

A susceptibility gene signature for ERBB2-driven mammary tumor development and metastasis in Collaborative Cross mice with human translational value

Hui Yang^{1,2}, Xinzhi Wang^{1,3}, Adrián Blanco-Gómez^{4,5}, Li He^{1,6}, Natalia García-Sancha^{4,5}, Roberto Corchado-Cobos^{4,5}, Manuel Jesús Pérez-Baena^{4,5}, Alejandro Jiménez-Navas^{4,5}, Pin Wang^{1,7}, Jamie L Inman¹, Antoine M Snijders^{1,8}, David W Threadgill^{9,10}, Allan Balmain¹¹, Hang Chang^{1,8*}, Jesus Perez-Losada^{4,5*}, Jian-Hua Mao^{1,8*}

¹Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.

²New address: Department of Radiation and Medical Oncology, Hubei Key Laboratory of Tumor Biological Behaviors, Zhongnan Hospital of Wuhan University, Wuhan, Hubei 430071, China.

³New address: College of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210046, China.

⁴Instituto de Biología Molecular y Celular del Cáncer (IBMCC-CIC), Universidad de Salamanca/CSIC, Salamanca, 37007, Spain.

⁵Instituto de Investigación Biosanitaria de Salamanca (IBSAL), Salamanca, 37007, Spain.

⁶Department of Hematology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei 430079, China.

⁷Department of Gastroenterology, Nanjing Drum Tower Hospital, Affiliated Hospital of Nanjing University Medical School, Nanjing, Jiangsu 210008, China.

⁸Berkeley Biomedical Data Science Center, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

⁹Department of Nutrition, Texas A&M University, College Station, Texas 77843, USA.

¹⁰Department of Molecular and Cellular Medicine and Department of Biochemistry & Biophysics, Texas A&M University, College Station, Texas 77843, USA.

¹¹Helen Diller Family Comprehensive Cancer Center, University of California San Francisco; San Francisco, 94158, USA.

* Correspondence: jperezlosada@usal.es, hchang@lbl.gov, JHMao@lbl.gov

Abstract

Background Deeper insights into ERBB2-driven cancers are essential to develop novel treatment avenues for ERBB2+ breast cancers (BCs). We employed Collaborative Cross (CC) mouse model, along with human translational evaluation, to unearth genetic factors underpinning *ErbB2*-driven mammary tumor development and metastasis.

Methods 732 F1 hybrid female mice between FVB/N MMTV-*ErbB2* and 30 CC strains were monitored for mammary tumor phenotypes. GWAS pinpointed SNPs that influence various tumor phenotypes. Clinical value of a mouse tumor susceptibility gene signature (mTSGS) was evaluated using public datasets, encompassing TCGA, METABRIC and I-SPY2 cohorts. The predictive power of mTSGS for response to chemotherapy was validated *in vivo* using genetically diverse MMTV-*ErbB2* mice.

Results Distinct variances in tumor onset, multiplicity, and metastatic patterns were observed across CC strains. Besides lung metastasis, liver and kidney metastases emerged in specific CC strains. GWAS identified 1525 SNPs, 800 SNPs, 568 SNPs, and 23 SNPs significantly associated with tumor onset, multiplicity, lung metastasis, and liver metastasis, respectively. Multivariate analyses flagged SNPs in 20 genes independently tied to various tumor characteristics, designated as mTSGS. These 20 genes were transcriptionally altered in human BCs. We then established mTSGS scores (mTSGSS) based on their transcriptional levels. The mTSGSS showed prognostic values, superseding clinical factors and PAM50 molecular subtype across cohorts. Moreover, mTSGSS predicted pathological complete response (pCR) to six of thirteen treatment regimens, including chemotherapy only, in I-SPY2 study. Importantly, the predictive value of the mTSGSS for pCR stood independent of the MammaPrint score. The power of mTSGSS for predicting chemotherapy response was validated in an *in vivo* MMTV-*ErbB2* model, showing that like findings in human patients, mouse tumors with low mTSGSS were most likely to respond to treatment.

Conclusion Our investigation has unveiled many novel genes predisposing individuals to ERBB2-driven cancer. Translational findings indicate that mTSGSS holds promise as a biomarker to refine treatment strategies for BC patients.

Introduction

Twenty to thirty percent of primary breast cancers (BCs) amplify/overexpress the epidermal growth factor receptor 2 (ERBB2, HER2, or NEU).^{1,2} These ERBB2+ tumors have more aggressive disease and poorer clinical outcome, and are more refractory to radiotherapy, chemotherapy, and hormone therapy.³⁻⁶ Although a humanized anti-ERBB2 monoclonal antibody (Herceptin) and the small molecule inhibitor of ERBB2 (Lapatinib) are effective for treating ERBB2+ BC patients, most ERBB2+ BCs do not respond to either Herceptin or Lapatinib (intrinsic resistance), and the majority of responders become resistant within 12 months of initial therapy (acquired or secondary resistance).⁷⁻¹¹ Therefore, new biological insights into HER2-driven cancers are still needed.

Our previous F1 backcross (F1Bx) study between the resistant C57BL/6J strain and FVB/N has shown a strong genetic effect on ERBB2-initiated tumor development and metastasis.¹² Moreover, our omics analysis of tumors revealed similarities between ERBB2 tumors in humans and those from F1Bx mice at clinical, genomic, expression, and signaling levels.¹² However, an obvious limitation of this F1Bx study is that we only found genetic variants between C57BL/6J and FVB/N and likely missed variants relevant to more diverse human populations. The Collaborative Cross (CC) mouse resource, established from 8-parental recombinant inbred mouse strains, contains uniformly distributed natural variants and a level of genetic diversity on a par with the human population.¹³⁻¹⁵ Moreover, large CC strain-dependent variations in many phenotypes, such as spontaneous tumor development, have been reported.¹⁶⁻²⁹

In this study, we identified host genetic variants that predispose *ErbB2*-driven tumor development and metastasis using the CC mouse resource. Additionally, we systematically evaluated the clinical value for prognosis and therapeutic responses of a mouse mammary tumor susceptibility gene signature in human BC using publicly available cohorts, including clinical trial cohorts. Our findings substantially increase biological insights into ERBB2-driven cancers, which may provide new strategies and define new targets for improving outcomes of ERBB2-targeted therapies.

Methods

CC mice experiments

All CC mice were purchased from the Systems Genetics Core Facility at the University of North Carolina, and FVB-Tg(MMTV-ErbB2)NK1Mul/J (FVB/N MMTV-*ErbB2*) mice were purchased from the Jackson Laboratory. F1 hybrid mice were generated by crossing FVB/N MMTV-*ErbB2* female mice with CC male mice from 30 CC strains. The number of female mice used in this study was summarized in Suppl. Table 1. 20 FVB/N MMTV-*ErbB2* female mice served as control. All mice were monitored for mammary tumor development by palpating with a maximum follow-up of 2 years. This study was approved by the Animal Welfare and Research Committee at Lawrence Berkeley National Laboratory.

Chemotherapy experiment in a genetically diverse MMTV-ErbB2 model

Genetically diverse F1 backcross (F1Bx) mice between C57BL/6J and FVB/N MMTV-*ErbB2* transgenic mice were generated as described in our previous study.¹² 50 *ErbB2*-positive F1Bx mice were housed at IBMCC-FICUS's Animal Research Facility and observed twice a week for tumor manifestation. Before starting treatment (when tumor volume reached 500mm³), two biopsies were collected under aseptic conditions, in a flow chamber, and with isoflurane anesthesia. One biopsy was frozen for transcriptional analysis, and the other was fixed for histological analysis. Two weeks after collection of the tumor biopsy, mice underwent chemotherapy consisting of 5 intraperitoneal injections of 25 mg/kg docetaxel with a recovery time of 8 days between injections.

