc*). Negative control (no transfection cell lysate) and positive control (posi-tag epitope tag control protein) were included. Positive and negative (not shown in figure) controls were also included for the right side of blot (probed for the HA tag). The band on the *Calm2*+ lane, below 24 kDa mark, indicated the presence of HA tagged *Calm2* protein (MW: 16.7 kDa) over-expression and the close to 76 kDa mark band on the *Acin1*+ lane indicated over-expression of HA tagged *Acin1* protein (MW: 67 kDa). The lack of prominent band around the 24 kDa mark in the *p21Arc*+ lane indicated no HA tagged *p21Arc* protein over-expression is this dual transfection.

The above result indicated that both *Sipa1* and *Calm2* were over-expressed and both *Sipa1* and *Acin1* were over-expressed in the dual transfection. This dual up-regulation would permit adequate amount of both proteins for optimal detection of protein interaction and binding in the immunoprecipitation assays. Therefore, the *Sipa1* + *Calm2* dual-up-regulated sample and *Sipa1* + *Acin1* dual-up-regulated sample were subsequently immunoprecipitated for HA tag, utilizing the ProFound HA Tag IP/Co-IP Kit and Application Set (Pierce, USA). The ProFound HA Tag IP/Co-IP Kit and Application Set pulled down the HA tagged protein (the

Calm2 protein and the *Acin1* protein) by trapping the proteins with the beads in the column provided in the kit (the *Calm2* and *Acin1* protein were assayed in two separated columns). Any protein that was physically bounded with *Calm2* or *Acin1* would remain in the column. The Western blot assays analysing the eluted content of the columns then revealed the binding partners of *Calm2* and *Acin1*. The Western blot assays on the eluted proteins were probed for HA tag (*Calm2* and *Acin1*) for confirmation of the IP experiment and probed for V5 tag for *Sipa1*.

The result of the Co-IP assays is presented in Figure 4.4. The results indicated that *Calm2* is physically interacting with *Sipa1*, while *Acin1* is not physically interacting with *Sipa1*. The HA blot (Figure 4.4, left side) was probed for HA tagged *Calm2* and *Acin1* proteins. The little amount of *Acin1* identified could be a reason for the failure in *Sipa1* detection. Alternative methods such as detecting the endogenous proteins was not a viable option as the *Sipa1* antibodies available have extremely high background and thus would be quite impedimentary to a Co-IP study. The prominent band in the *Calm2*+ lane, positioned around the 24 kDa mark, indicated that the Co-IP assay successfully immunoprecipitated the HA tagged *Calm2* proteins. The visible band located below the 76 kDa mark in the *Acin1*+ lane indicated immunoprecipitation of the HA tagged *Acin1* proteins. The V5 blot (Figure 4.4, right side) showed the presence of the V5 tagged *Sipa1* protein in the *Calm2*+ lane (Figure 4.4, right side, between 102 and 150 kDa) but not the *Acin1*+ lane, indicating HA tagged *Calm2* protein interacted with the V5 tagged *Sipa1* protein while the HA tagged *Acin1* protein didn't bind to the V5 tagged *Sipa1* protein.

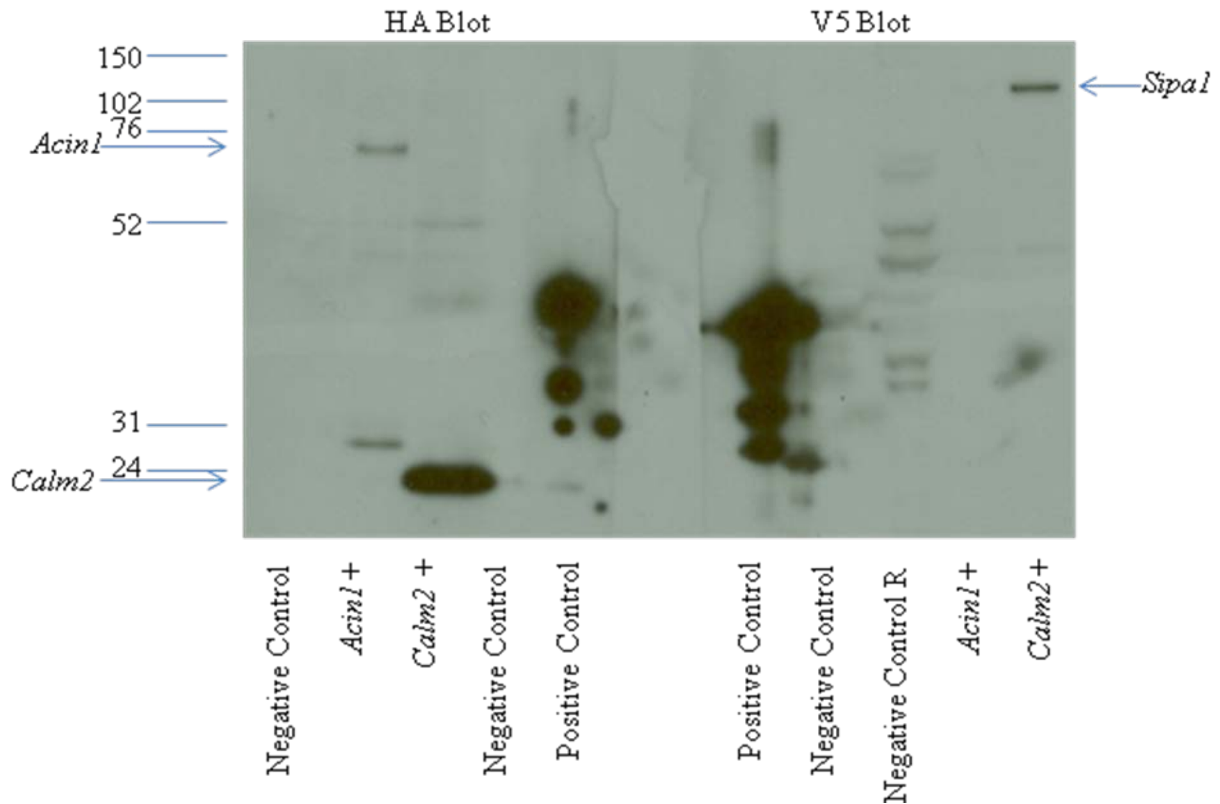


Figure 4.4: *Sipa1*, *Calm2* and *Acin1* Co-IP Assay. The left side of blot was probed for the HA tagged *Calm2* proteins and *Acin1* proteins, the band on the *Calm2*+ lane around the 24 kDa mark indicated the presence of HA tagged *Calm2* protein (MW: 16.7 kDa) in the elute of the Co-IP assay and the band in the *Acin1*+ lane located close to 76 kDa mark indicated the presence of HA tagged *Acin1* protein (MW: 67 kDa) in the elute of the Co-IP assay. Negative control from the Co-IP assay (Negative Co lane) and positive control (posi-tag epitope tag control protein, Covance, CA, USA) were included. The right side of blot probed for the V5 tagged *Sipa1* also included positive and negative (from cell lysate without Co-IP assay) controls. The right side of blot was probed for the V5 tagged *Sipa1* protein in the Co-IP elutes, the bands between the 102 kDa and 150 kDa indicated the presence of V5 tagged *Sipa1* proteins (MW: 112 kDa) in the *Calm2*+ Co-IP sample. No bands were shown in the *Acin1*+ lane indicating that no V5 tagged *Sipa1* protein was present after the Co-IP assay.

4.3.2 Calm 2 Up-Regulated Clone Selection

To determine if up-regulating the protein expression of *Calm2* by itself was sufficient to modulate breast cancer metastasis; *in vivo* assays on *Calm2* were conducted. The *in vivo* assay required cell lines with stable over-expression of *Calm2*; this was achieved by generating stable *Calm2* over-expressing clones in a mouse mammary metastatic cell line (Met-1). The Met-1 cell line is moderately aggressive in generating lung metastasis when implanted either orthotopically in the mouse mammary fat-pad or injected into the tail vein.

The control Met-1 clones, which were clones with stable over-expression of the control gene β-Galactosidase (β-Gal) were generated (with the pCMV-SPORT-β-Galactosidase vector (Invitrogen) and the identity of the vector was sequence verified before transfection. Stable *Calm2* up-regulated Met-1 clones were generated as described in the methods section and the up-regulation was confirmed by quantitative real-time PCR. Clones Q, T, K, and J showed 1.5 to 2 fold up-regulation of *Calm2* compared to the control β-Gal clones (Clone 7 and 8) (Figure 4.5) and were selected for the *in vivo* mammary fat-pad implantation assay, which is described in the next section.

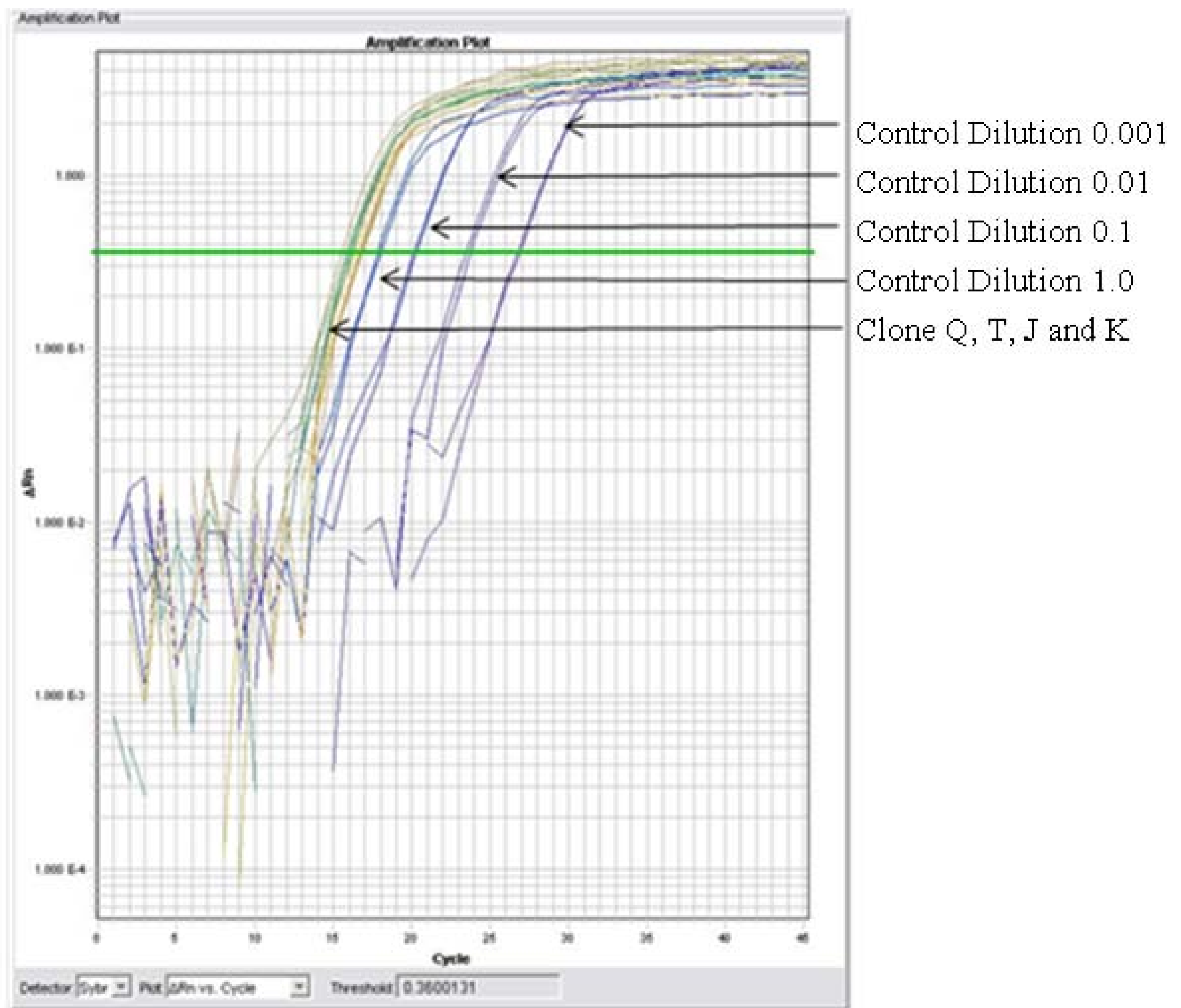


Figure 4.5: *Calm2* up-regulation qPCR plot. The standard curve was generated by serially diluting the control β-Gal cDNA (from Control 7 and 8) by 1/10, 1/100 and 1/1000 respectively. Those samples (clones Q, T, J and K) with expression significantly above the concentration of the base-line undiluted â-Gal (control 1), have over-expression of *Calm2*. The samples were analyzed in inter-run triplicates and intra-run duplicates, with the control samples having inter-run and intra-run duplicates.

4.3.3 Calm2 Up-Regulated Clones Implantation in Mouse Mammary Fat-Pad

The *Calm2* expressing and β-Gal control clones were orthotopically injected into the mammary fat pad of adult female virgin FVB. 250,000 cells from each Calm2 and control clones were injected. Five mice were allocated for each sample (Q, T, J and K) and control (clone 7 and 8) clone. The mice were sacrificed 6 weeks later due to humanitarian reasons; the primary tumour from control clone 7 mice 2 was ulcerated, however no other control clones or any *Calm2* clone mice had a palpable primary tumour.

Control clone 7 mice were the only set of mice presenting primary tumour, and the sizes (in gram) of the primary tumours were summarised in Table 4.1.

Table 4.1: Primary Tumour Weight of Control Clone 7.

Control Clone 7, Mouse Number	1	2	3	4	5
Primary Tumour Weight (g)	0.4	1.3	0.5	0.3	0.5

The lungs of the mice were harvested for sections (3 step sections, drop 20, H & E stain, snow coat slides) to observe the amount of lung metastasis. No lung metastasis was present in either the control clones or the sample clones. Since no metastasis or primary tumour was observed in the control clones, additional *in vivo* assays are needed to draw conclusion of the *in vivo* property of the *Calm2* gene.

4.4 Summary of the Novel *Sipa1* Protein Binding Partner Identification Results

The aim was to identify *Sipa1* protein binding partners from the list of potential interactors obtained from the yeast-2-hybrid assays. In order to have adequate protein expression of both *Sipa1* and its potential binding partners for Co-IP, dual over-expression of targeted proteins and *Sipa1* was performed. Epitope tagged constructs were generated by the Protein expression laboratory at NCI Fredrick using the Gateway system for rapid assembly and verification. Whilst it is known that epitope tagging of proteins can impair both function and interaction, the cost of employing native antibodies to all the native proteins of interest would not make this project feasible. Western blot assays checking the expression of targeted proteins and *Sipa1* protein revealed that *Calm2* and *Sipa1* were successfully dual over-expressed in a transient transfection in CHTG49 cells, and *Acin1* and *Sipa1* were also successfully dual over-expressed. Those two samples were subsequently taken for Co-IP assays. The Co-IP assays revealed that *Calm2* physically bound to *Sipa1*, while *Acin1* didn't bind to *Sipa1* in this assay.

The second aim of this part of the research was to investigate the properties of the newly identified *Sipa1* protein binding partner, Calm2, in the context of breast cancer metastasis modulation. In order to study Calm2 in breast cancer metastasis, stable *Calm2* up-regulated clones were generated. The stable over-expression of *Calm2* in the Met-1 cells were identified and confirmed by quantitative real-time PCR; with four clones selected for *in vivo* metastasis studies. The *in vivo* study of Calm2 was conducted with orthotopical implantation of the *Calm2* over-expressed cells from each clone into mouse mammary fat-pad and assessment of the lung metastasis six weeks post injection. The *Calm2* Met-1 clones were

compared to the control cells, which were clones transfected with the control β-Galactosidase gene. Only one control clone successfully generated primary tumours in mice, and no other clones, including *Calm2* over-expressed clones, gave rise to any visible primary tumour. None of the animals had lung metastasis in the subsequent pathology examination of their lung sections.

In summary, the *in vitro* assays on the *Sipa1* protein identified *Calm2* as a novel protein binding partner of the breast cancer metastasis promoter *Sipa1*. The *in vivo* metastasis assay on the *Calm2* had no lung metastasis in the control clones, therefore, no strong conclusions can be drawn as to whether *Calm2* is directly involved in breast cancer metastasis, with further assessments required in another/other cell line/s to determine what role, if any, *Calm2* protein expression plays in breast cancer progression.

Calmodulin 2 is a major calcium binding protein in the cytoplasm and has not been reported to be associated with cancer progression; only a recent study reported a difference in *Calm2* expression in anaplastic large cell lymphoma (Rust *et al*, 2005). Hence, the novel identification of partnership between *Calm2* and *Sipa1* may suggest a possible novel involvement of *Calm2* in breast cancer. However, more *in vivo* and *in vitro* analyses of *Calm2* in breast cancer and metastasis are needed to draw conclusions.

A previous study demonstrated that the localization of *Sipa1* protein binding partners dictates the localization of *Sipa1* and therefore, the identification of a novel cytoplasmic binding partner of *Sipa1* is of particular importance. This is in part due to the observation that *Sipa1*'s function alters according to its cellular localisation and it is possible that the *Calm2* protein

localises the *Sipa1* proteins to the cytoplasm and leads to other *Sipa1* involved breast cancer metastasis pathways and mechanisms.

Additionally, cellular calcium level modulation has been linked to cancer progression (Shirasaki *et al*., 2006). It is plausible that *Sipa1* and *Calm2* may be involved in breast cancer metastasis independently and/or dependent on calcium modulation, since *Calm2* is a major cytoplasmic calcium binding protein. it therefore, may be of critical importance to further characterise *Calm2* and *Sipa1* independent and dependent to each other in calcium regulation in breast cancer metastasis.

Chapter 5

Summary and Discussion of Results

5.1 Introduction

The aim of this research was to verify potential candidate breast cancer metastasis modulator genes, which were identified in the mouse model, in the human population. The hypothesis of this research is that subtle coding variations in putative breast cancer metastasis genes that were identified through literature review and microarray data from differential metastasis potency mice are breast cancer prognostic and survival indicators in human breast cancer. It has been established that subtle genetic variations in the mouse genome can lead to differential metastasis potency and thus this study theorises that subtle genomic variations in the human candidate genes can modulate human breast cancer characteristics and survival rates. This hypothesis is supported by the common disease-common variant hypothesis, which suggests complex traits, such as cancer metastasis, are most likely shaped by multiple genes that exert weak allelic effects. These weak effects have differential distributions of allele frequencies and effect sizes across different populations and therefore common traits are most likely due to common variants. Single nucleotide polymorphisms (SNPs), being the most common genetic variations in the genome, were therefore chosen as the tool for exploring the potential association of the candidate genes and human breast cancer metastasis. Significant and reproducible associations between SNPs from the candidate genes and clinical markers of human breast cancer metastasis and survival were shown in this research, which contributes to the validation of these candidate genes as novel human breast cancer metastasis modulator genes. In addition, a molecular study was performed on one of the breast cancer metastasis modulator genes (*Sipa1*) to identify novel protein binding partners, which led to the identification and characterisation of calmodulin 2 and its potential role in breast cancer metastasis.

Cancer metastasis, the spread of tumour cells from its primary site and colonization at secondary sites, is responsible for nearly 90% of cancer related mortality and understanding this lethal part of carcinogenesis is therefore an urgent task. Identifying and characterizing genes that modulate the metastatic process will facilitate better identification and treatment of patients who are at risk of developing cancer metastasis.

Recent research demonstrated that inherited genetic variation is one of the intrinsic modulators for metastasis and its susceptibility (Park *et al.*, 2005). The Hunter lab showed that germline polymorphisms had significant influence on transgene-induced mouse metastasis capability. The model they described used the highly metastatic Polyoma middle-T transgenic mouse bred to mice from the AKXD panel, which have diverse, yet well mapped genetic backgrounds, with the resultant F1 progenies having a wide range of metastasis potency. Since all cancerous dissemination in the F1 progenies were introduced by the same level of the oncogenic antigen PyMT, the difference in metastasis efficiency observed is best explained by germline variation in the F1 progenies, which they obtained from the different paternal inbred strains (Park *et al.*, 2005). Subsequent research into the genetics underlining the difference in metastasis strength via quantitative trait locus analyses and literature review led to the identification of several candidate metastasis modulator genes.

To translate these findings, association studies exploring the possible involvement between the candidate genes and breast cancer clinical markers, survival and metastasis were conducted to examine whether the potential metastasis modulators identified in the mouse model are associated with human breast cancer metastasis. In addition, molecular characterization of the high priority *Sipa1* gene (precedence ranked by preliminary molecular

and epidemiology studies) was conducted to gain further insight into the functionality of this gene in breast cancer metastasis.

5.2 Summary and Discussion of the Epidemiology Studies

The aim of the epidemiology study was to validate potential breast cancer metastasis modulator genes in human breast cancer cohorts that possess clinical follow-up data. Reproducible preliminary associations between single nucleotide polymorphisms (SNPs) in the candidate genes and breast cancer clinical markers and survival were observed in this research, which provide additional evidence supporting the candidate genes as novel human breast cancer metastasis susceptibility genes.

To address the hypothesis that candidate genes are involved in human breast cancer and metastasis, single nucleotide polymorphisms, the most common genetic variation in the genome, from the implicated candidate genes, were studied in relation to clinical breast cancer markers and survival. Polymorphism association studies provide additional benefit in exploring cancer metastasis, which doesn't show a simple Mendelian inheritance pattern and is most likely shaped by multiple genes that exert weak allelic effects, with different distributions of allele frequencies in a population. Studying the most frequent genetic variation (i.e SNPs) gives greater power in observing the possible association between the complex phenotype of cancer metastasis and its genetics (Erichsen and Chanock, 2004). This research demonstrated numerous and various associations between the SNPs in novel metastasis modulating genes and breast cancer clinical features, metastasis and survival, supporting a role for these candidate genes as novel cancer metastasis modulating genes and providing evidence that subtle genetic variation can systematically impact upon metastasis severity and susceptibility.

This research showed that not only are these candidate genes engaged in breast cancer at the inherited genetic level, but the methodology employed is also advantageous for association studies when sample resources are limited. The genotyping assays were carried out utilizing ABI genotyping methodology (Applied Biosystems, USA) and DNA from multiple populations, to identify and confirm genotypes for the implicated candidate genes. Assays were conducted in the 384-well plate platform with dried down DNA, which gives several advantages; the 384-well plate platform allowed processing of large amounts of samples in a short period of time, the amount of DNA required in each reaction can be as low as 10ng per reaction and half the amount of the recommended Genotyping Assay kit reagents was required in each reaction in this research, which gave good financial advantage.

Single nucleotide polymorphisms that strongly represent the candidate genes, were selected based on their LD block and haplotype tagging; this yielded 49 candidate SNPs from 16 mouse study implicated candidate genes to be screened in two independent pilot cohorts. The results of the pilot study, along with molecular studies results gathered in the lab and the literature, yielded a list of candidate genes and their candidate SNPs with high to low priority for additional genotyping assays. Those SNPs that showed association with breast cancer survival in both pilot populations and that possessed clearly identifiable metastasis modulator functions were ranked as higher priority. This elucidated four candidate genes and seven candidate SNPs for analyses in the six additional larger independent cohorts.

5.2.1 The *ARAP3* gene and its candidate SNPs (SNP rs440279 and rs3763120)

5.2.1.1 Introduction of the *ARAP3* Gene

ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 (*ARAP3*) gene was previously identified as a candidate breast cancer progression and metastasis modifier gene in the AKXD/PyMT mouse model, by its high mRNA transcript correlation with extra-cellular matrix (ECM) genes signature, which is a marker signature for metastasis. The *ARAP3* gene encodes a phosphoinositide binding protein, containing ARF-GAP, RHO-GAP, *RAS*-associating, and pleckstrin homology domains. It is known to participate in cell cytoskeleton rearrangement and cell shape change, which have been linked to cell migration and cancer metastasis previously (Schneider *et al*., 2008). *ARAP3* has recently been shown to interact with PI(3,4,5)P3 phosphatase SHIP2 (inositol polyphosphate phosphatase-like 1) (Raaijmakers *et al*., 2007), which was shown in recently published research to be a positive regulator for cytoskeleton remodelling, cell adhesion and spreading (Prasad, 2009). These molecular research data support *ARAP3* protein's engagement in cytoskeleton remodelling and cell migration, spreading mechanisms.

