

## RESEARCH ARTICLE

## The genomic landscape of metastasis in treatment-naïve breast cancer models

Christina Ross<sup>1</sup>, Karol Szczepanek<sup>1</sup>, Maxwell Lee<sup>2</sup>, Howard Yang<sup>2</sup>, Tinghu Qiu<sup>1</sup>, Jack D. Sanford<sup>1</sup>, Kent Hunter<sup>1\*</sup>

**1** Laboratory of Cancer Biology and Genetics, Metastasis Susceptibility Section, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland, United States of America, **2** Laboratory of Cancer Biology and Genetics, High-Dimension Data Analysis Group, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland, United States of America

\* [hunterk@mail.nih.gov](mailto:hunterk@mail.nih.gov)

## OPEN ACCESS

**Citation:** Ross C, Szczepanek K, Lee M, Yang H, Qiu T, Sanford JD, et al. (2020) The genomic landscape of metastasis in treatment-naïve breast cancer models. *PLoS Genet* 16(5): e1008743. <https://doi.org/10.1371/journal.pgen.1008743>

**Editor:** Kornelia Polyak, Dana-Farber Cancer Institute/Harvard Medical School, UNITED STATES

**Received:** November 13, 2019

**Accepted:** March 28, 2020

**Published:** May 28, 2020

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All exome-, whole genome-, and RNA-sequencing data has been uploaded to the NCBI Gene Expression Omnibus under accession number GSE142387.

**Funding:** Funding for this work was provided to KH by the National Cancer Institute Project ZIA BC 011255. This research was also supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Metastasis remains the principle cause of mortality for breast cancer and presents a critical challenge because secondary lesions are often refractory to conventional treatments. While specific genetic alterations are tightly linked to primary tumor development and progression, the role of genetic alteration in the metastatic process is not well-understood. The theory of tumor evolution postulated by Peter Nowell in 1976 has yet to be proven in the context of metastasis. Therefore, in order to investigate how somatic evolution contributes to breast cancer metastasis, we performed exome, whole genome, and RNA sequencing of matched metastatic and primary tumors from pre-clinical mouse models of breast cancer. Here we show that in a treatment-naïve setting, recurrent single nucleotide variants and copy number variation, but not gene fusion events, play key metastasis-driving roles in breast cancer. For instance, we identified recurrent mutations in *Kras*, a known driver of colorectal and lung tumorigenesis that has not been previously implicated in breast cancer metastasis. However, in a set of *in vivo* proof-of-concept experiments we show that the *Kras* G12D mutation is sufficient to significantly promote metastasis using three syngeneic allograft models. The work herein confirms the existence of metastasis-driving mutations and presents a novel framework to identify actionable metastasis-targeted therapies.

## Author summary

The majority of breast cancer-associated deaths are due to metastatic disease, the process where cancerous cells leave the primary tumor and spread to a new location in the body, because metastatic tumors often become insensitive to the same therapies that were successful in treating the primary tumor. To date, this complex process has been attributed to dynamic changes in tumor cell gene expression regulated by epigenetic factors. Interestingly, while genomic alterations are accepted drivers of neoplastic transformation, it is unknown if such events contribute to metastatic spread. One reason for this is the limited availability of matched and treatment-naïve primary tumor and metastatic tumor samples from patients for comparative genomic testing. Here we use two pre-clinical mouse

**Competing interests:** The authors have declared that no competing interests exist.

models of metastatic breast cancer to test if genomic alterations can drive metastatic capacity. We identified metastasis-specific events of single nucleotide variation and gene amplification in well-known oncogenic genes, as well as lesser known factors. We also show that expression of these mutant factors can drive metastasis of weakly and non-metastatic mouse mammary cancer cell lines when implanted in mice. Crucially, by observing and reporting this untested etiology of metastatic disease, specific genomic events can now be included in efforts to develop targets for metastasis-specific therapies.

## Introduction

Metastatic breast cancer remains the leading cause of cancer-related death among women [1,2]. Of the 1.7 million new cases of breast cancer diagnosed annually worldwide, approximately 30% of patients diagnosed with localized disease eventually present with metastatic lesions in distant organs. While non-metastatic breast cancer has a 5-year survival rate of 99%, metastatic disease reduces 5-year survival to only 25% [2]. Therapeutic strategies to treat localized disease, such as molecular profiling and targeted therapy, have been increasingly successful, but patients with disseminated disease continue to face much worse outcomes, as metastases are largely insensitive to such treatments [3,4]. Therefore, to improve outcome for patients with advanced cancer, specific metastasis-targeted strategies will need to be developed, as will a deeper understanding of the unique biological processes that occur during disease progression [4–6].

Despite the importance of this process, relatively little is known regarding somatic genomic events that drive the metastatic cascade. The most commonly accepted hypothesis of tumor progression postulates that mutations are acquired over time, resulting in heterogeneous primary tumor tissue composed of distinct subclones [7]. According to this hypothesis, metastatic capability is induced when a subclone acquires all of the necessary secondary genomic alterations to intravasate into the circulation, survive in circulation, arrest and extravasate at a distant site, and colonize that distant organ [7,8]. However, while the primary tumor genetic heterogeneity predicted by this model is widely accepted, evidence of somatic metastasis-driving mutations is lacking [9–11]. Large-scale projects such as The Cancer Genome Atlas (TCGA) [12] have provided a detailed inventory of oncogenic driver mutations, but no equivalent data set currently exists for metastatic disease. Furthermore, data from our lab and the work of others suggest that metastasis may instead be driven by dynamic epigenetic variation of gene expression programs [13–17] and there is increasing evidence to favor a model of early metastatic dissemination, which is inconsistent with the somatic evolutionary hypothesis [18–20]. However, the difficulty of targeting such dynamic gene expression programs and preventing dissemination of cells from clinically undetectable tumors obliges a deeper assessment of the existence of targetable metastasis-driving mutations. Using genetically engineered mouse models of mammary cancer, we have shown that recurrent single nucleotide variants and copy number variants, but not gene fusion events, occur spontaneously in the absence of therapeutic pressures and drive breast cancer metastasis. Of note, we identified recurrent mutations in several members of the Ras signaling pathway, and as a proof-of-concept experiment show that such somatic events can significantly and specifically promote metastasis *in vivo*.

## Results

### Unique SNVs are enriched in metastases

To complement ongoing human tissue studies and understand the metastatic genomic landscape in a treatment-naïve setting, we performed next-generation sequencing on metastatic

tissue from pre-clinical mouse models of metastatic breast cancer (S1 Fig). We focused on the luminal-like MMTV-PyMT (PyMT) and MMTV-Her2 (Her2) genetically engineered mouse models (GEMMs) of metastatic breast cancer, which model the PI3K activation or the *HER2* amplification seen in 42% or 20% of breast cancer patients, respectively [12,21]. The PyMT model produces palpable primary tumors with a mean latency of 60 days and pulmonary metastases at 100 days in 85% of mice [22]. The Her2 model produces mammary tumors with a mean latency of 100 days and pulmonary metastases at 200 days in 60% of mice [23–25]. We crossed the PyMT model (FVB/NJ background) with five mouse strains (FVB/NJ, C57BL/6J, C57BL/10J, CAST/EiJ, MOLF/EiJ) to more closely recapitulate the genetic heterogeneity of human populations. Due to the latency and high variability of the Her2 (FVB/NJ background) model disease progression compared with PyMT, we only crossed the Her2 model with FVB/NJ for this analysis; however, sequencing of tissues from additional strains is ongoing (S1A Fig).

To investigate whether somatic mutations might contribute to metastatic progression, we performed exome sequencing (exome-seq) of 53 and 12 paired primary tumor (PT) and lung metastases (LM) from the PyMT and Her2 GEMMs, respectively (S1A Fig). Single nucleotide variant (SNV) analysis was performed using three independent SNV-calling algorithms. Genes predicted to harbor SNVs by each algorithm were intersected to identify a common set of potentially mutated genes. We identified 1202 exon SNVs in PT tissue and 1768 in LM tissue compared with normal strain-specific gDNA sequence (S2A and S2B Fig).

