

Research article

Open Access

Distinct inherited metastasis susceptibility exists for different breast cancer subtypes: a prognosis studySzu-Min Hsieh^{1*}, Maxime P Look^{2*}, Anieta M Sieuwerts², John A Foekens² and Kent W Hunter¹¹Laboratory of Cancer Biology and Genetics, NCI, NIH, Room 5046, 37 Convent Drive, Bethesda, Maryland, USA 20892-4264²Department of Medical Oncology, Erasmus Medical Center, Josephine Nefkens Institute and Cancer Genomics Centre, Rotterdam, 3015 GE, The Netherlands

* Contributed equally

Corresponding author: Kent W Hunter, hunterk@mail.nih.gov

Received: 16 Jun 2009 Revisions requested: 27 Aug 2009 Revisions received: 17 Sep 2009 Accepted: 13 Oct 2009 Published: 13 Oct 2009

Breast Cancer Research 2009, **11**:R75 (doi:10.1186/bcr2412)This article is online at: <http://breast-cancer-research.com/content/11/5/R75>© 2009 Hsieh *et al.*; licensee BioMed Central Ltd.**Abstract**

Introduction Previous studies in mouse models and pilot epidemiology studies have demonstrated that inherited polymorphisms are associated with inherited risk of tumor progression and poor outcome in human breast cancer. To extend these studies and gain better understanding of the function of inherited polymorphism in breast cancer progression, a validation prognosis study was performed in a large independent breast cancer patient population.

Methods The study population consisted of 1863 Dutch patients with operable primary breast cancer from Rotterdam, The Netherlands. Genomic DNA was genotyped for the missense Pro436Leu RRP1B single nucleotide polymorphism (SNP) rs9306160 and the intronic SIPA1 SNP rs2448490 by SNP-specific PCR.

Results A significant association of variants in RRP1B with metastasis-free survival was observed ($P = 0.012$), validating the role of RRP1B with inherited metastatic susceptibility. Stratification of patients revealed that association with patients'

survival was found to be specifically restricted to estrogen receptor positive, lymph node-negative (ER+/LN-) patients ($P = 0.011$). The specific association with metastasis-free survival only in ER+/LN- patients was replicated for SIPA1, a second metastasis susceptibility gene known to physically interact with RRP1B ($P = 0.006$). Combining the genotypes of these two genes resulted in the significant ability to discriminate patients with poor metastasis-free survival (HR: 0.40, 95% CI: 0.24 to 0.68, $P = 0.001$).

Conclusions These results validate SIPA1 and RRP1B as metastasis susceptibility genes and suggest that genotyping assays may be a useful supplement to other clinical and molecular indicators of prognosis. The results also suggest that lymphatic and hematogeneous metastases are genetically distinct that may involve different mechanisms. If true, these results suggest that metastatic disease, like primary breast cancer, may be multiple diseases and that stratification of late stage patients may therefore be required to fully understand breast cancer progression and metastasis.

Introduction

Cancer mortality can be attributed mostly to metastatic disease, with an estimated 90% of deaths associated with solid tumors resulting from the pathophysiological impact of secondary disease. Despite many advances in both basic science and applied clinical research over recent years, advanced disseminated disease remains an incurable condition. Further investigations into the myriad of factors associated with metastatic disease are therefore warranted to identify critical molecular nodes and targets in this complex process that will enable development and deployment of new or improved clin-

ical tools for combating the effects of advanced disseminated disease.

One of these factors for breast cancer is inherited metastatic susceptibility. Recently, using a mouse model system, it was demonstrated that germline polymorphisms have significant effects on the ability of a transgene-induced mouse mammary tumor to metastasize [1-3]. Subsequently, using small pilot clinical cohorts, significant associations with markers of poor outcome were observed, consistent with the presence of metastasis susceptibility in the human populations [4,5].

CI: confidence interval; DFS: disease-free survival; ER: estrogen receptor; HER: human epidermal growth factor receptor; HR: hazard ratio; LN: lymph node; MFS: metastasis-free survival; OS: overall survival; PCR: polymerase chain reaction; PR: progesterone receptor; SNP: single nucleotide polymorphism.

Descriptive epidemiology studies further support this hypothesis, demonstrating familial clustering of outcome in a variety of different cancer types [6-11].

The current study builds on and extends the previous studies of the first two identified metastasis efficiency modifier genes, *SIPA1* [3] and *RRP1B* [4]. Using a much larger cohort, significant associations between polymorphisms in these genes and advanced disease were identified replicating earlier studies. Unexpectedly, however, these associations were restricted to subgroups of patients after stratification by the estrogen receptor (ER) and lymph node (LN) status. The results suggest that at least for inherited metastatic susceptibility in breast cancer that these subpopulations could be biologically distinct with different pathways leading to the metastatic disease.