The *ARAP3* gene is also heavily associated with breast cancer at the transcriptional level; with *ARAP3* gene expression shown to be predictive of breast cancer survival. When comparing survival benefit in a well characterized Dutch breast cancer data (van de Vijver *et al*, 2002), significant survival differences were observed in gene expression signatures induced by activation of *ARAP3* (Crawford *et al*., 2007), with its expression found to be lower in breast cancer tissues compared to normal mammary tissues in another independent study (Saal *et al*., 2007) and in oestrogen receptor negative breast tumours compared to

oestrogen receptor positive breast tumours in a different study (Ivshina *et al.*, 2006). In addition, *ARAP3* expression was also linked to breast cancer histology grade, with lower expression of *ARAP3* observed in high-grade 3 tumours compared to lower grade 1 and 2 tumours (Richardson *et al.*, 2006).

5.2.1.2 Summary of the *ARAP3* Gene SNPs Genotyping Results

This research was undertaken to determine whether the *ARAP3* gene is engaged in breast cancer and metastasis at the intrinsic genetic level and investigated the relation of single nucleotide polymorphisms in this gene and breast cancer clinical markers and survival. There were seven SNPs chosen to represent the large *ARAP3* gene, candidate SNPs rs440279 and rs3763120 showed association with breast cancer patient survival in both pilot cohorts in the multivariate adjusted analyses and were assayed in additional independent larger cohorts. Two additional SNPs (SNP rs6891143 and rs4912610) showed association with survival in only one of the pilot cohorts, since this observation was not reproduced; these two SNPs were not assayed in other cohorts.

Of particular interest, the candidate SNPs rs440279 and rs3763120 (in the same LD block) from the *ARAP3* gene were found to be associated with various clinical markers in some of the larger tested cohorts. The GG genotype from candidate SNP rs440279 was found to be associated with smaller tumour size, negative progesterone receptor tumours, negative oestrogen receptor tumours (Umea, Sweden cohort) and non-localized tumours (Southern California, USA cohort). The AA genotype from candidate SNP rs3763120 was found to be associated with larger tumour size (Greater Baltimore, USA cohort), with the variant A allele associated with positive oestrogen receptor tumours, negative progesterone receptor tumours, and higher tumour grade (FHCRC, USA cohort). The A allele or the AA genotype from SNP

rs3763120 was associated with larger tumour size found in the Southern California cohort, which is consistent with the result of higher tumour grade, found in the FHCRC cohort. The GG genotype from SNP rs440279, on the other hand, did not give consistent results; this genotype was found to be associated with smaller tumour size in the Umea, Sweden population, yet with non-localized tumour (higher stage) in the Southern California cohort. Both candidate SNPs were found to be associated with hormonal status, yet these results were not confirmed by other populations. These results, with some marginally significant un-adjusted *p*-values, showed promising association of the SNPs and clinical markers, yet no firm associations could be derived from these results.

The two candidate SNPs from the *ARAP3* gene were not found to be associated with lymph node metastasis in any of the populations screened; this may be explained by the low records of such information in some of the datasets, for example, the Umea, Sweden cohort only has 14 (out of 801) records with distance metastasis recording. However, those SNPs were found to be associated with breast cancer patient survival across different independent cohorts. The *ARAP3* gene SNP rs3763120 GG genotype was found to be associated with poor survival rate (Southern California, USA and Greater Baltimore, USA pilot cohorts); however, rs440279 GG genotype was found to be associated with better survival rate (Southern California, USA and Greater Baltimore, USA pilot cohorts). It could be that the G alleles of each SNP are in disequilibrium. Since I can't get access to the information of whether the G allele of each SNP are in disequilibrium or not, I can't confirm this hypothesis. The association of the SNP rs440279 GG genotype with better survival was confirmed in the Rotterdam, Netherlands population, with better survival rate of the variant G allele identified in metastasis-free survival, disease-free survival and overall survival. The heterozygous CG genotype of this SNP was found to be associated with poor survival rate (FHCRC, USA cohort). These

consistent associations with survival across different independent cohorts, which has diverse patient selections, with different statistical analysis methods, provide confidence of the association observed. Even though some associations in some cohorts may only be marginally significant in statistical terms, the multiple positive cohort results and the same direction of association adds confidence to a role of this gene in breast cancer.

To further analyse the survival associations observed in the Rotterdam, Netherland population, the patient population was subsequently stratified by lymph node metastasis status and oestrogen receptor metastasis status, with, the variant G allele found to be associated with favourable disease-free survival and metastasis-free survival in lymph node metastasis positive, with oestrogen receptor positive, patient subgroup and not in any other subgroup. These findings need to be interpreted with caution and would be best viewed as promising associations, since the events for each subgroup, especially when combined into the two stratifications, are rather small.

No association with the clinical markers was observed in the German population, with the DNA samples extracted from the bone marrow of breast cancer patients with bone metastasis. This finding and the limitation of this experiment was the poor DNA quality used in this population; the DNA was supplied at low concentration (lower than the requirement for the whole genome amplification process: 10ng/ul) and the reconcentrating step (drying and rehydrating sample DNA) may have further damaged the already weak sample DNA. The whole genome amplification process can introduce genotyping errors as the amplification process doesn't have a accurate proof reading mechanism. The low concentration of DNA may further disrupt the amplification process as the kinetics of the process may be disrupted by the insufficient amount of DNA. The genotyping assay on this population yielded a very low call rate of samples (<10% call rate), which led to low amounts of samples for analysis.

No significant association between the candidate SNPs and breast cancer incidence was identified in the Australian population as well (a case-control population). The call rate was lower in this cohort (~ 40%) compared to other cohorts (>95%), which might also lead to insufficient amount of samples for analysis or low power of results. Similar to the German cohort, the poor DNA quality may be the limitation of the experiment.

The Minnesota, USA cohort which allowed an investigation of the possible associations between the candidate SNPs and mammographic dense area, an independent marker for breast cancer risk (Reeves *et al*., 2009), showed preliminary indication of association with the *ARAP3* gene candidate SNPs. Increased mammary density was an indicator of increased risk of breast cancer (Reeves *et al*., 2009), candidate SNP rs440279 from *ARAP3* was found to be associated with mammary dense area, in that the GG genotype (homozygous variant) appeared to be associated with larger dense area compared to the other two genotypes. However, this result was not sustained in further analyses and was regarded as a preliminary indication of putative association of the candidate SNP with mammary dense area.

Inconsistent findings between different cohorts were observed and this inconsistency can be explained by the limitations of this research. Since the epidemiology studies are based on a serious of collaborations with different research labs across the world, each population has its unique patient selection and different methods of reporting important clinical markers, for example, some populations have tumour grade reported, yet the grading system is different between populations; similar situations can also be found in oestrogen receptor and progesterone receptor detection that different cohorts used different methods and threshold to detect the presence of these receptors. The various methods used in analyses different cohorts

and diverse nature of each population may facilitate slight inconsistency in the findings, when making comparison across independent cohorts.

In the interest of discovering potential prognostic association, the *ARAP3* gene and its SNPs, which to the best of the author's knowledge, have not been investigated before; the patients from Rotterdam, Netherlands were stratified with treatment. The AA genotype of SNP rs3763120 is found to be associated with poor response to first line chemotherapy. Patients carrying the GG genotype of SNP rs440279 have worse response for tamoxifen therapy; first line chemotherapy receipt patients carrying the same genotype of the same SNP have better progression-free survival. This could once again be explained by false positive results due to small sample size and result can be best viewed as an indicator of potential associations. It is also possible that *ARAP3* could mediate drug responses, however, little molecular function of this gene is know and no evidence suggests that this gene maybe involved in drug metabolism. If these data are proven to be true, with additional populations and samples, the biological explanation of this observation would be that patients with this genotype have worse initial response for tamoxifen therapy, but with more rapid relapse or resistance to other first line treatment. This SNP could be a potential SNP to help delineate what treatment a patient should receive. The idea is that by combining SNPs, such as the SNPs in the *SIPA1* and *RRP1B* SNP combination study, it could facilitate the development of a panel that has high specificity in delineating what treatment a patient should receive and apply this information in the clinic.

5.2.1.3 Discussion of *ARAP3* Gene SNPs Genotyping Results

In summary, the *ARAP3*, SNP rs3763120 was not found to be associated with survival in the larger independent cohorts however SNP rs440279 was found to be associated with survival

in additional larger populations, as well as the original pilot cohorts. These survival association data, even though not ideally reproduced, add evidence to support a role for the *ARAP3* gene in breast cancer survival and metastasis at the basic genetic level and support the view that it could be a novel breast cancer metastasis/survival modulating gene. It also further supports the hypothesis that subtle inherited genetic variation is directly associated with outcome. The result of SNP rs440279 associating with poor progression-free survival in patient receipt of first line chemotherapy in the Rotterdam, Netherlands population is especially interesting, since this is the first time that the *ARAP3* gene and its SNP have been reported to be associated with treatment related response. This suggests plausible molecular roles for the *ARAP3* gene and its protein product in tumour biology, in areas of drug response in relation to breast tumour metastasis and survival.

If the *ARAP3* association is proven to be true, additional biological rational may need to be offered to explain the findings that the rs440279 and possibly the SNP rs3763120 appear to be the SNPs causing the observed effects and they are intronic SNPs. The explanation of these intronic SNPs located in LD with other important functional SNPs would be the most logical explanation, however, these intronic SNPs has been check by the collaborators to determine whether they are located in the same LD with another functional SNP. Unfortunately, none of these intronic SNPs were in the same LD of other known functional SNPs. An alternative explanation could be that since intronic-mediated microRNA mediates mRNA interference, which is a mechanism for post-transcriptional gene silencing and modulating (Lin *et al*., 2008), it is possible that the polymorphisms in this intronic region of the *ARAP3* gene may alter or affect the microRNA sequence, therefore changing the expression of the target gene, and subsequently changing tumour cell biology. However it is not known whether these SNPs are located at such microRNA locations or not and which

gene/genes they potentially modulate. No other evidence suggests that these intronic SNPs are part of miRNAs. However, this possibility can't be ruled out, as the correlation of these SNPs with clinical characteristics was quite consistent and the database for legitimate miRNAs expands on a regular basis. The miRbase program showed that some sequences that contain valid SNPs, such as rs440279, might be part of miRNA mature sequences. The sequence that contains rs440279 is part of an evolutionary conserved region and has been shown to express mature miRNAs in various other organisms, such as cte-miR-210 from Capitella teleta. However, no evidence has been published pertaining to a miRNA sequence located in this region for *Homo sapiens* and the SNP itself is only present in humans. It is also possible that the observations obtained in this research are caused by other SNPs located in the same LD block where SNP rs440279 and rs3763120 reside. However, this still implies that the *ARAP3* gene is involved in breast cancer biology and survival in the inherited genetic level, which adds validation of the *ARAP3* gene as novel breast cancer metastasis and survival modulating gene. Interestingly, a recent publication reported that extra-nuclear signalling of oestrogen receptor mediates breast cancer cytoskeleton remodelling, with enhancement of breast cancer cell migration and invasion (Giretti *et al*., 2008). This provides plausible linkage between the observed association of the *ARAP3* gene SNPs and oestrogen receptor stratified breast cancer and its metastasis.

In addition, the associations observed between the oestrogen receptor status and SNPs rs440279 and rs3763120 in this study suggested a possible biological interaction between the transcription factor and the intronic sequences of the *ARAP3* gene. Programs such as MatInspector and DBISS are freely available on the internet and can identify potential transcription factor binding sites within a given sequence. I have checked these two website, unfortunately, I can't get access to the program in both program. MatInspector has a program

for investigating SNPs on possible transcription binding site. I can't utilize that program. DBISS required payment for analyses, without the proper support, I can't get access to this program either. To examine whether if this section of the *ARAP3* gene has functional significance for a theoretical binding to the oestrogen receptor (*ESR1*), an EMSA (electrophoretic mobility shift assay) was conducted for the rs440279 intronic sequence (data not presented in this thesis) to determine if any binding of these two transcripts (*ARAP3* and *ESR1)* exists and if it did, does the polymorphism alter its binding efficiency? However, due to technical difficulties with biotinylating the intronic sequences, no valid data could be generated despite many attempts at optimising. Alternative methods such as resequencing of the gene and functional analyses of potentially implicated SNPs, for example, the Electrophoretic Mobility Shift Assay (EMSA), would be important in identifying the true SNP.

These epidemiology results plus other published results support a role for the *ARAP3* gene in breast cancer carcinogenesis at the genetic, expression (expression signature predicts outcome) and proteomic (*ARAP3* protein in cytoskeleton remodelling) level; supporting the ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 gene as a novel breast cancer metastasis predisposition gene.

5.2.2 The *SIPA1* gene and its candidate SNPs (SNP rs746429, rs931127, rs2448490 and rs3741378)

5.2.2.1 Introduction of the *SIPA1* Gene

In the attempt to identify potential metastasis modulators, gene expression profiles were compared between mice with high metastasis potency and low metastasis potency; loci in the mouse genome that substantially influence the metastatic efficiency of mammary tumours were then mapped. The metastasis efficiency modifier locus *Mtes1* was identified and signal-induced proliferation-associated 1 (*Sipa1*) was found to be a candidate gene underlying this locus; analysis of the mouse *Sipa1* gene identified a nonsynonymous amino acid polymorphism, which affects the *Sipa1* Rap-GAP function and changes lung metastasis in the mouse model, which supports *Sipa1* as novel metastasis modulator (Park *et al.*, 2005). Spontaneous metastasis assays using ectopically expressed *Sipa1* cells or cells with knocked-down *Sipa1* expression demonstrated that metastatic capacity was correlated with cellular *Sipa1* levels. Molecular studies revealed that *Sipa1* is a mitogen-inducible GTPase activating (GAP) protein and is specific for members of the Ras-related proteins: Rap1 and Rap2 (Kurachi *et al.*, 1997). *In vitro* studies revealed that abnormally and/or prematurely expressing the *Sipa1* protein severely impedes the mitogen-induced cell cycle progression (Hattori *et al.*, 1995). The *Sipa1* protein was also found to play a role in cell adhesion via its Rap1GTP domain; with *Sipa1* suppression (via RNAi) inducing increased cell adhesion and over-expression reducing cell attachment (Park *et al.*, 2005). The protein binding partners of *Sipa1* include highly important metastasis modulating candidate genes, such as *Rrp1b* and *Brd4*, which are both candidate metastasis predisposition genes. These data suggest mouse *Sipa1* is strongly engaged in mouse carcinogenesis and metastasis. This research aimed to translate the association of *Sipa1* and breast cancer metastasis into the human population.

Candidate SNPs of the *SIPA1* gene were genotyped in multiple independent patient populations to determine whether the *SIPA1* gene is engaged in human breast cancer and metastasis at the inherited genetic level. Initially there were three SNPs selected to represent this gene; SNP rs746429 (synonymous Ala to Ala) and rs3741378 (missense) are exonic SNPs, rs931127 is a 5' (near gene) SNP, within the promoter region of the *SIPA1* gene.

5.2.2.2 Summary of the *SIPA1* Gene SNPs Genotyping Results

The *SIPA1* SNPs were found to be associated with breast cancer clinical markers. In one of the pilot populations (Southern California, USA cohort), the variant T allele from SNP rs3741378 was found to be associated with negative ER and negative PR, which suggest this allele as an indicator of more aggressive disease (Crawford *et al*., 2006). A related finding was found in the large population (Rotterdam, Netherlands), in that the TC genotype from SNP rs3741378 is more prevalent in patients with PR negative breast tumours. The T allele appears to be associated with PR status. The Netherlands result didn't directly reproduce the findings in the Southern California cohort. However, different analyses were utilized in these two population and this could explain the seemly related but not directly reproducible observation. No association was observed with candidate SNP rs931127 for lymph node status, which showed association with positive lymph node metastasis in the pilot cohort (Southern California, USA cohort), suggesting that either the pilot result were due to a cohort specific effect or a false positive result due to limited number of samples. Once again, the various differences in cohort enrolment between the cohorts may create inconsistency of the results. To take advantage of the larger Rotterdam, Netherlands population, an additional SNP rs2448490 (intronic SNP) was utilised for analyses, to clarify the observations and the four SNPs together encompass greater coverage of the *SIPA1* gene. Significant prevalence in

ER, as well as PR positive tumour patients were observed with SNP rs2448490 in this larger dataset, increasing the probability of a true association of *SIPA1* with PR status in human breast cancer.

The prognosis analyses, which are the main focus of this research, across different populations, revealed that *SIPA1* SNPs were associated with breast cancer survival. Candidate SNP rs746429 was found to be associated with poor outcome in numerous independent populations. Initially the variant A allele was found to be associated with positive lymph node metastasis and advanced disease, with non-localized tumours in the pilot cohort (Crawford *et al*., 2006). The AA genotype was then found to be associated with poor survival rate in the FHCRC, USA cohort, which agrees with the findings of advanced disease obtained in the pilot cohort. Univariate survival analyses on 1864 patients in the Rotterdam, Netherlands cohort showed no trends of survival association and the AA genotype of candidate SNP rs746429. However, more specified analysis with patient stratification of lymph node metastasis status and oestrogen receptor status show association of AA genotype with poor disease-free survival, metastasis-free survival and overall survival in lymph node metastasis positive/oestrogen receptor negative patients. Patients homozygous for the variant A allele had a 2-fold higher risk of developing distant metastases compared with the combined GA and GG genotypes. Interestingly, the GA genotype from the same SNP showed a trend of association with poor metastasis-free survival in lymph node metastasis negative/oestrogen receptor positive patients, indicating this SNP is associated with survival in more than one category of patient subgroup. Since the actual value of the trend of each association was not available and the strength of the association was not explored, these data can be best viewed as trends of significant association over survival. In addition, there are only 16 lymph node metastasis positive/oestrogen receptor negative patients carrying the AA

genotype, the power of survival analysis over this set of patients may not be sufficient. This may explain the observation that SNP rs746429 showed no association of survival in the larger dataset of 1864 samples, yet showed trends of association after stratification with important clinical parameters. The association of AA genotype and particular class of breast cancer patient may be a false positive result, due to small dataset. However, if this association is proven true with additional samples and analysis, one plausible speculation for the specificity in patient class would be that the four classes of patients: lymph node (LN) metastasis positive/ER positive, LN metastasis positive/ ER negative, LN metastasis negative/ER positive and LN metastasis negative/ER negative patients, are representing four distinct phenotypes of breast cancer and are fundamentally biologically different. Therefore, the SNP may delineate association in distinct classes but not in mixed population as a whole. In non-strict loss comparison, the association of variant allele of SNP rs746429 and poor survival is reproduced across different independent cohorts, even with the adverse condition of diverged nature of sample population selections, inconsistent statistical methodology employed and rather small sample set in certain sub-population. However, more datasets and uniformed analysis are needed to add confidence over the finding.

No significant associations with clinical markers or patient survival rate were identified in the German population (most probably due to the low calling rate caused by poor DNA quality); Umea, Sweden and Minnesota, USA populations. However, a trend of association was found in the Australian population with the TT genotype of SNP rs3741378 was found to be associated with increased breast cancer incidence in the Queensland, Australia population. This data is marginally significant as the number of events of the TT genotype carrier in the breast cancer group is only 6 and the control group is 1. Even though this result was reproduced with another statistical analysis designed for small events dataset, this potential

association still needs to be approached with caution. More samples and cohorts are needed to confirm this potential association; however, if the association with breast cancer incidence is proven true, then the *SIPA1* gene, may act as a breast cancer metastasis modulating gene, as well as breast cancer predisposition marker, with potential participation in breast cancer initiation.

The *SIPA1* gene SNPs rs931127 and rs3741378 showed no association with disease-free survival, metastasis-free survival or overall survival in the total patient population, prognostic subgroup or in the patients stratified by lymph node and ER status. This is consistent with other cohort findings, since no survival association was observed of these two SNPs in any of the cohorts.

The fourth SNP rs2448490, which was only assayed in the Rotterdam, Netherlands cohort, showed association with survival. Patients homozygous for the A allele of this SNP had a favourable overall survival (OS) in the total prognostic-available patient sub-population (n=1864). When stratified patients by lymph node and ER status to evaluate associations in good (lymph node negative, ER-positive; LNN/ER+) and worst classes (lymph node positive, ER-; LNP/ER-), as determined by standard clinical parameters, the AA genotype of SNP rs2448490 was associated with a favourable prognosis only in the lymph node negative, oestrogen receptor positive subgroup for both metastasis-free survival and overall survival. The variant A allele of candidate SNP rs2448490 was then proposed as a better survival indicator in the best patient classes.

This result has the same limitation as the SNP rs746429 result; they both have small events (18 patients of the AA genotype/lymph node positive/oestrogen receptor negative) and

therefore the data needs to be interpreted with caution. The low events recorded seriously hinder the power of the survival analysis and larger datasets are needed to confirm this association, however, if this preliminary finding is proven true with additional datasets and analyses, it has important impact in the clinic. First, since currently the majority of lymph node metastasis negative breast cancer patients receive systemic adjuvant therapy, due to the inability to accurately determine an individual's relapse risk. There is only a minority of node-negative patients will actually develop recurrent disease, therefore it is crucial to stratify lymph node metastasis negative patients with respect to breast cancer relapse risk (Urquidi and Goodison, 2007).

Secondly, this result suggests that oestrogen receptor positive tumours that do or don't involve the lymph nodes have a distinct biological mechanism; implying that genetic susceptibility of hematogenous and lymphatic metastasis are distinct entities and not merely different stages in a linear breast cancer progression pathway. No model of this phenomenon has been proposed, however, one possible explanation might be differences in chemokine or cytokine receptor profiles in the genetically different tumour subtypes. Differences in these receptor profiles might dictate the ability of tumour cells to either home to or grow in the local regional lymphatics by modulating the ability of the cells to respond to the local lymphatic microenvironment.

To the best of the author's knowledge, this is the first indication of an inherited polymorphism with the capability to discriminate patient outcome in specific clinical tumour types. Previous findings from other researchers also support this phenomenon that distinct mechanisms exist for different types of breast cancer progression and metastasis. For example, analysis of breast cancer subtypes as defined by gene expression profiles (Perou *et*

al., 2000) demonstrated preferential sites of relapse (Smid *et al*., 2008), suggesting different mechanisms of colonization. Women with triple negative breast cancers (ER negative, PR negative and HER2 negative) are less likely to experience a local recurrence before developing a distant recurrence (Dent *et al*., 2007).

Since preliminary associations were identified, other important factors that may influence survival rate were considered for analyses, in order to gain a more in depth insight of the observations. Since the association with overall survival observed in this population might be partially influenced by response to therapy after relapse, association with response to clinical intervention was subsequently examined. No significant associations of any of the SNPs with the efficacy of first-line tamoxifen therapy in 315 ER positive recurrent breast cancer patients were observed, irrespective of the endpoint studied. However, due to the small dataset, the power of the analysis is low and a potential association can't be ruled out. In the 226 first line chemotherapy treated patients, the AA genotype of *SIPA1* SNP rs2448490 was found to be associated with a better response to first line chemotherapy.

In the analysis for PFS in first line chemotherapy receipt patients, the AA genotype of *SIPA1* candidate SNP rs746429 and the TT genotype of SNP rs3741378 were both associated with poor progression-free survival in this subset of patients. When stratifying this subset of population with oestrogen receptor status, which is an important marker in determining chemotherapy responses, the variant A allele from SNP rs746429 was still associated with poor progression-free survival in oestrogen receptor negative/first line chemotherapy treated patients. However, the TT genotype from *SIPA1* candidate SNP rs3741378 was found to be associated with better progression-free survival in oestrogen receptor positive/first line chemotherapy treated patients. SNP rs2448490 was able to discriminate between patient PFS

after first-line chemotherapy. The AA genotype was found to be associated with better progression-free survival, as well as ER-negative patients/first line chemotherapy; the AA genotype of SNP rs2448490 predicted a significantly delayed disease progression in ER-negative patients, with a median time to progression of 9.0 months compared with 5.1 months for the GA genotype and 4.5 months for the GG genotype.

Due to the limited number of treated patients available in this cohort, these results must be interpreted with caution. The data however would suggest that the SNPs in *SIPA1*, while predictive of the development of metastasis, are likely to be associated with therapeutic response. Candidate SNP rs2448490 showed an association with progression after chemotherapy treatment in ER-negative patients, yet not the ER-positive classes, in which it was associated with metastasis-free survival. If this association is proven true, with additional datasets, it would suggest the metastasis efficiency modifier gene *SIPA1* is engaged in tumour progression biology as well as tumour cell drug response biology. A potential biological linkage between the metastasis susceptibility modulation function and drug response function of *SIPA1* may be explored.

5.2.2.3 Discussion of *SIPA1* Gene SNPs Genotyping Results

To explain the preliminary associations observed between the *SIPA1* SNPs and various breast cancer clinical markers and survival, the properties of the SNPs may offer a logical rationale. The four SNPs selected to cover the *SIPA1* gene include intronic (SNP rs2448490), exonic (SNP rs746429 (synonymous Ala to Ala); rs3741378 (nonsysnonymous Ser to Phe)) and 5' near gene (SNP rs931127) SNP.