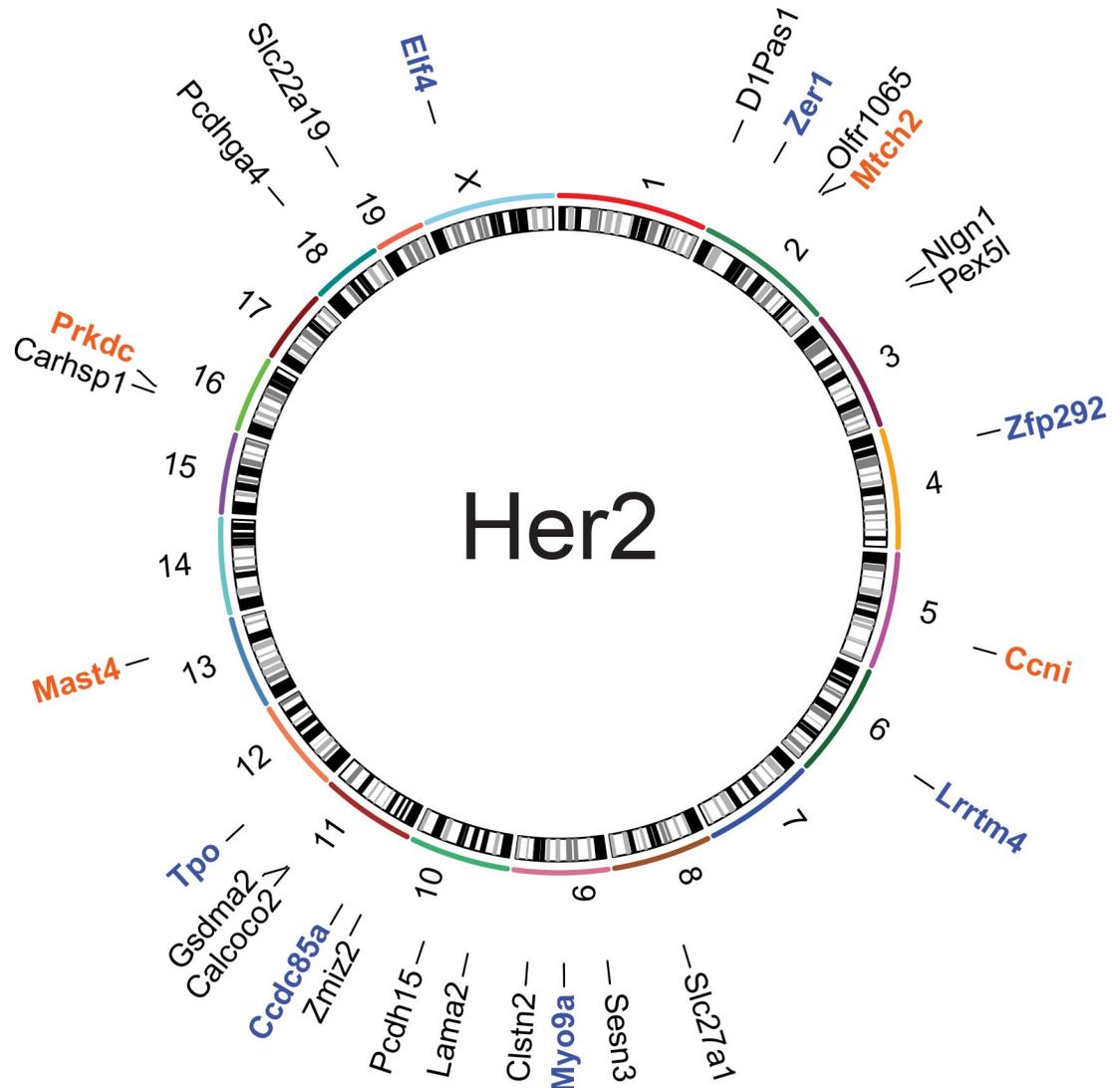
Candidate metastasis-driving SNVs were defined by two criteria: 1) enrichment in metastatic tissue but not in primary tumor and 2) presence within the metastatic seeding cell. To identify SNVs likely present in the metastatic seeding cell, only those SNVs found in at least 60% of the metastatic lesion (variant allele frequency between 0.3 to 0.5) were considered as potential metastasis-driver events. This cut-off accounts for the infiltration of non-tumor cells but still requires a heterozygous mutation to be present in at least 60% of the cells within the lesion. The comparison of LM sequences to matched PT identified 196 SNVs in 164 genes common to all 3 SNV calling algorithms (Figs 1 and 2, S2C Fig and S1 Table). The 164 genes were then screened to identify recurrently mutated codons. Five genes (*Kras*, *Shc1*, *Ccni*, *Mtch2*, *Snrk*) were identified with recurrent mutations in the same codon, and again were common to all three algorithms (S1 Table, Fig 3A–3D).

The SNV calls were then analysed for genes that were recurrently mutated in different codons. A total of 147 genes were identified with mutations observed in only a single animal by all three algorithms. Twelve genes were identified as having been recurrently mutated in different codons in independent animals (S1 Table). Finally, to prioritize genes for future analysis, RNA sequencing (RNA-seq) data was examined to validate SNVs in tissues analyzed by both analyses. An additional instance of the *Shc1* mutated codon was identified, and a potential recurrent mutation in *Rfx8* was found (S1 Table).

Two genes, *Kras* and *Shc1*, had recurring metastasis-specific mutations in independent animals of the PyMT cohort (Fig 3A and 3C). Two PyMT animals carried the oncogenic activating *Kras* G12D mutations (Fig 3A), and an additional three PyMT animals carried different nucleotide substitutions all within the codon of proline 561 or 451 (isoform a or b respectively) of *Shc1*, resulting in P561S, P561T, or P561H (Fig 3C). For those samples with gDNA or RNA still available (4 out of 7), Sanger sequencing validated the presence of *Kras* and *Shc1* SNVs in the original metastatic tissues analyzed by exome- and RNA-seq, and in some cases in additional metastases from the same animal (S2 Table). The presence of recurrent metastasis-enriched SNVs in several animals in the absence of therapeutic pressures suggests that specific coding mutations may play a fundamental role in driving metastasis.

To expand the search for additional genes with recurrent metastasis-enriched mutations, we screened RNA-seq data from a cohort of 42 matched primary tumors and metastatic pairs