Materials and methods

Patient population

The protocol to study biological markers associated with disease outcome was approved by the medical ethics committee of the Erasmus Medical Center Rotterdam, The Netherlands (MEC 02.953). This retrospective study used coded primary tumor tissue, in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands [12] and, as much as possible, was reported in line with the REMARK guidelines [13]. The single nucleotide polymorphisms (SNPs) were determined in 1863 tumor tissues. ER levels were missing for nine patients and progesterone receptor (PR) levels for 104 patients. Data for one of the two SNPs were not available for 25 tumors. The final study includes breast tumor tissue specimens of 1725 female Dutch patients with primary operable breast cancer (990 patients underwent a mastectomy, 735 patients underwent breast-conserving lumpectomy) who entered the clinic in Rotterdam between 1979 and 2002 with ER and PR levels known as well as both SNPs rs2448490 and rs9306160. Radiotherapy was given to 1162 patients as part of primary treatment. Adjuvant therapy was not performed as part of the primary treatment for LN-patients. Of the LN+ patients, 24% (187 of 766) were received systemic adjuvant therapy. The median follow up of alive patients was 90 months (range, 4 to 231 months). The clinical questions addressed in the present study include the associations of the various SNP frequencies with patient and tumor characteristics, and prognosis in primary breast cancer.

Tumor ER and PR levels were determined in cytosolic extracts by routine ligand binding assay or by enzyme immunoassay [14]. The cut point to classify primary breast tumors as ER and/or PR positive was 10 fmol/mg cytosolic protein. None of the patients had received neo-adjuvant therapy. Details on patient and tumor characteristics are presented in Table 1.

DNA isolation and whole genome amplification

Genomic DNA was isolated from two to ten 30 μ m cryostat sections (5 to 20 mg) with the NucleoSpin[®]Tissue kit (Macherey-Nagel; Bioké, Leiden, The Netherlands) according to the protocol provided by the manufacturer. The quantity and quality of the isolated DNA was established by ultraviolet spectroscopy, by examination of the product size after agarose gel electrophoresis, and by the ability of the sample to be linearly amplified by real-time PCR in a serial dilution with a set of primers located in an intron of the hydroxymethylbilane synthase on chromosome 11 and thymidine kinase on chromosome 17. Samples not showing a DNA band of at least 20 kb or at 5 to 25 ng DNA not amplifiable by both real-time PCR assays were excluded. Prior to SNP genotyping, 10 ng aliquots of genomic DNA were amplified with the GenomiPhi V2 DNA amplification kit (GE Healthcare, Piscataway, NJ, USA) according to the protocol provided by the manufacturer, typically yielding 4 μ g amplifiable genomic DNA with the 20 kb band still visible on gel.

SNP selection and genotyping

SIPA1 and *RRP1B* polymorphisms were characterized using allele-specific PCR. PCR primers were designed using Vector NTI 9.0 software (Invitrogen, Carlsbad, CA, USA) according to parameters described elsewhere [15] or purchased from Applied Biosystems (Foster City, CA, USA). Each probe was labeled with a reporter dye (either VIC[®] (a proprietary fluorescent dye produced by Applied Biosystems) or FAM (5-(6)-carboxyfluorescein)) specific for wildtype and variant alleles of each SNP.

The *RRP1B* SNP rs9306160 was previously described [4]. Briefly, it encodes a Pro436Leu missense mutation in the *RRP1B* protein, and tags an approximately 200 kb haplotype block encompassing both *RRP1B* and the adjacent *HSF2BP* gene. The primers and probes for the *RRP1B* were as follows, 5'-3': forward, TGGACGTGGCCTCTGCAC; reverse, CAC-CACCTGCAGCCTGAAA; Vic labeled, AGGGCTTTCG-GCCCAG; FAM labeled AGGGCTTTCAGCCCAGAG. The *SIPA1* SNP rs2448490 was genotyped using the ABI assay C__15797548_10.

Reaction mixtures consisted of 300 nM of each oligonucleotide primer, 100 nM fluorogenic probes 8 ng template DNA, and 2 \times TaqMan Universal PCR Master Mix (Applied Biosystems, Foster City, CA, USA) in a total volume of 10 μ l. The amplification reactions were performed in a MJ Research DNA Engine thermocycler (Bio-Rad, Hercules, CA, USA) with two initial hold steps (50°C for 2 minutes, followed by 95°C for 10 minutes) and 40 cycles of a two-step PCR (92°C for 15 seconds, 60°C for 1 minute). The fluorescence intensity of each sample was measured post-PCR in an ABI Prism 7900 HT sequence detection system (Applied Biosystems, Foster City, CA, USA), and genotypes were determined by the fluorescence ratio of the nucleotide-specific fluorogenic probes. The