We evaluated the local tumor by assessing changes in the tumor growth. Tumor volume was estimated each week using the formula: Tumor volume = length x width² x 0.5. We quantified tumor volume changes and growth rate. We calculated the tumor growth rate by estimating the linear regression slope of the logarithm of tumor volume in mm³ onto time in days. We defined (a) complete response (nonpalpable mammary tumor); (b) partial response (tumor volume significantly reduced at the end of treatment in comparison to the volume at the beginning of treatment); (c) tumor stabilization (no change in tumor volume during treatment in comparison to the volume at beginning of treatment); and (d) early resistance (increase in tumor volume during treatment in comparison to the volume at beginning of treatment) to therapy. All mice were housed at the Animal Research Facility of the University of Salamanca for mouse chemotherapy. All the procedures were approved by the Institutional Animal Care and Bioethical Committee of the University of Salamanca.

Genome-wide Association Study (GWAS)

GWAS analysis has been described previously.^{19,26-28} At each SNP, Cox regression was used to assess the significance of associations between tumor onset and allele types; the Mann-Whitney test was used to assess the significance of associations between tumor multiplicity and allele types; while the Chi-square test was used to assess the significance of associations between tumor metastasis (overall, lung, liver, and kidney metastasis) and allele types. Putative candidate genes were defined as those genes containing a significant SNP within the boundaries of the gene sequence (<http://www.informatics.jax.org/>). KEGG pathway enrichment analysis was performed on candidate genes using WebGestalt (<https://www.webgestalt.org/>).³⁰

RNA Extraction from Tumors

The Qiagen miRNeasy Mini Kit-50 was used for RNA extraction, preserving miRNA populations for further studies. The protocol followed was as previously described.¹² Global RNA expression was assessed using Affymetrix chips at the University of Salamanca's Cancer Research Center's Genomics Unit.

Gene Expression Profiling and Analysis

RNA integrity was evaluated using the Agilent 2100 Bioanalyzer. RNA samples (100-300 ng) were labeled and amplified using the Ambion Expression Kit. The Affymetrix GeneChip system was used for washing and scanning procedures. The [MoGene-2_0-st] Affymetrix Mouse Gene 2.0 ST Array platform was employed for expression array studies. Microarray signal data normalization across chips utilized the Robust Multichip Analysis (RMA) algorithm (Affymetrix Expression Console v. 1.4.1) as described in our previous study.¹² The gene expression data for mouse breast tumors is available in the Gene Expression Omnibus (GEO) (GSE252001).

Polygenic risk score (PGS) and mouse tumor susceptibility gene signature score (mTSGSS)

Multivariate Cox regression, multivariate linear regression, and multivariate logistic regression were used on significant SNPs from GWAS for the identification of independent and significant SNPs for tumor onset, tumor multiplicity, and tumor metastasis, respectively. The PGS was then constructed as following the formula:

$$PGS^{phenotype} = \sum_{k=0}^{N^{phenotype}} Coefficient_k^{phenotype} * SNP_k^{phenotype}$$

where $phenotype \in (tumor\ onset, tumor\ multiplicity, tumor\ metastasis)$, $N^{phenotype}$ refers to the number of independent and significant SNPs associated with specific phenotype, $Coefficient_k^{phenotype}$ refers to the coefficient of k^{th} SNP associated with a specific phenotype ($SNP_k^{phenotype}$) derived from multivariate analysis.

The mTSGSS was established in human BC cohorts using the transcriptional levels based on the combination of genes associated with pre-identified SNPs during PGS construction for all three different phenotypes. Specifically, mTSGSS was defined as follows:

$$mTSGSS = \sum_{k=0}^N Coefficient_k * Gene\ Expression_k$$

where $Coefficient_k$ is the coefficient of k^{th} gene derived from Cox regression analysis in prognosis study and from logistic regression analysis in drug response study. The risk groups (i.e., low, intermediate, and high) of PGS and mTSGSS were defined as the tertiles (top, middle, and bottom) of PGS and mTSGSS, respectively.

Human public cohorts

The Cancer Genome Atlas (TCGA) (TCGA-BRCA) and METABRIC breast cancer transcriptome and clinical data including PAM50-based molecular subtypes³¹ were downloaded from the cBioPortal (<https://www.cbioportal.org/>).^{32,33} The GSE96058 and I-SPY2 (GSE194040) cohorts were downloaded from Gene Expression Omnibus (GEO) database. The list of genes for human BCs identified in human GWAS was downloaded from the GWAS Catalog database (<https://www.ebi.ac.uk/gwas/search?query=rs6928864>).³⁴ There was no additional modification in the downloaded data during our analyses.

Statistical analysis

TNMplot (<https://tnmplot.com/analysis/>) was used to compare gene transcriptional expression in normal and BC tissues based on RNA-seq data.³⁵ The difference in overall survival (OS) was assessed by Kaplan-Meier analysis (survminer package in R, version 0.4.8) and log-rank test (survival package in R, version 3.2-3). The p value < 0.05 was taken as statistically significant. All

data analysis was performed, and plots were generated using R software (version 3.5.0) or IBM SPSS (version 24).

Results

Variation in mammary tumor onset, multiplicity, and metastasis across CC strains

A total of 732 female F1 hybrid mice were generated from a cross between FVB/N MMTV-*ErbB2* and 30 CC strains and monitored for tumor development over two years. We observed large differences in mammary tumor onset and multiplicity (number of tumors per mouse) across CC strains (Fig 1A and 1B; Data Supplement, Table S1). The median age at tumor onset ranged from 166 to 497 days (Fig 1A; Data Supplement, Table S1). F1 hybrid MMTV-*ErbB2* mice from CC001, CC007, CC013, CC015, CC019, CC021, CC30, CC033, CC036, and CC42 strains had similar onset, while F1 hybrid MMTV-*ErbB2* mice from the remaining CC strains had significantly later onset in comparison to FVB/N MMTV-*ErbB2* mice (Fig 1A; Data Supplement, Table S1). Interestingly, about 20% of CC038 F1 mice did not develop any tumors within the two-year follow-up (Fig 1A, left panel; Data Supplement, Table S1). F1 hybrid mice from CC001, CC007, CC013, and CC042 strains developed significantly more tumors, while F1 hybrid mice from CC038, CC080, CC051, and CC012 strains developed significantly less tumors than FVB/N MMTV-*ErbB2* mice (Fig 1B; Data Supplement, Table S1). Additionally, we observed large variation in metastatic incidence across CC strains. Although the most frequent metastatic site was the lungs in all strains, we observed an increased frequency of liver metastasis in the CC024 and CC037 strains and an increased frequency of kidney metastasis in the CC013 strain (Fig 1C; Data Supplement, Fig S1, Table S1). We also found that mice with younger age onset developed significantly more tumors in comparison to those with older age onset ($p < 0.0001$, Fig 1D). Moreover, we found that mice with younger age onset also developed significantly more metastatic tumors ($p = 0.01$, Fig 1E). These findings indicate that host genetics significantly influences *ErbB2*-driven tumor development and progression.