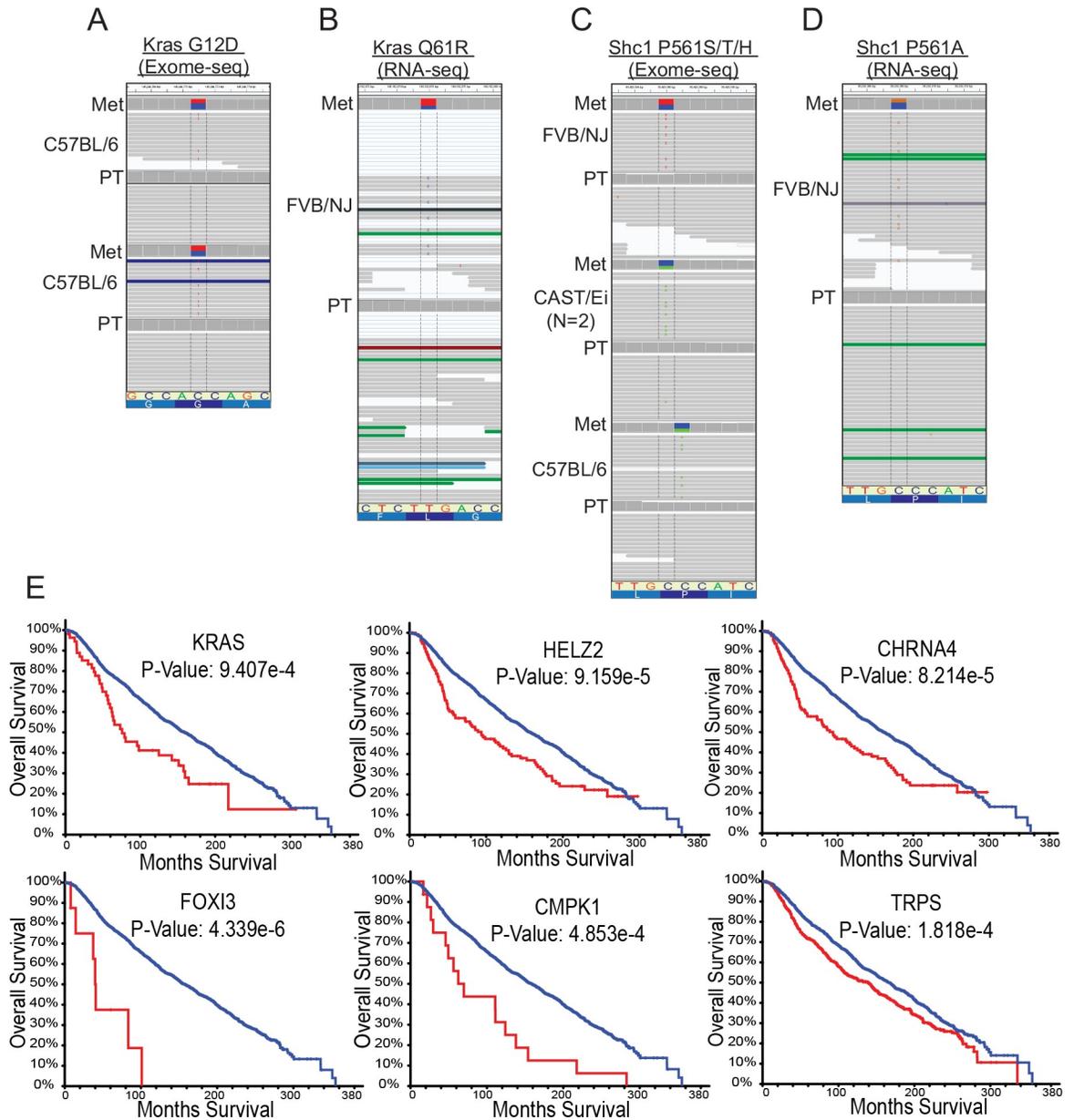




**Fig 2. Metastasis-specific SNVs can be found at high allele frequency in the Her2 model.** B-Circos plots showing names and chromosomal location of genes identified by exome- and RNA-seq with metastasis-specific SNVs (orange: recurrent single strain, blue: recurrent in both strains, and black: singly mutated) from the Her2 model.

<https://doi.org/10.1371/journal.pgen.1008743.g002>

additional animals with mutations in these genes (S1 Table and S2 Table). For animals used in both analyses, the predicted SNVs identified by exome-seq were validated by the RNA-seq data given the gene was transcribed, suggesting that the exome-seq filtering criteria correctly identified bona fide SNVs. One additional metastasis-enriched *Kras* mutation (Q61R; Fig 3B) and an additional mutation in *Shc1* P561 (P561A; Fig 3D) were observed in two animals from the RNA-seq-only cohort. Unexpectedly, in addition to a single metastasis-enriched *Cttnb1* oncogenically-activating SNV (S45P) [26] observed in exome-seq data, RNA-seq data revealed three more independent animals possessing metastasis-specific *Cttnb1* S45F SNVs and another two with *Cttnb1* point mutations that are also often observed in human tumors (K335N, N387K) (S1 Table). Sanger sequencing validated the presence of the SNVs for *Cttnb1* exome-seq S45P and RNA-seq N387K mutations (S2 Table). Insufficient sample material prevented Sanger validation of the remaining RNA-seq-identified SNVs. Overall, for samples



**Fig 3. Metastasis-specific mutations are recurrent and stratify patient outcome.** IGV screen shots showing allele frequency in the blue and red boxes, representative SNV reads, mutated codon letter, tissue type, and mouse strain. **A**, Exome-seq identified a C (blue) to T (red) SNV within *Kras* resulting in the G12D amino acid substitution in metastases from two C57BL/6J x PyMT animals. **B**, RNA-seq identified a T (red) to C (blue) SNV within *Kras* resulting in the Q61R amino acid substitution in a metastatic lesion from one FVB/NJ x PyMT animal. **C**, Exome-seq identified C (blue) to T (red) and C (blue) to A (green) SNVs in metastases from three animals (FVB/NJ, CAST/Ei (2 mets from 1 mouse with same SNV), and C57BL/6J x PyMT) resulting in the P561S/T or H amino acid substitutions, respectively. **D**, RNA-seq identified a C (blue) to G (orange) SNV resulting in the P561A amino acid substitution in one FVB/NJ x PyMT animal. **E**, Kaplan-Meier plots generated using METABRIC for the top six genes identified with metastasis-driver SNVs by exome-seq in mice that most significantly stratify patient survival when altered in primary tumor tissue (blue = no CNV, red = CNV present). PT, primary tumor; Met, metastasis.

<https://doi.org/10.1371/journal.pgen.1008743.g003>

with available material, the rate of validation by Sanger sequencing for metastasis-specific SNVs was approximately 80% (8 out of 10) (S2 Table and S1 Fig). Finally, Sanger sequencing of matched pairs of tumors and metastases from an additional 22 animals not used for RNA-

or exome-seq was used to survey for 11 of the SNVs identified by exome- and RNA-seq. This analysis identified yet another *Shc1* P561H mutation, but no additional samples containing the other 10 mutations were found. Taken together, these data indicate that the putative metastasis-driving recurrently mutated genes are mutually exclusive, as we did not observe the presence of mutations in two or more of the recurrently mutated genes within a single metastasis.

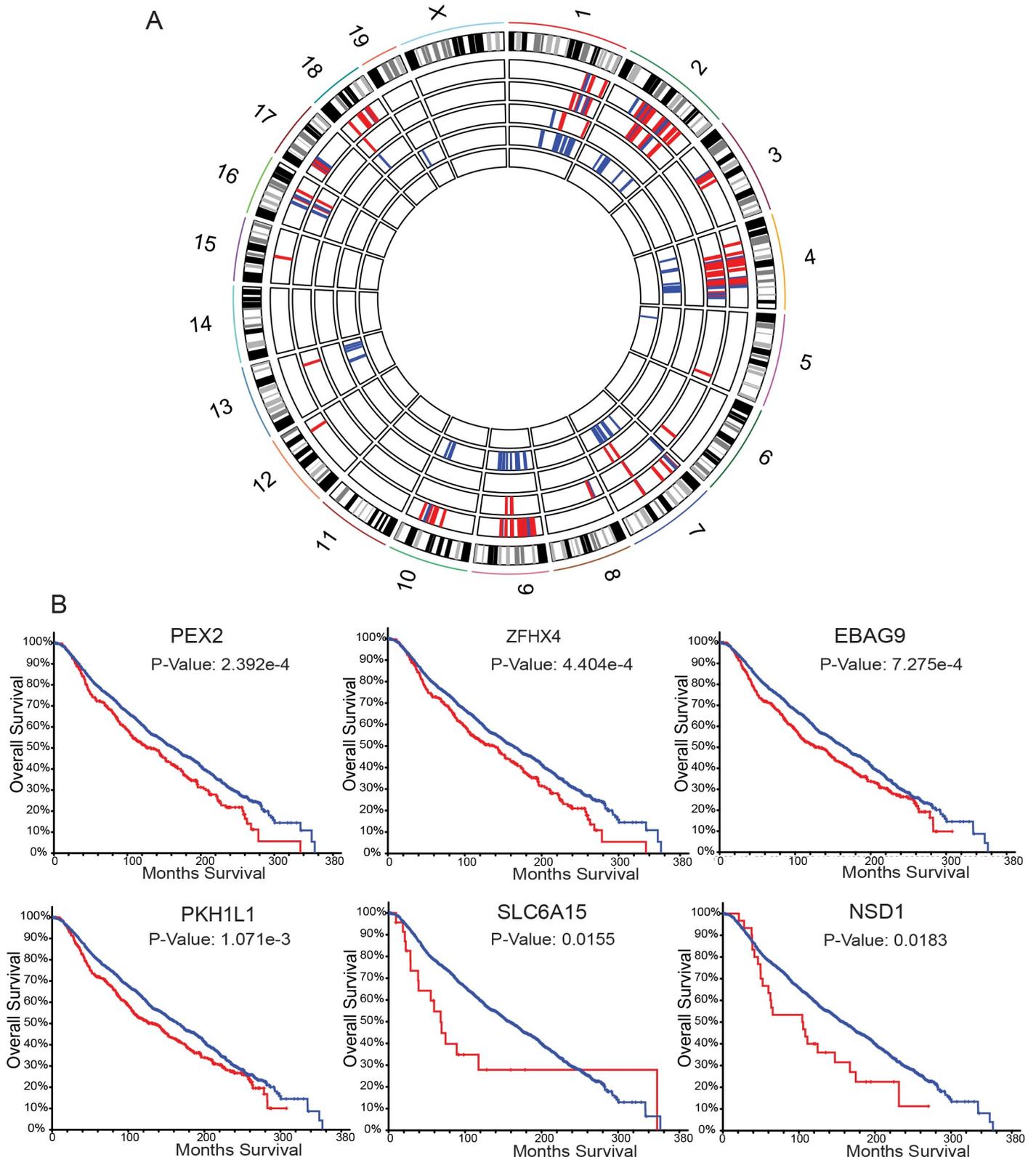
### Copy number variation drives metastasis and reduces overall survival among human breast cancer patients