Table 1**Genotype distributions by patient and tumor characteristics**

Characteristic	Patients	SIPA1 rs2448490						RRP1B rs9306160					
		GG		AG		AA		CC		CT		TT	
		n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Total	1863 ¹	748	(40)	807	(43)	293	(16)	616	(33)	904	(49)	333	(18)
Age (years)													
≤40	239	98	(41)	98	(41)	43	(18)	82	(34)	123	(51)	34	(14)
41-55	736	301	(41)	323	(44)	103	(14)	235	(32)	359	(49)	139	(19)
56-70	597	228	(39)	265	(45)	98	(17)	193	(33)	285	(48)	114	(19)
>70	291	121	(42)	121	(42)	49	(17)	106	(37)	137	(47)	46	(16)
				<i>P</i> = 0.661				<i>P</i> = 0.481					
Menopausal status													
Premenopausal	823	334	(41)	364	(45)	117	(14)	268	(33)	419	(51)	135	(16)
Postmenopausal	1040	414	(40)	443	(43)	176	(17)	348	(34)	485	(47)	198	(19)
				<i>P</i> = 0.288				<i>P</i> = 0.165					
Lymph nodes involved													
0	1095	441	(41)	459	(42)	183	(17)	358	(33)	535	(49)	192	(18)
1-3	350	134	(38)	165	(47)	50	(14)	107	(31)	180	(51)	63	(18)
>3	418	173	(42)	183	(44)	60	(14)	151	(36)	189	(45)	78	(19)
				<i>P</i> = 0.438				<i>P</i> = 0.459					
Tumor size													
pT1	686	280	(41)	289	(43)	109	(16)	209	(31)	340	(50)	131	(19)
pT2	977	390	(40)	435	(45)	148	(15)	333	(34)	461	(47)	180	(18)
pT3/4	200	78	(40)	83	(42)	36	(18)	74	(37)	103	(52)	22	(11)
				<i>P</i> = 0.789				<i>P</i> = 0.049					
Grade													
Poor	1007	403	(40)	437	(44)	162	(16)	326	(33)	508	(51)	168	(17)
Good/moderate	282	113	(41)	120	(43)	44	(16)	101	(36)	122	(44)	56	(20)
Unknown	574	232	(41)	250	(44)	87	(15)	189	(33)	274	(48)	109	(19)
				<i>P</i> = 0.994				<i>P</i> = 0.280					
ER status													
Positive	1367	529	(39)	606	(45)	219	(16)	445	(33)	666	(49)	249	(18)
Negative	487	217	(45)	194	(40)	74	(15)	169	(35)	233	(48)	82	(17)
				<i>P</i> = 0.087				<i>P</i> = 0.625					
PR status													
Positive	1149	442	(39)	521	(46)	174	(15)	369	(32)	563	(49)	212	(19)
Negative	601	267	(45)	232	(39)	99	(17)	207	(35)	290	(49)	99	(17)
				<i>P</i> = 0.018				<i>P</i> = 0.459					

¹Due to missing values the numbers do not always add up to 1863.²All *P* values derived from Pearson's chi-squared statistics.

xER = oestrogen receptor; PR = progesterone receptor.

genotyping success rate for rs2448490 was 99.2% (2491 of 2511 samples, controls and duplicates). The success rate for rs9306160 was 99.3%. The concordance rate for rs2448490 was 98.3% (404 of 411 duplicates) and 98.1% (404 of 411) for rs9306160.

Statistical analysis

Pearson's chi-squared statistic was used to study the relation of the variant SNP alleles with patient and tumor characteristics. The hazard ratios (HRs) for SNPs and traditional prognostic factors were determined with Cox proportional hazards models for both univariate (disease-free survival (DFS), metastasis-free survival (MFS), and overall survival (OS)) and multivariate regression analyses (with backward elimination) in 1725 patients. The assumption of proportional hazards was checked using Schoenfeld residuals. We stratified for ER because the assumption of proportionality was violated for ER. MFS was considered the major endpoint for the prognostic study. The endpoint for DFS was defined as any recurrence of the disease (958 events) including secondary breast cancer in the contralateral breast. Metastasis was defined as any distant recurrence (772 events) not including secondary breast cancer or local or regional recurrences. For OS, death from any cause was considered an event ($n = 684$). The HRs are represented with their 95% confidence intervals (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 log-rank test for trend. Computations were performed with the STATA statistical package, release 10.0 (STATA Corp, College Station, TX, USA). All P values were two-sided, and $P < 0.05$ was considered statistically significant.

Results

SNP selection

Previous studies with *SIPA1* revealed some potential associations with LN metastasis [5], but no significant associations with distant MFS or patient survival [5,16]. Recent examination of the HapMap database [17], however, indicated that the previous SNPs did not tag all of the haplotype blocks for this gene. As a result, the tagged *SIPA1* SNP rs2448490 was selected to provide more comprehensive *SIPA1* coverage in addition to the previously metastasis associated *RRP1B* rs9306160 for genotyping in this cohort.

Analysis of SNP frequencies and tumor characteristics

The SNPs were genotyped in 1863 breast cancer cases. The frequencies of the variant alleles were 37.5% for rs2448490 A and 42.5% for rs9306160 T. rs9306160 was found to be in Hardy-Weinberg equilibrium, while rs2448490 was not in this population. No significant or strong relations of any of the SNP genotypes with patient's age or menopausal status, tumor size, LN status or hormone receptor status were observed (Table 1). Tumors from homozygote carriers of the *RRP1B*

rs9306160 T allele tended to be smaller tumors in this cohort ($P = 0.049$). *SIPA1* SNP rs2448490 was found to be associated with the tumor steroid hormone receptor status (ER, $P = 0.087$; PR, $P = 0.018$; Table 1).