Genetic determinants of mammary tumor onset, multiplicity, and metastasis across CC strains

To investigate the contribution of genetic variants to mammary tumor onset, tumor multiplicities, and metastasis, GWAS analysis was performed with 70,273 SNPs across 30 CC strains. We

identified 1,525 SNPs significantly associated with tumor onset ($p < 1.00E-30$) corresponding to 275 known genes (Fig 2A; Data Supplement, Fig S2A, Tables S2 and S3), 800 SNPs significantly associated with the number of tumors ($p < 1.00E-15$) corresponding to 194 known genes (Fig 2B; Data Supplement, Fig S2B, Tables S2 and S3), 588 SNPs significantly associated with overall tumor metastasis ($p < 1.00E-4$) corresponding to 171 known genes (Fig 2C; Data Supplement Fig 2C, Tables S2 and S3), 568 SNPs significantly associated with lung metastasis ($p < 1.00E-4$) corresponding to 168 known genes (Fig 2D, left panel; Data Supplement, Tables S2 and S3) and 23 SNPs significantly associated with liver metastasis ($p < 1.00E-4$) corresponding to 12 known genes (Fig 2E, middle panel; Data Supplement, Tables S2 and S3). We did not find any SNPs significantly associated with kidney metastasis (Fig 2F, right panel; Data Supplement, Table S2). Given the limited number of mice with kidney metastasis, our study may have lacked the statistical power to detect SNPs.

To elucidate the mechanisms underlying tumor susceptibility, we used WebGestalt to evaluate functional enrichment analysis of candidate susceptibility genes for each phenotype using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. For tumor onset, the genes were predominantly enriched in pathways such as Ras ($p = 0.0033$), Hedgehog signaling ($p = 0.0061$), and ECM-receptor interaction ($p = 0.0091$), among others (Data Supplement, Fig S3A). In the context of tumor multiplicity, there was significant enrichment in pathways like ECM-receptor interaction ($p = 0.0087$) and transcriptional misregulation in cancer ($p = 0.010$) (Data Supplement, Fig S3B). For metastasis, pathways such as gap junction ($p = 0.0078$) and regulation of lipolysis in adipocytes ($p = 0.0013$) were predominantly represented (Data Supplement, Fig S3C).

Establishment of polygenic risk scores for tumor onset, multiplicity, and metastasis

We used multivariate analysis to determine the most critical SNPs for each phenotype. Multivariate Cox regression analysis identified SNPs in 8 genes (*Stx6*, *Ramp1*, *Traf3ip1*, *Nckap5*, *Pfkfb2*, *Trmt11*, *Rprd1b*, and *Rer1*) that were independently associated with age of tumor onset (Data Supplement, Fig S4A). The risk score of SNPs in these 8 genes was significantly associated with age of tumor onset (Fig 3A). Multivariate linear regression analysis identified SNPs in 11 genes (*Sepsecs*, *Rhobtb1*, *Tsen15*, *Abcc3*, *Arid5b*, *Tnr*, *Dock2*, *Tti1*, *Fam81a*, *Stx6*, and *Oxr1*) that were independently associated with the tumor multiplicities (Data Supplement, Fig S4B). The risk

score of SNPs in these 11 genes was significantly associated with the number of tumors (Fig 3B). Multivariate logistic regression analysis identified SNPs in 2 genes (*Plxna2* and *Tbc1d31*) that were independently associated with tumor metastasis (Data Supplement, Fig S4C). The risk score of SNPs in these 2 genes was significantly associated with tumor metastasis (Fig 3C). We pooled all 20 genes (*Stx6*, *Ramp1*, *Traf3ip1*, *Nckap5*, *Pfkfb2*, *Trmt1l*, *Rprd1b*, *Rer1*, *Sepsecs*, *Rhobtb1*, *Tsen15*, *Abcc3*, *Arid5b*, *Tnr*, *Dock2*, *Tti1*, *Fam81a*, *Oxr1*, *Plxna2* and *Tbc1d31*) together as the mouse tumor susceptibility gene signature (mTSGS).

mTSGS score (mTSGSS) is significantly associated with prognosis of human BC

To evaluate the impact of mTSGS on human breast cancer (BC), we first used TNMplot to examine their transcriptional expression in BC and found that all genes transcriptionally altered. The expression levels of *ABCC3*, *ARD5B*, *OXR1*, *PLXNA2*, *RHOBTB1* and *SEPSECS* gene were significantly reduced while the expression levels of the remaining genes were significantly elevated in comparison to the normal mammary tissues (Data Supplement, Fig S5). We established a mTSGS score (mTSGSS) based on their transcriptional levels (details in the method). We found that mTSGSS was significantly associated with different clinical outcomes, such as overall survival (OS), disease-free survival (DFS), and progression-free survival (PFS) in multiple datasets (Fig 5A-5C; Data Supplement, Tables S4-6). Patients with low mTSGSS have a favorable prognosis (Fig 5A-5C). Moreover, the prognostic impact of mTSGSS is independent of clinical factors (such as age, ER, and PR) and PAM50 molecular subtypes (Fig 5D-5F).

mTSGSS predicts responses to different treatments in human breast cancer

Using the I-SPY2 datasets, we discovered that mTSGSS is significantly correlated with pathological complete response (pCR) to different treatment regimens (Fig 5A; Data Supplement, Table S7). Overall, patients with low mTSGSS have a higher pCR rate in comparison to those with high mTSGSS for 6 of 13 treatment regimens (Fig 5A).

Taxanes are still highly active chemotherapy agents used in metastatic BC. To evaluate the predictive value of mTSGSS for the responses to taxane, 50 genetically diverse F1Bx MMTV-*ErbB2* mice were treated with docetaxel when the tumor volume reached 500mm³, and the treatment responses for each mouse was assessed (detail see method and material section). As we found in human studies, mTSGSS generated from the transcript levels measured in the pre-

treatment biopsy was able to predict the response to docetaxel treatment, and the tumors with low mTSGSS were more likely to respond to docetaxel treatment (Fig 5B).

Finally, multivariate logistic regression analysis showed that the predictive value of mTSGSS in pCR is independent of the MammaPrint (MP) score (Fig 6). These findings indicate that the mTSGSS is equal or better than MP in all treatment groups except those containing pembrolizumab.

Discussion

Even with substantial progress in ERBB2-targeted therapies, resistance - whether acquired or intrinsic - remains a formidable challenge. This resistance is thought to arise from a range of mechanisms, including the activation of alternative signaling pathways, ERBB2 gene mutations, and tumor heterogeneity.^{36,37} Given these challenges, identifying patients who stand to benefit the most from a particular treatment is imperative, as this enhances therapeutic efficacy and reduces potential toxicity. Therefore, a better understanding of the biology of ERBB2-driven cancer supports the development of new treatment options for patients. The goal of this study was to identify genetic factors that control *ErbB2*-driven mammary tumor development and metastasis using a large cohort of genetically diverse CC mice. Our findings demonstrate that there is a large strain-dependent variation in *ErbB2*-initiated tumorigenic phenotypes, and analysis of such variability in CC mice can reveal the underlying genetic basis in human BCs.