Using the METABRIC sequencing dataset of human breast cancer primary tumors, we assessed the relevance to human disease of genes possessing metastasis-enriched SNVs identified in the PyMT and Her2 mouse models. This analysis revealed that many genes possessing SNVs in our preclinical mouse models were amplified in human tissue, and 18% (23 out of 123 human orthologs) of such alterations were significantly associated with poor patient outcome (Fig 3E, S4 and S5 Figs).

Since human primary tumors possessed gene amplification or deletion in addition to point mutations, we analyzed the PyMT primary tumor/metastasis exome-seq pairs for potential copy number variations (CNVs) that may contribute to metastatic spread. To determine allele-specific gain or loss we used only those samples generated from progeny of the FVB/NJ-PyMT outcross, as their genomes contained an FVB and non-FVB (C57BL/6J, C57BL/10J, CAST/EiJ, or MOLF/EiJ) allele that could be readily differentiated. Therefore, pure FVB/NJ-PyMT and FVB/NJ-Her2 tumor/metastasis pairs were not analyzed due to the lower confidence of CNV calling on a homozygous genetic background. Based on these criteria, we performed CNV analysis using exome-seq data from 30 of the primary tumor/metastasis pairs in our cohort. We first identified CNVs in primary tumors and metastases compared to normal, and then analyzed this data for metastasis-specific CNVs (S3–S7 Tables). Amplification and deletion events were primarily restricted to the F1 animals of crosses with C57BL/6J and MOLF/EiJ, with recurrent metastasis-specific CNVs on chromosomes 2, 4, 9, and 10 (Fig 4A). Examination of the putative CNVs suggested that the C57BL/6J alleles were specifically lost in the F1 hybrids, while FVB/NJ alleles were under-represented in most of the CNVs in the MOLF/EiJ F1 hybrids. In addition, the chromosome 4 and 9 CNVs overlapped with regions of the genome we previously demonstrated to harbor inherited metastasis susceptibility genes [27–29], suggesting these genomic intervals may contain important metastasis-associated factors. The Genomic Regions Enrichment of Annotations Tool (GREAT) [30] was used to identify the genes associated with recurrent metastasis-specific CNVs. 371 genes were associated with recurrent regions of CNV in PyMT metastatic lesions, and 46 of these genes possessed CNVs in two or more outcross strains (S8 Table and S6 Fig). The METABRIC dataset was then screened again to assess if CNVs within any of the 46 genes were significantly associated with breast cancer patient survival. The amplification of approximately 17% (8 out of 46) of the genes queried was significantly associated with worse patient outcome, including *EBAG9* and *PKHD1L1*, located on mouse chromosome 15, and *PEX2*, and *ZFH4*, located on mouse chromosome 2 (Fig 4B), which are all located on human chromosome 8 between q21.13 and q23.2. This analysis reveals that in addition to metastasis driver SNVs, metastatic cells may also acquire specific CNVs to support their spread to a secondary site.

### Metastases can originate from multiple clones in the MMTV-PyMT model

Curation of the RNA-seq data revealed several instances of discordant SNVs between independent metastases within the same animal (e.g., *Ctnnb1* K335N in animal 974). The presence of a high probability SNV in only a subset of metastatic lesions within an animal may be due to the



**Fig 4. Metastasis-specific CNVs in the PyMT and Her2 mouse models.** **A.** Circos plot representing metastasis-specific genomic alterations. The outermost track shows the mouse chromosomes and established G-banding. The remaining five inner tracks show deleted (and amplified for MOLF/Eij only) genomic regions in F1 PyMT mice resulting in enrichment of the non-FVB/NJ allele (red) or enrichment of FVB/NJ allele (blue). From the outermost track the strain order is as follows:

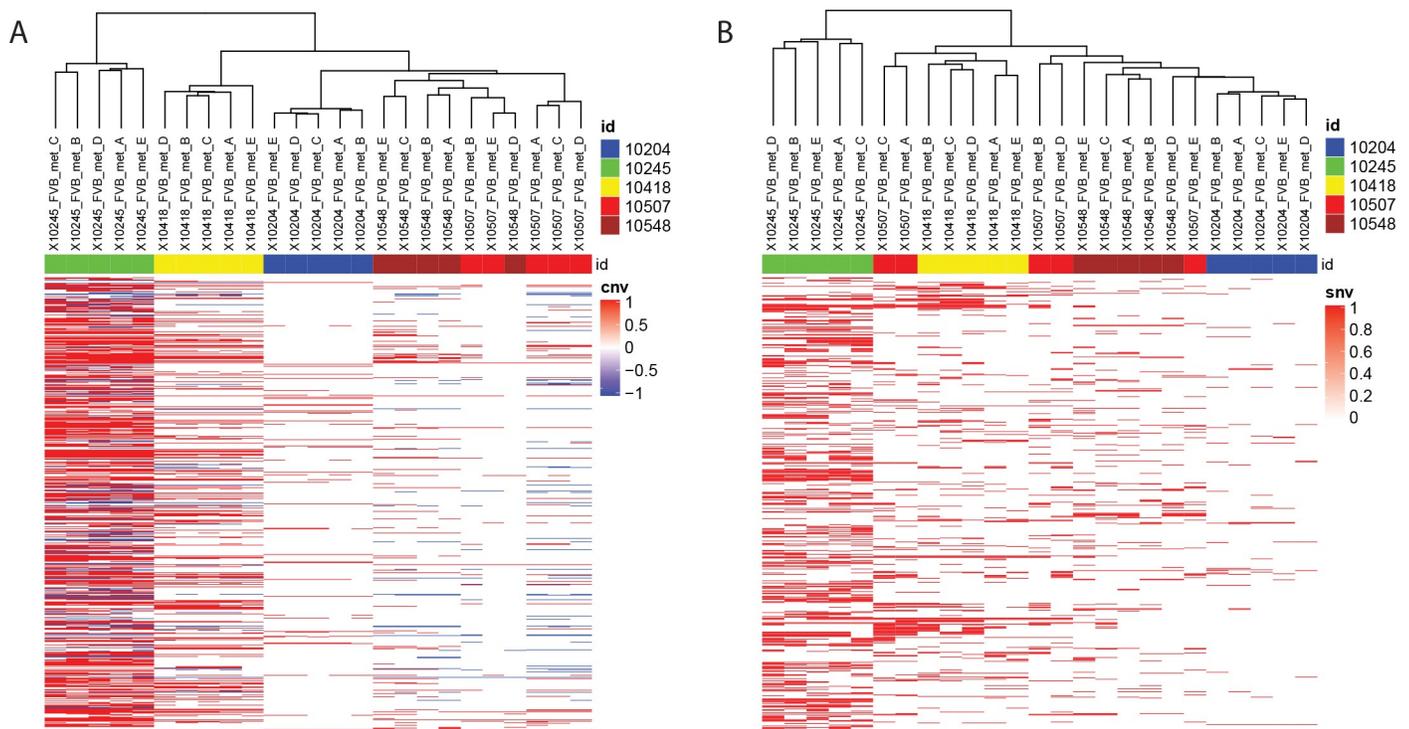
MOLF/EiJ amplification, MOLF/EiJ deletion, CAST/EiJ deletion, C57BL/6J deletion, C57BL/10J deletion. **B.** Kaplan-Meier plots generated using METABRIC for six out of eight genes identified as metastasis drivers by exome-seq CNV analysis in mice that significantly stratify patient survival when altered in primary tumor tissue and reside on human chromosome 8 (blue = no CNV, red = CNV present).