Intronic SNPs, as described in the *ARAP3* gene section, could participate in microRNA functions; exonic synonymous SNP, even though they may not change protein structure, have been shown to alter protein function. A recent study reported that a synonymous SNP from the Multidrug Resistance 1 (*MDR1*) gene produced a gene product (P-pg protein) with altered drug and inhibitor interactions. Similar mRNA and protein levels of the *MDR1* gene were observed in this study, yet altered conformations of the polymorphic P-pg were found. The authors hypothesize that the presence of the rare codon, which is marked by the synonymous polymorphism, may have altered mRNA secondary structure and mRNA stability and affect the timing of cotranslational folding and insertion of P-pg into the membrane, hence altering the structure of substrate and inhibitor interaction sites (Kimchi-Sarfaty *et al*., 2007; Wang *et al*., 2005). These data provide a possible explanation for the association with the *SIPA1* synonymous polymorphism rs746429 (exon 13, outside of the putative Smc: chromosome segregation ATPases domain, this domain has not been confirmed yet) and poor survival rate across independent populations, in that the *SIPA1* protein is involved in metastasis mechanics, and the polymorphism of rs746429 may change the structure of the protein and its interaction site via the usage of rare codon.

The non-synonymous SNP rs3741378 was found to be associated with hormonal receptors in the pilot study and progesterone receptor status in the 1922 Rotterdam, Netherlands dataset. However, no survival association was established. Nonsynonymous SNPs are known to change protein functions, for example, the two common nonsynonymous SNPs ($Ile^{105}Val$ and $Ala^{114}Val$) alterations in the Glutathione S-transferase P1 (*GSTP1*) gene, which is heavily involved in chemotherapy drug metabolizing and cell cycle, are known to be associated with variations in cancer risk and clinical response to anti-neoplastic drug therapy. The Val^{105} substitution, for instance, was found to result in steric restriction of the H-site due to shifts in

the side chain of several amino acids and subsequently being able to accommodate less bulky substrates than the Ile^{105} allele and displays substrate specificities that differ from the wild-type allozyme (Moyer *et al*., 2008). Signal-induced proliferation-associated 1 gene SNP rs3741378 (exon 2, outside of the Rap-Gap domain) has significant correlation to both the oestrogen receptor status and the progesterone receptor status, it is possible that this exonic missense SNP, a Ser (S) to Phe (F) change, which goes from a hydrophilic residue (S) to a hydrophobic residue (F), changes the protein folding and functions of *SIPA1*. This may have functional alteration in the potential interaction of the *SIPA1* protein and the hormonal receptors, which subsequently change hormonal status biology of breast carcinogenesis. In addition, SNP rs3741378 showed a trend of association with breast cancer incidence, while most other SNPs were reported to be associated with metastasis and patient survival. it is likely that since these SNPs are in LD, they are also in LD with a functional SNP, which may function to influence multiple events such as breast cancer development and worse prognosis via the same biological mechanism.

A recent case-control study from Gaudet *et al*. reported a similar trend, though the difference in their population did not reach statistical significance. The Gaudet *et al*. group explored the same SNPs as this research in a Polish (1995 cases, 2296 control) and a British cohort (2142 cases, 2257 controls). They reported a suggestion of an increased risk of breast cancer associated with the TT genotype of the SNP rs3741378. Both the Gaudet group and this research identified a potential relationship between this SNP and breast cancer risk. The inconsistency of the significance of the association between this study and the Gaudet is likely best explained by the diverse nature of the populations studied, through either environmental or genetic variance. The ratio of the TT genotype in the control samples has a wider range of diversity between the Gaudet study and this investigation, with the TT

genotype comprising 0.62% of the control group for this study compared to 1.8% in the pooled study from the Gaudet group, both of which are less than the CAUC 1 population, indicating that this SNP may be highly variable in different areas. The difference in this rare allele frequency may explain the slight inconsistency in strength of association identified between the researches. However, the two independent groups did identifying a similar relationship, only differing in the strength and significance of the relationship identified. This, as well as previous functional work, adds support to the involvement of the *SIPA1* gene and its SNPs in breast cancer susceptibility.

Single nucleotide polymorphisms in the promoter region have been linked to altered gene expression, as demonstrated by the promoter SNP rs16949646 from the NME1 promoter region, which altered the nuclear proteins binding capacity and reduced promoter activity (Qu *et al*., 2008). Single nucleotide polymorphism rs931127 is located at 5' near gene region, which can be part of the promoter region of the *SIPA1* gene. However, no reproducible association was found between this SNP and various breast cancer clinical markers, such as lymph node metastasis (in the pilot cohort). This weak association requires additional analyses to elucidate its putative relation to the *SIPA1* gene and clinical markers.

Among all four SNPs screened in the *SIPA1* gene across independent populations, polymorphism variants functioning in opposite directions (SNP rs746429 was associated with poor survival, while SNP rs2448490 was associated with better survival rate) in different subclasses of breast cancer patients was observed. Since these variants tag different ends of the *SIPA1* locus, which has been demonstrated to potentially bind many different putative protein partners (Crawford *et al*., 2007; Farina *et al*., 2004), both cytoplasmic and nuclear, it is plausible that the opposite effects observed may be due to the function of different parts of

the *SIPA1* protein. Studies have identified roles for *SIPA1* stem cell maintenance (Lee *et al.*, 2008) and in nuclear complexes, which are associated with control of transcriptional elongation (Farina *et al.*, 2004; Wu *et al.*, 2007), in additional to its well-characterized RAP1GAP functions (Tsukamoto *et al.*, 1999; Ishida *et al.*, 2003). *SIPA1* is therefore likely to be a complex, multifunctional protein with different roles in different tissue types and in different stages of breast carcinogenesis. It is therefore conceivable that one polymorphism might be associated with a change in function in lymphoid tissue, for example, which might impact the mechanisms unique to lymphatic tumour spread, but not play a role in the hematogenous pathway, and vice versa. This may explain the complex observation from this research, with *SIPA1* SNPs showing putative association with both breast cancer incidence and metastasis/survival.

In conclusion of the *SIPA1* epidemiology studies, it is suggested that the *SIPA1* gene is engaged in human breast cancer biology and metastasis from the intrinsic inherited subtle genetic level as well as other biological levels (transcriptional correlation: the *Sipa1* expression is highly differentially regulated between and low metastasis potency mice strains and mouse lung metastasis potency is correlated to cellular *Sipa1* levels; translational correlation: the *Sipa1* protein is associated with roles in stem cell maintenance and other cell cycle/survival functions). This adds evidence supporting *SIPA1* as a novel breast cancer metastasis susceptibility gene. Additionally, this research provides data of the possible association between the *SIPA1* gene and chemotherapy response, implying a potential participation of the *SIPA1* gene in drug response and metabolism.

In addition, the preliminary data suggesting *SIPA1* SNPs distinguishing oestrogen receptor positive tumours colonizing to the lymph node from oestrogen receptor positive tumours not

homing to the lymph node propose a novel hypothesis that lymph node metastasis is not merely a step in the breast cancer progression time line, but a distinct phenotype of breast malignant tumours. More studies are required to confirm this novel hypothesis.

5.2.3 The *RRP1B* gene and its candidate SNP (SNP rs9306160)

5.2.3.1 Introduction of the *RRP1B* Gene

The ribosomal RNA processing 1 homolog B gene was identified by the same method used to identify the *ARAP3* gene as a novel metastasis modulator gene; it has high expression correlation with extra-cellular matrix (ECM) genes, which have been proposed to be markers for multiple independent metastasis predictive gene expression signatures, and has high differential expression between the high and low metastasis strength mouse strain. The ribosomal RNA processing 1 homolog B (*Rrp1b*) protein is a binding partner of breast cancer metastasis gene *SIPA1* and *in vitro* expression of *Rrp1b* was confirmed to altering extracellular matrix expression. *In vitro* expression of *RRP1B* reduces primary breast tumour growth and lung metastasis capacity in the mouse model, with a gene expression signature generated by indicative ectopic expression of *Rrp1b* predicts breast cancer patient survival (Crawford *et al.*, 2007). These data indicate that *Rrp1b* is engaged in breast cancer progression in the transcription and translational level. This part of the research was designed to explore to potential involvement of this gene with breast cancer and metastasis at the genetic level.

5.2.3.2 Summary of the *RRP1B* Gene SNP Genotyping Results

Single nucleotide polymorphism rs9306160 (missense; nonsynonymous: Pro to Leu) was selected to represent the *RRP1B* gene, it was genotyped in both pilot cohorts and found to be associated with various breast cancer clinical markers; the variant A allele was found to be associated with negative lymph node metastasis, positive oestrogen receptor and progesterone receptor, localized disease and well/moderately differentiated tumours (Southern California, USA pilot cohort). The protective feature of the variant A allele of this SNP continued to show in the other pilot cohort (Greater Baltimore, USA pilot cohort), with lower disease stage, lower tumour grade, negative lymph node metastasis, positivity in oestrogen receptor and progesterone receptor and better survival rate identified. Further genotyping assays were carried out in six larger independent populations; no associations were found in the Minnesota, USA; Queensland, Australia; Umea, Sweden and the Germany populations, implying this SNP is associated with neither mammographic dense area nor breast cancer incidence. However, this SNP was found to be associated with smaller tumour size in the FHCRC, USA cohort, yet this association was not reproduced in other cohorts (only a trend in the larger Rotterdam, Netherlands cohort) and regarded as a preliminary indication of possible association - however, this protective feature agrees with the observations found in the pilot studies.

In an attempt to reproduce the survival association identified in the pilot cohort, survival analysis was performed in the additional independent populations. Association with survival rate was identified; patients carrying the CT genotype of *RRP1B* candidate SNP rs9306160 showed a significant favourable prognosis in the larger Rotterdam, Netherlands population for metastasis-free survival and a borderline statistical association with better disease-free survival and overall compared to the homozygous CC genotype. Most often associations are

identified with the homozygotes genotype rather than the hetrozygote genotype. However, it is a very rare event that hetrogygote genotypes have a biological function and this result doesn't suggest that the CT genotype, which is a synonymous polymorphism, has a biological function. However, it is possible that homozygous genotypes are associated with rare tRNA usage and that heterozygous genotypes give better availability for all tRNA species, which may impact upon the quaternary structure of the mature RRP1B protein by reducing the free-energy necessary for folding. How differences in the mature quaternary structure of RRP1B could impact upon breast cancer progression is not yet known, as there have not been any x-ray crystallography or protein modelling work performed that focuses on RRP1B.

When adjusting the assay model (assuming a dominant or additive model), associations with better outcome were observed for the variant A allele for all three endpoints (disease-free survival, metastasis-free survival and overall survival). The result for metastasis-free survival remained significant in multivariate analysis, independent of the traditional prognostic factors, while the associations of the A allele with disease-free survival and overall survival were not significant in the multivariate analysis. Addition of adjuvant chemotherapy to the multivariate model did not affect the estimated coefficients of the *RRP1B* SNP rs9306160 genotype. This is the only population, other than one of the pilot cohorts, that this SNP showed association with survival. The limitation of different statistical methods used and different criteria applied for different populations may explain the lack of association identified in other cohorts. However, the Rotterdam, Netherlands data reproduced the survival association found in the pilot cohort adding confidence in the result.

To further explore the property of the survival association identified, the dataset was stratified by clinical lymph node metastasis status and oestrogen receptor status; an association of variant allele A of *RRP1B* candidate SNP rs9306160 with a favourable prognosis was

observed and only in the subgroup of lymph node metastasis negative/oestrogen receptor positive (the better condition class) patients, but not in the other patient subgroups. This protective association remained statistically significant for disease-free survival, metastasis-free survival and overall survival after correction for the traditional prognostic factors in multivariate analysis.

This result has great interest as it reproduced the finding observed in the *ARAP3* study as well as the *SIPA1* study, in that the SNPs showed association with survival rate, or trends of significant association, in specific classes of patients stratified by lymph node metastasis status and oestrogen receptor status. This reinforced the novel hypothesis that these results suggest different and distinct biology and mechanisms between ER-positive tumours that do and do not colonize the lymphatic system. It is plausible, that the ER-positive/negative tumours, which do and do not colonize the lymphatic system, are genetically distinct subtypes of breast cancer. These unique subtypes of breast cancer might have different chemokine or cytokine receptor profiles, which modulating the ability of the cells to respond to the local lymphatic microenvironment. The result of this research is the first example, to the best of the author's knowledge, of an inherited polymorphism capable of discriminating patient outcome in specific clinical tumour types. There are literature reports suggesting distinct mechanisms for breast cancer progression and metastasis, for example, *BRCA1* carriers have been shown to be less likely to have positive axillary lymph nodes at diagnosis than non-hereditary breast cancers (Foulkes *et al*., 2003). Furthermore, the rate of *TRP53* mutation varies significantly between *BRCA1* tumours of the luminal subtypes compared to either *BRCA1* or sporadic basal-like tumours (Manie *et al*., 2009).

Since molecular studies demonstrated that the *SIPA1* and *RRP1B* gene products physically interact (Crawford *et al*., 2007) and the *SIPA1* SNP rs2448490 and *RRP1B* SNP rs9306160 were both associated with a favourable prognosis in lymph node metastasis negative/oestrogen receptor positive patients, the possibility that the combination of both SNP genotypes might show additional prognostic power was explored. In Cox multivariate analysis, *SIPA1* SNP rs2448490 and *RRP1B* SNP rs9306160 remained independent factors predicting a prolonged metastasis-free survival with Hazard ratio of 0.61 and 0.69, respectively. Kaplan-Meier analysis for metastasis-free survival as a function of the combined genotypes in the lymph node metastasis negative with oestrogen receptor positive patients demonstrated that the combination of the homozygous AA genotype of *SIPA1* SNP rs2448490 and at least one copy of the T allele (TC+TT) of *RRP1B* SNP rs9306160 was associated with the best prognosis. The risk for developing distant metastasis was more than 2.5-fold lower compared with that observed for the GG+GA/CC combination. This result implies that genetic susceptibility, hematogenous and lymphatic metastases are distinct entities and not simply different stages in a linear breast cancer progression pathway, which supports the results observed in other sections of this research.

Once again, since the survival association observed may be affected by response to chemotherapy, the efficacy of the tamoxifen treatment and first line chemotherapy were determined. No associations of the *RRP1B* SNP rs9306160 and either Tamoxifen therapy or first-line chemotherapy response and patient survival were identified. This indicates that the survival associations identified are independent of the affects of treatments.

5.2.3.3 Discussion of *RRP1B* Gene SNP Genotyping Results

To sum up, the protective effect of the variant A allele of SNP rs9306160 of the *RRP1B* gene is supported by multiple independent cohorts, as showed in the pilot study and the Rotterdam, Netherlands data. These data also suggests that there are distinct molecular mechanisms for lymphatic and hemotogenous dissemination and the investigations into the genetic basis of breast cancer progression, stratification of cohorts by oestrogen receptor and lymph node status appears to be necessary to accurately assess the inherited components of metastatic capacity. In addition, the *RRP1B* SNP rs9306160 and *SIPA1* SNP rs2448490 combined analysis gave the best prognosis values, illustrating that these two genes not only interact in the proteomic level, but also at the genetic level. The combined analysis adds evidence that complex phenotypes, such as cancer metastasis, are shaped by multiple genes, which might have interactions at the molecular level, although being individually weak genetic variants; joint analysis of the metastasis modulating genes and their genetic variants, gives greater power of prediction, as evident by the combined analysis of the *SIPA1* SNP and the *RRP1B* SNP. The ribosomal RNA processing 1 homolog B gene is like the *ARAP3* gene and *SIPA1* gene, which all are linked to breast cancer metastasis from the genetic level to the proteomic level. The inherited genetic variance of the *RRP1B* gene was shown to be associated with breast cancer metastasis, and supports the *RRPB* gene as a novel metastasis modulating gene.

5.2.4 The BRD4 gene and its candidate SNPs (SNP rs4809130 and rs4808272)

5.2.4.1 Introduction of the *BRD4* Gene

The bromodomain containing 4 (*Brd4*) gene was first identified as a potential metastasis modulating gene when attempting to characterize genetic loci modulating metastasis-predictive ECM gene expression, via expression QTL (eQTL) mapping in recombinant inbred mice. The *Brd4* protein is a binding partner of metastasis modulators *Sipa1* and *Rrp1b*, supporting *BRD4* as a metastasis modulator (Crawford, *et al*., 2008). The *Brd4* protein is from the bromodomains and extraterminal (BET) family, which is involved in multiple cell activities, including transcription, replication, signal transduction pathways and cell cycle progression. The *Brd4* protein, with two bromodomains, conducts its functions mostly via protein association with acetylated chromatin to induce transcriptional activation (Liu *et al*., 2008). *In vitro* analysis of ectopically expressed *Brd4* in a highly metastatic mouse mammary tumour cell line demonstrates significant reduction of invasiveness, without altering intrinsic growth rate, yet a dramatic reduction of tumour growth and pulmonary metastasis was observed after implantation into mice, implying that activation of *Brd4* may alter response to tumour microenvironment *in vivo*. Additional *in vitro* analysis showed that *Brd4* modulates extracellular matrix gene expression, a class of genes frequently present in metastasis-predictive gene signatures. Microarray analysis of the mammary tumour cell lines identified a *Brd4* activation signature, which robustly predicted progression and/or survival in multiple human breast cancer datasets analysed on different microarray platforms; in addition, the *Brd4* signature closely matches a molecular classifier of low-grade tumours. These data suggest that dysregulation of *Brd4*-associated pathways may be engaged heavily in breast

cancer progression and underlies multiple common prognostic signatures (Crawford *et al.*, 2008).

5.2.4.2 Summary and Discussion of the *BRD4* Gene SNPs Genotyping Results

In order to determine whether the *BRD4* gene is engaged in human breast cancer at the inherited genetic level, single nucleotide polymorphisms of this gene were assayed in the two pilot studies. The pilot cohorts identified two different *BRD4* SNPs showing association with breast cancer patient survival, yet this result wasn't reproduced between the pilot populations (SNP rs4808272 (intronic) was found to be associated with survival in the Greater Baltimore pilot cohort; rs4809130 (intronic) in the Southern pilot California cohort). Since the Greater Baltimore cohort has a smaller collection of Caucasian population, SNP rs4809130 wasn't selected for further analyses.

No significant associations between SNP rs4809130 and any clinical markers and survival were found in any of the larger independent populations, other than the Rotterdam, Netherlands cohort. The other SNP rs4808272 was subsequently screened in this population as well; trends of significant associations with breast cancer survival were identified for both SNPs in this population. The AG genotype of SNP rs4808272 was associated with poor preogression-free survival in the 1st line Anthracycline receipt patients and the CT genotype of SNP rs4809130 is associated with better progression-free outcome in first line chemotherapy receipt patients with negative oestrogen receptor tumours (the analysis of these two SNPs in the entire 1922 samples and prognosis group were not available to the author, therefore, it is not certain whether the results obtained can be reproduced or not). More analyses of the Rotterdam, Netherlands cohort and other large breast cancer cohorts are needed before a confident conclusion can be drawn in relation to this gene. The present data

merely suggests that the *BRD4* gene SNPs may have association with breast cancer survival. However, these preliminary data supports *BRD4* gene as a participant in breast cancer metastasis and survival.

5.3 Future Directions of the Epidemiology Studies

The major finding of this research is identifying novel associations of single nucleotide polymorphisms of novel metastasis susceptibility genes and breast cancer markers and survival. Additionally, the data suggests the possibility of a novel breast cancer-subtype stratified by lymph node metastasis status.

The hypothesis for the logical progression of inherited genetic variance leading up to differential metastasis strength and survival rate is that subtle inherited genetic variances alter tumour cell biology via changing regulation and function of metastasis susceptibility genes. This, along with other mechanisms modulating tumour progression, modifies metastasis susceptibility and potency, which ultimately facilitates differential survival rate across the population. This research provides evidence of subtle genetic variance, in the form of single nucleotide polymorphisms, engaging in breast cancer progression; at the same time, the data supports the involvement of a number of a novel metastasis susceptibility genes and a novel pathway for future studies.

To add confidence to these findings, the strength of the associations must be determined. Additional datasets and epidemiology studies testing these SNPs would help to elucidate the nature of some of the weak and trends of association observed, as well as giving more confidence in the already established associations. Once the observations identified in this research are confirmed and reproduced in other independent studies, the characteristics of the SNPs could then be further explored. For the intronic SNPs, the possibility of these SNPs as part of microRNA sequences could be examined; a direct *in vitro* qPCR experiment on *SIPA1* expression in cells with different genotype of 5’ near gene SNP rs931127, may actually

elucidate whether this SNP modulates *SIPA1* expression. Several exonic SNPs showed associations with breast cancer metastasis and survival, as well as trends of treatment response; for the missense SNP rs3741378 of the *SIPA1* gene, it is important to examine to potential protein folding alteration caused by this SNP. The potential drug metabolism association between the *SIPA1* and *ARAP3* genes required further confirmation; however, simple *in vitro* experiments could examine the potential response of cell lines with elevated/decreased *SIPA1*/*ARAP3* gene expression treated with different types of therapy medicine, and this may be helpful in elucidating the potential association between these genes and treatment response.

It is important to note that no samples were followed up with genotyping of non-amplified DNA, as those samples were precious and only sufficient in amount for the whole genome amplification process. However, the expected error rate of the whole genome amplification process can be obtained from the company's website http://www.gelifesciences.com/APTRIX/upp00919.nsf/content/27A64F7F2801041AC125712E000DC140?OpenDocument&Path=Catalog&Hometitle=Catalog&entry=8&newrel&LinkParent=C1256FC4003AED4031BCBD969DF2B097C12571170042170A_RelatedLinksNew1BC6C3F1E0C8AF6D852573DE006A999C&newrel&hidesearchbox=yes&moduleid=166713.

The company's research indicates that the error rate for the Phi29 DNA polymerase is 1 in 107. This was determined using a Tom Kunkel type, lacZ mutation test assay.

The ultimate goal of this research was to develop diagnostic tools that could discriminate patients with high metastasis-development risk from the low risk group; prognostic and

treatment strategies that would lead to personalized cancer patient management. The diagnostic tool could take advantage of the non-tissue specific, inherited variance property of single nucleotide polymorphisms; this would facilitate low risk accessing of viable DNA for genotyping, with the inert nature of DNA enabling speedy and accurate analyses of the genotypes. The growing understanding of the metastasis susceptibility genes functions would also help in improved patient management and better quality of life. It is hoped that in the future identification of all the genes involved in breast cancer incidence, metastasis and survival will greatly aid in early diagnosis, treatment choice, management and disease prognosis.

5.4 Discussion of the Novel *Sipa1* Protein Binding Partner Identification Results

Signal-induced proliferation-associated 1 (*Sipa1*) is a regulator for Rap, which is involved in cell survival, proliferation, differentiation, adhesion and migration. *Sipa1* was identified as a novel breast cancer metastasis modulator, in that it promotes breast cancer metastasis with increased *Sipa1* expression and ablated metastasis with reduced *Sipa1* expression (Park *et al*., 2005). Additional independent research supports the metastasis enhancer role of *Sipa1*; for example, it was found that the *Sipa1* expression level was higher in primary human prostate tumours that metastasized than those didn't and the up-regulation of *Sipa1* in LNCaP cell line produced more metastasis in an orthotopical implantation *in vivo* assay with no changes in primary tumour sizes. In addition, knock down of *Sipa1* in the PC3 cell line reduced mouse lung metastasis in the *in vivo* assay (Minato and Hattori, 2009).

The mechanism of *Sipa1*'s metastasis modulating effect is largely unknown, hence the Hunter lab carried out a series of experiments aiming to understand the role of *Sipa1* in breast cancer metastasis. It was suggested that identifying and studying the protein binding partners of *Sipa1* may provide insight into the functions of *Sipa1* and therefore, yeast-2-hybrid assays using *Sipa1* as baits were conducted. Numerous genes were identified in the Y2H assay as potential protein binding partners of *Sipa1* and the focus of immediate research was centred on the genes identified around the PDZ domain of the *Sipa1* protein. This is due to the observation that the PDZ domain may hold an important functional role as a polymorphism in the PDZ domain changes the *Sipa1* protein binding efficiency for *Aqp2* and also its RapGAp activity. The polymorphism is an Alanine (A) to Theonine (T) change at amino acid position 741 of the PDZ domain and it was found that the mice strains with 741A had higher amount

of lung metastases, while the 741T strain mice had less dissemination. Another important aspect of this polymorphism is that cells with the 741A genotype had higher RapGAP activity compared to the 741T cells (Park *et al.*, 2005).