This study confirmed the significance of many loci that were previously identified using the F1 backcross approach,¹² but as expected, because of the increased genetic divergence in the CC mice, we identified many additional genetic loci that were strongly associated with tumorigenic phenotypes. Using all tumor phenotypes, we discovered a total of 551 candidate genes, human orthologs for which are shown in Table S8. Twenty-three of these genes (*RTKN2*, *PHF20*, *CPEB3*, *BCL2*, *NIPSNAP1*, *TENM2*, *PBX1*, *ITPR2*, *WWOX*, *HORMAD2*, *DNM3*, *PTPRN2*, *PRKG1*, *IQCA1*, *GPR161*, *SORCS3*, *PCMI*, *EBF2*, *JMJD1C*, *TGFBR2*, *SLC39A11*, *SEC14L4*, and *NYAP2*) have been found by human GWAS for BCs based on the GWAS Catalog database.³⁴ There are 87 overlapping susceptibility genes between tumor onset and multiplicity, but only seven overlapping susceptibility genes between overall tumor metastasis and onset and only five overlapping susceptibility genes between overall tumor metastasis and multiplicity (Data Supplement, Fig S6). Only two susceptibility genes (*Nckap5* and *Ptprt*) overlap among all tumor phenotypes (onset,

multiplicity, and metastasis) (Data Supplement, Fig S6). *PTPRT*, a member of the protein tyrosine phosphatase (PTP) family, has been reported to be a tumor suppressor gene in BC and other cancers.³⁸⁻⁴⁴ *NCKAP5*, potentially functioning in microtubule bundle formation and microtubule depolymerization, is less studied, and polymorphisms in this gene are reported to be associated with the clinical outcome of gastric cancer patients in a recent study.⁴⁵ Overall, our study suggests different genetic factors controlling tumor onset, multiplicity, metastasis, as well as the site of metastasis.

In the contemporary landscape of personalized medicine, the identification of biomarkers capable of forecasting treatment responses is of paramount importance. These predictive markers further tailor therapeutic interventions, circumventing unneeded drug exposure in patients unlikely to experience clinical advantages. As our comprehension of the molecular underpinnings of ERBB2-positive breast cancer expands, novel avenues will open for treatments that are even more patient-specific. Of note, genomic tests, especially those centered on gene expression signatures, are becoming increasingly prominent.^{46,47} In this study, we identified a mouse tumor susceptibility gene signature (mTSGS) comprised of 20 genes (*Stx6*, *Ramp1*, *Traf3ip1*, *Nckap5*, *Pfkfb2*, *Trmt11*, *Rprd1b*, *Rer1*, *Sepsecs*, *Rhobtb1*, *Tsen15*, *Abcc3*, *Arid5b*, *Tnr*, *Dock2*, *Tti1*, *Fam81a*, *Oxr1*, *Plxna2*, and *Tbc1d31*), and showed that transcriptional expression of mTSGS in human BC can be used to predict prognosis and response to different cancer treatments in patients. Moreover, we demonstrated that our signature stands as a prognostic indicator, distinct from other recognized signatures like the PAM50 molecular subtype³¹ and MammaPrint⁴⁸. The integration of our signature with different BC treatment regimens might enhance the precision of adjuvant treatment decisions for BC patients. Our study further indicates that the CC mouse model can serve as an invaluable pre-clinical model for genetic understanding of drug resistance.

In conclusion, we have identified many novel susceptible genes for ERBB2-driven cancer. Translational studies indicate that mTSGS may serve as a biomarker for tailoring treatment to BC patients.

Acknowledgements

The authors thank the staff in the animal facility for their skillful help with animal maintenance and health evaluation.

Authors' contributions

JHM, HC, and JPL conceived and designed the overall study; JHM and JPL acquired funding; HY, HC and JHM performed experimental study in a large cohort of CC mice; LH, XW and JLI made some contributions to CC study; MJPB, AJN and JPL performed mouse therapy study; PW, HC and JHM performed data analysis; JLI, PW, AMS, HC, JPL and JHM were involved in interpretation of results; JHM, JPL and HC wrote the manuscript; DW and AB provided suggestions and made substantial manuscript editing; all authors have read and edited the manuscript.

Funding

This work was supported by the Department of Defense (DoD) BCRP, No. BC190820. Lawrence Berkeley National Laboratory (LBNL) is a multi-program national laboratory operated by the University of California for the DOE under contract DE AC02-05CH11231. JPL's lab is sponsored by Grant PID2020-118527RB-I00 funded by MCIN/AEI/10.13039/501100011039; Grant PDC2021-121735-I00 funded by MCIN/AEI/10.13039/501100011039 and by the "European Union Next Generation EU/PRTR," the Regional Government of Castile and León (CSI144P20).

Availability of data and materials

The datasets generated in this study are publicly available.

Declarations

Ethics approval and consent to participate

Animal experiments were reviewed and approved by the Animal Welfare and Research Committee at Lawrence Berkeley National Laboratory or by the Institutional Animal Care and Bioethical Committee of the University of Salamanca.

Consent for publication

The authors consent to publish this work.

Competing interests

The authors declare no competing interests.

References

1. Slamon DJ, Clark GM, Wong SG, et al: Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235:177-82, 1987
2. Slamon DJ, Godolphin W, Jones LA, et al: Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 244:707-12, 1989
3. Jukkola A, Bloigu R, Soini Y, et al: c-erbB-2 positivity is a factor for poor prognosis in breast cancer and poor response to hormonal or chemotherapy treatment in advanced disease. *Eur J Cancer* 37:347-54, 2001
4. Kallioniemi OP, Holli K, Visakorpi T, et al: Association of c-erbB-2 protein over-expression with high rate of cell proliferation, increased risk of visceral metastasis and poor long-term survival in breast cancer. *Int J Cancer* 49:650-5, 1991
5. Muss HB, Thor AD, Berry DA, et al: c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *N Engl J Med* 330:1260-6, 1994
6. Ross JS, Fletcher JA: The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy. *Stem Cells* 16:413-28, 1998
7. Callahan R, Hurvitz S: Human epidermal growth factor receptor-2-positive breast cancer: Current management of early, advanced, and recurrent disease. *Curr Opin Obstet Gynecol* 23:37-43, 2011
8. Dean-Colomb W, Esteva FJ: Her2-positive breast cancer: herceptin and beyond. *Eur J Cancer* 44:2806-12, 2008
9. Gajria D, Chandralapaty S: HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. *Expert Rev Anticancer Ther* 11:263-75, 2011
10. Garrett JT, Arteaga CL: Resistance to HER2-directed antibodies and tyrosine kinase inhibitors: mechanisms and clinical implications. *Cancer Biol Ther* 11:793-800, 2011
11. Geuna E, Milani A, Redana S, et al: Hitting multiple targets in HER2-positive breast cancer: proof of principle or therapeutic opportunity? *Expert Opin Pharmacother* 12:549-65, 2011
12. Castellanos-Martin A, Castillo-Lluva S, Saez-Freire Mdel M, et al: Unraveling heterogeneous susceptibility and the evolution of breast cancer using a systems biology approach. *Genome Biol* 16:40, 2015
13. Churchill GA, Airey DC, Allayee H, et al: The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* 36:1133-7, 2004
14. Threadgill DW, Hunter KW, Williams RW: Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm Genome* 13:175-8, 2002
15. Zou F, Gelfond JA, Airey DC, et al: Quantitative trait locus analysis using recombinant inbred intercrosses: theoretical and empirical considerations. *Genetics* 170:1299-311, 2005