<https://doi.org/10.1371/journal.pgen.1008743.g004>

acquisition of the mutation early during the establishment of individual metastases or metastatic seeding by multiple primary tumor subclones within an animal. To address this, exome-seq was performed for primary tumors and matched metastatic tissue samples from five additional PyMT animals. Hierarchical clustering was then performed using both SNV and CNV analysis of those sequenced tissues. This analysis revealed that the metastases from three of the five animals (10204, 10245, and 10418) were highly related and likely originated from the same primary tumor subclone (Fig 5A and 5B). In contrast, two of the mice (10507 and 10548) produced heterogeneous metastases (Fig 5 and 5B), indicating that these nodules may have originated from different primary tumor subclones.

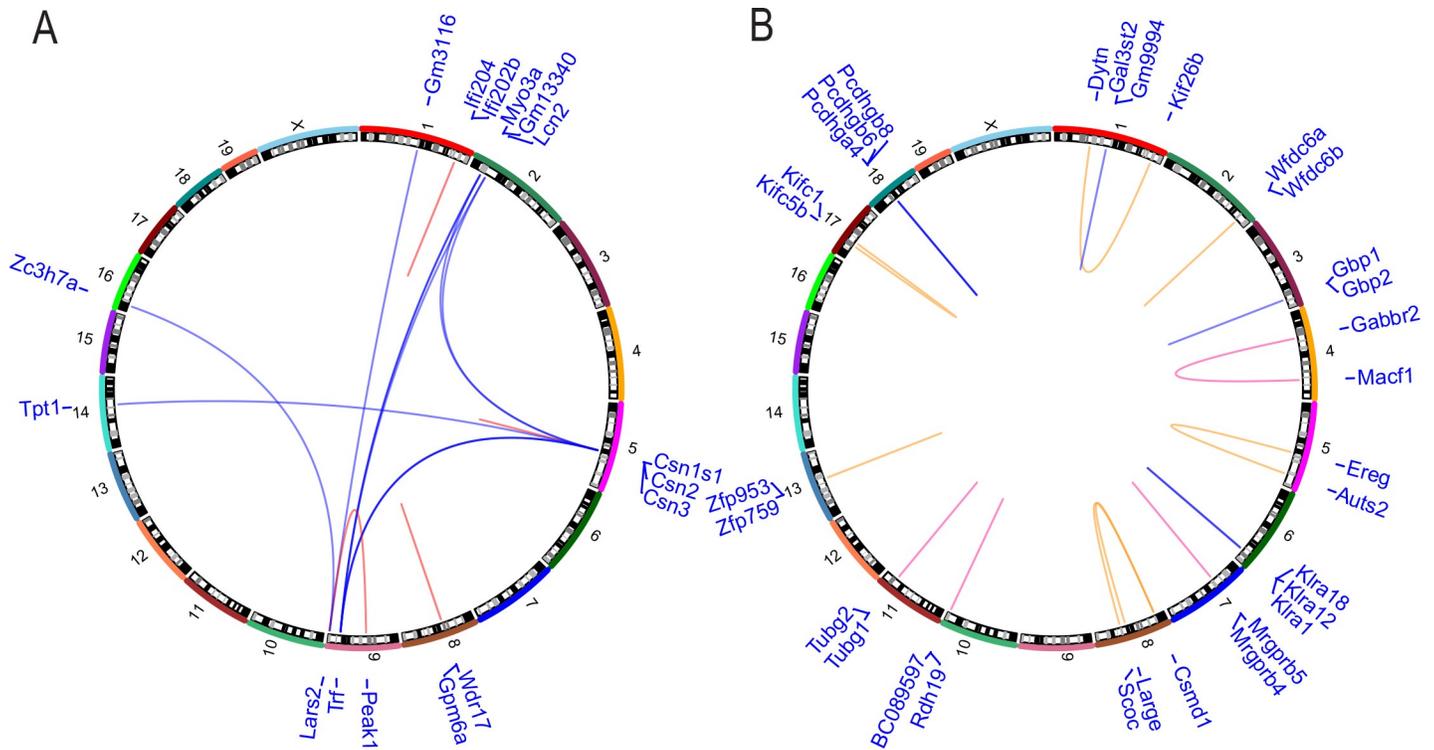
**Fusion genes are not enriched in metastases**

In a large-scale analysis of 560 human breast cancer samples, Nik-Zainal et al. discovered frequent variation in genomic structure in primary tumor tissue [31]. To identify any potential targetable gene fusion events that may contribute to metastatic progression, we analyzed the RNA-seq data (S1 Fig) using the deFuse algorithm to identify discordant paired-end alignments and thus putative gene fusion events specific to metastatic tissue [32]. To reduce false positive results, putative alternative splice events from adjacent genes were excluded from the



**Fig 5. Metastases can originate from multiple primary tumour subclones.** Heat maps for hierarchical clustering of genomic **a.** CNV and **b.** SNV fingerprints from five metastases isolated from five mice and compared to matched primary tumors. Chromosomes are represented along the Y axis, and sample names across the X axis. Colored bands across the top of the heatmap indicate mouse ID (Blue: 10204, Green: 10245, Yellow: 10418, Red: 10507, burgundy: 10548). Within the heat maps, red indicates amplification (A), or presence of an SNV (B), and blue indicates a deletion (A).

<https://doi.org/10.1371/journal.pgen.1008743.g005>



**Fig 6. Metastasis-specific structural variants do not drive metastasis in pre-clinical breast cancer mouse models.** **A.** Circos plots representing deFuse2 analysis of RNA-seq data for the 41 matched primary and metastatic PyMT tumors. **B.** BreakDancer algorithm analysis of 10x WGS data from 20 matched pairs of primary and metastatic lesions from FVB/NJ PyMT x MOLF/EiJ and FVB/NJ PyMT x CAST/EiJ animals. For both plots, the names of the genes associated with predicted structural variations are listed in blue in the outermost track by their corresponding chromosome numbers, which are found in the second track. The inner loops represent breakpoints within repetitive elements at both sites (pink loops), repetitive element at one site (tan loops), or within single copy sequence at both sites (blue loops).