SNP associations with prognosis in primary breast cancer

SNP allele frequencies were subsequently examined for their associations with prognosis. Survival analyses were conducted for 1,725 patients with both SNPs determined and ER and PR status known. Since *SIPA1* and *RRP1B* had been previously associated with risk of metastatic progression, metastasis-free survival (MFS) was considered the primary endpoint for this study. Disease-free survival (DFS) and overall survival (OS) were included to examine the associations of these genes with disease relapse and overall outcome.

Consistent with our previous observations [4], patients carrying T allele (CT or TT) of *RRP1B* SNP rs9306160 showed a favorable prognosis for MFS and the same trends for DFS and OS, when compared with the wildtype CC. Based on these results, the CT and TT patients were grouped together assuming either an additive or dominant effect of the T allele and the univariate analysis repeated. Under this model significant associations of the T allele with better outcome were observed for MFS and OS (DFS (HR: 0.88, 95% CI: 0.77 to 1.01, $P = 0.063$); MFS (HR: 0.80, 95% CI: 0.69 to 0.92, $P = 0.002$); OS (HR: 0.85, 95% CI: 0.73 to 1.00, $P = 0.046$)). Multivariate analysis was then performed to determine whether the T allele was an independent prognostic factor when compared with standard clinical factors. Only the association with MFS was significant in multivariate analysis (Table 2) while the associations of the T allele with DFS and OS were partly confounded in the multivariate analysis and not statistically significant (data not shown). Adjuvant chemotherapy did not affect the estimated coefficients of the *RRP1B* SNP rs9306160 genotype.

As the proportional hazards assumption was violated by ER, we stratified the patients. LN- patients were not treated with systemic adjuvant therapy while LN+ patients did receive systemic treatment. This enabled us to evaluate associations in four subgroups, LN-/ER+ with good prognosis and LN-/ER-, LN+/ER+ and LN+/ER- with an expected poor prognosis. Surprisingly, the association of the variant T allele in *RRP1B* rs9306160 with a favorable prognosis was significant only for the subgroup of LN-/ER+ patients (Figure 1a), but not in the other patient subgroups (LN+/ER+ (Figure 1b), LN-/ER-, LN+/ER- (not shown)). This association was statistically significant for DFS and MFS in multivariate analysis (DFS (HR: 0.75, 95% CI: 0.60 to 0.94, $P = 0.013$), MFS (HR: 0.70, 95% CI: 0.54 to 0.91, $P = 0.009$) and OS (HR: 0.78, 95% CI: 0.59 to 1.03, $P = 0.082$)).

The variant A allele of *SIPA1* rs2448490 was also found to be associated with better outcome after stratification by ER and

Table 2**Final model for multivariate analysis of MFS in all patients¹**

Characteristic	HR (95% CI)	P
Age (years)		
41-55 vs. ≤ 40	0.88 (0.71-1.10)	0.265
56-70 vs. ≤ 40	0.70 (0.5-0.98)	0.036
>70 vs. ≤ 40	0.62 (0.43-0.90)	0.013
Menopausal status		
Post vs. premenopausal	1.30 (1.01-1.67)	0.045
Lymph nodes involved		
1-3 vs. 0	1.80 (1.49-2.18)	< 0.001
>3 vs. 0	2.73 (2.29-3.25)	< 0.001
Tumor size		
pT2 vs. pT1	1.56 (1.32-1.86)	< 0.001
pT3/4 vs. pT1	1.84 (1.44-2.36)	< 0.001
Grade		
Good/moderate vs. poor	0.56 (0.43-0.72)	< 0.001
Unknown vs. poor	0.91 (0.77-1.06)	0.221
ER		
Positive vs. negative	0.79 (0.65-0.96)	0.016
PR		
Positive vs. negative	0.82 (0.69-0.98)	0.032
<i>RRP1B</i> rs9306160		
CT+TT vs. CC	0.83 (0.71-0.96)	0.012
<i>SIPA1</i> rs2448490		
GG+GA vs. AA	0.85 (0.69-1.05)	0.129

¹1725 patients

CI = confidence interval; ER = estrogen receptor; HR = hazard ratio; MFS = metastasis-free survival; PR = progesterone receptor.

LN status. Patients homozygous for the A allele had a favorable MFS (Figure 1c) and OS (HR: 0.57, 95%CI: 0.38 to 0.86, $P = 0.007$ HR: 0.64, 95%CI: 0.43 to 0.97, $P = 0.036$). A similar trend, although not statistically significant, was found for DFS. The associations were significant in the multivariate analysis for DFS (HR: 0.73, 95% CI: 0.54 to 0.99, $P = 0.042$), MFS (HR: 0.56, 95% CI: 0.37 to 0.84, $P = 0.005$) and OS (HR: 0.59, 95% CI: 0.39 to 0.89, $P = 0.012$).