16. Dorman A, Binenbaum I, Abu-Toamih Atamni HJ, et al: Genetic mapping of novel modifiers for Apc(Min) induced intestinal polyps' development using the genetic architecture power of the collaborative cross mice. *BMC Genomics* 22:566, 2021
17. Lawley KS, Rech RR, Elenwa F, et al: Host genetic diversity drives variable central nervous system lesion distribution in chronic phase of Theiler's Murine Encephalomyelitis Virus (TMEV) infection. *PLoS One* 16:e0256370, 2021
18. Luo YS, Cichocki JA, Hsieh NH, et al: Using Collaborative Cross Mouse Population to Fill Data Gaps in Risk Assessment: A Case Study of Population-Based Analysis of Toxicokinetics and Kidney Toxicodynamics of Tetrachloroethylene. *Environ Health Perspect* 127:67011, 2019
19. Mao JH, Kim YM, Zhou YX, et al: Genetic and metabolic links between the murine microbiome and memory. *Microbiome* 8:53, 2020
20. Milhem A, Abu Toamih-Atamni HJ, Karkar L, et al: Studying host genetic background effects on multimorbidity of intestinal cancer development, type 2 diabetes and obesity in response to oral bacterial infection and high-fat diet using the collaborative cross (CC) lines. *Animal Model Exp Med* 4:27-39, 2021
21. Mosedale M, Cai Y, Eaddy JS, et al: Human-relevant mechanisms and risk factors for TAK-875-Induced liver injury identified via a gene pathway-based approach in Collaborative Cross mice. *Toxicology* 461:152902, 2021
22. Noll KE, Ferris MT, Heise MT: The Collaborative Cross: A Systems Genetics Resource for Studying Host-Pathogen Interactions. *Cell Host Microbe* 25:484-498, 2019
23. Wang P, Wang Y, Langley SA, et al: Diverse tumour susceptibility in Collaborative Cross mice: identification of a new mouse model for human gastric tumourigenesis. *Gut* 68:1942-1952, 2019
24. Zeiss CJ, Gatti DM, Toro-Salazar O, et al: Doxorubicin-Induced Cardiotoxicity in Collaborative Cross (CC) Mice Recapitulates Individual Cardiotoxicity in Humans. *G3 (Bethesda)* 9:2637-2646, 2019
25. He L, Wang P, Schick SF, et al: Genetic background influences the effect of thirdhand smoke exposure on anxiety and memory in Collaborative Cross mice. *Sci Rep* 11:13285, 2021
26. Jin X, Zhang Y, Celniker SE, et al: Gut microbiome partially mediates and coordinates the effects of genetics on anxiety-like behavior in Collaborative Cross mice. *Sci Rep* 11:270, 2021
27. Snijders AM, Langley SA, Kim YM, et al: Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. *Nat Microbiol* 2:16221, 2016
28. Zhong C, He L, Lee SY, et al: Host genetics and gut microbiota cooperatively contribute to azoxymethane-induced acute toxicity in Collaborative Cross mice. *Arch Toxicol* 95:949-958, 2021
29. Noll KE, Whitmore AC, West A, et al: Complex Genetic Architecture Underlies Regulation of Influenza-A-Virus-Specific Antibody Responses in the Collaborative Cross. *Cell Rep* 31:107587, 2020
30. Wang J, Vasaikar S, Shi Z, et al: WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* 45:W130-W137, 2017
31. Parker JS, Mullins M, Cheang MC, et al: Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27:1160-7, 2009

32. Cerami E, Gao J, Dogrusoz U, et al: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2:401-4, 2012
33. Gao J, Aksoy BA, Dogrusoz U, et al: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6:p11, 2013
34. Sollis E, Mosaku A, Abid A, et al: The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* 51:D977-D985, 2023
35. Bartha A, Gyorffy B: TNMplot.com: A Web Tool for the Comparison of Gene Expression in Normal, Tumor and Metastatic Tissues. *Int J Mol Sci* 22, 2021
36. Hanker AB, Kaklamani V, Arteaga CL: Challenges for the Clinical Development of PI3K Inhibitors: Strategies to Improve Their Impact in Solid Tumors. *Cancer Discov* 9:482-491, 2019
37. Loibl S, Gianni L: HER2-positive breast cancer. *Lancet* 389:2415-2429, 2017
38. Chen C, Liu H, Xu Q, et al: Association of PTPRT Mutations with Cancer Metastasis in Multiple Cancer Types. *Biomed Res Int* 2022:9386477, 2022
39. Hsu HC, Lapke N, Chen SJ, et al: PTPRT and PTPRD Deleterious Mutations and Deletion Predict Bevacizumab Resistance in Metastatic Colorectal Cancer Patients. *Cancers (Basel)* 10, 2018
40. Lee JW, Jeong EG, Lee SH, et al: Mutational analysis of PTPRT phosphatase domains in common human cancers. *APMIS* 115:47-51, 2007
41. Li L, Xu F, Xie P, et al: PTPRT Could Be a Treatment Predictive and Prognostic Biomarker for Breast Cancer. *Biomed Res Int* 2021:3301402, 2021
42. Lui VW, Peyser ND, Ng PK, et al: Frequent mutation of receptor protein tyrosine phosphatases provides a mechanism for STAT3 hyperactivation in head and neck cancer. *Proc Natl Acad Sci U S A* 111:1114-9, 2014
43. Scott A, Wang Z: Tumour suppressor function of protein tyrosine phosphatase receptor-T. *Biosci Rep* 31:303-7, 2011
44. Sen M, Kindsfather A, Danilova L, et al: PTPRT epigenetic silencing defines lung cancer with STAT3 activation and can direct STAT3 targeted therapies. *Epigenetics* 15:604-617, 2020
45. Liu ZX, Zhang XL, Zhao Q, et al: Whole-Exome Sequencing Among Chinese Patients With Hereditary Diffuse Gastric Cancer. *JAMA Netw Open* 5:e2245836, 2022
46. Curtis C, Shah SP, Chin SF, et al: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486:346-52, 2012
47. Sestak I, Cuzick J, Dowsett M, et al: Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol* 33:916-22, 2015
48. Cardoso F, van't Veer LJ, Bogaerts J, et al: 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* 375:717-29, 2016