Due to the polymorphism modulating the functionality of the PDZ domain, the potential *Sipa1* protein binding partners of the PDZ domain were selected for analyses to confirm the associations. Six genes; actin related protein 2/3 complex, subunit 3 (*p21Arc*), Niemann Pick type C1 (*Npc1*), general transcription factor II H, polypeptide 2 (*Gtf2H2*), phosphoribosylglycinamide formyltransferase (*Gart*), Calmodulin 2 (*Calm2*) and apoptotic chromatin condensation inducer 1 (*Acin1*), were selected for investigation in this part of the research. To ensure adequate amount of proteins of both *Sipa1* and its potential binding partners, dual-transfection were conducted to increase expression of the genes. Four genes, *p21Arc*, *Gart*, *Npc1* and *Gtf2H2*, didn't have successful dual up-regulation and were not taken for further binding confirmation assays. Another member from the Hunter lab later revisited these four genes and had successful up-regulation on both the target genes and *Sipa1*, with subsequent Co-IP assays finding that none of the genes showed physical interactions with *Sipa1*.

Co-immunoprecipitation assays aiming to co-immunoprecipitate potential Sipa1 binding partners and Sipa1 were performed on *Acin1* + *Sipa1* dual transfected cell lysate and *Calm2* + *Sipa1* dual transfected cell lysate. It was found that Calm2 was physically interacting with Sipa1 while Acin1 didn't show such interaction. This part of the research identified Calm2 as a novel Sipa1 binding protein. Calmodulin 2 is a major cytoplasmic calcium binding protein that mediates the calcium signal in cells and is known to regulate numerous pathways important to neural function (Shirasaki *et al.*, 2006). Little is known about the *Calm2* gene

characteristics in cancer and cancer metastasis; only recent research has found differentially expressed *Calm2* in anaplastic large cell lymphoma (Rust *et al*., 2005) and in differentiating human IMR-32 neuroblastoma cells (Toutenhoofd and Strehler, 2002).

The identification of novel cytoplasmic binding partner of Sipa1 is important, in that the localization of Sipa1 in the cell is dictated by the localizations of its binding proteins. For example, Sipa1 is localized more in the cytoplasm than the nucleus, however, when Sipa1 is bound to Brd4, which is localized in the nucleus, it was found to be localised in the nucleus and Brd4 enhanced the RapGAP function of Sipa1 (Farina *et al*., 2004). Therefore, the cytoplasmic localization of Calm2 may influence the localization of Sipa1; in addition, this change of localization of Sipa1 may in turn change the Rap signalling. It was found that the localization of RapGAPs, such as Sipa1, is important in regulating Rap functions (Minato and Hattori, 2009).

In addition to the RapGAP function of Sipa1, Sipa1 may be involved in breast cancer metastasis in other pathways and mechanisms. The physical binding of Sipa1 and its novel binding partner Calm2 may suggest the possibility that Sipa1 affects metastasis via calcium modulation, as Calm2 responds to intracellular calcium concentration changes (Shirasaki *et al*., 2006) and the intracellular calcium concentration is closely related to cellular proliferation, differentiation and apoptosis (Sergeev and Rhoten, 1998). Whether intracellular calcium level plays a role in cancer metastasis is not known, yet it is plausible that Sipa1 may mediate part of its metastasis modulating effect via binding to Calm2, hence modulating its calcium binding affinity and intracellular calcium concentration.

Since it is possible that Calm2 may affect breast cancer metastasis via its calcium association, it is important to investigate and clarify the Sipa1's metastasis modulating effect as an effect independent of Calm2 or a syngeneic effect with Calm2. Therefore, *in vivo* assay on ectopically expressed *Calm2* clones was carried out. The *in vivo* assay aimed to investigate whether *Calm2* up-regulation alone would produce an altered metastasis effect, as seen with the *Sipa1* clones. The assay was terminated prematurely as one of the control mice had to be sacrificed due to humane reasons before the due course of the experiment. The early scarifice on all the mice produced results where only one group of the control mice had primary tumour, while none of the other groups of control mice or *Calm2* clones mice had any primary tumour, with no lung metastasis observed in any of the mice experimented in this *in vivo* assay. A modification on the experimental design and the repetition of the *in vivo* assay may be required before conclusions can be drawn about the specific relation of *Calm2* and *Sipa1* in breast cancer metastasis.

It is important to note that it was found in our lab's experience that often candidate genes that pass most or all of the common *in vitro* assays, such as the soft agar assays, invasion assays, etc, failed to show relevance to metastasis in the *in vivo* assays. Therefore, it is recommended by my onsite supervisor to jump directly from the *in vitro* assay to the *in vivo* assay. This is the most time saving fashion of identifying the genes that has relevance in metastasis. If a gene proves to be involved in metastasis, such as *SIPA1* and other genes studies in this study, other *in vitro* assays were then employed to find out the mechanism and pathways that these genes are involved in.

In summary, the identification of a novel *Sipa1* protein binding partner is important in understanding the mechanism of *Sipa1*'s metastasis promoting and modulating effect. A

novel *Sipa1* protein binding partner in *Calm2* was identified in this part of the research. The cytoplasmic localization and calcium binding properties of *Calm2* may offer areas of interests for understanding the specific mechanisms of *Sipa1* in breast cancer metastasis, which ultimately would lead to better understanding of the metastasis machinery and improved diagnosis, prognosis and treatment with patients with metastatic breast cancer.

5.5 Future Direction of *Calm2* and *Sipa1* Association Research

The results of this part of the research identified Calm2 as a novel protein binding partner of Sipa1. *In vitro* analyses may be designed towards investigating the change of binding efficiency, if any, between the two polymorphisms of the PDZ domain, to Calm2. This would repeat the assay performed on Aqp2 and determine whether this change of binding efficiency is universal. Subsequent assay on the RapGAP activity of Sipa1 while bounded to Calm2 may be important, since little is known about Sipa1's activity while bounded to Calm2.

The *in vivo* assay investigating *Calm2* up-regulation in breast cancer metastasis required modification. In the *in vivo* assay performed in this part of the research, up-regulation of *Calm2* was performed in the Met-1 cell line, with the *Calm2* and the control β-Galactosidase gene driven by the CMV promoter in a vector construct. The Met-1 cell line has moderate metastasis potency, with relatively better transfection rate compared to the previously used highly metastatic mouse mammary cell line Mvt-1 (Park *et al.*, 2005). The moderate metastasis potency of the Met-1 cell line gave inconsistent metastasis rates in the control cases and this was experienced by other members of the Hunter lab as well. Additionally, the CMV promoter and vector construct were found to be less efficient in up-regulation of the genes of interest, in comparison to the lentivirus transfection system, as experienced by an associated lab. Therefore, it was decided by the Hunter lab to change the transfection system to the lentivirus system and perform the up or down regulation in the Mvt-1 cell line. The modified experimental design in the transfection system and cell line used was expected to produce tighter control over the modulated expression level of the genes of interest, with a more consistent metastasis rate *in vivo* for the control β-Gal clones.

The immediate aim of the future work of this part of the research is to establish *Calm2* up-regulated clones with the lentivirus system in the highly metastatic mouse mammary carcinoma cell line, Mvt-1. All the subsequent *in vivo* and *in vitro* experiments regarding the *Calm2* and *Sipa1* associations would be performed on those stable *Calm2* over-expressed clonal cells.

5.6 Discussion

5.6.1 The Diasporin Pathway

To sum up, the data from this thesis support the use of a novel metastasis candidate gene identification method, which compared expression signatures derived from mouse tumours with high and low metastasis strength. In addition, molecular analyses revealed that most of the implicated candidate genes may act together in a novel metastasis regulating pathway - the Diasporin pathway.

The bromodomain containing 4 (*Brd4*) gene, along with the *Rrp1b* gene, are proposed as a central node within the Diasporin Pathway, a novel pathway important in tumour progression in both mice and humans. The Diasporin Pathway (Figure 5.1) was initially constructed by gene expression analyses, which facilitated the construction of the Diasporin transcriptional network; this pathway contains the seven candidates included in this research (*Ndn, Pi16, Luc7l, Rrp1b, Brd4, Arap3* and *Csf1r*), as well as metastasis-predictive ECM genes and metastasis suppressors. Microarray gene expression signatures induced by activation of candidate genes in the Mvt-1 cell line can accurately predict survival in a well-characterized human breast cancer cohort. Most of the high priority candidate genes studied in this research and those that showed association, or trends of association with survival, belong to the Diasporin pathway, *Sipa1*, *Arap3*, *Rrp1b* and *Brd4*, are at the centre of this network. Other candidate genes, such as *Pi16*, *Luc7l* and *Csf1r*, for example, did not showed association with survival, yet with important clinical markers. Since more datasets and more SNPs are required to have broader coverage and analyses over these genes, they are not excluded from being metastasis modulating genes.

This research supports the Diasporin pathway as an important novel metastasis regulating pathway. Research in this thesis has shown that core genes in this pathway, at the intrinsic genetic level, have shown association with human breast cancer survival in multiple independent cohorts.

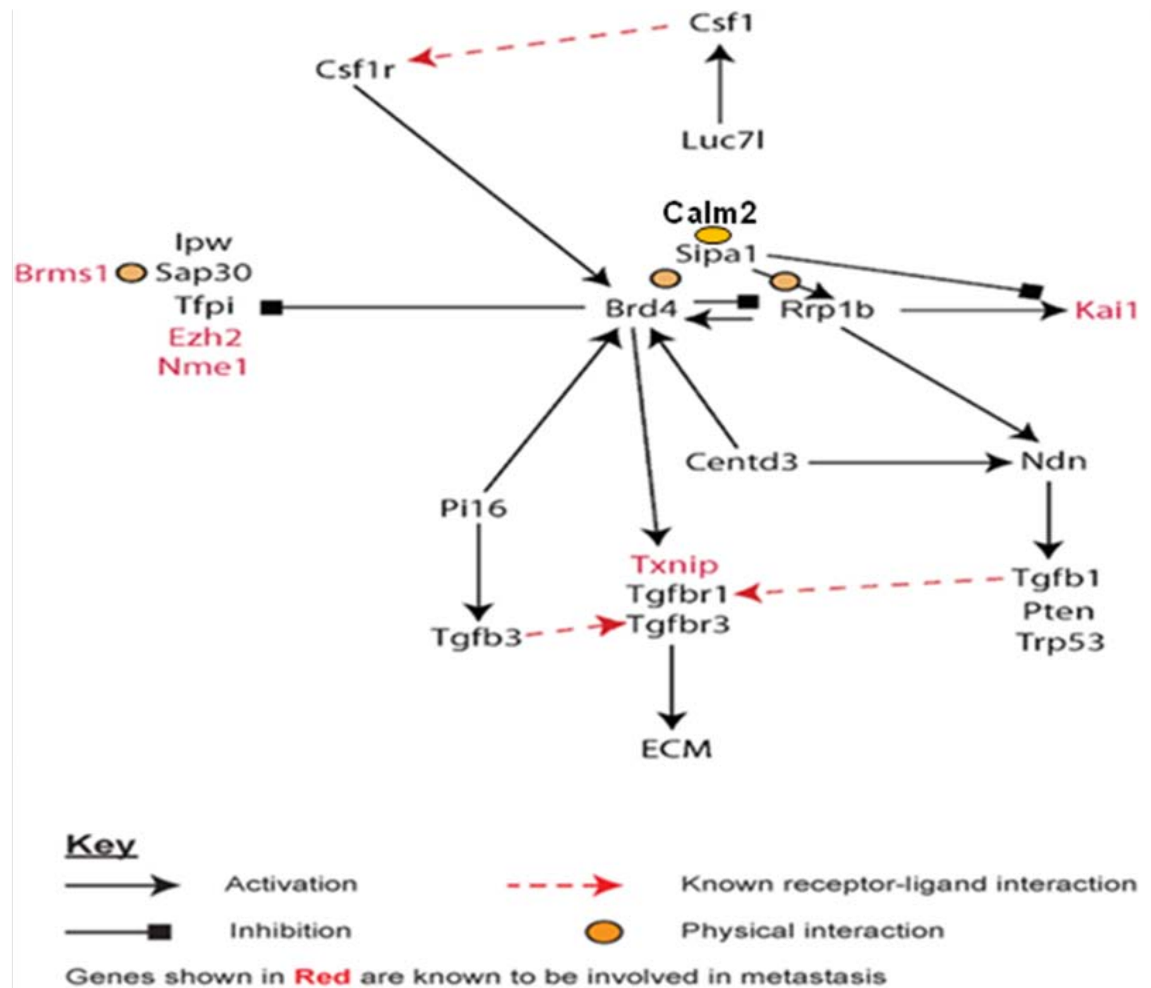


Figure 5.1: The Diasporin Pathway. This novel network describes the expression regulation of novel metastasis modulators and their relation to each other; this putative transcriptional network was constructed by applying the Occam's Razor principle, i.e. the simplest explanation is likely correct, by minimizing the number of potential interactions that would still explain the results. The pathway consists primarily of the seven novel metastasis efficiency modifier candidate genes, as well as a number of factors that are known to be modulators of metastasis (such as *Pten* and *Tgfb3*). The transcriptional relationship between each gene was determined by qPCR and microarray expression analysis, of the highly metastatic Mvt-1 cell line, stably transfected with one of the seven candidate genes. At this stage of analysis, *Brd4* and *Rrp1b* appear to be at the heart of this network, while both physically interact with the previously described metastasis efficiency modifier *Sipa1*. (The *Centd3* gene is also known as the *Arap3* gene) (Crawford *et al*., 2008).

5.6.2 Inherited genetic variance in breast cancer and cancer metastasis

The data gathered in this research contributes to an understanding of candidate metastasis susceptibility genes as well as in the broader scope of breast cancer and cancer metastasis modelling. The results support the genetic predisposition model; in that subtle inherited genetic variance influences metastasis potency.

Cancer metastasis involves proliferation of the primary tumour, tumour cells detach from the primary tumour, with subsequent invasion through adjacent tissues and basement membranes; the malignant cells then enter into blood and/or lymphatic systems and are carried via the blood or lymph to a distant target organ. To complete the metastatic process, the travelling tumour cells arrest in small vessels within the distant organ, extravasate into the surrounding tissue and proliferate at the secondary site (Figure 5.2) (Hunter *et al*., 2008).

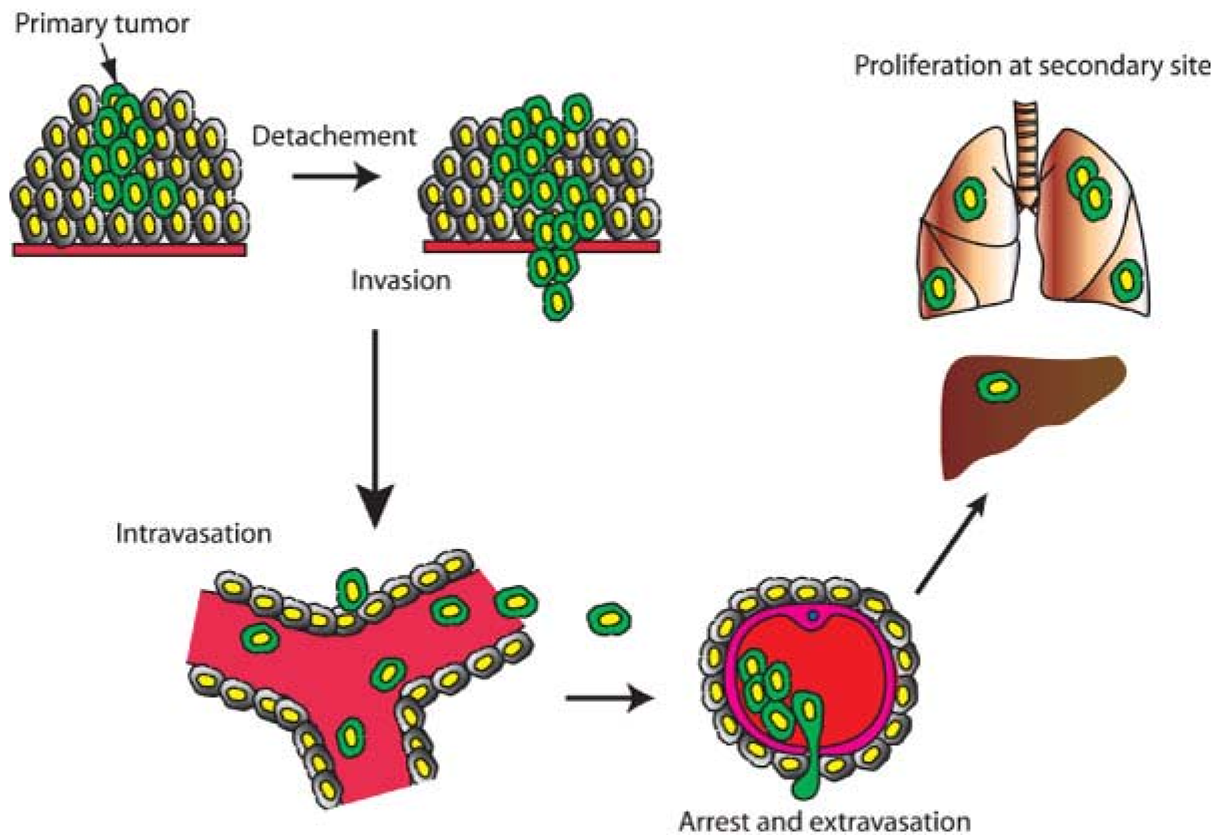


Figure 5.2: The Metastasis process. The metastatic process starts with malignant cells detaching from the primary tumour mass and invading the adjacent tissues and basement membranes. The invading process continues, with the tumour cells reaching the blood and/or lymphatic channels; the malignant cells are then carried via the blood and/or lymph to a distant organ. After reaching the secondary site, the tumour cells arrest in small vessels within the distant organ, extravasate into the surrounding tissue and proliferate at the secondary site, completing the metastatic process (Hunter *et al*., 2008).

Many hypotheses were generated in an attempt to understand the mechanisms of metastasis and explain the inefficiency of the metastatic process, but none of these theories completely explains the current biological and clinical observations (Hunter *et al.*, 2008). The progression model, one of the most commonly accepted models, emphasises the mutational events that occurred in malignant cells, either in primary tumours or in disseminated cells; a fraction of cells then acquire full metastatic potential. This model fails to explain the unknown-primary cancer metastatic disease, which constitutes approximately 5% of solid tumour-related cases and has disseminated disease, with no clinically detectable primary tumour, or only a small, well differentiated lesion found at autopsy (Riethmuller & Klein,

2001). This model, which expects the metastasis capacity stably inherited throughout generations, doesn't explain the observation that variant clones with high metastatic capacity often revert to a low-metastatic capacity after several generations (Chambers *et al*., 1984; Harris *et al*., 1982). The dynamic heterogeneity model (Harris *et al*., 1982), with the subsequently extended transient metastatic compartment model (Weiss, 1990), proposed to explain the lack of consistent increases in metastatic capacity of secondary tumours, compared to primary tumours, can't be explained by the progression model. The transient metastatic compartment model suggests that due to positional and/or random epigenetic events, only a small fraction of malignant cells are capable of completing the process at a given moment in time (Weiss, 1990), yet this model doesn't explain the clonal nature of metastases (Nakayama *et al*., 2001; Chambers *et al*., 1988; Cheung *et al*., 2002).

The early oncogenesis model, which is derived from microarray observations, is a variation of the transient compartment model; recent studies using microarrays to quantify global gene expression patterns in bulk primary human tumour tissue found that gene signature profiles can distinguish metastatic and non-metastatic tumours (van't Veer *et al*., 2002; Ramaswamy *et al*., 2003). These findings challenge the progression model, which proposed that only a small subpopulation of primary tumour cells will acquire the complete phenotype necessary to successfully colonize distant organs. Since the primary tumours' gene signature can distinguish metastatic and non-metastatic tumours, the majority of cells within a primary tumour would possess an inherent metastatic capacity; a pro-metastatic gene signature expressed within a small subpopulation of cells within the primary tumour, as proposed by the progression hypothesis, would be masked by the larger bulk of the tumour (Hunter *et al*., 2008). The early oncogenesis model proposed that metastatic propensity is established early in oncogenesis, potentially by the same sets of activation/inactivation events forming the

primary tumour, which is contradictory to the somatic evolution model. The early establishment of the metastatic state in the primary tumour cells would harbour the metastatic gene expression signature; with the same oncogenic events drove metastasis, small tumours are capable for dissemination and colonization of distant sites, explaining the metastatic disease of unknown primary origin (Hunter *et al.*, 2008). Since this model proposed that metastatic behaviour is primarily determined by early oncogenic events, therefore the majority of tumour epithelial cells would possess the ability to metastasize; this model would predict much higher colonization efficiency than observed in clinical practice (Hunter *et al.*, 2008).

An alternative explanation to the observation of gene expression predicting tumour metastasis potency is the genetic predisposition model, which proposed that the metastasis predictive gene expression signatures are not just an indication of somatic mutations driving progression, but may also be a measure of inherited metastasis susceptibility segregating throughout the human population. This model may be able to reconcile the various hypothesis proposed previously; metastasis susceptibility, which is derived from subtle inherited genetic variance (as predicted by genetic predisposition model), is reflected by the bulk tumour gene expression patterns and established early or before oncogenic transformation (as predicted by the early oncogenesis model). The metastatic process is followed by somatic mutations, with tumour evolution until subsets of cells acquire the necessary capabilities to complete the metastatic program (as predicted by the progression model). The genetic predisposition model also suggests that genetic background has an impact not only on the primary tumour, but also all of the tissues of the body, which would influence the establishment of the microenvironment of both primary and metastatic tumour cells. Microenvironment in metastasis has been recognized as a potential major factor in progression, for example, data

suggests that macrophages play an important role in inducing tumour cell motility at the primary site and may play an important role in the entry of tumour cells into the vasculature (Wyckoff *et al*., 2007).

This research supports the genetic predisposition hypothesis, in that candidate metastasis regulatory genes identified in the mouse model, which utilized gene expression data from genetic polymorphic mice with different metastasis susceptibilities; were shown to be associated with metastasis, as the inherited genetic variation, the single nucleotide polymorphisms, in the candidate genes was associated with breast cancer clinical markers, metastasis and survival in multiple independent patients populations. These novel metastasis susceptibility modulating genes form a novel transcriptional network, the Diasporin pathway, with data gathered from this research supporting this as a novel metastasis regulating pathway. Several of the novel metastasis modulators proteins have been shown to physically interact with each other and are involved in modulating metastasis potency *in vivo*. These data taken together supports the genetic predisposition hypothesis that subtle genetic variations have effects over metastasis potency in the human population.

The incorporation of genetic background into metastasis susceptibility models facilitates assessment of prognosis possible in using non-tumour tissue, potentially even before cancer develops. If a sufficient fraction of metastasis risk is encoded by germline polymorphisms, then any tissue in the body should carry some reflection of that risk and since the underlying susceptibility polymorphisms are ubiquitous across different tissues, unlike gene products which vary in different tissues and time point of the cell cycle, it would be theoretically possible to use any tissue to interrogate an individual's susceptibility state, at any time point of the cell cycle. Preliminary results to define risk groups in mice using mouse salivary gland

protein profiles have been demonstrated (Yang *et al*., 2005), and further support the genetic predisposition hypothesis and the application it offers.

In conclusion, the epidemiology results support the hypothesis of this research, in that putative breast cancer metastasis genes identified based on comparing expression signatures over differential metastasis potency mice, in the mouse model, are breast cancer prognosis and survival indicators in human breast cancer. In the broader context of metastasis modelling, the data provide evidence supporting the genetic predisposition model, in that subtle genetic variation contributes to differential metastasis susceptibility. Additionally, the observations of this research have proposed a novel hypothesis that oestrogen receptor positive/negative breast tumours do and do not metastasize to the lymph node system are distinct subtypes of breast cancer and may require different management.

Chapter 6

References

Adeyinka A, Kytola S, Mertens F, Pandis N, Larsson C (2000). Spectral karyotyping and chromosome banding studies of primary breast carcinomas and their lymph node metastases. *Int J Mol Med* **5:** 235-40.

Agnarsson BA, Jonasson JG, Bjornsdottir IB, Barkardottir RB, Egilsson V, Sigurdsson H (1998). Inherited BRCA2 mutation associated with high grade breast cancer. *Breast Cancer Res Treat* **47:** 121-7.

Ahnstrom M, Nordenskjold B, Rutqvist LE, Skoog L, Stal O (2005). Role of cyclin D1 in ErbB2-positive breast cancer and tamoxifen resistance. *Breast Cancer Res Treat* **91:** 145-51.