SIPA1 SNP rs2448490/*RRP1B* SNP rs9306160 combination

Previous studies demonstrated that the *SIPA1* and *RRP1B* gene products physically interact [4]. As the *SIPA1* SNP rs2448490 and *RRP1B* SNP rs9306160 were both associated with a favorable prognosis in LN-/ER+ patients, we

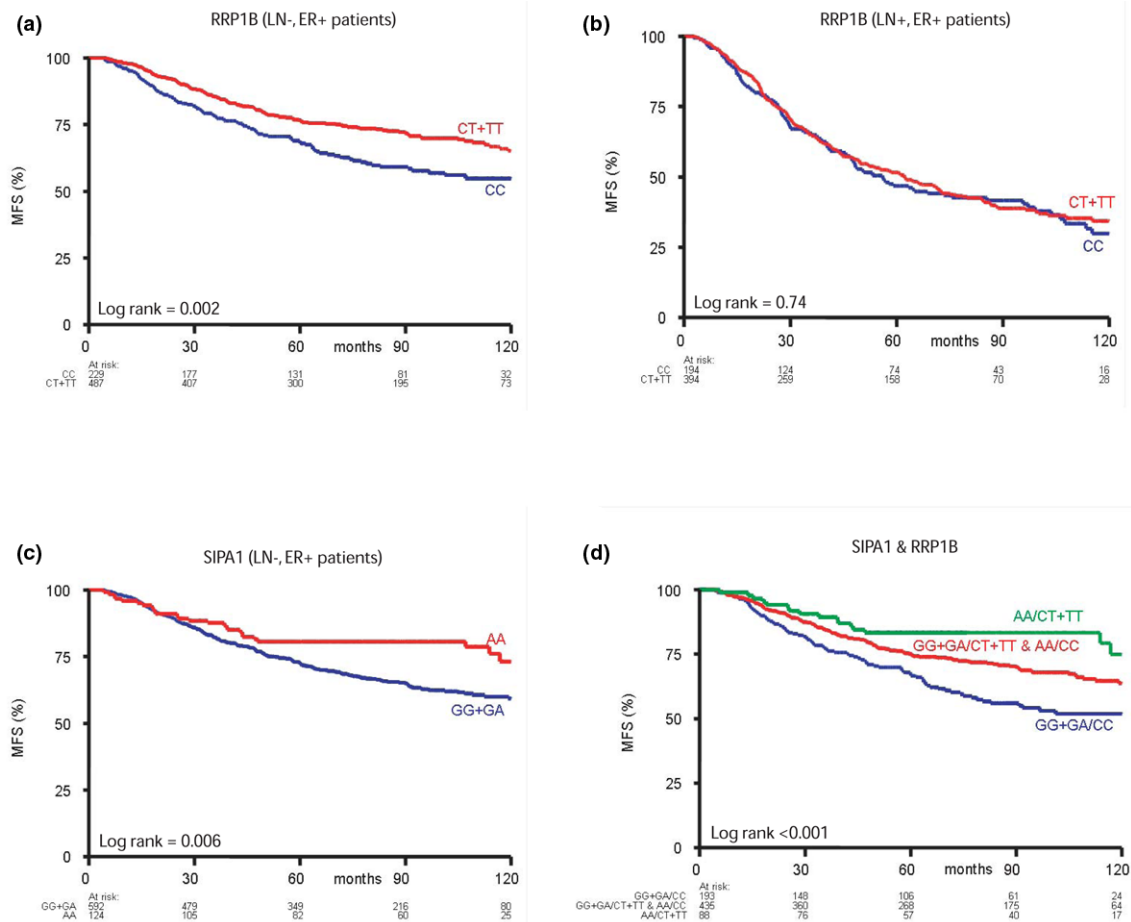
explored the possibility that they were independent predictors of MFS and that the combination of both SNP genotypes might show increased prognostic power. Indeed, *SIPA1* SNP rs2448490 and *RRP1B* SNP rs9306160 remained independent predictors of MFS if both SNPs were included in the final multivariate model (Table 3). Kaplan-Meier analysis for MFS as a function of the combined genotypes in the LN-/ER+ patients showed that the combination of the homozygous AA variant allele of *SIPA1* SNP rs2448490 and the T variant allele (CT+TT) of *RRP1B* SNP rs9306160 was the best prognosticator (Figure 1d). The risk for developing distant metastasis was about 2.5-fold lower for patients with this genotype combination compared with carriers of the GG and GA/CC genotype combination (HR: 0.40, 95% CI: 0.24 to 0.68, $P = 0.001$).

Discussion

Significant advances have been made in the understanding of breast cancer in the past decade. It is now understood that there are at least four molecular subtypes: luminal A, luminal B, basal and human epidermal growth receptor (HER)2-positive tumors [18]. Furthermore, a variety of studies have demonstrated that gene expression profiles can discriminate between patients of differing outcome. As a result, a number of different commercial assays are currently available [19] to aid patients and clinicians in their decisions for therapeutic intervention, two of which are currently in prospective clinical trials [20,21]. Despite the importance of these findings, the origins of the gene expression signatures are unclear. Based on the prevailing model, it was presumed gene expression signatures would be the result of an accumulation of somatic mutations during the evolution of the tumor. However, the ability to discriminate patient outcome based on bulk tumor expression data was considered inconsistent with that hypothesis, because only a small fraction of the tumor would be predicted to express the appropriate signature, as predicted by the progression model [22]. These observations have led to a renewed discussion into the molecular mechanisms of breast cancer metastasis [23,24].