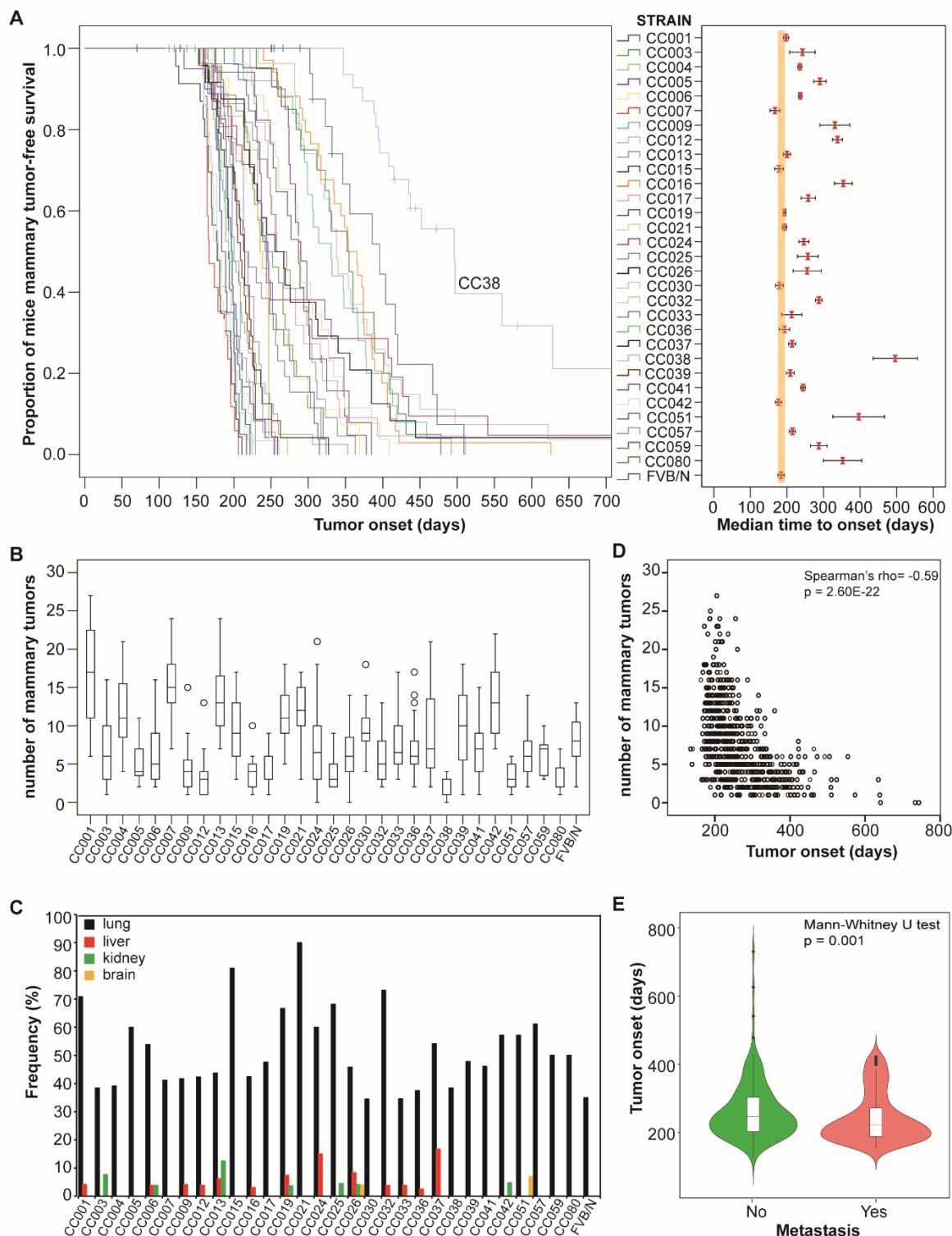


Figure 1. Variations in *Erbb2*-initiated tumor phenotypes across 30 Collaborative Cross (CC) strains. (A) Tumor onset. Left panel: The Kaplan-Meier curve for tumor-free survival in F1 hybrids between each CC strain and FVB/N MMTV-*Erbb2* mice. Right Panel: median time to tumor onset in F1 hybrids between each CC strain and FVB/N MMTV-*Erbb2* mice. The bars show the 95% confidence interval for median time. (B) Multiplicities. Box plot for number of tumors

in each CC strain. The low edge of the box represents the lower quartile, while the upper edge of the box represents the upper quartile. The open circles on the diagram show the outliers. **(C)** Frequencies of metastasis in different sites. **(D)** Correlation between tumor onset and multiplicities. **(E)** Correlation between tumor onset and metastasis.

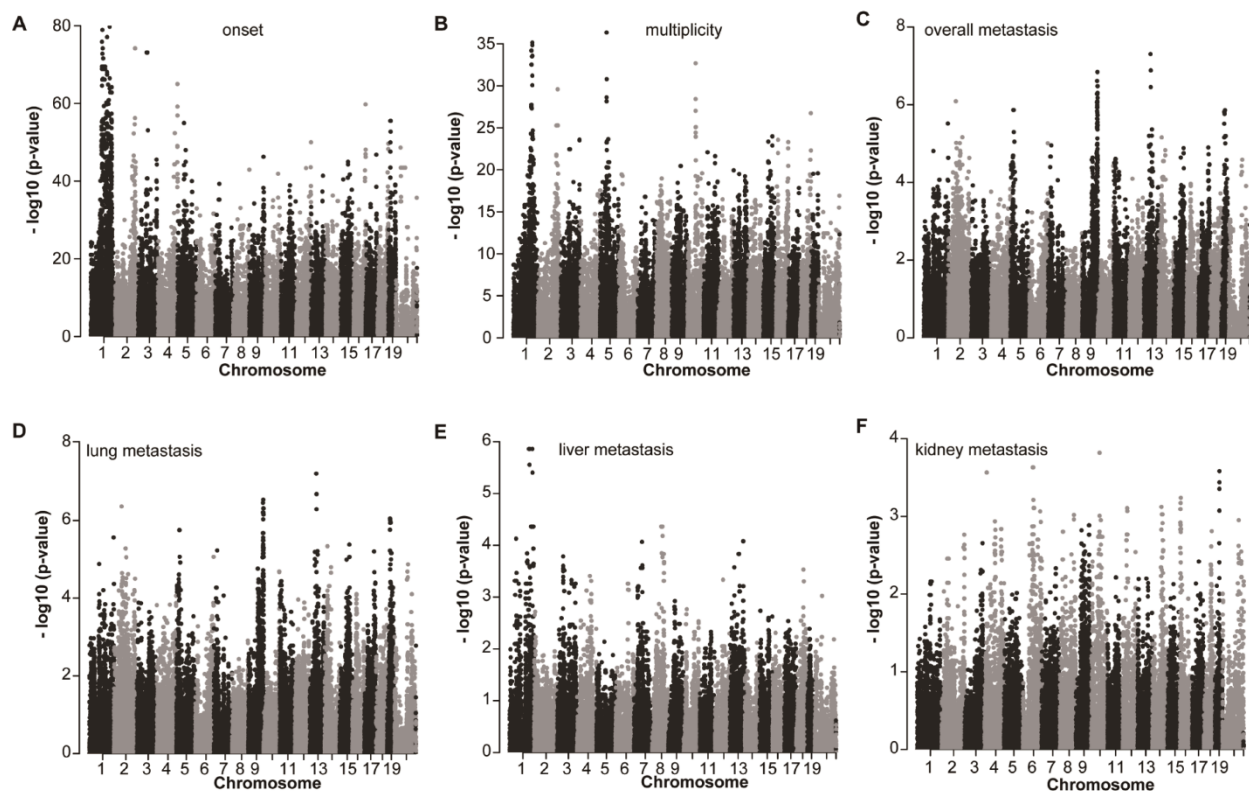


Figure 2. Genome-wide association study of ErBB2-driven tumor phenotypes. The Manhattan plot for **(A)** tumor onset, which was assessed by the Kaplan-Meier method and log-rank test; **(B)** tumor multiplicities, which was assessed by the Mann-Whitney test; **(C)** overall metastasis (any metastasis in any sites); **(D)** lung metastasis; **(E)** liver metastasis; and **(F)** kidney metastasis. All metastatic phenotypes were assessed by the Chi-square test.