Akiyama T, Sudo C, Ogawara H, Toyoshima K, Yamamoto T (1986). The product of the human c-erbB-2 gene: a 185-kilodalton glycoprotein with tyrosine kinase activity. *Science* **232:** 1644-6.

Albasri A, Aleskandarany M, Benhasouna A, Powe DG, Ellis IO, Ilyas M *et al* (2010). CTEN (C-terminal tensin-like), a novel oncogene overexpressed in invasive breast carcinoma of poor prognosis. *Breast Cancer Res Treat*. **14**.

Albiges L, Andre F, Balleyguier C, Gomez-Abuin G, Chompret A, Delaloge S (2005). Spectrum of breast cancer metastasis in BRCA1 mutation carriers: highly increased incidence of brain metastases. *Ann Oncol* **16:** 1846-7.

Alimandi M, Romano A, Curia MC, Muraro R, Fedi P, Aaronson SA *et al* (1995). Cooperative signaling of ErbB3 and ErbB2 in neoplastic transformation and human mammary carcinomas. *Oncogene* **10:** 1813-21.

Alpy F, Tomasetto C (2006). MLN64 and MENTHO, two mediators of endosomal cholesterol transport. *Biochem Soc Trans* **34:** 343-5.

Ankrapp DP, Bennett JM, Haslam SZ (1998). Role of epidermal growth factor in the acquisition of ovarian steroid hormone responsiveness in the normal mouse mammary gland. *J Cell Physiol* **174:** 251-60.

Aranda V, Haire T, Nolan ME, Calarco JP, Rosenberg AZ, Fawcett JP *et al* (2006). Par6-aPKC uncouples ErbB2 induced disruption of polarized epithelial organization from proliferation control. *Nat Cell Biol* **8:** 1235-45.

Ardlie KG, Kruglyak L, Seielstad M (2002). Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* **3:** 299-309.

Armes JE, Egan AJ, Southey MC, Dite GS, McCredie MR, Giles GG *et al* (1998). The histologic phenotypes of breast carcinoma occurring before age 40 years in women with and without BRCA1 or BRCA2 germline mutations: a population-based study. *Cancer* **83:** 2335-45.

Armes JE, Trute L, White D, Southey MC, Hammet F, Tesoriero A *et al* (1999). Distinct molecular pathogeneses of early-onset breast cancers in BRCA1 and BRCA2 mutation carriers: a population-based study. *Cancer Res* **59:** 2011-7.

Aubele M, Auer G, Braselmann H, Nahrig J, Zitzelsberger H, Quintanilla-Martinez L *et al* (2002). Chromosomal imbalances are associated with metastasis-free survival in breast cancer patients. *Anal Cell Pathol* **24:** 77-87.

Balasubramanian L, Sham JS, Yip KP (2008). Calcium signaling in vasopressin-induced aquaporin-2 trafficking. *Pflugers Arch* **456:** 747-54.

Balleine RL, Earl MJ, Greenberg ML, Clarke CL (1999). Absence of progesterone receptor associated with secondary breast cancer in postmenopausal women. *Br J Cancer* **79:** 1564-71.

Banerjee S, Reis-Filho JS, Ashley S, Steele D, Ashworth A, Lakhani SR *et al* (2006). Basal-like breast carcinomas: clinical outcome and response to chemotherapy. *J Clin Pathol* **59:** 729-35.

Barletta F, Wong CW, McNally C, Komm BS, Katzenellenbogen B, Cheskis BJ (2004). Characterization of the interactions of estrogen receptor and MNAR in the activation of cSrc. *Mol Endocrinol* **18:** 1096-108.

Barnes CJ, Kumar R (2004). Biology of the epidermal growth factor receptor family. *Cancer Treat Res* **119:** 1-13.

Bates RC, Edwards NS, Yates JD (2000). Spheroids and cell survival. *Crit Rev. Oncol. Hematol.* **36:** 61–74.

Bazley LA, Gullick WJ (2005). The epidermal growth factor receptor family. *Endocr Relat Cancer* **12 Suppl 1:** S17-27.

Bedard PL, Freedman OC, Howell A, Clemons M (2008). Overcoming endocrine resistance in breast cancer: are signal transduction inhibitors the answer? *Breast Cancer Res Treat* **108:** 307-17.

Beer S, Zetterberg A, Ihrie RA, McTaggart RA, Yang Q, Bradon N *et al* (2004). Developmental context determines latency of MYC-induced tumorigenesis. *PLoS Biol* **2:** e332.

Beerli RR, Wels W, Hynes NE (1994). Intracellular expression of single chain antibodies reverts ErbB-2 transformation. *J Biol Chem* **269:** 23931-6.

Bedigian HG, Copeland NG, Jenkins NA, Salvatore K, Rodick S. (1983). Emv-13 (Akv-3): a noninducible endogenous ecotropic provirus of AKR/J mice. *J Virol.* **46:** 490-7.

Belsches-Jablonski AP, Biscardi JS, Peavy DR, Tice DA, Romney DA, Parsons SJ (2001). Src family kinases and HER2 interactions in human breast cancer cell growth and survival. *Oncogene* **20:** 1465-75.

Benlimame N, He Q, Jie S, Xiao D, Xu YJ, Loignon M *et al* (2005). FAK signaling is critical for ErbB-2/ErbB-3 receptor cooperation for oncogenic transformation and invasion. *J Cell Biol* **171:** 505-16.

Benusiglio PR, Lesueur F, Luccarini C, Conroy DM, Shah M, Easton DF *et al* (2005). Common ERBB2 polymorphisms and risk of breast cancer in a white British population: a case-control study. *Breast Cancer Res* **7:** R204-9.

Bergamaschi A, Tagliabue E, Sorlie T, Naume B, Triulzi T, Orlandi R *et al* (2008). Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *J Pathol* **214:** 357-67.

Bergers G, Benjamin LE (2003). Tumourgenesis and angiogenic switch. *Nat. Rev. Cancer* **3:** 401-410.

Bernards R, Weinberg RA (2002). A progression puzzle. *Nature* **418:** 823.

Bernig T, Chanock SJ (2006). Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer. *Expert Rev Mol Diagn* **6:** 319-31.

Bertram JS (2000). The molecular biology of cancer. *Mol Aspects Med* **21:** 167-223.

Betapudi V, Licate LS, Egelhoff TT (2006). Distinct roles of nonmuscle myosin II isoforms in the regulation of MDA-MB-231 breast cancer cell spreading and migration. *Cancer Res* **66:** 4725-33.

Betsill WL, Jr., Rosen PP, Lieberman PH, Robbins GF (1978). Intraductal carcinoma. Long-term follow-up after treatment by biopsy alone. *JAMA* **239:** 1863-7.

Bishop JM (1981). Enemies within: the genesis of retrovirus oncogenes. *Cell* **23:** 5-6.

Blegen H, Will JS, Ghadimi BM, Nash HP, Zetterberg A, Auer G *et al* (2003). DNA amplifications and aneuploidy, high proliferative activity and impaired cell cycle control characterize breast carcinomas with poor prognosis. *Anal Cell Pathol* **25:** 103-14.

Bogenrieder T, Herlyn M (2003). Axis of evil: molecular mechanisms of cancer metastasis. *Oncogene* **22:** 6524-36.

Boller K, Vestweber D, Kemler R (1985). Cell-adhesion molecule uvomorulin is localized in the intermediate junctions of adult intestinal epithelial cells. *J Cell Biol* **100:** 327-32.

Boxer RB, Jang JW, Sintasath L, Chodosh LA (2004). Lack of sustained regression of c-MYC-induced mammary adenocarcinomas following brief or prolonged MYC inactivation. *Cancer Cell* **6:** 577-86.

Bratthauer GL, Moinfar F, Stamatakos MD, Mezzetti TP, Shekitka KM, Man YG *et al* (2002). Combined E-cadherin and high molecular weight cytokeratin immunoprofile differentiates lobular, ductal, and hybrid mammary intraepithelial neoplasias. *Hum Pathol* **33:** 620-7.

Broustas CG, Ross JS, Yang Q, Sheehan CE, Riggins R, Noone AM *et al* (2010). The Proapoptotic Molecule BLID Interacts with Bcl-XL and Its Downregulation in Breast Cancer Correlates with Poor Disease-Free and Overall Survival. *Clin Cancer Res.* **25:** 2939.

Brown D.M, Ruoslahti E (2004). Metadherin, a cell surface protein in breast tumors that mediates lung metastasis. *Cancer Cell* **5:** 365-374.

Bur ME, Zimarowski MJ, Schnitt SJ, Baker S, Lew R (1992). Estrogen receptor immunohistochemistry in carcinoma in situ of the breast. *Cancer* **69:** 1174-81.

Byron SA, Horwitz KB, Richer JK, Lange CA, Zhang X, Yee D (2006). Insulin receptor substrates mediate distinct biological responses to insulin-like growth factor receptor activation in breast cancer cells. *Br J Cancer* **95:** 1220-8.

Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K *et al* (2006). Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* **295:** 2492-502.

Carlson CS, Eberle MA, Kruglyak L, Nickerson DA (2004). Mapping complex disease loci in whole-genome association studies. *Nature* **429:** 446-52.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* **74:** 106-20.

Carraway KL, 3rd, Rossi EA, Komatsu M, Price-Schiavi SA, Huang D, Guy PM *et al* (1999). An intramembrane modulator of the ErbB2 receptor tyrosine kinase that potentiates neuregulin signaling. *J Biol Chem* **274:** 5263-6.

Castiglioni F, Tagliabue E, Campiglio M, Pupa SM, Balsari A, Menard S (2006). Role of exon-16-deleted HER2 in breast carcinomas. *Endocr Relat Cancer* **13:** 221-32.

Cavalli LR, Liang BC (1998). Mutagenesis, tumorigenicity, and apoptosis: are the mitochondria involved? *Mutat Res* **398:** 19-26.

Chalbos D, Chambon M, Ailhaud G, Rochefort H (1987). Fatty acid synthetase and its mRNA are induced by progestins in breast cancer cells. *J Biol Chem* **262:** 9923-6.

Chambers AF, Groom AC, MacDonald IC (2002). Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* **2:** 563-72.

Chambers AF, Harris JF, Ling V, Hill RP (1984). Rapid phenotype variation in cells derived from lung metastases of KHT fibrosarcoma. *Invasion Metastasis* **4:** 225-37.

Chambers AF, Wilson S (1988). Use of NeoR B16F1 murine melanoma cells to assess clonality of experimental metastases in the immune-deficient chick embryo. *Clin Exp Metastasis* **6:** 171-82.

Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T *et al* (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* **102:** 3738-43.

Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K *et al* (2004). Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* **2:** E7.

Chanock S (2001). Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis Markers* **17:** 89-98.

Chappuis PO, Nethercot V, Foulkes WD (2000). Clinico-pathological characteristics of BRCA1- and BRCA2-related breast cancer. *Semin Surg Oncol* **18:** 287-95.

Chelouche Lev D, Price JE (2002). Therapeutic intervention with breast cancer metastasis. *Crit Rev Eukaryot Gene Expr* **12:** 137-50.

Chen GK, Jorgenson E, Witte JS (2007). An empirical evaluation of the common disease-common variant hypothesis. *BMC Proc* **1 Suppl 1:** S5.

Cheung ST, Chen X, Guan XY, Wong SY, Tai LS, Ng IO *et al* (2002). Identify metastasis-associated genes in hepatocellular carcinoma through clonality delineation for multinodular tumor. *Cancer Res* **62:** 4711-21.

Chia KS, Du WB, Sankaranarayanan R, Sankila R, Wang H, Lee J *et al* (2004). Do younger female breast cancer patients have a poorer prognosis? Results from a population-based survival analysis. *Int J Cancer* **108:** 761-5.

Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL *et al* (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10:** 529-41.

Chlebowski RT, Aiello E, McTiernan A (2002). Weight loss in breast cancer patient management. *J Clin Oncol* **20:** 1128-43.

Cho HS, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, Jr. *et al* (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421:** 756-60.

Choudhury A, Charo J, Parapuram SK, Hunt RC, Hunt DM, Seliger B *et al* (2004). Small interfering RNA (siRNA) inhibits the expression of the Her2/neu gene, upregulates HLA class I and induces apoptosis of Her2/neu positive tumor cell lines. *Int J Cancer* **108:** 71-7.

Christofori G (2006). New signals from the invasive front. *Nature* **441:** 444–450.

Claus EB, Stowe M, Carter D (2003). Family history of breast and ovarian cancer and the risk of breast carcinoma in situ. *Breast Cancer Res Treat* **78:** 7-15.

Cobleigh MA, Tabesh B, Bitterman P, Baker J, Cronin M, Liu ML *et al* (2005). Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clin Cancer Res* **11:** 8623-31.

Coleman WB, Tsongalis GJ (2006). Molecular mechanisms of human carcinogenesis. *EXS***:** 321-49.

Collins LC, Tamimi RM, Baer HJ, Connolly JL, Colditz GA, Schnitt SJ (2005). Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. *Cancer* **103:** 1778-84.

Colomer R, Lupu R, Bacus SS, Gelmann EP (1994). erbB-2 antisense oligonucleotides inhibit the proliferation of breast carcinoma cells with erbB-2 oncogene amplification. *Br J Cancer* **70:** 819-25.

Conne B, Stutz A, Vassalli JD (2000). The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? *Nat Med* **6:** 637-41.

Cowen D, Jacquemier J, Houvenaeghel G, Viens P, Puig B, Bardou VJ *et al* (1998). Local and distant recurrence after conservative management of "very low-risk" breast cancer are dependent events: a 10-year follow-up. *Int J Radiat Oncol Biol Phys* **41:** 801-7.

Crawford NP, Qian X, Ziogas A, Papageorge AG, Boersma BJ, Walker RC *et al* (2007). Rrp1b, a new candidate susceptibility gene for breast cancer progression and metastasis. *PLoS Genet* **3:** e214.

Crawford NP, Walker RC, Lukes L, Officewala JS, Williams RW, Hunter KW (2008). The Diasporin Pathway: a tumor progression-related transcriptional network that predicts breast cancer survival. *Clin Exp Metastasis* **25:** 357-69.

Crawford NP, Ziogas A, Peel DJ, Hess J, Anton-Culver H, Hunter KW (2006). Germline polymorphisms in SIPA1 are associated with metastasis and other indicators of poor prognosis in breast cancer. *Breast Cancer Res* **8:** R16.

Crowe JP, Patrick RJ, Rybicki LA, Escobar PF, Weng D, Thomas Budd G *et al* (2006). A data model to predict HER2 status in breast cancer based on the clinical and pathologic profiles of a large patient population at a single institution. *Breast* **15:** 728-35.

Cui D, Jin G, Gao T, Sun T, Tian F, Estrada GG *et al* (2004). Characterization of BRCAA1 and its novel antigen epitope identification. *Cancer Epidemiol Biomarkers Prev* **13:** 1136-45.

Cui X, Schiff R, Arpino G, Osborne CK, Lee AV (2005). Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. *J Clin Oncol* **23:** 7721-35.

Dal Lago L, Durbecq V, Desmedt C, Salgado R, Verjat T, Lespagnard L *et al* (2006). Correction for chromosome-17 is critical for the determination of true Her-2/neu gene amplification status in breast cancer. *Mol Cancer Ther* **5:** 2572-9.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001). High-resolution haplotype structure in the human genome. *Nat Genet* **29:** 229-32.

de Mascarel I, MacGrogan G, Mathoulin-Pelissier S, Soubeyran I, Picot V, Coindre JM (2002). Breast ductal carcinoma in situ with microinvasion: a definition supported by a long-term study of 1248 serially sectioned ductal carcinomas. *Cancer* **94:** 2134-42.

De Vivo I, Hankinson SE, Colditz GA, Hunter DJ (2003). A functional polymorphism in the progesterone receptor gene is associated with an increase in breast cancer risk. *Cancer Res* **63:** 5236-8.

Dellapasqua S, Castiglione-Gertsch M (2005). Adjuvant endocrine therapies for postmenopausal women with early breast cancer: standards and not. *Breast* **14:** 555-63.

Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA *et al* (2007). Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res* **13:** 4429-34.

Derksen PW, Liu X, Saridin F, van der Gulden H, Zevenhoven J, Evers B *et al* (2006). Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma through induction of anoikis resistance and angiogenesis. *Cancer Cell* **10:** 437–449.

Deshane J, Grim J, Loechel S, Siegal GP, Alvarez RD, Curiel DT (1996). Intracellular antibody against erbB-2 mediates targeted tumor cell eradication by apoptosis. *Cancer Gene Ther* **3:** 89-98.

Diel IJ, Kaufmann M, Costa SD, Holle R, von Minckwitz G, Solomayer EF *et al* (1996). Micrometastatic breast cancer cells in bone marrow at primary surgery: prognostic value in comparison with nodal status. *J Natl Cancer Inst* **88:** 1652-8.

Dorval M, Maunsell E, Deschenes L, Brisson J, Masse B (1998). Long-term quality of life after breast cancer: comparison of 8-year survivors with population controls. *J Clin Oncol* **16:** 487-94.

Doyle T, Schultz DJ, Peters C, Harris E, Solin LJ (2001). Long-term results of local recurrence after breast conservation treatment for invasive breast cancer. *Int J Radiat Oncol Biol Phys* **51:** 74-80.

Duan J, Sanders AR, Molen JE, Martinolich L, Mowry BJ, Levinson DF *et al* (2003). Polymorphisms in the 5'-untranslated region of the human serotonin receptor 1B (HTR1B) gene affect gene expression. *Mol Psychiatry* **8:** 901-10.

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J *et al* (2000). The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat Genet* **25:** 320-3.

Eccles SA, Welch DR (2007). Metastasis: recent discoveries and novel treatment strategies. *Lancet* **369:** 1742–1757.

Eisinger F, Nogues C, Guinebretiere JM, Peyrat JP, Bardou VJ, Noguchi T *et al* (1999). Novel indications for BRCA1 screening using individual clinical and morphological features. *Int J Cancer* **84:** 263-7.

El Khoury M, Cherel P, Becette V, De Maulmont C, Costes V, Talma V *et al* (2004). Unusual soft-tissue metastasis of an invasive lobular carcinoma mimicking fasciitis. *AJR Am J Roentgenol* **182:** 745-7.

Elder EE, Kennedy CW, Gluch L, Carmalt HL, Janu NC, Joseph MG *et al* (2006). Patterns of breast cancer relapse. *Eur J Surg Oncol* **32:** 922-7.

Ellis MJ, Coop A, Singh B, Tao Y, Llombart-Cussac A, Janicke F *et al* (2003). Letrozole inhibits tumor proliferation more effectively than tamoxifen independent of HER1/2 expression status. *Cancer Res* **63:** 6523-31.

Erichsen HC, Chanock SJ (2004). SNPs in cancer research and treatment. *Br J Cancer* **90:** 747-51.

Evan GI, Wyllie AH, Gilbert CS, Littlewood TD, Land H, Brooks M *et al* (1992). Induction of apoptosis in fibroblasts by c-myc protein. *Cell* **69:** 119-28.

Fadare O, Tavassoli FA (2008). Clinical and pathologic aspects of basal-like breast cancers. *Nat Clin Pract Oncol* **5:** 149-59.

Faltus T, Yuan J, Zimmer B, Kramer A, Loibl S, Kaufmann M *et al* (2004). Silencing of the HER2/neu gene by siRNA inhibits proliferation and induces apoptosis in HER2/neu-overexpressing breast cancer cells. *Neoplasia* **6:** 786-95.

Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB *et al* (2006). Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* **355:** 560-9.

Farina A, Hattori M, Qin J, Nakatani Y, Minato N, Ozato K (2004). Bromodomain protein Brd4 binds to GTPase-activating SPA-1, modulating its activity and subcellular localization. *Mol Cell Biol* **24:** 9059-69.

Fedi P, Pierce JH, di Fiore PP, Kraus MH (1994). Efficient coupling with phosphatidylinositol 3-kinase, but not phospholipase C gamma or GTPase-activating protein, distinguishes ErbB-3 signaling from that of other ErbB/EGFR family members. *Mol Cell Biol* **14:** 492-500.

Felsher DW, Bishop JM (1999). Reversible tumorigenesis by MYC in hematopoietic lineages. *Mol Cell* **4:** 199-207.

Feng Y, Manka D, Wagner KU, Khan SA (2007). Estrogen receptor-alpha expression in the mammary epithelium is required for ductal and alveolar morphogenesis in mice. *Proc Natl Acad Sci U S A* **104:** 14718-23.

Fidler IJ (1986). Rationale and methods for the use of nude mice to study the biology and therapy of human cancer metastasis. *Cancer Metastasis Rev* **5:** 29-49.

Fidler IJ (1991). Cancer metastasis. *Br Med Bull* **47:** 157-77.

Fidler IJ (2003). The pathology of cancer metastasis: the ‘seed and soil’ hypothesis revisited. *Nat. Rev. Cancer.* **3:** 453-458.

Fidler IJ, Kripke ML (1977). Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197:** 893-5.

Filardo EJ, Thomas P (2005). GPR30: a seven-transmembrane-spanning estrogen receptor that triggers EGF release. *Trends Endocrinol Metab* **16:** 362-7.

Finkle D, Quan ZR, Asghari V, Kloss J, Ghaboosi N, Mai E *et al* (2004). HER2-targeted therapy reduces incidence and progression of midlife mammary tumors in female murine mammary tumor virus huHER2-transgenic mice. *Clin Cancer Res* **10:** 2499-511.

Fleishman SJ, Schlessinger J, Ben-Tal N (2002). A putative molecular-activation switch in the transmembrane domain of erbB2. *Proc Natl Acad Sci U S A* **99:** 15937-40.

Foekens JA, Portengen H, van Putten WL, Peters HA, Krijnen HL, Alexieva-Figusch J *et al* (1989). Prognostic value of estrogen and progesterone receptors measured by enzyme immunoassays in human breast tumor cytosols. *Cancer Res* **49:** 5823-8.

Font de Mora J, Brown M (2000). AIB1 is a conduit for kinase-mediated growth factor signaling to the estrogen receptor. *Mol Cell Biol* **20:** 5041-7.

Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P *et al* (1998). Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* **62:** 676-89.

Foulkes WD, Metcalfe K, Hanna W, Lynch HT, Ghadirian P, Tung N *et al* (2003). Disruption of the expected positive correlation between breast tumor size and lymph node status in BRCA1-related breast carcinoma. *Cancer* **98:** 1569-77.

Fowble B (1999). Ipsilateral breast tumor recurrence following breast-conserving surgery for early-stage invasive cancer. *Acta Oncol* **38 Suppl 13:** 9-17.

Frank TS, Deffenbaugh AM, Reid JE, Hulick M, Ward BE, Lingenfelter B *et al* (2002). Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals. *J Clin Oncol* **20:** 1480-90.

Frisch SM, Francis H (1994). Disruption of epithelial cell–matrix interactions induces apoptosis. *J. Cell Biol.* **124:** 619–626.
http://www.sciencedirect.com/science?_ob=RedirectURL&_method=outwardLink&_partnerName=3&_originPage=article&_zone=art_page&_targetURL=http%3A%2F%2Fdx.doi.org%2F10.1083%252Fjcb.124.4.619&_acct=C000006418&_version=1&_userid=79777&md5=20529dc97ebb27210e2e670108ce9f6e

Frisch SM, Ruoslahti E (1997). Integrins and anoikis. *Curr. Opin. Cell Biol.* **9:** 701–706.

Friedl P, Wolf K (2003). Tumour-cell invasion and migration: diversity and escape mechanisms. *Nat. Rev. Cancer* **3:** 362–374.

Fritz DT, Jiang S, Xu J, Rogers MB (2006). A polymorphism in a conserved posttranscriptional regulatory motif alters bone morphogenetic protein 2 (BMP2) RNA:protein interactions. *Mol Endocrinol* **20:** 1574-86.

Frykberg ER (1999). Lobular Carcinoma In Situ of the Breast. *Breast J* **5:** 296-303.

Frykberg ER, Bland KI (1993). In situ breast carcinoma. *Adv Surg* **26:** 29-72.

Fuchs E, Metastasenbildung., Sarcom des Uvealtractus, Wilhelm Braumueller (1882). Pp 197-206.

Fujii S, Ito K, Ito Y, Ochiai A (2008). Enhancer of zeste homologue 2 (EZH2) down-regulates RUNX3 by increasing histone H3 methylation. *J Biol Chem* **283:** 17324-32.

Fulford LG, Easton DF, Reis-Filho JS, Sofronis A, Gillett CE, Lakhani SR *et al* (2006). Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast. *Histopathology* **49:** 22-34.