Studies in our laboratory have suggested that one of the previously unknown factors contributing to breast cancer metastasis is genetic background. Using an animal model system, we demonstrated that the genetic background had a significant impact on its ability to form pulmonary metastases [2]. Subsequently, systems genetics approaches have identified a number of polymorphic metastasis efficiency genes [3,4,25,26]. These results therefore suggest that the prognostic gene expression signatures currently in clinical trials may be in part due to inherited polymorphism rather than somatic mutation, and may be a surrogate for inherited metastasis susceptibility segregating in the human population. This interpretation is strengthened by the recent demonstration that prognostic gene expression signatures pre-exist between normal tissues of animals of high- or low-metastatic genotypes

Figure 1



Kaplan-Meier analysis of MFS for *SIPA1* and *RRP1B* SNPs on stratified patient populations. (a) *RRP1B* rs9306160 in node-negative, estrogen receptor (ER)-positive patients; (b) *RRP1B* rs9306160 in node-positive, ER-positive patients; (c) *SIPA1* rs2448490 in node-negative, ER-positive patients; (d) Combination of *SIPA1* rs2448490 and *RRP1B* rs9306160 in node-negative, ER-positive patients. LN = lymph node; MFS = metastasis-free survival.

[27]. Taken together, these data support the hypothesis that genotype-based assays may be a valuable complement or supplement to clinical and gene expression-based prognostic tools.

This study therefore builds on the preliminary epidemiology studies of two of our previously described metastasis efficiency genes, *SIPA1* [5] and *RRP1B*, a chromatin associated protein of unknown function [4,28]. Initial investigations of *SIPA1* did not reveal associations with MFS [5,16]. However, recent analysis of HapMap database [17] indicated that the SNPs investigated in these previous studies did not completely haplotype-tag this locus. Therefore an additional SNP was investigated in this study to improve coverage. In contrast, evidence for an association with a polymorphism in *RRP1B* and MFS had been previously observed in two small pilot cohorts [4]. This study therefore sought to replicate these results in a larger cohort, as well as to investigate whether

there was a genetic, in addition to the physical, interaction between *RRP1B* and *SIPA1* [4].

The results of these studies suggest a number of important points. First, as predicted by the mouse genetic and pilot epidemiology studies, genetic background is likely to be an important factor for human breast cancer progression because significant associations were observed for both genes. Although the results are consistent with these associations resulting from inherited predisposition an important caveat of this study is that it is also formally possible that these results stem from copy number variation in the tumor DNA used, which is the only material available from this unique cohort. We believe, however, that this is unlikely for the following reasons. First, the results are consistent with the previous studies which were performed in constitutional DNA from normal lymphocytes. For *RRP1B* at least, this is unlikely to be a false-positive result because the same association has now been observed in three independent patient populations. Sec-

Table 3**Final model for multivariate analysis of MFS in ER+ lymph node-negative patients¹**

Characteristic	HR (95% CI)	P
Age (years)		
41-55 vs. ≤ 40	0.72 (0.49-1.06)	0.101
56-70 vs. ≤ 40	0.43 (0.28-0.65)	< 0.001
>70 vs. ≤ 40	0.44 (0.28-0.70)	0.001
Tumor size		
pT2 vs. pT1	1.41 (1.07-1.84)	0.013
pT3/4 vs. pT1	1.82 (0.95-3.51)	0.073
Grade		
Good/moderate vs. Poor	0.58 (0.38-0.87)	0.009
Unknown vs. Poor	1.05 (0.79-1.39)	0.729
<i>SIPA1</i> rs2448490		
AA vs. AG+GG	0.56 (0.38-0.85)	0.006
<i>RRP1B</i> rs9306160		
CT+TT vs. CC	0.71 (0.54-0.92)	0.011

¹761 patients

CI = confidence interval; ER = estrogen receptor; HR = hazard ratio; MFS = metastasis-free survival.

ond, the allele frequencies of the SNPs does not vary between tumor types and subgroups, as might be expected if there was a preferential copy number change in a subset of tumors. Thus although at this time we can not formally rule out a contribution of somatic evolution we favor the hypothesis that these effects are likely due to inherited factors. Future replication in an independent cohort based on constitutional DNA will resolve this possibility.

The second major point, as suggested by the physical interaction of the gene products, is that the combination of the *SIPA1* and *RRP1B* SNPs is an independent predictor of MFS when compared with standard clinical parameters, capable of discriminating high-risk, intermediate-risk and low-risk individuals. This combination SNP assay may therefore provide a valuable addition to current methods. This SNP assay would have a number of advantages over current gene expression based assays. As it is based on constitutional DNA it can be performed from routinely collected peripheral blood, rather than tumor tissue, which require more invasive procedures. In addition, because DNA is more stable than RNA, there are fewer constraints on collection, handling and processing procedures. Furthermore, genotyping methods are relatively inexpensive, robust and rapid, and thus would likely be significantly less expensive than expression array based methods.