<https://doi.org/10.1371/journal.pgen.1008743.g006>

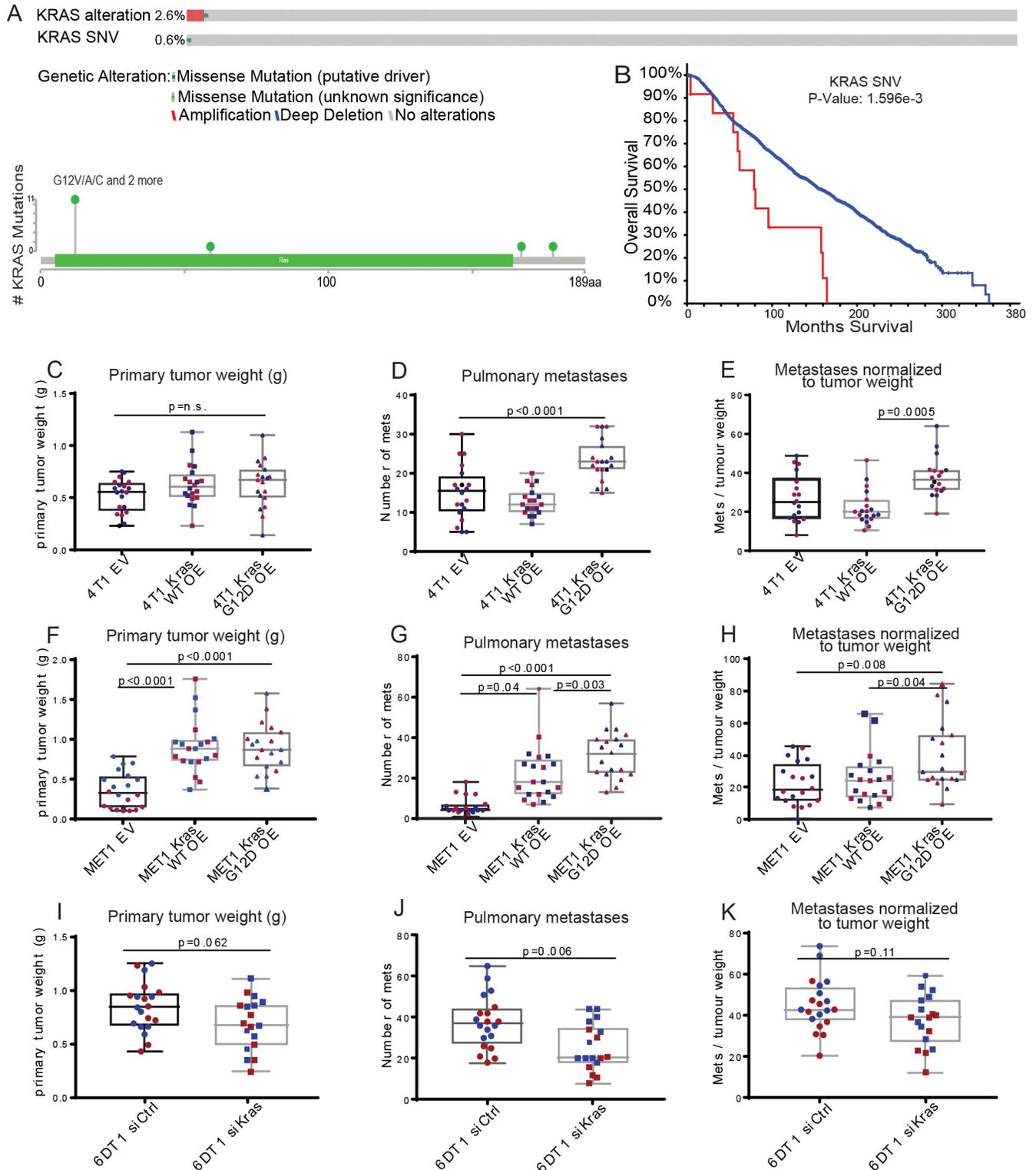
analysis to eliminate rare transcriptional read-through products. Forty-four putative fusion transcripts were detected in the primary tumors by this analysis. Seventy-one fusion transcripts were detected in the metastases, with 17 fusions in common with the primary tumors and 54 unique to the metastatic lesions. The 54 metastasis-specific putative fusion events involved a total of 85 genes. Sixteen genes were associated with putative fusion events in more than one animal or involved genes with multiple fusion partners (Fig 6A). However, manual curation of the putative recurrent fusion events revealed that 7 of the 15 breakpoints occurred in regions of repetitive elements within introns or UTRs, suggesting potential alignment artefacts. Moreover, for the highly expressed genes (*Csn3*, *Trf*), the variant transcripts were represented at <1% of the total read count, suggesting that these transcripts were either not required for maintenance of the metastatic lesions, were rare aberrant transcripts, or were artefacts of the RNA-seq/deFuse analysis.

Additionally, we performed 10x whole genome sequencing (WGS) on 20 matched pairs of primary and metastatic lesions from FVB/NJ-PyMT x MOLF/EiJ and FVB/NJ-PyMT x CAST/EiJ animals. This data set was comprised of tissue from 16 of the animals represented in the exome-seq dataset and 4 additional animals (S1A Fig). MOLF/EiJ and CAST/EiJ F1 animals were used for this analysis as their genomes are significantly polymorphic compared to FVB/NJ, allowing for allele-specific identification. SNVs predicted by exome-seq were confirmed in all of those animals also included in the WGS analysis (S2 Table). Putative coding-related structural variants such as insertion-deletions (indels), inversions, and translocations were identified using the BreakDancer algorithm [33] by limiting the analysis to within transcripts

and requiring a minimum of three variant reads per sample (S9 Table). The resulting gene fusions were predominantly intrachromosomal and tended to cluster within regions of gene families, suggesting potential alignment artefacts. To reduce this potential artefact, a threshold of a minimum of 10 kb between putative fusion partners was introduced, resulting in 18 putative fusion events involving 31 genes observed in either more than one animal or with multiple fusion partners (Fig 6B). No overlap between the putative structural variants observed from the deFuse and BreakDancer analysis was observed. All structural variations predicted by BreakDancer under these conditions were intrachromosomal, and 11 of the 18 predicted recurrent structural variants occurred between 2 members of the same gene family. Of the remaining seven, two of the remaining predicted structural variation breakpoints were within repetitive elements at both sites (Fig 6B, pink loops). Two additional predicted structural variations contained a repetitive element at one of the breakpoints (Fig 6B, tan loops), while three were within single copy sequence at both sites (Fig 6B, blue loops). These results suggest that structural variation resulting in fusion transcripts does not play a significant role in the metastatic progression of the MMTV-PyMT mammary tumor model.

### Single nucleotide variants drive metastasis

The presence of recurrent oncogenic mutations in *Kras* and *Ctnnb1*, in addition to the previously unreported recurrent mutations in *Shc1*, are consistent with a potential metastasis-driving function of these genes in the PyMT mouse model. Interestingly, previous analysis of a panel of metastatic mouse mammary tumor cell lines revealed recurrent mutation of *Kras* in several cell lines [34]. Moreover, SNVs within *KRAS* in the primary tumors of the METABRIC data set, although rare (0.6% of patients), were significantly associated with poor survival in human breast cancer (Fig 7A and 7B). *CTNNB1* and *SHC1* were amplified (0.3% and 22% of patients, respectively), but were not included in the targeted sequencing performed on 173 cancer-related genes in METABRIC. Therefore, the mutational burden for these genes in patient primary tumors within this data set is unknown. As METABRIC does not contain sequencing data from metastatic tissue, differential gene expression analysis of matched PDX primary mammary gland tumor and lung metastases from a recent study of PDX models [35] was examined by Ingenuity pathway Analysis (IPA). This analysis revealed that in 2 out of the 3 PDX models, *KRAS* signalling was significantly elevated in metastatic PDX tissue relative to primary tumor (S10–S12 Tables), consistent with the mutational activation observed in the exome-sequence analysis described here. Furthermore, *CTNNB1* signalling was also significantly altered in 2 of the 3 PDX models, however it was elevated in 1 and decreased in another (S10–S12 Tables). We thus focused on *Kras* as a potential breast cancer metastasis driver gene and performed a proof-of-concept experiment to determine if *Kras* mutation in the primary tumor alters metastatic potential. Wildtype (WT) or *Kras* G12D constructs were expressed in two independent *Kras* WT mouse mammary tumor cell lines: 4T1, which is derived from a spontaneous BALB/c mammary tumor, and MET1, derived from the MMTV-PyMT model (S7A Fig) [34]. *Kras* WT-, *Kras* G12D-, or control empty vector (EV)-transduced cells were implanted orthotopically in syngeneic mice and primary tumor weight and the number of pulmonary metastases were assessed after four weeks. Expression of *Kras* WT and G12D increased tumor burden compared to EV in the MET1 line and a small but non-significant increase in tumor growth was also observed with the 4T1 line (Fig 7C and 7F). Therefore, to more accurately assess metastatic capacity, the number of metastatic nodules (Fig 7D and 7G) was normalized to primary tumor weight for each mouse. This analysis revealed a significant increase in metastatic capacity of *Kras* G12D-expressing cells compared to both *Kras* WT and EV cells in both lines (Fig 7E and 7H). To further test how modulation of *Kras* G12D affects



**Fig 7. The *Kras* G12D mutation drives breast cancer metastasis.** A. METABRIC sample data showing that *KRAS* is altered in 2.6% and mutated in 0.6% of primary tumors from breast cancer patients. B. Kaplan-Meier analysis of patients with SNVs (red) in *KRAS* vs (wildtype) WT (blue). C-K. Primary tumor weights, numbers of metastases, and metastases normalized to primary tumor weight from orthotopic injection of 4T1 cells transduced with empty vector (EV), *Kras* WT, or *Kras* G12D (C-E), MET1 cells transduced with EV, *Kras* WT, or *Kras* G12D (F-H), or 6DT1 cells transfected with siControl or si*Kras* (I-K). Each assay was performed twice with 10 mice per group (red and blue points), box whiskers represent min and max points, box boundaries represent the 25th to 75th percentiles, and the horizontal line within the box represents the median. p<0.05 = significant.