Fulford LG, Reis-Filho JS, Ryder K, Jones C, Gillett CE, Hanby A *et al* (2007). Basal-like grade III invasive ductal carcinoma of the breast: patterns of metastasis and long-term survival. *Breast Cancer Res* **9:** R4.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B *et al* (2002). The structure of haplotype blocks in the human genome. *Science* **296:** 2225-9.

Gambaletta D, Marchetti A, Benedetti L, Mercurio AM, Sacchi A, Falcioni R (2000). Cooperative signaling between alpha(6)beta(4) integrin and ErbB-2 receptor is required to promote phosphatidylinositol 3-kinase-dependent invasion. *J Biol Chem* **275:** 10604-10.

Gandhi A, Holland PA, Knox WF, Potten CS, Bundred NJ (1998). Evidence of significant apoptosis in poorly differentiated ductal carcinoma in situ of the breast. *Br J Cancer* **78:** 788-94.

Garrett TP, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO *et al* (2003). The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Mol Cell* **11:** 495-505.

Gatenby RA, Gillies RJ (2008). A microenvironmental model of carcinogenesis. *Nat. Rev. Cancer* **8:** 56–61.

Giagounidis AA, Heinsch M, von Barany RU, Erlemannb R, Aul C (2005). Breast cancer metastasis to the aortic vessel wall. *Onkologie* **28:** 369.

Gimond C, van der Flier A, van Delft S, Brakebusch C, Kuikman I, Collard JG et al (1999). Induction of cell scattering by expression of beta1 integrins in beta1-deficient epithelial cells requires activation of members of the rho family of GTPases and downregulation of cadherin and catenin function. *J. Cell Biol.* **147:** 1325–1340.

Giretti MS, Fu XD, De Rosa G, Sarotto I, Baldacci C, Garibaldi S *et al* (2008). Extra-nuclear signalling of estrogen receptor to breast cancer cytoskeletal remodelling, migration and invasion. *PLoS ONE* **3:** e2238.

Geiger TR, Peeper DS (2005). The neurotrophic receptor TrkB in anoikis resistance and metastasis: a perspective. *Cancer Res.* **65:** 7033–7036.

Geiger TR, Peeper DS (2009). Metastasis mechanisms. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1796:** 293-308

Goel A, Janknecht R (2003). Acetylation-mediated transcriptional activation of the ETS protein ER81 by p300, P/CAF, and HER2/Neu. *Mol Cell Biol* **23:** 6243-54.

Goldstein NS, Vicini FA, Hunter S, Odish E, Forbes S, Kraus D *et al* (2005). Molecular clonality determination of ipsilateral recurrence of invasive breast carcinomas after breast-conserving therapy: comparison with clinical and biologic factors. *Am J Clin Pathol* **123:** 679-89.

Gonzalez-Angulo AM, McGuire SE, Buchholz TA, Tucker SL, Kuerer HM, Rouzier R *et al* (2005). Factors predictive of distant metastases in patients with breast cancer who have a pathologic complete response after neoadjuvant chemotherapy. *J Clin Oncol* **23:** 7098-104.

Goodearl A, Viehover A, Vartanian T (2001). Neuregulin-induced association of Sos Ras exchange protein with HER2(erbB2)/HER3(erbB3) receptor complexes in Schwann cells through a specific Grb2-HER2(erbB2) interaction. *Dev Neurosci* **23:** 25-30.

Goss PE, Ingle JN, Martino S, Robert NJ, Muss HB, Piccart MJ *et al* (2005). Randomized trial of letrozole following tamoxifen as extended adjuvant therapy in receptor-positive breast cancer: updated findings from NCIC CTG MA.17. *J Natl Cancer Inst* **97:** 1262-71.

Graus-Porta D, Beerli RR, Daly JM, Hynes NE (1997). ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO J* **16:** 1647-55.

Grigoriadis A, Mackay A, Reis-Filho JS, Steele D, Iseli C, Stevenson BJ *et al* (2006). Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. *Breast Cancer Res* **8:** R56.

Gross GE, Clark GM, Chamness GC, McGuire WL (1984). Multiple progesterone receptor assays in human breast cancer. *Cancer Res* **44:** 836-40.

Gumbiner B, Stevenson B, Grimaldi A (1988). The role of the cell adhesion molecule uvomorulin in the formation and maintenance of the epithelial junctional complex. *J Cell Biol* **107:** 1575-87.

Guo W, Giancotti FG (2004). Integrin signalling during tumour progression. *Nat. Rev. Mol. Cell Biol.* **5:** 816–826.

Gupta GP, Massague DJ (2006). Cancer metastasis: building a framework, *Cell* **127:** 679–695.

Gusterson BA, Ross DT, Heath VJ, Stein T (2005). Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res* **7:** 143-8.

Haber D, Harlow E (1997). Tumour-suppressor genes: evolving definitions in the genomic age. *Nat Genet* **16:** 320-2.

Han W, Han MR, Kang JJ, Bae JY, Lee JH, Bae YJ *et al* (2006). Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer. *BMC Cancer* **6:** 92.

Hanahan D, Weinberg RA (2000). The hallmarks of cancer. *Cell* **100:** 57-70.

Hanahan D, Folkman J (1996). Patterns and emerging mechanisms of the angiogenic switch during tumourgenesis. *Cell* **86:** 353-364.

Harrington EA, Fanidi A, Evan GI (1994). Oncogenes and cell death. *Curr Opin Genet Dev* **4:** 120-9.

Harris JF, Chambers AF, Hill RP, Ling V (1982). Metastatic variants are generated spontaneously at a high rate in mouse KHT tumor. *Proc Natl Acad Sci U S A* **79:** 5547-51.

Harvey JM, Clark GM, Osborne CK, Allred DC (1999). Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* **17:** 1474-81.

Haslam SZ, Counterman LJ, Nummy KA (1993). Effects of epidermal growth factor, estrogen, and progestin on DNA synthesis in mammary cells in vivo are determined by the developmental state of the gland. *J Cell Physiol* **155:** 72-8.

Haslam SZ, Woodward TL (2003). Host microenvironment in breast cancer development: epithelial-cell-stromal-cell interactions and steroid hormone action in normal and cancerous mammary gland. *Breast Cancer Res* **5:** 208-15.

Hattori M, Tsukamoto N, Nur-e-Kamal MS, Rubinfeld B, Iwai K, Kubota H *et al* (1995). Molecular cloning of a novel mitogen-inducible nuclear protein with a Ran GTPase-activating domain that affects cell cycle progression. *Mol Cell Biol* **15:** 552-60.

Hawkins RA, Tesdale AL, Ferguson WA, Going JJ (1987). Oestrogen receptor activity in intraduct and invasive breast carcinomas. *Breast Cancer Res Treat* **9:** 129-33.

Hawks J, Hunley K, Lee SH, Wolpoff M (2000). Population bottlenecks and Pleistocene human evolution. *Mol Biol Evol* **17:** 2-22.

Haylock BJ, Coppin CM, Jackson J, Basco VE, Wilson KS (2000). Locoregional first recurrence after mastectomy: prospective cohort studies with and without immediate chemotherapy. *Int J Radiat Oncol Biol Phys* **46:** 355-62.

Hayward JL (1977). The Guy's trial of treatments of "early" breast cancer. *World J Surg* **1:** 314-6.

Hellyer NJ, Kim MS, Koland JG (2001). Heregulin-dependent activation of phosphoinositide 3-kinase and Akt via the ErbB2/ErbB3 co-receptor. *J Biol Chem* **276:** 42153-61.

Hewitt SC, Korach KS (2000). Progesterone action and responses in the alphaERKO mouse. *Steroids* **65:** 551-7.

Hicks DG, Short SM, Prescott NL, Tarr SM, Coleman KA, Yoder BJ *et al* (2006). Breast cancers with brain metastases are more likely to be estrogen receptor negative, express the basal cytokeratin CK5/6, and overexpress HER2 or EGFR. *Am J Surg Pathol* **30:** 1097-104.

Hildenbrand R, Arens N (2004). Protein and mRNA expression of uPAR and PAI-1 in myoepithelial cells of early breast cancer lesions and normal breast tissue. *Br J Cancer* **91:** 564-71.

Hinz S, Kempkensteffen C, Christoph F, Hoffmann M, Krause H, Schrader M *et al* (2008). Expression of the polycomb group protein EZH2 and its relation to outcome in patients with urothelial carcinoma of the bladder. *J Cancer Res Clin Oncol* **134:** 331-6.

Hirakawa T, Ruley HE (1988). Rescue of cells from ras oncogene-induced growth arrest by a second, complementing, oncogene. *Proc Natl Acad Sci U S A* **85:** 1519-23.

Hiratsuka S, Watanabe A, Aburatani H, Maru Y (2006). Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. *Nat. Cell Biol.* **8:** 1369–1375.

Hiratsuka S, Watanabe A, Sakurai Y, Kashi-Takamura S, Ishibashi S, Miyake K *et al* (2008). The S100A8-serum amyloid A3-TLR4 paracrine cascade establishes a pre-metastatic phase. *Nat. Cell Biol.* **10:** 1349–1355.

Hiyama E, Hiyama K (2003). Telomerase as tumor marker. *Cancer Lett* **194:** 221-33.

Hoadley KA, Weigman VJ, Fan C, Sawyer LR, He X, Troester MA *et al* (2007). EGFR associated expression profiles vary with breast tumor subtype. *BMC Genomics* **8:** 258.

Hoffmann MJ, Engers R, Florl AR, Otte AP, Muller M, Schulz WA (2007). Expression changes in EZH2, but not in BMI-1, SIRT1, DNMT1 or DNMT3B are associated with DNA methylation changes in prostate cancer. *Cancer Biol Ther* **6:** 1403-12.

Holash J, Maisonpierre PC, Compton D, Boland P, Alexander CR, Zagzag D *et al* (1999). Vessel cooption, regression and growth in tumors mediated by angioproteins and VEGF. *Science* **284:** 1994-1998.

Holbro T, Beerli RR, Maurer F, Koziczak M, Barbas CF, 3rd, Hynes NE (2003). The ErbB2/ErbB3 heterodimer functions as an oncogenic unit: ErbB2 requires ErbB3 to drive breast tumor cell proliferation. *Proc Natl Acad Sci U S A* **100:** 8933-8.

Holling TM, Bergevoet MW, Wilson L, Van Eggermond MC, Schooten E, Steenbergen RD *et al* (2007). A role for EZH2 in silencing of IFN-gamma inducible MHC2TA transcription in uveal melanoma. *J Immunol* **179:** 5317-25.

Holmgren L, O'Reilly MS, Folkman J (1995). Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression. *Nat Med* **1:** 149-53.

Hood JD, Cheresh DA (2002). Role of integrins in cell invasion and migration. *Nat. Rev. Cancer* **2:** 91–100.

Hoogendoorn B, Coleman SL, Guy CA, Smith K, Bowen T, Buckland PR *et al* (2003). Functional analysis of human promoter polymorphisms. *Hum Mol Genet* **12:** 2249-54.

Hopp TA, Weiss HL, Hilsenbeck SG, Cui Y, Allred DC, Horwitz KB *et al* (2004). Breast cancer patients with progesterone receptor PR-A-rich tumors have poorer disease-free survival rates. *Clin Cancer Res* **10:** 2751-60.

Hoque A, Sneige N, Sahin AA, Menter DG, Bacus JW, Hortobagyi GN *et al* (2002). Her-2/neu gene amplification in ductal carcinoma in situ of the breast. *Cancer Epidemiol Biomarkers Prev* **11:** 587-90.

Hsieh SM, Lintell NA, Hunter KW (2006). Germline polymorphisms are potential metastasis risk and prognosis markers in breast cancer. *Breast Dis* **26:** 157-62.

Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF *et al* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7:** 96.

Hueber AO, Evan GI (1998). Traps to catch unwary oncogenes. *Trends Genet* **14:** 364-7.

Hugo H, Ackland ML, Blick T, Lawrence MG, Clements JA, Williams ED *et al* (2007). Epithelial--mesenchymal and mesenchymal--epithelial transitions in carcinoma progression. *J Cell Physiol.* **213:** 374-83.

Hunter K (2005). The intersection of inheritance and metastasis: the role and implications of germline polymorphism in tumor dissemination. *Cell Cycle* **4:** 1719-21.

Hunter K (2006). Host genetics influence tumour metastasis. *Nat Rev Cancer* **6:** 141-6.

Hunter KW (2003). Allelic diversity in the host genetic background may be an important determinant in tumor metastatic dissemination. *Cancer Lett* **200:** 97-105.

Hunter KW, Crawford NP (2006). Germ line polymorphism in metastatic progression. *Cancer Res* **66:** 1251-4.

Hunter KW, Crawford NP (2008). The future of mouse QTL mapping to diagnose disease in mice in the age of whole-genome association studies. *Annu Rev Genet* **42:** 131-41.

Hunter KW, Crawford NP, Alsarraj J (2008). Mechanisms of metastasis. *Breast Cancer Res* **10 Suppl 1:** S2.

Hwang ES, DeVries S, Chew KL, Moore DH, 2nd, Kerlikowske K, Thor A *et al* (2004). Patterns of chromosomal alterations in breast ductal carcinoma in situ. *Clin Cancer Res* **10:** 5160-7.

Hwang ES, McLennan JL, Moore DH, Crawford BB, Esserman LJ, Ziegler JL (2007). Ductal carcinoma in situ in BRCA mutation carriers. *J Clin Oncol* **25:** 642-7.

I ST, Nie Z, Stewart A, Najdovska M, Hall NE, He H *et al* (2004). ARAP3 is transiently tyrosine phosphorylated in cells attaching to fibronectin and inhibits cell spreading in a RhoGAP-dependent manner. *J Cell Sci* **117:** 6071-84.

Ignatoski KM, Maehama T, Markwart SM, Dixon JE, Livant DL, Ethier SP (2000). ERBB-2 overexpression confers PI 3' kinase-dependent invasion capacity on human mammary epithelial cells. *Br J Cancer* **82:** 666-74.

Igney FH, Krammer PH (2002). Death and anti-death: tumour resistance to apoptosis. *Nat Rev Cancer* **2:** 277-88.

Ilyas M, Straub J, Tomlinson IP, Bodmer WF (1999). Genetic pathways in colorectal and other cancers. *Eur J Cancer* **35:** 1986-2002.

Ingraham CA, Cox ME, Ward DC, Fults DW, Maness PF (1989). c-src and other proto-oncogenes implicated in neuronal differentiation. *Mol Chem Neuropathol* **10:** 1-14.

Irvine T, Fentiman IS (2007). Biology and treatment of ductal carcinoma in situ. *Expert Rev Anticancer Ther* **7:** 135-45.

Işeri OD, Kars MD, Arpaci F, Gündüz U (2010). Gene expression analysis of drug-resistant MCF-7 cells: implications for relation to extracellular matrix proteins. *Cancer Chemother Pharmacol.* **65:** 447-55.

Ishida D, Kometani K, Yang H, Kakugawa K, Masuda K, Iwai K *et al* (2003). Myeloproliferative stem cell disorders by deregulated Rap1 activation in SPA-1-deficient mice. *Cancer Cell* **4:** 55-65.

Ishizawar RC, Miyake T, Parsons SJ (2007). c-Src modulates ErbB2 and ErbB3 heterocomplex formation and function. *Oncogene* **26:** 3503-10.

Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J *et al* (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* **66:** 10292-301.

Jackson DG (2004). Biology of the lymphatic marker LYVE-1 and applications in research into lymphatic trafficking and lymphangiogenesis. *APMIS* **112:** 526-38.

Jacob J, Sebastian KS, Devassy S, Priyadarsini L, Farook MF, Shameem A *et al* (2006). Membrane estrogen receptors: genomic actions and post transcriptional regulation. *Mol Cell Endocrinol* **246:** 34-41.

Jacobsen BM, Richer JK, Schittone SA, Horwitz KB (2002). New human breast cancer cells to study progesterone receptor isoform ratio effects and ligand-independent gene regulation. *J Biol Chem* **277:** 27793-800.

Janes PW, Lackmann M, Church WB, Sanderson GM, Sutherland RL, Daly RJ (1997). Structural determinants of the interaction between the erbB2 receptor and the Src homology 2 domain of Grb7. *J Biol Chem* **272:** 8490-7.

Janes SM, Watt FM (2006). New roles for integrins in squamous-cell carcinoma. *Nat. Rev. Cancer* **6:** 175–183.

Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ (2007). Cancer statistics, 2007. *CA Cancer J Clin* **57:** 43-66.

James P, Halladay J, Craig EA (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid seletion in yeast. *Genetics* **144:** 1425-1436.

Jiang P, Enomoto A, Takahashi M (2009). Cell biology of the movement of breast cancer cells: Intracellular signalling and the actin cytoskeleton. *Cancer Letters* **284:** 122-130.

Jézéquel P, Campone M, Roché H, Gouraud W, Charbonnel C, Ricolleau G *et al* (2009). A 38-gene expression signature to predict metastasis risk in node-positive breast cancer after systemic adjuvant chemotherapy: a genomic substudy of PACS01 clinical trial. *Breast Cancer Res Treat.* Aug;116(3):509-20.

Johannsson OT, Idvall I, Anderson C, Borg A, Barkardottir RB, Egilsson V *et al* (1997). Tumour biological features of BRCA1-induced breast and ovarian cancer. *Eur J Cancer* **33:** 362-71.

Johnson AJ, Mone AP, Abhyankar V, Byrd JC (2003). Advances in the therapy of chronic lymphocytic leukemia. *Curr Opin Hematol* **10:** 297-305.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G *et al* (2001). Haplotype tagging for the identification of common disease genes. *Nat Genet* **29:** 233-7.

Jones C, Ford E, Gillett C, Ryder K, Merrett S, Reis-Filho JS *et al* (2004). Molecular cytogenetic identification of subgroups of grade III invasive ductal breast carcinomas with different clinical outcomes. *Clin Cancer Res* **10:** 5988-97.

Jonkers J, Berns A (2004). Oncogene addiction: sometimes a temporary slavery. *Cancer Cell* **6:** 535-8.

Juhl H, Downing SG, Wellstein A, Czubayko F (1997). HER-2/neu is rate-limiting for ovarian cancer growth. Conditional depletion of HER-2/neu by ribozyme targeting. *J Biol Chem* **272:** 29482-6.

Kadivar M, Joulaee A, Kashkouli MB, Kharazi HH, Kalantari M, Kumar PV (2006). Orbital metastasis as the first presentation of nonpalpable invasive lobular carcinoma of the breast. *Breast J* **12:** 75-6.

Kahlert S, Nuedling S, van Eickels M, Vetter H, Meyer R, Grohe C (2000). Estrogen receptor alpha rapidly activates the IGF-1 receptor pathway. *J Biol Chem* **275:** 18447-53.

Kalluri R (2003). Basement membranes: structure, assembly and role in tumour angiogenesis. *Nat. Rev. Cancer* **3:** 422-433.

Kamby C, Sengelov L (1997). Pattern of dissemination and survival following isolated locoregional recurrence of breast cancer. A prospective study with more than 10 years of follow up. *Breast Cancer Res Treat* **45:** 181-92.

Karp SE, Tonin PN, Begin LR, Martinez JJ, Zhang JC, Pollak MN *et al* (1997). Influence of BRCA1 mutations on nuclear grade and estrogen receptor status of breast carcinoma in Ashkenazi Jewish women. *Cancer* **80:** 435-41.

Kastner P, Krust A, Turcotte B, Stropp U, Tora L, Gronemeyer H *et al* (1990). Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. *EMBO J* **9:** 1603-14.

Kauffman EC, Robinson VL, Stadler WM, Sokoloff MH, Rinker-Schaeffer CW (2003). Metastasis suppression: the evolving role of metastasis suppressor genes for regulating cancer cell growth at the secondary site. *J Urol* **169:** 1122-33.

Kauraniemi P, Barlund M, Monni O, Kallioniemi A (2001). New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res* **61:** 8235-40.

Kim M, Gans JD, Nogueira C, Wang A, Paik JH, Feng B *et al* (2006). Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene. *Cell* **125:** 1269–1281.

Kim H, Chan R, Dankort DL, Zuo D, Najoukas M, Park M *et al* (2005). The c-Src tyrosine kinase associates with the catalytic domain of ErbB-2: implications for ErbB-2 mediated signaling and transformation. *Oncogene* **24:** 7599-607.

Kim NW (1997). Clinical implications of telomerase in cancer. *Eur J Cancer* **33:** 781-6.

Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV *et al* (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315:** 525-8.

Knaup KX, Roemer K (2004). Cell type-specific regulation of calmodulin 2 expression by mutant p53. *FEBS Lett* **569:** 70-4.

Knudson AG, Jr. (1978). Retinoblastoma: a prototypic hereditary neoplasm. *Semin Oncol* **5:** 57-60.

Komoike Y, Akiyama F, Iino Y, Ikeda T, Akashi-Tanaka S, Ohsumi S *et al* (2006). Ipsilateral breast tumor recurrence (IBTR) after breast-conserving treatment for early breast cancer: risk factors and impact on distant metastases. *Cancer* **106:** 35-41.

Korsching E, Packeisen J, Agelopoulos K, Eisenacher M, Voss R, Isola J *et al* (2002). Cytogenetic alterations and cytokeratin expression patterns in breast cancer: integrating a new model of breast differentiation into cytogenetic pathways of breast carcinogenesis. *Lab Invest* **82:** 1525-33.

Kraus WL, Montano MM, Katzenellenbogen BS (1993). Cloning of the rat progesterone receptor gene 5'-region and identification of two functionally distinct promoters. *Mol Endocrinol* **7:** 1603-16.

Kremer NE, D'Arcangelo G, Thomas SM, DeMarco M, Brugge JS, Halegoua S (1991). Signal transduction by nerve growth factor and fibroblast growth factor in PC12 cells requires a sequence of src and ras actions. *J Cell Biol* **115:** 809-19.

Kripke ML, Fidler IJ (1980). Enhanced experimental metastasis of ultraviolet light-induced fibrosarcomas in ultraviolet light-irradiated syngeneic mice. *Cancer Res* **40:** 625-9.

Kroemer G (1998). The mitochondrion as an integrator/coordinator of cell death pathways. *Cell Death Differ* **5:** 547.

Krugmann S, Williams R, Stephens L, Hawkins PT (2004). ARAP3 is a PI3K- and rap-regulated GAP for RhoA. *Curr Biol* **14:** 1380-4.

Krusekopf S, Chauchereau A, Milgrom E, Henderson D, Cato AC (1991). Co-operation of progestational steroids with epidermal growth factor in activation of gene expression in mammary tumor cells. *J Steroid Biochem Mol Biol* **40:** 239-45.

Kryj M, Maciejewski B, Withers HR, Taylor JM (1997). Incidence and kinetics of distant metastases in patients with operable breast cancer. *Neoplasma* **44:** 3-11.

Kurachi H, Wada Y, Tsukamoto N, Maeda M, Kubota H, Hattori M *et al* (1997). Human SPA-1 gene product selectively expressed in lymphoid tissues is a specific GTPase-activating protein for Rap1 and Rap2. Segregate expression profiles from a rap1GAP gene product. *J Biol Chem* **272:** 28081-8.

Kwong KY, Hung MC (1998). A novel splice variant of HER2 with increased transformation activity. *Mol Carcinog* **23:** 62-8.

Kzhyshkowska J, Marciniak-Czochra A, Gratchev A (2007). Perspectives of mathematical modelling for understanding of intracellular signalling and vesicular trafficking in macrophages. *Immunobiology* **212:** 813-25.

Lacenere CJ, Sternberg PW (2000). Regulation of EGF receptor signaling in the fruitfly D. melanogaster and the nematode C. elegans. *Breast Dis* **11:** 19-30.

Lakhani SR (1999). The pathology of familial breast cancer: Morphological aspects. *Breast Cancer Res* **1:** 31-5.

Lakhani SR, Gusterson BA, Jacquemier J, Sloane JP, Anderson TJ, van de Vijver MJ *et al* (2000). The pathology of familial breast cancer: histological features of cancers in families not attributable to mutations in BRCA1 or BRCA2. *Clin Cancer Res* **6:** 782-9.

Lakhani SR, Jacquemier J, Sloane JP, Gusterson BA, Anderson TJ, van de Vijver MJ *et al* (1998). Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *J Natl Cancer Inst* **90:** 1138-45.

Lakhani SR, O'Hare MJ (2001). The mammary myoepithelial cell--Cinderella or ugly sister? *Breast Cancer Res* **3:** 1-4.

Lakhani SR, Reis-Filho JS, Fulford L, Penault-Llorca F, van der Vijver M, Parry S *et al* (2005). Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. *Clin Cancer Res* **11:** 5175-80.