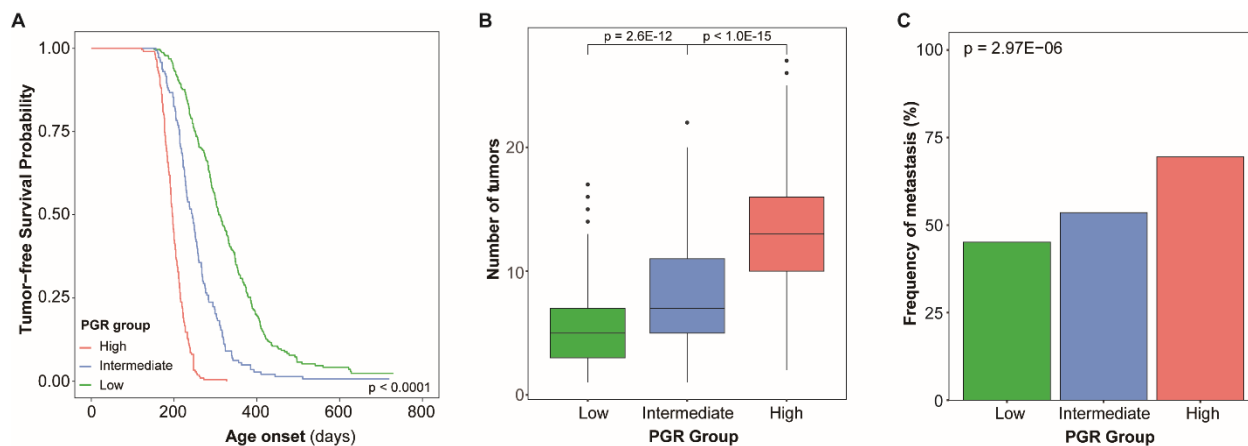


Figure 3. Polygenic risk score for ErbB2-driven tumor phenotypes. (A) The polygenic risk score for tumor onset. The Kaplan-Meier curve for tumor-free survival among different polygenic risk groups. The p-value was obtained from the log-rank test. **(B)** The polygenic risk score for tumor multiplicities. Box plot for number of tumors among different polygenic risk groups. The p-value was obtained from the Mann-Whitney test. **(C)** The polygenic risk score for overall metastasis. Frequencies of metastasis among different polygenic risk groups. The p-value was obtained from the Chi-square test.

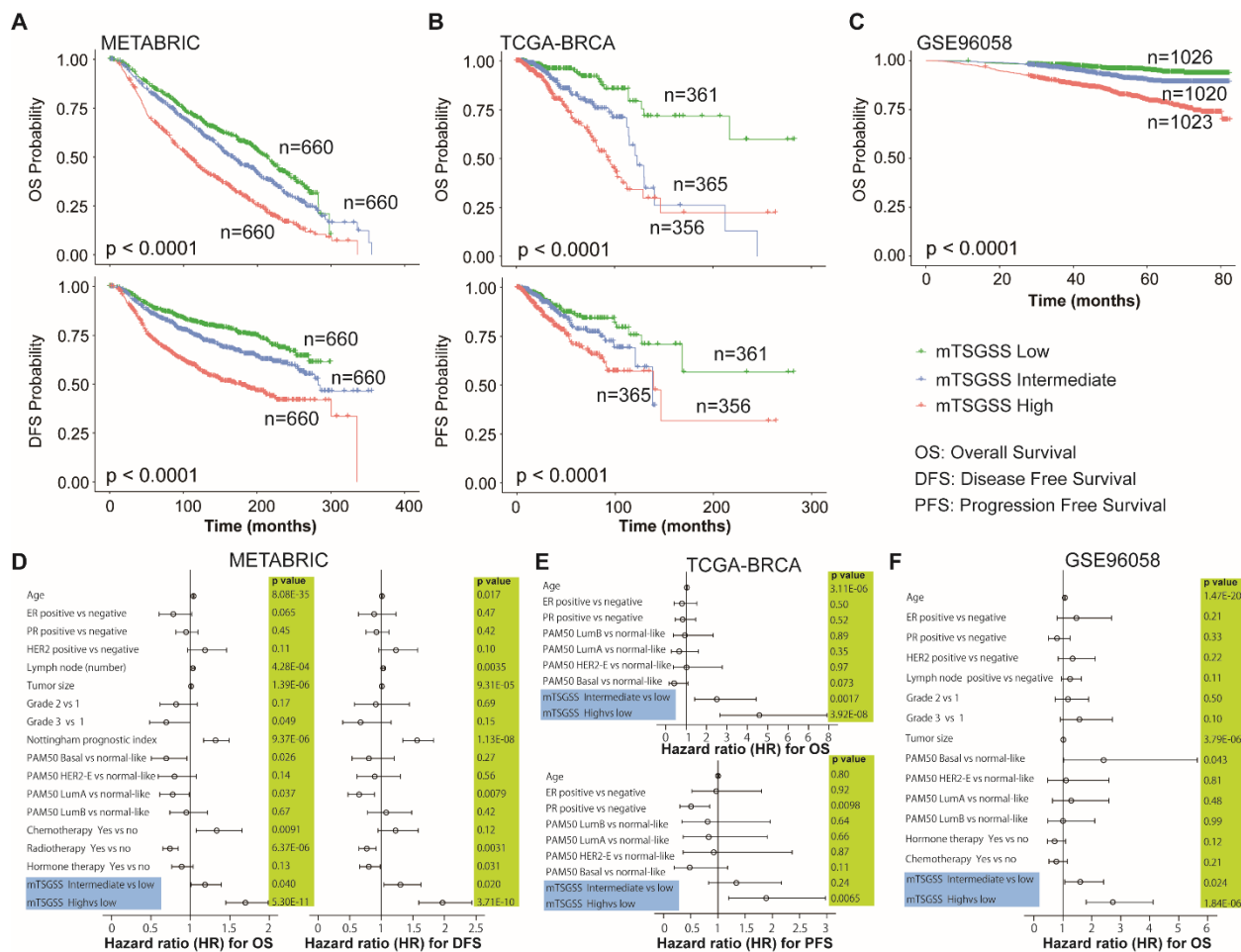


Figure 4. Association of the mouse tumor susceptibility gene signature score (mTSGSS) with prognosis in human breast cancer. mTSGSS was created based on transcriptional expression by multivariate Cox regression. The patients were divided into three groups based on mTSGSS (top, intermediate, and bottom tertile). mTSGSS was a significant and independent prognostic factor. **(A)** Kaplan-Meier survival curves for disease-free (DFS) and overall (OS) survival are presented in METABRIC dataset. **(B)** Kaplan-Meier survival curves for progression-free (PFS) and overall (OS) survival in TCGA-BRCA dataset. **(C)** Kaplan-Meier survival curves for overall survival (OS) in GSE96058 dataset. The P-values shown were obtained from a log-rank test. **(D)** The forest plot shows results of the multivariate Cox regression model for exploring clinical factors, PAM50, and mTSGSS for OS (left panel) and DFS (left panel) in the METABRIC dataset. **(E)** The forest plot shows results of the multivariate Cox regression model for exploring clinical factors, PAM50 and mTSGSS for OS (top panel) and PFS (bottom panel) in the TCGA-BRCA dataset. **(F)** The forest plot shows the results of the multivariate Cox regression model for exploring clinical factors, PAM50 and mTSGSS for OS in the GSE96058 dataset. The bars show the 95% confidence interval for the hazard ratio. The hazard ratios and p-values were obtained from multivariate Cox regression.