In addition to the potential clinical benefit, this study has important implications for our understanding of the mechanisms of metastatic progression. The fact that these polymorphisms are predictive of MFS in LN-, ER+, but not other subgroups suggest that at least for inherited metastatic susceptibility, there must be at least two pathways for metastatic progression. The lack of association in the LN+ samples indicates that these individuals are not simply diagnosed at a later time along a linear progression pathway. Instead, it suggests that those tumors that spread through the vasculature and those that seed the lymphatics likely use distinct molecular pathways during dissemination. This interpretation is consistent with previous observations in the literature. Analysis of breast cancer subtypes as defined by gene expression profiles [18] demonstrated preferential sites of relapse [29], suggesting different mechanisms of colonization. In addition, women with triple-negative breast cancers (ER-, PR-, HER2-) are less likely to experience a local recurrence before developing a distant recurrence [30]. Similarly, *BRCA1* carriers have been shown to be less likely to have positive axillary LNs at diagnosis than non-hereditary breast cancers [31]. To our knowledge, however, this is the first example of the ability of common allelic variants to discriminate patient outcome in specific clinical tumor types.

Although these results are consistent with constitutional predisposition to metastatic disease and suggest that the inherited susceptibilities for tumors that disseminate to the LNs is different than those that metastasize directly distant organs, the mechanisms used are currently unknown. It is possible that the allelic variants of these two genes might significantly alter the likelihood of tumors activating different pathways; for example, angiogenesis versus lymphangiogenesis, which would be expected to help direct tumor cells away from or toward sentinel LNs. At present, however, neither of these genes have been directly implicated in these pathways. *SIPA1* is a RAPGAP signaling molecule [32] and *RRP1B* is a chromatin associated protein of unknown function [28]. Both molecules have been previously implicated in the expression of extracellular matrix genes, which in and of themselves have been associated with metastatic progression. As these variants are present in constitutional DNA the effect of the different alleles on metastatic disease could be due to modulation of tumor cells, the microenvironments the cells encounter or a combination of both. At present it is not clear which of these possibilities is most applicable. Because of these multiple possibilities and the complexity of each component, the exact biological mechanism by which these molecules operate is therefore likely to be complex and require significant additional efforts to unravel the exact mechanistic details.

Conclusions

In summary, this study replicates the previous association of *RRP1B* with MFS and establishes a previously unknown association with *SIPA1*. Combination of these SNPs enables effec-

tive discrimination of patients into high-risk, intermediate-risk and low-risk categories for MFS, independent of standard clinical parameters. Furthermore, the association of genetic susceptibility for MFS only in specific clinical subgroups indicates that multiple molecular mechanisms for metastatic progression are likely to be involved in breast cancer progression. Further investigations into the utility of these polymorphisms, including validation in additional retrospective cohorts, analysis in prospective clinical trials, and analysis of the effect of the variants on the molecular biology of the tumor and host are clearly warranted to provide further insights into the role of inherited polymorphism in breast cancer dissemination and metastasis.

Competing interests

JA Foekens received research support from Veridex LLC. Patent applications filed by NCI for RRP1B genotyping.

Authors' contributions

SH performed the genotyping and interpreted the data. ML performed the statistical analysis. AS prepared the samples. JF and KH designed the experiments and interpreted the data. SH, ML, JF and KH wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Drs Nigel Crawford and Stefan Ambs for critical comments on this manuscript. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research to KWH and in part by The Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) to JAF.