<https://doi.org/10.1371/journal.pgen.1008743.g007>

metastatic potential, we utilized the 6DT1 mouse mammary cancer cell line that bears an endogenous homozygous *Kras* G12D mutation [34]. Despite the short-lived effects of siRNA, knock down of *Kras* in 6DT1 cells reduced metastasis to the lungs in spontaneous metastasis assays compared to a non-targeting siRNA control (Fig 7I–7K and S7B Fig). This result was significant before controlling for primary tumor weight, and trending after normalization ( $p = 0.1$ ) (Fig 7J and 7K). These results are consistent with a function for constitutively activated *Kras* as a metastasis driver gene in breast cancer, as well as significant evidence for the existence of spontaneous metastasis-driving somatic mutations.

Several studies have reported an association between oncogenic *Kras* expression, and a more mesenchymal-like phenotype associated with breast cancer progression [36–38]. Using the available RNA-seq data from two of the PyMT mice with spontaneous metastasis-specific *Kras* G12D mutations, we assessed the mRNA levels of several epithelial-to-mesenchymal transition (EMT)-related factors in the metastatic nodules [39]. There was no difference in the expression of these factors between tissues with WT or mutant *Kras* (S8A Fig). Additional RNA-seq was then performed on the 4T1 *Kras* WT, G12D, and EV cells. Again, no difference in expression of EMT-related factors was observed between the three cell lines (S8B Fig). These data suggest that a potentially novel mechanism, independent of EMT, may be responsible for the *Kras* G12D metastasis-driving function observed in this study.

## Discussion

Breast cancer metastasis is the primary cause of breast cancer-related mortality, and due to the disparate biology of metastatic lesions compared with the original tumor there are few options for clinical intervention [4,5,40–43]. In this study, we hypothesized that somatic evolution of the tumor cell genome may drive metastasis and could also provide novel therapeutic targets. However, no robust sequencing datasets of paired treatment-naïve primary and metastatic tumor samples that are large enough to identify recurrent mutations or genomic events currently exist. A key obstacle in this progress is the difficulty of obtaining appropriate tissue samples. Metastatic lesions are typically not surgically removed, and data obtained from resected lesions are often confounded by the application of neo-adjuvant and adjuvant therapies, which may select for events associated with treatment resistance rather than metastasis [44]. Our utilization of mouse models in the present study overcomes these limitations and serves as a hypothesis generation and testing platform for subsequent validation and characterization in human experimental systems.

Here we have performed exome, whole genome, and RNA sequencing on a total of 94 pairs of treatment-naïve primary and metastatic tumors [21,45] of the luminal-like PyMT and Her2 GEMMs. Moreover, sequencing of matched primary tumors and metastases from five additional MMTV-PyMT animals was also performed. Several important observations were made from these studies. First, similar to human breast cancer [40,46–48], both mono- and polyclonal seeding of metastases was observed in MMTV-PyMT animals. MMTV-PyMT animals develop multiple primary tumors in independent mammary glands, so it is possible that the polyclonal metastases arise from independent tumors, rather than different tumor subclones, as is the case in human breast cancer. Resolution of this question, however, would require comprehensive sequencing of all of the tumors from individual animals, which is beyond the scope of this study.

The second important observation from this study was the identification of multiple recurring metastasis-enriched SNVs. The recurrence and high enrichment of these SNVs within metastatic lesions strongly suggest roles as metastasis drivers. The vast majority of the metastasis-enriched SNVs were validated by orthogonal sequencing methods, 96% (47 out of 49; S2

Table), indicating that the analytical strategy was robust. The fact that MMTV-PyMT animals develop multiple primary tumors is not likely to confound this interpretation, since the metastasis-enriched SNVs did not appear as primary tumor drivers in any of the animals assayed. Many of the mutated genes were also found to be associated with poor outcome in human breast cancer data sets, consistent with a potential role in metastatic breast cancer. Moreover, we performed a proof-of-concept experiment to show that the recurrent *Kras* G12D mutation observed in metastatic lesions from two PyMT mice increases the metastatic efficiency of mouse mammary cancer cell lines in an orthotopic implantation model.

Importantly, the samples sequenced in this study are from therapy-naïve animals. Recent large-scale sequencing efforts focused on metastatic lesions and matched normal tissue have become available for human breast cancer patients. However, several of these studies were performed primarily with small targeted gene sets [49,50]. In addition, the vast majority of the patient samples in these studies [49–53] had previously been exposed to systemic treatment, which was found to be a major contributor to tumor mutational burden [51] and to metastasis mutational burden [49]. Moreover, all of these studies have been performed on heterogeneous human populations, with different primary tumor driver mutations on diverse genetic backgrounds that likely select for different metastatic driver mutations. Taken together, these caveats suggest that very large human sample sets may be necessary to begin to identify recurrent metastasis-driver mutations in human patients.

The use of therapy-naïve, genetically reproducible animals with defined primary tumor drivers provides an important complementary platform to identify somatic mutations that contribute to metastatic progression. This may be particularly important if, as we observed in the mouse models presented here, no single putative metastasis driver is mutated at a high frequency. This result suggests that there may be many metastasis drivers, each contributing a small fraction of the overall metastasis-driving function within a population. We did, however, observe several SNVs located in genes functioning within the Ras signaling pathway. These SNVs were observed in ~15% (14 (3-*Kras*, 5-*Shc1*, 6-*Ctnnb1*) out of 94) of the animals in this study, which implicates the *Kras* pathway as an important metastasis signaling axis in this model. The *KRAS* G12D mutation is a well-studied oncogenic driver leading to the development of aggressive lung and colorectal cancers [54–56]. However, *KRAS* is not considered a driver of human breast cancer development and its metastasis-specific recurrence in this model was unexpected [57]. Additionally, the exact role of the *Shc1* P561X mutation is unknown, but the localization of this mutation to the SH2 domain of *Shc1* is consistent with activating function and warrants further investigation [58].

While *KRAS* is not considered a major driver of breast cancer, recent work implicates this factor as an important player for some subtypes. In 2001, D’Cruz et al. described sustained oncogenic transformation following spontaneous mutagenesis of *Kras2* in an inducible *c-Myc* mammary tumorigenesis model [59]. In 2019, Campbell et al. reported spontaneous alterations in *Kras* expression and activity in the aggressive ER $\alpha$ <sup>+</sup> NRL-PRL postmenopausal murine model, that drove neoplastic transformation [60]. Additionally, the *Kras* Q61H mutation was identified in primary tumors from a triple negative murine model by Liu et. al, who also report mutations and amplifications of *KRAS* and the *KRAS* pathway in 11% and 93%, respectively, of triple negative breast cancer patients in the TCGA data set [61]. Hollern et al. recently reported that murine mammary tumors with EMT histopathology, which are similar to human claudin-low breast tumors, also showed increased activity of *KRAS* signaling [38]. Interestingly, we did not observe any mesenchymal shift in gene expression in any of the mutant *Kras* tissues analyzed. However, we did also observe *KRAS* activation using pathway analysis of RNA-seq data collected from metastatic PDX lung tumors compared to matched primary mammary gland tumors [35], suggesting that *KRAS* pathway activation may be an

important mechanism for at least some breast cancer patients. While these studies bring to light a potentially important role for KRAS signaling in breast cancer initiation, the work presented herein describes the first instance of *Kras* mutagenesis as a spontaneous driver of mammary tumor metastasis.

The SNV data reported here also suggest that there may be many ways within metastasis-associated pathways to activate a pro-metastatic function. Again, the identification of *Kras* and *Ctnnb1* as recurrently mutated metastasis drivers was unanticipated due to the lack of association with primary tumour biopsies of human breast cancer. The presence of known, well-studied activating mutations in these genes within metastasis-founding cells, however, strongly implicates the KRAS-CTNNB1 signaling axis as a critical component of metastatic progression, at least in the PyMT model system. The activation of the CTNNB1 pathway has been previously linked to breast cancer progression, and thus our data corroborates the importance of this network [62] and suggests an additional therapeutic opportunity to intervene in metastasis establishment and progression.