LaMarca HL, Dash PR, Vishnuthevan K, Harvey E, Sullivan DE, Morris CA *et al* (2008). Epidermal growth factor-stimulated extravillous cytotrophoblast motility is mediated by the activation of PI3-K, Akt and both p38 and p42/44 mitogen-activated protein kinases. *Hum Reprod* **23:** 1733-41.

Lamerato L, Havstad S, Gandhi S, Jones D, Nathanson D (2006). Economic burden associated with breast cancer recurrence: findings from a retrospective analysis of health system data. *Cancer* **106:** 1875-82.

Lancaster M, Rouse J, Hunter KW (2005). Modifiers of mammary tumor progression and metastasis on mouse chromosomes 7, 9, and 17. *Mamm Genome* **16:** 120-6.

Lange CA (2008). Challenges to defining a role for progesterone in breast cancer. *Steroids* **73:** 914-21.

Lauffenburger DA, Horwitz AF (1996). Cell migration: a physically integrated molecular process. *Cell* **84:** 359–369.

Laughner E, Taghavi P, Chiles K, Mahon PC, Semenza GL (2001). HER2 (neu) signaling increases the rate of hypoxia-inducible factor 1alpha (HIF-1alpha) synthesis: novel mechanism for HIF-1-mediated vascular endothelial growth factor expression. *Mol Cell Biol* **21:** 3995-4004.

Le Beau JM, Tedeschi B, Walter G (1991). Increased expression of pp60c-src protein-tyrosine kinase during peripheral nerve regeneration. *J Neurosci Res* **28:** 299-309.

Le MG, Arriagada R, Spielmann M, Guinebretiere JM, Rochard F (2002). Prognostic factors for death after an isolated local recurrence in patients with early-stage breast carcinoma. *Cancer* **94:** 2813-20.

Lee JM, Bernstein A (1993). p53 mutations increase resistance to ionizing radiation. *Proc Natl Acad Sci U S A* **90:** 5742-6.

Lee YJ, Nah HY, Hong SH, Lee JW, Jeon I, Pak JH *et al* (2008). Spa-1 regulates the maintenance and differentiation of human embryonic stem cells. *Int J Dev Biol* **52:** 43-53.

Lemasters JJ, Qian T, Elmore SP, Trost LC, Nishimura Y, Herman B *et al* (1998). Confocal microscopy of the mitochondrial permeability transition in necrotic cell killing, apoptosis and autophagy. *Biofactors* **8:** 283-5.

Levin ER (2005). Integration of the extranuclear and nuclear actions of estrogen. *Mol Endocrinol* **19:** 1951-9.

Li CI, Anderson BO, Porter P, Holt SK, Daling JR, Moe RE (2000). Changing incidence rate of invasive lobular breast carcinoma among older women. *Cancer* **88:** 2561-9.

Li YM, Pan Y, Wei Y, Cheng X, Zhou BP, Tan M *et al* (2004). Upregulation of CXCR4 is essential for HER2-mediated tumor metastasis. *Cancer Cell* **6:** 459-69.

Liang J, Zubovitz J, Petrocelli T, Kotchetkov R, Connor MK, Han K *et al* (2002). PKB/Akt phosphorylates p27, impairs nuclear import of p27 and opposes p27-mediated G1 arrest. *Nat Med* **8:** 1153-60.

Lin SL, Kim H, Ying SY (2008). Intron-mediated RNA interference and microRNA (miRNA). *Front Biosci* **13:** 2216-30.

Liotta LA, Kohn E (2004). Anoikis: cancer and the homeless cell. *Nature* **430:** 973–974.

Liu Y, Wang X, Zhang J, Huang H, Ding B, Wu J *et al* (2008). Structural basis and binding properties of the second bromodomain of Brd4 with acetylated histone tails. *Biochemistry* **47:** 6403-17.

Loden M, Perris F, Nielsen NH, Emdin SO, Landberg G (2003). C-erbB2, p27 and G1/S aberrations in human primary breast cancer. *Anticancer Res* **23:** 2053-61.

Low YL, Wedren S, Liu J (2006). High-throughput genomic technology in research and clinical management of breast cancer. Evolving landscape of genetic epidemiological studies. *Breast Cancer Res* **8:** 209.

Lowe SW (1999). Activation of p53 by oncogenes. *Endocr Relat Cancer* **6:** 45-8.

Luo J, Manning BD, Cantley LC (2003). Targeting the PI3K-Akt pathway in human cancer: rationale and promise. *Cancer Cell* **4:** 257-62.

Lydon JP, DeMayo FJ, Conneely OM, O'Malley BW (1996). Reproductive phenotpes of the progesterone receptor null mutant mouse. *J Steroid Biochem Mol Biol* **56:** 67-77.

Maitra A, Wistuba, II, Washington C, Virmani AK, Ashfaq R, Milchgrub S *et al* (2001). High-resolution chromosome 3p allelotyping of breast carcinomas and precursor lesions demonstrates frequent loss of heterozygosity and a discontinuous pattern of allele loss. *Am J Pathol* **159:** 119-30.

Manie E, Vincent-Salomon A, Lehmann-Che J, Pierron G, Turpin E, Warcoin M *et al* (2009). High frequency of TP53 mutation in BRCA1 and sporadic basal-like carcinomas but not in BRCA1 luminal breast tumors. *Cancer Res* **69:** 663-71.

Mansi JL, Gogas H, Bliss JM, Gazet JC, Berger U, Coombes RC (1999). Outcome of primary-breast-cancer patients with micrometastases: a long-term follow-up study. *Lancet* **354:** 197-202.

Marr N, Bichet DG, Hoefs S, Savelkoul PJ, Konings IB, De Mattia F *et al* (2002). Cell-biologic and functional analyses of five new Aquaporin-2 missense mutations that cause recessive nephrogenic diabetes insipidus. *J Am Soc Nephrol* **13:** 2267-77.

McDonnell DP, Clemm DL, Hermann T, Goldman ME, Pike JW (1995). Analysis of estrogen receptor function in vitro reveals three distinct classes of antiestrogens. *Mol Endocrinol* **9:** 659-69.

McLean GW, Carragher NO, Avizienyte E, Evans J, Brunton VG, Frame MC (2005). The role of focal-adhesion kinase in cancer—a new therapeutic opportunity. *Nat. Rev. Cancer* **5:** 505–515.

Meredith JE, Fazeli B, Schwartz MA (1993). The extracellular matrix as a cell survival factor. *Mol. Biol. Cell* **4:** 953–961.

Minato N and Hattori M (2009). Spa-1 (Sipa1) and Rap signaling in leukemia and cancer metastasis. *Cancer Sci* **100:** 17-23.

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD *et al* (2005). Genes that mediate breast cancer metastasis to lung. *Nature* **436:** 518-24.

Mitsumoto M, Kamura T, Kobayashi H, Sonoda T, Kaku T, Nakano H (1998) Emergence of higher levels of invasive and metastatic properties in the drug-resistant cancer cell lines after the repeated administration of cisplatin in tumor-bearing mice. *J Cancer Res Clin Oncol* **124:** 607–614

Moghal N, Sternberg PW (2003). The epidermal growth factor system in Caenorhabditis elegans. *Exp Cell Res* **284:** 150-9.

Montgomery KG, Gertig DM, Baxter SW, Milne RL, Dite GS, McCredie MR *et al* (2003). The HER2 I655V polymorphism and risk of breast cancer in women < age 40 years. *Cancer Epidemiol Biomarkers Prev* **12:** 1109-11.

Moore MR, Zhou JL, Blankenship KA, Strobl JS, Edwards DP, Gentry RN (1997). A sequence in the 5' flanking region confers progestin responsiveness on the human c-myc gene. *J Steroid Biochem Mol Biol* **62:** 243-52.

Moran MS, Haffty BG (2002). Local-regional breast cancer recurrence: prognostic groups based on patterns of failure. *Breast J* **8:** 81-7.

Morin PJ (2003) Drug resistance and the microenvironment: nature and nurture. *Drug Resist Updat* **4:** 169–172

Morrison LE, Jewell SS, Usha L, Blondin BA, Rao RD, Tabesh B *et al* (2007). Effects of ERBB2 amplicon size and genomic alterations of chromosomes 1, 3, and 10 on patient response to trastuzumab in metastatic breast cancer. *Genes Chromosomes Cancer* **46:** 397-405.

Moyer AM, Salavaggione OE, Wu TY, Moon I, Eckloff BW, Hildebrandt MA *et al* (2008). Glutathione s-transferase p1: gene sequence variation and functional genomic studies. *Cancer Res* **68:** 4791-801.

Mulac-Jericevic B, Conneely OM (2004). Reproductive tissue selective actions of progesterone receptors. *Reproduction* **128:** 139-46.

Muller A, Homey B, Soto H, Ge N, Catron D, Buchanan ME *et al* (2001). Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410:** 50-56.

Murdoch C, Muthana M, Coggelt SB, Lewis CE (2008). The role of myeloid cells in the promotion of tummour angiogenesis. *Nat. Rev. Cancer* **8:** 618-631.

Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG *et al* (2007). A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene* **26:** 1507-16.

Nakayama T, Taback B, Turner R, Morton DL, Hoon DS (2001). Molecular clonality of in-transit melanoma metastasis. *Am J Pathol* **158:** 1371-8.

Nicolini A, Giardino R, Carpi A, Ferrari P, Anselmi L, Colosimo S *et al* (2006). Metastatic breast cancer: an updating. *Biomed Pharmacother* **60:** 548-56.

Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z *et al* (2004). Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* **10:** 5367-74.

Nilsson S, Makela S, Treuter E, Tujague M, Thomsen J, Andersson G *et al* (2001). Mechanisms of estrogen action. *Physiol Rev* **81:** 1535-65.

Noda Y, Sasaki S (2008). The role of actin remodeling in the trafficking of intracellular vesicles, transporters, and channels: focusing on aquaporin-2. *Pflugers Arch* **456:** 737-45.

Normanno N, Di Maio M, De Maio E, De Luca A, de Matteis A, Giordano A *et al* (2005). Mechanisms of endocrine resistance and novel therapeutic strategies in breast cancer. *Endocr Relat Cancer* **12:** 721-47.

Northouse LL, Mood D, Kershaw T, Schafenacker A, Mellon S, Walker J *et al* (2002). Quality of life of women with recurrent breast cancer and their family members. *J Clin Oncol* **20:** 4050-64.

Nowell PC (1976). The clonal evolution of tumor cell populations. *Science* **194:** 23-8.

Oh S, Heflin L, Meyerowitz BE, Desmond KA, Rowland JH, Ganz PA (2004). Quality of life of breast cancer survivors after a recurrence: a follow-up study. *Breast Cancer Res Treat* **87:** 45-57.

Onder TT, Gupta PB, Mani SA, Yang J, Lander ES, Weinberg RA (2008). Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Res.* **68:** 3645–3654.

Osborne CK, Schiff R (2005). Estrogen-receptor biology: continuing progress and therapeutic implications. *J Clin Oncol* **23:** 1616-22.

Overall CM, Kleifeld O (2006). Tumour microenvironment-opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* **6:** 227-239.

Paez J, Sellers WR (2003). PI3K/PTEN/AKT pathway. A critical mediator of oncogenic signaling. *Cancer Treat Res* **115:** 145-67.

Pagani F, Baralle FE (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* **5:** 389-96.

Page MJ, Amess B, Townsend RR, Parekh R, Herath A, Brusten L *et al* (1999). Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties. *Proc Natl Acad Sci U S A* **96:** 12589-94.

Page S (1889). The distribution of secondary growths in cancer of the breast. *Lancet* **1:** 571-573.

Palacios J, Honrado E, Osorio A, Cazorla A, Sarrio D, Barroso A *et al* (2005). Phenotypic characterization of BRCA1 and BRCA2 tumors based in a tissue microarray study with 37 immunohistochemical markers. *Breast Cancer Res Treat* **90:** 5-14.

Palacios J, Honrado E, Osorio A, Cazorla A, Sarrio D, Barroso A *et al* (2003). Immunohistochemical characteristics defined by tissue microarray of hereditary breast cancer not attributable to BRCA1 or BRCA2 mutations: differences from breast carcinomas arising in BRCA1 and BRCA2 mutation carriers. *Clin Cancer Res* **9:** 3606-14.

Pallis L, Wilking N, Cedermark B, Rutqvist LE, Skoog L (1992). Receptors for estrogen and progesterone in breast carcinoma in situ. *Anticancer Res* **12:** 2113-5.

Parada LF, Tabin CJ, Shih C, Weinberg RA (1982). Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene. *Nature* **297:** 474-8.

Park YG, Clifford R, Buetow KH, Hunter KW (2003). Multiple cross and inbred strain haplotype mapping of complex-trait candidate genes. *Genome Res* **13:** 118-21.

Park YG, Zhao X, Lesueur F, Lowy DR, Lancaster M, Pharoah P *et al* (2005). Sipa1 is a candidate for underlying the metastasis efficiency modifier locus Mtes1. *Nat Genet* **37:** 1055-62.

Parker B, Sukumar S (2003). Distant metastasis in breast cancer: molecular mechanisms and therapeutic targets. *Cancer Biol Ther* **2:** 14-21.

Paulin FE, O'Neill M, McGregor G, Cassidy A, Ashfield A, Ali CW *et al* (2008). MDM2 SNP309 is associated with high grade node positive breast tumours and is in linkage disequilibrium with a novel MDM2 intron 1 polymorphism. *BMC Cancer* **8:** 281.

Payne SJ, Bowen RL, Jones JL, Wells CA (2008). Predictive markers in breast cancer--the present. *Histopathology* **52:** 82-90.

Pedraza-Farina LG (2006). Mechanisms of oncogenic cooperation in cancer initiation and metastasis. *Yale J Biol Med* **79:** 95-103.

Pelengaris S, Khan M, Evan GI (2002). Suppression of Myc-induced apoptosis in beta cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression. *Cell* **109:** 321-34.

Pellikainen JM, Ropponen KM, Kataja VV, Kellokoski JK, Eskelinen MJ, Kosma VM (2004). Expression of matrix metalloproteinase (MMP)-2 and MMP-9 in breast cancer with a special reference to activator protein-2, HER2, and prognosis. *Clin Cancer Res* **10:** 7621-8.

Pereira H, Pinder SE, Sibbering DM, Galea MH, Elston CW, Blamey RW *et al* (1995). Pathological prognostic factors in breast cancer. IV: Should you be a typer or a grader? A comparative study of two histological prognostic features in operable breast carcinoma. *Histopathology* **27:** 219-26.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA *et al* (2000). Molecular portraits of human breast tumours. *Nature* **406:** 747-52.

Peters LL, Robledo RF, Bult CJ, Churchill GA, Paigen BJ, Svenson KL (2007). The mouse as a model for human biology: a resource guide for complex trait analysis. *Nat Rev Genet* **8:** 58-69.

Pietras RJ (2006). Biologic basis of sequential and combination therapies for hormone-responsive breast cancer. *Oncologist* **11:** 704-17.

Politz O, Gratchev A, McCourt PA, Schledzewski K, Guillot P, Johansson S *et al* (2002). Stabilin-1 and -2 constitute a novel family of fasciclin-like hyaluronan receptor homologues. *Biochem J* **362:** 155-64.

Poller DN, Snead DR, Roberts EC, Galea M, Bell JA, Gilmour A *et al* (1993). Oestrogen receptor expression in ductal carcinoma in situ of the breast: relationship to flow cytometric analysis of DNA and expression of the c-erbB-2 oncoprotein. *Br J Cancer* **68:** 156-61.

Polyak K (2007). Breast cancer: origins and evolution. *J Clin Invest* **117:** 3155-63.

Ponder BA (2001). Cancer genetics. *Nature* **411:** 336-41.

Popescu NC, Zimonjic DB (2002). Chromosome and gene alterations in breast cancer as markers for diagnosis and prognosis as well as pathogenetic targets for therapy. *Am J Med Genet* **115:** 142-9.

Poste G, Fidler IJ (1980). The pathogenesis of cancer metastasis. *Nature* **283:** 139-46.

Prasad NK (2009). SHIP2 phosphoinositol phosphatase positively regulates EGFR-Akt pathway, CXCR4 expression, and cell migration in MDA-MB-231 breast cancer cells. *Int J Oncol* **34:** 97-105.

Pratt WB, Toft DO (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp Biol Med (Maywood)* **228:** 111-33.

Prenzel N, Fischer OM, Streit S, Hart S, Ullrich A (2001). The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. *Endocr Relat Cancer* **8:** 11-31.

Preston TJ, Abadi A, Wilson L, Singh G (2001). Mitochondrial contributions to cancer cell physiology: potential for drug development. *Adv Drug Deliv Rev* **49:** 45-61.

Prigent SA, Gullick WJ (1994). Identification of c-erbB-3 binding sites for phosphatidylinositol 3'-kinase and SHC using an EGF receptor/c-erbB-3 chimera. *EMBO J* **13:** 2831-41.

Pritchard JK, Przeworski M (2001). Linkage disequilibrium in humans: models and data. *Am J Hum Genet* **69:** 1-14.

Qu S, Long J, Cai Q, Shu XO, Cai H, Gao YT *et al* (2008). Genetic polymorphisms of metastasis suppressor gene NME1 and breast cancer survival. *Clin Cancer Res* **14:** 4787-93.

Raaijmakers JH, Deneubourg L, Rehmann H, de Koning J, Zhang Z, Krugmann S *et al* (2007). The PI3K effector Arap3 interacts with the PI(3,4,5)P3 phosphatase SHIP2 in a SAM domain-dependent manner. *Cell Signal* **19:** 1249-57.

Raaphorst FM, Meijer CJ, Fieret E, Blokzijl T, Mommers E, Buerger H *et al* (2003). Poorly differentiated breast carcinoma is associated with increased expression of the human polycomb group EZH2 gene. *Neoplasia* **5:** 481-8.

Rabbani SA, Mazar AP (2007). Evaluating distant metastases in breast cancer: from biology to outcomes. *Cancer Metastasis Rev* **26:** 663-74.

Ram TG, Ethier SP (1996). Phosphatidylinositol 3-kinase recruitment by p185erbB-2 and erbB-3 is potently induced by neu differentiation factor/heregulin during mitogenesis and is constitutively elevated in growth factor-independent breast carcinoma cells with c-erbB-2 gene amplification. *Cell Growth Differ* **7:** 551-61.

Ramaswamy S, Ross KN, Lander ES, Golub TR (2003). A molecular signature of metastasis in primary solid tumors. *Nat Genet* **33:** 49-54.

Ramsauer VP, Carraway CA, Salas PJ, Carraway KL (2003). Muc4/sialomucin complex, the intramembrane ErbB2 ligand, translocates ErbB2 to the apical surface in polarized epithelial cells. *J Biol Chem* **278:** 30142-7.

Ray PS, Wang J, Qu Y, Sim MS, Shamonki J, Bagaria SP *et al* (2010). FOXC1 is a potential prognostic biomarker with functional significance in basal-like breast cancer. *Cancer Res.* **70:** 3870-6.

Reed JC, Green DR (2002). Remodeling for demolition: changes in mitochondrial ultrastructure during apoptosis. *Mol Cell* **9:** 1-3.

Reed SI (2003). Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover. *Nat Rev Mol Cell Biol* **4:** 855-64.

Reeves JR, Xuan JW, Arfanis K, Morin C, Garde SV, Ruiz MT *et al* (2005). Identification, purification and characterization of a novel human blood protein with binding affinity for prostate secretory protein of 94 amino acids. *Biochem J* **385:** 105-14.

Reeves KW, Stone RA, Modugno F, Ness RB, Vogel VG, Weissfeld JL *et al* (2009). Longitudinal association of anthropometry with mammographic breast density in the Study of Women's Health Across the Nation. *Int J Cancer* **124:** 1169-77.

Reiners J, Nagel-Wolfrum K, Jurgens K, Marker T, Wolfrum U (2006). Molecular basis of human Usher syndrome: deciphering the meshes of the Usher protein network provides insights into the pathomechanisms of the Usher disease. *Exp Eye Res* **83:** 97-119.

Reis-Filho JS, Tutt AN (2008). Triple negative tumours: a critical review. *Histopathology* **52:** 108-18.

Richardsen E, Uglehus RD, Due J, Busch C, Busund LT (2008). The prognostic impact of M-CSF, CSF-1 receptor, CD68 and CD3 in prostatic carcinoma. *Histopathology* **53:** 30-8.

Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A *et al* (2006). X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* **9:** 121-32.

Richer JK, Jacobsen BM, Manning NG, Abel MG, Wolf DM, Horwitz KB (2002). Differential gene regulation by the two progesterone receptor isoforms in human breast cancer cells. *J Biol Chem* **277:** 5209-18.

Riethmuller G, Klein CA (2001). Early cancer cell dissemination and late metastatic relapse: clinical reflections and biological approaches to the dormancy problem in patients. *Semin Cancer Biol* **11:** 307-11.

Riethmuller G, Klein CA (2001). Early cancer cell dissemination and late metastatic relapse: clinical reflections and biological approaches to the dormancy problem in patients. *Semin Cancer Biol* **11:** 307-11.

Robertson JF (1996). Oestrogen receptor: a stable phenotype in breast cancer. *Br J Cancer* **73:** 5-12.

Robinson GW, Hennighausen L, Johnson PF (2000). Side-branching in the mammary gland: the progesterone-Wnt connection. *Genes Dev* **14:** 889-94.

Robson ME, Chappuis PO, Satagopan J, Wong N, Boyd J, Goffin JR *et al* (2004). A combined analysis of outcome following breast cancer: differences in survival based on BRCA1/BRCA2 mutation status and administration of adjuvant treatment. *Breast Cancer Res* **6:** R8-R17.

Roh H, Pippin J, Drebin JA (2000). Down-regulation of HER2/neu expression induces apoptosis in human cancer cells that overexpress HER2/neu. *Cancer Res* **60:** 560-5.

Rosen PP (1991). Pathological assessment of nonpalpable breast lesions. *Semin Surg Oncol* **7:** 257-60.

Ruan W, Monaco ME, Kleinberg DL (2005). Progesterone stimulates mammary gland ductal morphogenesis by synergizing with and enhancing insulin-like growth factor-I action. *Endocrinology* **146:** 1170-8.

Rudolph B, Hueber AO, Evan GI (2000). Reversible activation of c-Myc in thymocytes enhances positive selection and induces proliferation and apoptosis in vitro. *Oncogene* **19:** 1891-900.

Rugo HS (2008). The importance of distant metastases in hormone-sensitive breast cancer. *Breast* **17 Suppl 1:** S3-8.

Ruoslahti E (1996). How cancer spreads. *Sci Am* **275:** 72-7.

Ruoslahti E, Rajotte D (2000). An address system in the vascularture of normal tissues and tumors. *Annu. Rev. Immunol.* **18**: 813-827.

Rust R, Visser L, van der Leij J, Harms G, Blokzijl T, Deloulme JC *et al* (2005). High expression of calcium-binding proteins, S100A10, S100A11 and CALM2 in anaplastic large cell lymphoma. *Br J Haematol* **131:** 596-608.

Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She QB, Maurer M *et al* (2007). Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci U S A* **104:** 7564-9.

Saikumar P, Dong Z, Mikhailov V, Denton M, Weinberg JM, Venkatachalam MA (1999). Apoptosis: definition, mechanisms, and relevance to disease. *Am J Med* **107:** 489-506.

Sanz-Moreno V, Gadea G, Ahn J, Paterson H, Marra P, S. Pinner S *et al* (2008). Rac activation and inactivation control plasticity of tumor cell movement. *Cell* **135:** 510–523.

Sasaki M, Ikeda H, Itatsu K, Yamaguchi J, Sawada S, Minato H *et al* (2008). The overexpression of polycomb group proteins Bmi1 and EZH2 is associated with the progression and aggressive biological behavior of hepatocellular carcinoma. *Lab Invest* **88:** 873-82.

Sassone-Corsi P, Der CJ, Verma IM (1989). ras-induced neuronal differentiation of PC12 cells: possible involvement of fos and jun. *Mol Cell Biol* **9:** 3174-83.

Savage K, Leung S, Todd SK, Brown LA, Jones RL, Robertson D *et al* (2008). Distribution and significance of caveolin 2 expression in normal breast and invasive breast cancer: an immunofluorescence and immunohistochemical analysis. *Breast Cancer Res Treat* **110:** 245-56.

Savage SA, Chanock SJ (2004). Using germ-line genetic variation to investigate and treat cancer. *Drug Discov Today* **9:** 610-8.

Savanis G, Simatos G, Tzaida O, Tsikkinis C, Ammari S, Mylonas A *et al* (2006). Gastrointestinal tract metastasis as first presentation of breast cancer. *J BUON* **11:** 79-81.