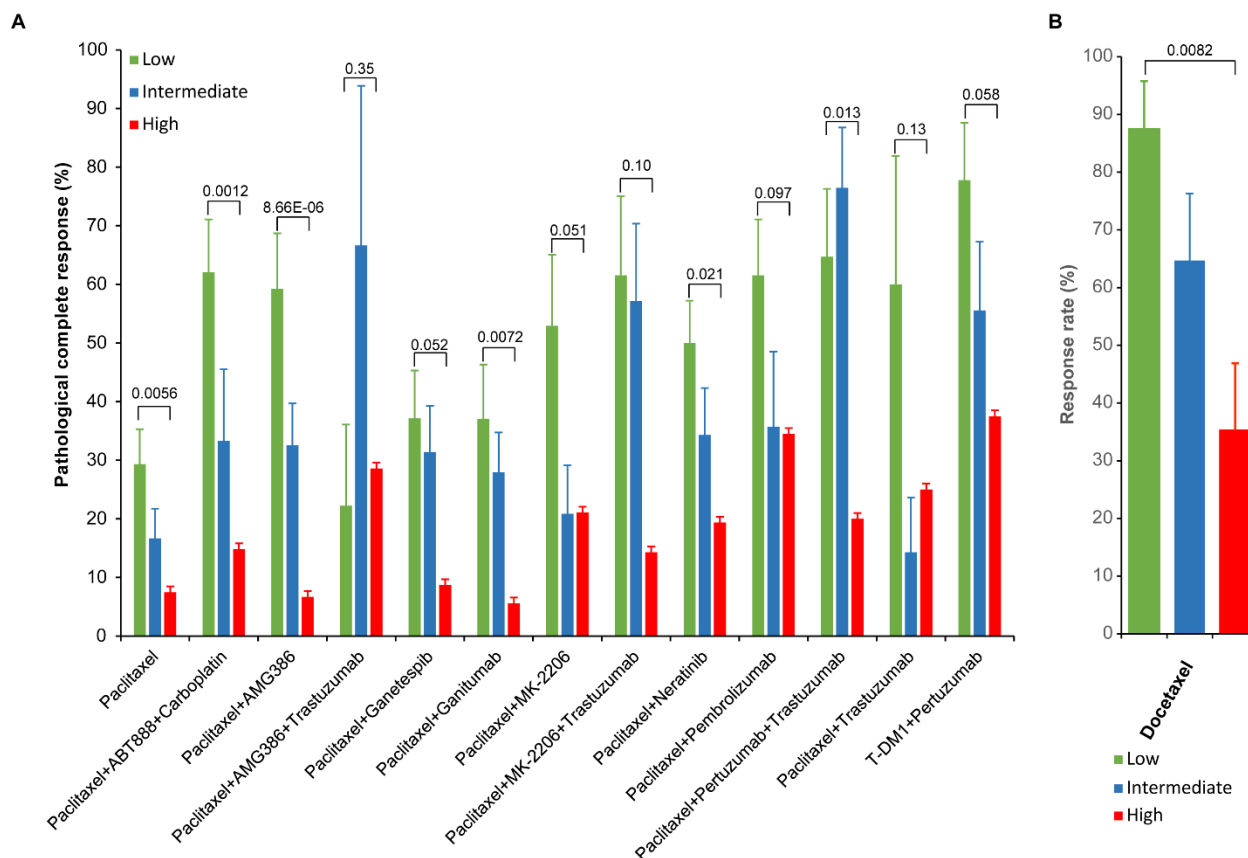


Figure 5. mTSGSS is a predictor for response to different treatment regimens in human breast cancer. (A) mTSGSS significantly correlated with pathological complete response (pCR) to different treatment regimens in the I-SPY2 dataset (GSE194040). (B) mTSGSS significantly correlated with response to docetaxel treatment in genetically diverse F1Bx MMTV-ErbB2 mice. The p-values were obtained from the Chi-square test.

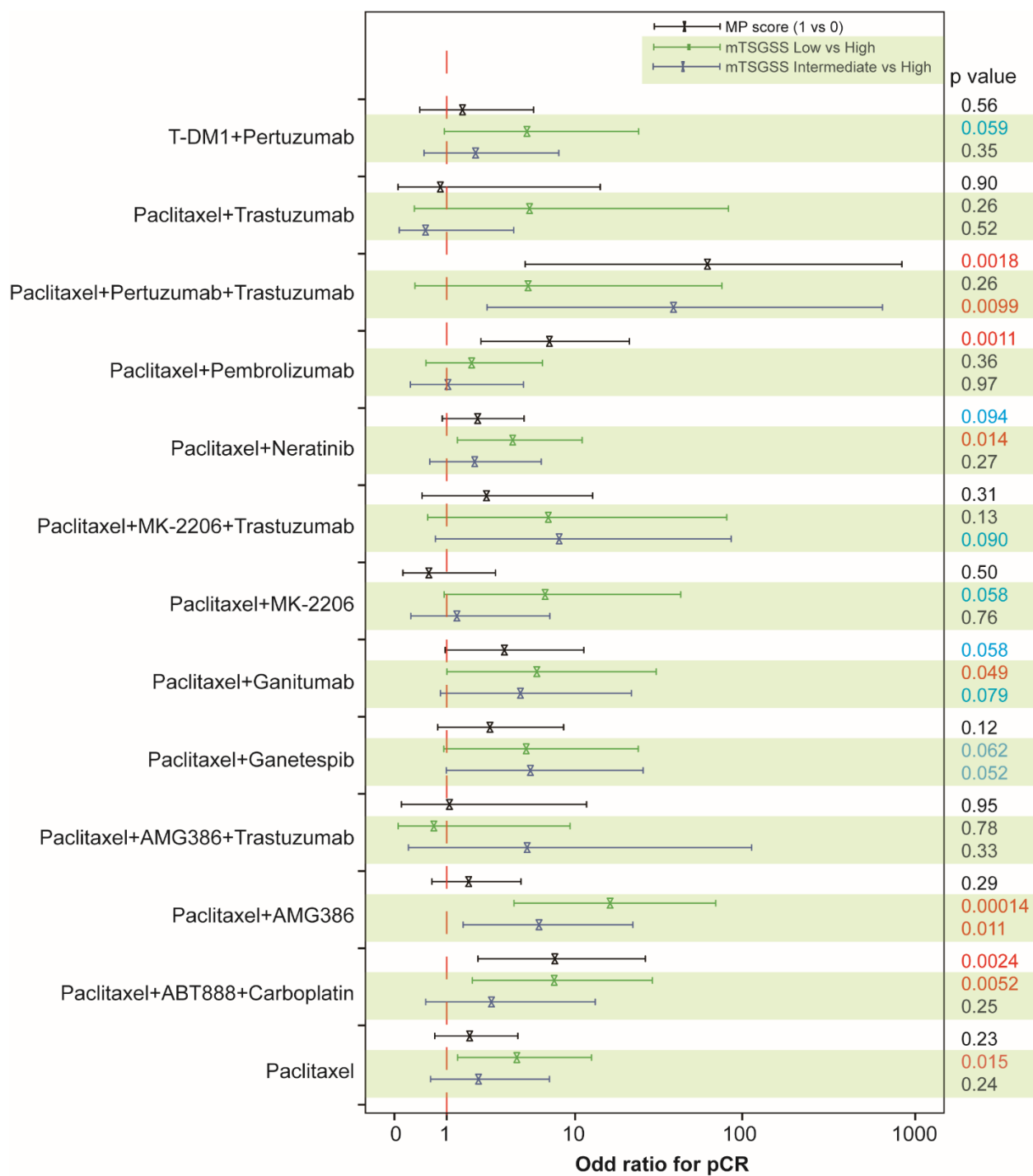


Figure 6. Predictive value of mTSGSS for response to different treatment regimen is independent of MammaPrint in human breast cancer. The forest plot shows results of the multivariate logistic regression model for exploring mTSGSS and MammaPrint for pCR. The bars show the 95% confidence interval for the odd ratio. The odd ratios and p-values were obtained from multivariate logistic regression.

Data Supplement Figures and Tables

Data Supplement Figure S1-S6

Data Supplement Table S1-S8