References

- Hunter KW, Broman KW, Voyer TL, Lukes L, Cozma D, Debies MT, Rouse J, Welch DR: **Predisposition to efficient mammary tumor metastatic progression is linked to the breast cancer metastasis suppressor gene Brms1.** *Cancer Res* 2001, **61**:8866-8872.
- Lifsted T, Le Voyer T, Williams M, Muller W, Klein-Szanto A, Buetow KH, Hunter KW: **Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression.** *Int J Cancer* 1998, **77**:640-644.
- Park YG, Zhao X, Lesueur F, Lowy DR, Lancaster M, Pharoah P, Qian X, Hunter KW: **Sipa1 is a candidate for underlying the metastasis efficiency modifier locus Mtes1.** *Nat Genet* 2005, **37**:1055-1062.
- Crawford NP, Qian X, Ziogas A, Papageorge AG, Boersma BJ, Walker RC, Lukes L, Rowe WL, Zhang J, Ambs S, Lowy DR, Anton-Culver H, Hunter KW: **Rrp1b, a new candidate susceptibility gene for breast cancer progression and metastasis.** *PLoS Genetics* 2007, **3**:e214.
- Crawford NP, Ziogas A, Peel DJ, Hess J, Anton-Culver H, Hunter KW: **Germline polymorphisms in SIPA1 are associated with metastasis and other indicators of poor prognosis in breast cancer.** *Breast Cancer Res* 2006, **8**:R16.
- Hartman M, Lindstrom L, Dickman PW, Adami HO, Hall P, Czene K: **Is breast cancer prognosis inherited?** *Breast Cancer Res* 2007, **9**:R39.
- Hemminki K, Ji J, Forsti A, Sundquist J, Lenner P: **Concordance of survival in family members with prostate cancer.** *J Clin Oncol* 2008, **26**:1705-1709.
- Hemminki K, Ji J, Forsti A, Sundquist J, Lenner P: **Survival in breast cancer is familial.** *Breast Cancer Res Treat* 2008, **110**:177-182.
- Ji J, Forsti A, Sundquist J, Lenner P, Hemminki K: **Survival in bladder and renal cell cancers is familial.** *J Am Soc Nephrol* 2008, **19**:985-991.
- Lindstrom LS, Hall P, Hartman M, Wiklund F, Gronberg H, Czene K: **Familial concordance in cancer survival: a Swedish population-based study.** *Lancet Oncol* 2007, **8**:1001-1006.
- Qu S, Long J, Cai Q, Shu XO, Cai H, Gao YT, Zheng W: **Genetic polymorphisms of metastasis suppressor gene NME1 and breast cancer survival.** *Clin Cancer Res* 2008, **14**:4787-4793.
- Freter RR, Alberta JA, Hwang GY, Wrentmore AL, Stiles CD: **Platelet-derived growth factor induction of the immediate-early gene MCP-1 is mediated by NF-kappaB and a 90-kDa phosphoprotein coactivator.** *J Biol Chem* 1996, **271**:17417-17424.
- McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM: **REporting recommendations for tumor MARKer prognostic studies (REMARK).** *Breast Cancer Res Treat* 2006, **100**:229-235.
- Foekens JA, Portengen H, van Putten WL, Peters HA, Krijnen HL, Alexieva-Figusch J, Klijn JGM: **Prognostic value of estrogen and progesterone receptors measured by enzyme immunoassays in human breast tumor cytosols.** *Cancer Res* 1989, **49**:5823-5828.
- Crawford NP, Colliver DW, Funke AA, Young MN, Kelley S, Cobbs GA, Petras RE, Galandiuk S: **Characterization of genotype-phenotype relationships and stratification by the CARD15 variant genotype for inflammatory bowel disease susceptibility loci using multiple short tandem repeat genetic markers.** *Hum Mutat* 2005, **25**:156-166.
- Gaudet MM, Hunter K, Pharoah P, Dunning AM, Driver K, Lisowska J, Sherman M, Peplonska B, Brinton LA, Chanock S, Garcia-Closas M: **Genetic variation in SIPA1 in relation to breast cancer risk and survival after breast cancer diagnosis.** *Int J Cancer* 2009, **124**:1716-1720.
- The International HapMap Project.** *Nature* 2003, **426**:789-796.
- Perou CM, Sorlie T, Eisen MB, Rijn M van de, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
- Sotiriou C, Pusztai L: **Gene-expression signatures in breast cancer.** *N Engl J Med* 2009, **360**:790-800.
- Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ: **Clinical application of the 70-gene profile: the MIND-ACT trial.** *J Clin Oncol* 2008, **26**:729-735.
- Sparano JA: **TAILORx: trial assigning individualized options for treatment (Rx).** *Clin Breast Cancer* 2006, **7**:347-350.
- Bernards R, Weinberg RA: **A progression puzzle.** *Nature* 2002, **418**:823.
- Fidler IJ, Kripke ML: **Genomic analysis of primary tumors does not address the prevalence of metastatic cells in the population.** *Nat Genet* 2003, **34**:23.
- Sherley JL: **Metastasis: objections to the same-gene model.** *Nature* 2002, **419**:560. author reply 560
- Crawford NP, Alsarraj J, Lukes L, Walker RC, Officewala JS, Yang HH, Lee MP, Ozato K, Hunter KW: **Bromodomain 4 activation predicts breast cancer survival.** *Proc Natl Acad Sci USA* 2008, **105**:6380-6385.
- Crawford NP, Walker RC, Lukes L, Officewala JS, Williams RW, Hunter KW: **The Diasporin Pathway: a tumor progression-related transcriptional network that predicts breast cancer survival.** *Clin Exp Metastasis* 2008, **25**:357-369.
- Lukes L, Crawford NP, Walker R, Hunter KW: **The origins of breast cancer prognostic gene expression profiles.** *Cancer Res* 2009, **69**:310-318.
- Crawford NP, Yang H, Mattaini KR, Hunter KW: **The metastasis efficiency modifier Ribosomal RNA Processing 1 Homolog B (RRP1B) is a chromatin-associated factor.** *J Biol Chem* 2009, **284**:28660-28673.
- Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG, Foekens JA, Martens JW: **Subtypes of breast cancer show preferential site of relapse.** *Cancer Res* 2008, **68**:3108-3114.
- Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA: **Triple-negative breast cancer: clinical features and patterns of recurrence.** *Clin Cancer Res* 2007, **13**:4429-4434.

31. Foulkes WD, Metcalfe K, Hanna W, Lynch HT, Ghadirian P, Tung N, Olopade O, Weber B, McLennan J, Olivotto IA, Sun P, Chapuis PO, Begin LR, Brunet JS, Narod SA: **Disruption of the expected positive correlation between breast tumor size and lymph node status in BRCA1-related breast carcinoma.** *Cancer* 2003, **98**:1569-1577.
32. Tsukamoto N, Hattori M, Yang H, Bos JL, Minato N: **Rap1 GTPase-activating protein SPA-1 negatively regulates cell adhesion.** *J Biol Chem* 1999, **274**:18463-18469.