Schindlbeck C, Jeschke U, Schulze S, Karsten U, Janni W, Rack B *et al* (2007). Prognostic impact of Thomsen-Friedenreich tumor antigen and disseminated tumor cells in the bone marrow of breast cancer patients. *Breast Cancer Res Treat* **101:** 17-25.

Schneider S, Weydig C, Wessler S (2008). Targeting focal adhesions:Helicobacter pylori-host communication in cell migration. *Cell Commun Signal* **6:** 2.

Schulze WX, Deng L, Mann M (2005). Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol Syst Biol* **1:** 2005 0008.

Sergeev IN and Rhoten WB (1998). Regulation of intracellular calcium in human breast cancer cells. *Endocrine* **9:** 321-327.

Seton-Rogers SE, Lu Y, Hines LM, Koundinya M, LaBaer J, Muthuswamy SK *et al* (2004). Cooperation of the ErbB2 receptor and transforming growth factor beta in induction of migration and invasion in mammary epithelial cells. *Proc Natl Acad Sci U S A* **101:** 1257-62.

Shachaf CM, Kopelman AM, Arvanitis C, Karlsson A, Beer S, Mandl S *et al* (2004). MYC inactivation uncovers pluripotent differentiation and tumour dormancy in hepatocellular cancer. *Nature* **431:** 1112-7.

Sham PC, Curtis D (1995). Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* **59:** 97-105.

Shay JW, Gazdar AF (1997). Telomerase in the early detection of cancer. *J Clin Pathol* **50:** 106-9.

Shelly W, Draper MW, Krishnan V, Wong M, Jaffe RB (2008). Selective estrogen receptor modulators: an update on recent clinical findings. *Obstet Gynecol Surv* **63:** 163-81.

Shepherd TG, Kockeritz L, Szrajber MR, Muller WJ, Hassell JA (2001). The pea3 subfamily ets genes are required for HER2/Neu-mediated mammary oncogenesis. *Curr Biol* **11:** 1739-48.

Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA *et al* (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95:** 927-37.

Shim JY, An HJ, Lee YH, Kim SK, Lee KP, Lee KS (2003). Overexpression of cyclooxygenase-2 is associated with breast carcinoma and its poor prognostic factors. *Mod Pathol* **16:** 1199-204.

Shim V, Gauthier ML, Sudilovsky D, Mantei K, Chew KL, Moore DH *et al* (2003). Cyclooxygenase-2 expression is related to nuclear grade in ductal carcinoma in situ and is increased in its normal adjacent epithelium. *Cancer Res* **63:** 2347-50.

Shin I, Yakes FM, Rojo F, Shin NY, Bakin AV, Baselga J *et al* (2002). PKB/Akt mediates cell-cycle progression by phosphorylation of p27(Kip1) at threonine 157 and modulation of its cellular localization. *Nat Med* **8:** 1145-52.

Shirasaki Y, Kanazawa Y, Morishima Y, Makino M (2006). Involvement of calmodulin in neuronal cell death. *Brain Res* **1083:** 189-95.

Shyamala G, Yang X, Silberstein G, Barcellos-Hoff MH, Dale E (1998). Transgenic mice carrying an imbalance in the native ratio of A to B forms of progesterone receptor exhibit developmental abnormalities in mammary glands. *Proc Natl Acad Sci U S A* **95:** 696-701.

Siegel PM, Ryan ED, Cardiff RD, Muller WJ (1999). Elevated expression of activated forms of Neu/ErbB-2 and ErbB-3 are involved in the induction of mammary tumors in transgenic mice: implications for human breast cancer. *EMBO J* **18:** 2149-64.

Sierke SL, Cheng K, Kim HH, Koland JG (1997). Biochemical characterization of the protein tyrosine kinase homology domain of the ErbB3 (HER3) receptor protein. *Biochem J* **322 (Pt 3):** 757-63.

Simoncini T, Hafezi-Moghadam A, Brazil DP, Ley K, Chin WW, Liao JK (2000). Interaction of oestrogen receptor with the regulatory subunit of phosphatidylinositol-3-OH kinase. *Nature* **407:** 538-41.

Singletary SE, Patel-Parekh L, Bland KI (2005). Treatment trends in early-stage invasive lobular carcinoma: a report from the National Cancer Data Base. *Ann Surg* **242:** 281-9.

Skulachev VP (1999). Mitochondrial physiology and pathology; concepts of programmed death of organelles, cells and organisms. *Mol Aspects Med* **20:** 139-84.

Sliwkowski MX (2003). Ready to partner. *Nat Struct Biol* **10:** 158-9.

Smeds J, Warnberg F, Norberg T, Nordgren H, Holmberg L, Bergh J (2005). Ductal carcinoma in situ of the breast with different histopathological grades and corresponding new breast tumour events: analysis of loss of heterozygosity. *Acta Oncol* **44:** 41-9.

Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG *et al* (2008). Subtypes of breast cancer show preferential site of relapse. *Cancer Res* **68:** 3108-14.

Smith CL, Nawaz Z, O'Malley BW (1997). Coactivator and corepressor regulation of the agonist/antagonist activity of the mixed antiestrogen, 4-hydroxytamoxifen. *Mol Endocrinol* **11:** 657-66.

Smith RA, Curran JE, Weinstein SR, Griffiths LR (2001). Investigation of glutathione S-transferase zeta and the development of sporadic breast cancer. *Breast Cancer Res* **3:** 409-11.

Solomayer EF, Diel IJ, Meyberg GC, Gollan C, Bastert G (2000). Metastatic breast cancer: clinical course, prognosis and therapy related to the first site of metastasis. *Breast Cancer Res Treat* **59:** 271-8.

Soltoff SP, Carraway KL, 3rd, Prigent SA, Gullick WG, Cantley LC (1994). ErbB3 is involved in activation of phosphatidylinositol 3-kinase by epidermal growth factor. *Mol Cell Biol* **14:** 3550-8.

Sommer A, Hoffmann J, Lichtner RB, Schneider MR, Parczyk K (2003). Studies on the development of resistance to the pure antiestrogen Faslodex in three human breast cancer cell lines. *J Steroid Biochem Mol Biol* **85:** 33-47.

Sonnenfeld MR, Frenna TH, Weidner N, Meyer JE (1991). Lobular carcinoma in situ: mammographic-pathologic correlation of results of needle-directed biopsy. *Radiology* **181:** 363-7.

Sonnenschein C, Soto AM (2008). Theories of carcinogenesis: an emerging perspective. *Semin Cancer Biol* **18:** 372-7.

Sorensen JB, Rossel P, Holm S (2004). Patient-physician communication concerning participation in cancer chemotherapy trials. *Br J Cancer* **90:** 328-32.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98:** 10869-74.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A *et al* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100:** 8418-23.

Sorlie T, Wang Y, Xiao C, Johnsen H, Naume B, Samaha RR *et al* (2006). Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics* **7:** 127.

Spangenberg C, Lausch EU, Trost TM, Prawitt D, May A, Keppler R *et al* (2006). ERBB2-mediated transcriptional up-regulation of the alpha5beta1 integrin fibronectin receptor promotes tumor cell survival under adverse conditions. *Cancer Res* **66:** 3715-25.

Spataro V, Price K, Goldhirsch A, Cavalli F, Simoncini E, Castiglione M *et al* (1992). Sequential estrogen receptor determinations from primary breast cancer and at relapse: prognostic and therapeutic relevance. The International Breast Cancer Study Group (formerly Ludwig Group). *Ann Oncol* **3:** 733-40.

Speirs V, Boyle-Walsh E, Fraser WD (1997). Constitutive co-expression of estrogen and progesterone receptor mRNA in human meningiomas by RT-PCR and response of in vitro cell cultures to steroid hormones. *Int J Cancer* **72:** 714-9.

Speirs V, Malone C, Walton DS, Kerin MJ, Atkin SL (1999). Increased expression of estrogen receptor beta mRNA in tamoxifen-resistant breast cancer patients. *Cancer Res* **59:** 5421-4.

Speirs V, Walton DS, Hall MC, Atkin SL (1999). In vivo and in vitro expression of steroid-converting enzymes in human breast tumours: associations with interleukin-6. *Br J Cancer* **81:** 690-5.

Steeg PS, Ouatas T, Halverson D, Palmieri D, Salerno M (2003). Metastasis suppressor genes: basic biology and potential clinical use. *Clin Breast Cancer* **4:** 51-62.

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE *et al* (2001). Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293:** 489-93.

Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE *et al* (2003). Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. *Hum Hered* **55:** 27-36.

Subbaramaiah K, Norton L, Gerald W, Dannenberg AJ (2002). Cyclooxygenase-2 is overexpressed in HER-2/neu-positive breast cancer: evidence for involvement of AP-1 and PEA3. *J Biol Chem* **277:** 18649-57.

Susin SA, Zamzami N, Kroemer G (1998). Mitochondria as regulators of apoptosis: doubt no more. *Biochim Biophys Acta* **1366:** 151-65.

Swede H, Moysich KB, Freudenheim JL, Quirk JT, Muti PC, Hurd TC *et al* (2001). Breast cancer risk factors and HER2 over-expression in tumors. *Cancer Detect Prev* **25:** 511-9.

Taillon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen T, Cao A *et al* (2000). Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat Genet* **25:** 324-8.

Takeichi M (1991). Cadherin cell adhesion receptors as a morphogenetic regulator. *Science* **251:** 1451-5.

Talmadge JE, Fidler IJ (1982). Cancer metastasis is selective or random depending on the parent tumour population. *Nature* **297:** 593-4.

Tamimi SO, Ahmed A (1986). Stromal changes in early invasive and non-invasive breast carcinoma: an ultrastructural study. *J Pathol.* **150:** 43-9.

Tan M, Li P, Klos KS, Lu J, Lan KH, Nagata Y *et al* (2005). ErbB2 promotes Src synthesis and stability: novel mechanisms of Src activation that confer breast cancer metastasis. *Cancer Res* **65:** 1858-67.

Tavoloni N, Inoue H, Sabe H, Hanafusa H (1994). v-src transformation of rat embryo fibroblasts. Inefficient conversion to anchorage-independent growth involves heterogeneity of primary cultures. *J Cell Biol* **126:** 475-83.

Timms JF, White SL, O'Hare MJ, Waterfield MD (2002). Effects of ErbB-2 overexpression on mitogenic signalling and cell cycle progression in human breast luminal epithelial cells. *Oncogene* **21:** 6573-86.

Tiwari VK, Cope L, McGarvey KM, Ohm JE, Baylin SB (2008). A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res* **18:** 1171-9.

Tokunaga E, Kimura Y, Oki E, Ueda N, Futatsugi M, Mashino K *et al* (2006). Akt is frequently activated in HER2/neu-positive breast cancers and associated with poor prognosis among hormone-treated patients. *Int J Cancer* **118:** 284-9.

Tonini G, Schiavon G, Fratto ME, Vincenzi B, Santini D (2008). Hormono-biological therapy in metastatic breast cancer: preclinical evidence, clinical studies and future directions. *Expert Opin Biol Ther* **8:** 221-34.

Toutenhoofd SL, Strehler EE (2002). Regulation of calmodulin mRNAs in differentiating human IMR-32 neuroblastoma cells. *Biochim Biophys Acta* **1600:** 95-104.

Tsukamoto N, Hattori M, Yang H, Bos JL, Minato N (1999). Rap1 GTPase-activating protein SPA-1 negatively regulates cell adhesion. *J Biol Chem* **274:** 18463-9.

Tufarelli C, Frischauf AM, Hardison R, Flint J, Higgs DR (2001). Characterization of a widely expressed gene (LUC7-LIKE; LUC7L) defining the centromeric boundary of the human alpha-globin domain. *Genomics* **71:** 307-14.

Tzahar E, Waterman H, Chen X, Levkowitz G, Karunagaran D, Lavi S *et al* (1996). A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. *Mol Cell Biol* **16:** 5276-87.

Urquidi V, Goodison S (2007). Genomic signatures of breast cancer metastasis. *Cytogenet Genome Res* **118:** 116-29.

Vachon CM, Brandt KR, Ghosh K, Scott CG, Maloney SD, Carston MJ, *et al* (2007). Mammographic Breast Density as a General Marker of Breast Cancer Risk. *Cancer Epidemiol Biomarkers Prev* **16:** 43-9.

Vadlamudi R, Mandal M, Adam L, Steinbach G, Mendelsohn J, Kumar R (1999). Regulation of cyclooxygenase-2 pathway by HER2 receptor. *Oncogene* **18:** 305-14.

Vadlamudi RK, Sahin AA, Adam L, Wang RA, Kumar R (2003). Heregulin and HER2 signaling selectively activates c-Src phosphorylation at tyrosine 215. *FEBS Lett* **543:** 76-80.

van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J *et al* (2002). Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol* **161:** 1991-6.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW *et al* (2002). A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347:** 1999-2009.

van de Vijver MJ, Peterse JL, Mooi WJ, Wisman P, Lomans J, Dalesio O *et al* (1988). Neu-protein overexpression in breast cancer. Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer. *N Engl J Med* **319:** 1239-45.

van den Broek MF, Hengartner H (2000). The role of perforin in infections and tumour surveillance. *Exp Physiol* **85:** 681-5.

van Gurp M, Festjens N, van Loo G, Saelens X, Vandenabeele P (2003). Mitochondrial intermembrane proteins in cell death. *Biochem Biophys Res Commun* **304:** 487-97.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M *et al* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530-6.

Vartanian T, Goodearl A, Lefebvre S, Park SK, Fischbach G (2000). Neuregulin induces the rapid association of focal adhesion kinase with the erbB2-erbB3 receptor complex in schwann cells. *Biochem Biophys Res Commun* **271:** 414-7.

Vazquez A, Bond EE, Levine AJ, Bond GL (2008). The genetics of the p53 pathway, apoptosis and cancer therapy. *Nat Rev Drug Discov* **7:** 979-87.

Vega S, Morales AV, Ocana OH, Valdes F, Fabregat I, Nieto MA (2004). Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev.* **18:** 1131–1143.

Velicer CM, Heckbert SR, Rutter C, Lampe JW and Malone K (2005). Association between Antibiotic use Prior to Breast Cancer Diagnosis and Breast Tumour Characteristics (United States). *Cancer Causes & Control* **10:** 0445-9.

Veronesi U, Marubini E, Del Vecchio M, Manzari A, Andreola S, Greco M *et al* (1995). Local recurrences and distant metastases after conservative breast cancer treatments: partly independent events. *J Natl Cancer Inst* **87:** 19-27.

Vicini FA, Kestin L, Huang R, Martinez A (2003). Does local recurrence affect the rate of distant metastases and survival in patients with early-stage breast carcinoma treated with breast-conserving therapy? *Cancer* **97:** 910-9.

Viglietto G, Motti ML, Bruni P, Melillo RM, D'Alessio A, Califano D *et al* (2002). Cytoplasmic relocalization and inhibition of the cyclin-dependent kinase inhibitor p27(Kip1) by PKB/Akt-mediated phosphorylation in breast cancer. *Nat Med* **8:** 1136-44.

Vijapurkar U, Kim MS, Koland JG (2003). Roles of mitogen-activated protein kinase and phosphoinositide 3'-kinase in ErbB2/ErbB3 coreceptor-mediated heregulin signaling. *Exp Cell Res* **284:** 291-302.

Vos CB, ter Haar NT, Rosenberg C, Peterse JL, Cleton-Jansen AM, Cornelisse CJ *et al* (1999). Genetic alterations on chromosome 16 and 17 are important features of ductal carcinoma in situ of the breast and are associated with histologic type. *Br J Cancer* **81:** 1410-8.

Wallasch C, Weiss FU, Niederfellner G, Jallal B, Issing W, Ullrich A (1995). Heregulin-dependent regulation of HER2/neu oncogenic signaling by heterodimerization with HER3. *EMBO J* **14:** 4267-75.

Wang D, Johnson AD, Papp AC, Kroetz DL, Sadee W (2005). Multidrug resistance polypeptide 1 (MDR1, ABCB1) variant 3435C>T affects mRNA stability. *Pharmacogenet Genomics* **15:** 693-704.

Wang SC, Lien HC, Xia W, Chen IF, Lo HW, Wang Z *et al* (2004). Binding at and transactivation of the COX-2 promoter by nuclear tyrosine kinase receptor ErbB-2. *Cancer Cell* **6:** 251-61.

Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F *et al* (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365:** 671-9.

Weigelt B, Glas AM, Wessels LF, Witteveen AT, Peterse JL, van't Veer LJ (2003). Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci U S A* **100:** 15901-5.

Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG *et al* (2005). Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res* **65:** 9155-8.

Weigelt B, Peterse JL, van 't Veer LJ (2005). Breast cancer metastasis: markers and models. *Nat Rev Cancer* **5:** 591-602.

Weinberg RA (1991). Tumor suppressor genes. *Science* **254:** 1138-46.

Weiss L, Ward PM (1990). Contributions of vascularized lymph-node metastases to hematogenous metastasis in a rat mammary carcinoma. *Int J Cancer* **46:** 452-5.

Weiss RB, Woolf SH, Demakos E, Holland JF, Berry DA, Falkson G *et al* (2003). Natural history of more than 20 years of node-positive primary breast carcinoma treated with cyclophosphamide, methotrexate, and fluorouracil-based adjuvant chemotherapy: a study by the Cancer and Leukemia Group B. *J Clin Oncol* **21:** 1825-35.

Wellings SR, Jensen HM, Marcum RG (1975). An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. *J Natl Cancer Inst* **55:** 231-73.

Wenger CR, Clark GM (1998). S-phase fraction and breast cancer--a decade of experience. *Breast Cancer Res Treat* **51:** 255-65.

Wilson A, Murphy MJ, Oskarsson T, Kaloulis K, Bettess MD, Oser GM *et al* (2004). c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev* **18:** 2747-63.

Wilson GR, Cramer A, Welman A, Knox F, Swindell R, Kawakatsu H *et al* (2006). Activated c-SRC in ductal carcinoma in situ correlates with high tumour grade, high proliferation and HER2 positivity. *Br J Cancer* **95:** 1410-4.

Woelfle U, Breit E, Zafrakas K, Otte M, Schubert F, Müller V *et al* (2005). Bi-specific immunomagnetic enrichment of micrometastatic tumour cell clusters from bone marrow of cancer patients. *J of Immunological Methods* **300:** 136-145.

Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ *et al* (2007). American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer. *Arch Pathol Lab Med* **131:** 18.

Wu SY, Chiang CM (2007). The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J Biol Chem* **282:** 13141-5.

Wyckoff JB, Wang Y, Lin EY, Li JF, Goswami S, Stanley ER *et al* (2007). Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors. *Cancer Res* **67:** 2649-56.

Xie D, Shu XO, Deng Z, Wen WQ, Creek KE, Dai Q *et al* (2000). Population-based, case-control study of HER2 genetic polymorphism and breast cancer risk. *J Natl Cancer Inst* **92:** 412-7.

Xu W, Yuan X, Beebe K, Xiang Z, Neckers L (2007). Loss of Hsp90 association up-regulates Src-dependent ErbB2 activity. *Mol Cell Biol* **27:** 220-8.

Yang C, Ionescu-Tiba V, Burns K, Gadd M, Zukerberg L, Louis DN *et al* (2004). The role of the cyclin D1-dependent kinases in ErbB2-mediated breast cancer. *Am J Pathol* **164:** 1031-8.

Yang H, Crawford N, Lukes L, Finney R, Lancaster M, Hunter KW (2005). Metastasis predictive signature profiles pre-exist in normal tissues. *Clin Exp Metastasis* **22:** 593-603.

Yang J, Zhu J, Williams RW (2007). Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics.* **23:** 1527-36.

Yang LQ, Fang DC, Wang RQ, Yang SM (2004). Effect of NF-kappaB, survivin, Bcl-2 and Caspase3 on apoptosis of gastric cancer cells induced by tumor necrosis factor related apoptosis inducing ligand. *World J Gastroenterol* **10:** 22-5.

Yao J, Weremowicz S, Feng B, Gentleman RC, Marks JR, Gelman R *et al* (2006). Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. *Cancer Res* **66:** 4065-78.

Yarden Y, Sliwkowski MX (2001). Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* **2:** 127-37.

Yin C, Knudson CM, Korsmeyer SJ, Van Dyke T (1997). Bax suppresses tumorigenesis and stimulates apoptosis in vivo. *Nature* **385:** 637-40.

Yokota J (2000). Tumor progression and metastasis. *Carcinogenesis* **21:** 497-503.

Yunes MJ, Neuschatz AC, Bornstein LE, Naber SP, Band V, Wazer DE (2003). Loss of expression of the putative tumor suppressor NES1 gene in biopsy-proven ductal carcinoma in situ predicts for invasive carcinoma at definitive surgery. *Int J Radiat Oncol Biol Phys* **56:** 653-7.

Zhan L, Xiang B, Muthuswamy SK (2006). Controlled activation of ErbB1/ErbB2 heterodimers promote invasion of three-dimensional organized epithelia in an ErbB1-dependent manner: implications for progression of ErbB2-overexpressing tumors. *Cancer Res* **66:** 5201-8.

Zhu K, Glaser R, Mrowietz U (2002). Vitamin D(3) and analogues modulate the expression of CSF-1 and its receptor in human dendritic cells. *Biochem Biophys Res Commun* **297:** 1211-7.

Zhu Z, Sanchez-Sweatman O, Huang X, Wiltrout R, Khokha R, Zhao Q *et al* (2001). Anoikis and metastatic potential of cloudman S91 melanoma cells. *Cancer Res.* **61:** 1707–1716.

AIHW (Australian Institute of Health and Welfare) & AACR (Australasian Association of Cancer Registries) 2008. Cancer in Australia: an overview, 2008. Cancer series no. 46. Cat. no. CAN 42. Canberra: AIHW.

AIHW (Australian Institute of Health and Welfare) & AACR (Australasian Association of Cancer Registries) 2006. Cancer in Australia: an overview, 2006. Cancer series no. 46. Cat. no. CAN 42. Canberra: AIHW.

Colburn NH, Moses HL, Stanbridge EJ. , editors. Growth factors, tumor promoters, and cancer genes. Triton Biosciences-UCLA Symposium; 1986 April 6-13; Steamboat Springs, Colorado. New York: Liss; 1986.

Feramisco J, Ozanne B, Stiles CD. , editors. Growth Factors and Transformation. Vol. 3, Cancer Cells. New York: Cold Spring Harbor Laboratories; 1985.

Guroff, G. Oncogenes, genes, and growth factors. New York: Wiley; 1987.

Kudlow, JE. Biology of growth factors : molecular biology, oncogenes, signal transduction, and clinical implications. New York: Plenum Press; 1988.

Parker, PJ; Katan, M. Molecular biology of oncogenes and cell control mechanisms. New York: E Horwood; 1990.

AJCC Cancer Staging Manual *by* Frederick L. Greene, David L. Page *(Editor),* Charles M. Balch *(Editor),* Irvin D. Fleming *(Editor),* Monica Morrow *(Editor)* Publisher: Springer-Verlag New York, LLC, Pub. Date: May 2002, ISBN-13: 9780387952710, Sales Rank: 222,371, 435pp, Edition Description: REV, Edition Number: 6

The Breast Cancer Linkage Consortium. Pathology of familial breast cancer: Differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. *Lancet* 349: 1505-1510, 1997.

Jackson Laboratory website http://jaxmice.jax.org/strain/002374.html

http://training.seer.cancer.gov/ss_module01_breast/unit02_sec01_anatomy. html

What is the prognosis of patients with operable breast cancer (BC) five years after diagnosis? G. N. Hortobagyi, S.-W. Kau, A. U. Buzdar, R. L. Theriault, D. J. Booser, K. Gwyn and V. Valero

J Mansell, IJ Monypenny and AI Skene *et al.*, Predictors of early recurrence in postmenopausal women with operable breast cancer, *Breast Cancer Res Treat* **100** (Suppl 1) (2006), p. S111 Abstract 2091..

http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=6494&ordinalpos=1&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008

http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=283237&ordinalpos=3&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008

http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermTo
Search=51742&ordinalpos=2&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene
_RVDocSum, 20/11/2008

http://www.ncbi.nlm.nih.gov/sites/entrez?
Db=gene&Cmd=ShowDetailView&TermToSearch=64411&ordinalpos=4&itool=EntrezSyst
em2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008

http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch
=23476&ordinalpos=2&itool=EntrezSystem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVD
ocSum, 20/11/2008

http://www.ncbi.nlm.nih.gov/sites/entrez?Db=
gene&Cmd=ShowDetailView&TermToSearch=1432&ordinalpos=1&itool=EntrezSystem2.P
Entrez.Gene.Gene_ResultsPanel.Gene_RVDocSum, 20/11/2008

http:/www.fmvv.nl