We have also identified recurrent metastasis-specific regions of CNV that may contribute to metastatic spread. Several studies have shown divergent CNV between primary and metastatic tissue from small patient cohorts [63–66], and through our use of the treatment-naïve PyMT animal model we have further shown that these genomic alterations occur through natural tumor progression and can drive metastasis. Additionally, our data identified several genes associated with metastasis-specific CNV that, while distributed across the mouse genome, are all localized within human chromosome 8q. Amplifications on chromosome 8q are associated with tumor progression and worse outcomes in many cancer types, suggesting that our analysis appropriately captures genetic alterations observed in human patients [67–72]. This association suggests that our strategy for the identification of metastasis-driving CNVs aligns with human biology, revealing a robust tool for the identification of clinically relevant metastatic drivers. We anticipate that with additional pairs, strain-specific and driver-specific metastasis-driver CNVs will be distinguished.

Understanding the etiology of metastasis has been an important goal in cancer research for many decades. The accumulation of chromosomal aberrations during tumor progression led Nowell to propose the progression model, where tumor cells evolve over time until a subclone acquires all of the somatic events necessary to confer metastatic potential [7]. However, cells derived from metastatic lesions were found to be no more efficient at metastasizing than primary tumor cells, suggesting that additional, potentially transient events are critical for acquisition of metastatic capacity [73]. More recently, the finding that bulk primary tumor gene expression can stratify breast cancer patient outcome [74] has led to the suggestion that metastatic capacity is encoded by the primary tumor drivers [75]. The results of our study are most consistent with the progression model originally proposed by Nowell [7], although it does not rule out important combined contributions of the transient epigenetic effects, microenvironment, and selection as proposed by Weiss in 1979 [73]. The presence of metastasis-specific driver mutations may also, in part, contribute to the difference in response to therapeutics observed between primary tumors and metastases, both in animal models [76] and in the clinic, and suggests different therapeutic strategies targeting metastasis drivers may improve patient outcomes.

In summary, by performing a survey of the genomic landscape of metastatic progression in two preclinical mouse models, we have for the first time identified recurrent, clinically relevant spontaneous mutations in genes and pathways that drive metastasis. Breast cancer patients with metastatic disease have an average 5-year survival of only 25%, underscoring the urgent need for more comprehensive molecular analyses of metastases. By using the strategy outlined here, the continued characterization of preclinical mouse models of metastasis will provide

further insight into the associations between clinical subtypes, primary tumor drivers, and the now newly confirmed metastasis-driving genomic alterations, ultimately creating new opportunities for the generation of metastasis-targeted therapies.

## Methods

### Ethics statement

The research described in this study was performed under the Animal Study Protocol LCBG-004 and LPG-002, approved by the National Cancer Institute (NCI) Animal Use and Care Committee. Animal euthanasia was performed by cervical dislocation after anaesthesia by Avertin.

### Genetically engineered mouse models

FVB/N-Tg(MMTV-PyVT)634Mul/J (PyMT) and FVB/N-Tg(MMTVneu)202Mul/J (Her2) male mice were obtained from Jackson Labs. Male PyMT mice were crossed with female wild type FVB/NJ, MOLF/EiJ, CAST/EiJ, C57BL/6J, and C57BL10/J mice also obtained from Jackson Labs. Male Her2 mice were crossed with female wild type FVB/NJ mice. All female F1 progeny were genotyped by the Laboratory of Cancer Biology and Genetics genotype core for the PyMT or Her2 gene and grown until humane endpoint. Mice were euthanized using intraperitoneal Avertin to anesthetize followed by cervical dislocation. All primary tumors generated by one animal were isolated weighed, randomly sampled, and combined into a single cryovial. Metastatic nodules, and normal (tail) tissue were also isolated immediately following euthanasia and snap frozen in liquid nitrogen. Tissue samples were then stored at  $-80^{\circ}\text{C}$ .

### gDNA isolation

The combined primary tumor tissue from one mouse was ground on dry ice and small fragments were taken for gDNA isolation. Whole metastases were used for gDNA isolation. Tissue was lysed using Tail Lysis Buffer (100 mM Tris-HCl pH 8.0, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, 0.4 mg/ml proteinase K) at  $55^{\circ}\text{C}$  overnight. Samples were then placed in a shaking (1400 rpm) heat block for 1 hour at  $55^{\circ}\text{C}$ . RNaseA (Thermo Fisher Scientific) was added (2 mg/ml final) and lysates were incubated on the bench for 2 minutes. gDNA was then isolated using the ZR-Duet DNA/RNA MiniPrep kit (Zymo Research).

### Sequencing and analysis

All analyses were carried out on the NIH Biowulf2 high performance computing environment. All analyses were performed using software default parameters if not otherwise specified.

### Exome sequencing

Exome sequencing was performed by the NCI Center for Cancer Research (CCR) Genomics Core and the NCI Illumina Sequencing Core. Exome libraries were prepared by the Genomics Lab using Agilent SureSelectXT Mouse All Exon target enrichment kit. Libraries were bar-coded and pooled before sequencing on an Illumina HiSeq3000 or HiSeq4000 to an average depth of 40x. Samples were trimmed of adapters using Trimmomatic software. The trimmed reads were aligned to the mm10 reference mouse genome using BWA (0.7.17) or Bowtie (2–2.3.5.1) mapping software. SAMtools (1.9) Mpileup, GATK (3.8–1) Mutect2, and Strelka (2.7.1) were used to identify potential variants and the variants were filtered for known polymorphisms from mcp.v5 ([http://ftp-mouse.sanger.ac.uk/current\\_snps/mcp.v5.merged.snps\\_all.dbSNP142.vcf.gz](http://ftp-mouse.sanger.ac.uk/current_snps/mcp.v5.merged.snps_all.dbSNP142.vcf.gz)) and variants with Phred-scaled quality scores of  $< 30$  were removed.

Annotation was performed by Annovar (2018-04-16). High probability SNVs were identified after manual curation of the data and select SNVs were validated by Sanger sequencing at the NCI CCR Genomics Core. Copy number variation was performed for the F1 tumors using the R (3.6.0) packages BubbleTree (2.15.0) and cn.mops (1.31.0), followed by filtering for metastasis-specific events and manual curation. We used BubbleTree to do allele-specific and strain-specific copy number analysis, which required genotype data, and cn.mops for general copy number analysis, which doesn't require F1 cross experiment data. The recurrent CNVs were defined as CNVs present in multiple samples.

### Whole genome sequencing

Library preparation was performed using the TruSeq DNA Sample Prep kit FC-121-1001. Samples were barcoded, pooled, and sequenced on an Illumina HiSeq4000 to a depth of ~10x per sample. Samples were trimmed of adapters using Trimmomatic (0.39) software before the alignment. The trimmed reads were aligned with the mm10 reference mouse genome using BWA alignment. SNV calls were performed as described for exome sequencing. Structural variation analysis was performed using BreakDancer (1.4.5), followed by filtering for metastasis-specific events and manual curation. For breakpoints located within or near a gene, we filtered structural variations (SVs) by limiting the breakpoint within 10 kb of a gene. CNV analysis was performed using cn.mops as described for exome sequencing.

### RNA sequencing

RNA fusion transcripts were identified using deFuse (0.8.1) software. We used the filtered fusion transcripts. The putative fusion transcripts were further analyzed by comparing with WGS BreakDancer data to identify fusion transcripts that were supported by genomic changes. Mutant alleles from RNA-seq data were identified using SAMtools (1.9). The sequence reads in pileup format were generated using SAMtools mpileup for BAM files spanning the SNVs. We extracted the reference allele and mutant allele counts for each sample using a custom-built Perl (v5.24.3) script and R (3.6.0) script.

### Comparative expression analysis of EMT transcripts

Using Partek Flow Genomic Analysis Software, normalized RNA-seq counts for EMT factors were compared between *Kras* WT and *Kras* G12D metastatic nodules, or between 4T1 cells transduced with empty vector, *Kras* WT, or *Kras* G12D using gene-specific analysis (GSA). Within the Partek Flow GSA report each factor was identified individually by searching with the official gene symbol, and then dot plots were created. Gene symbols used in search included *Kras* (for GEMM metastatic nodules only), *Fn1*, *Epcam*, *Cdh1*, *Vim*, *Zeb1*, *Snai1*, and *Snai2*. Factors that were not expressed did not produce dot plots following the Partek Flow GSA report search. Only those factors that were expressed produced dot plots that were then included in the final figure.

### Hierarchical clustering/heatmap and PCA analysis

CNV data were generated from cnmops as described in the exome sequencing section and were further processed using the DiffBind R package to generate a data matrix for the 25 lung metastases. Each genomic region is 4000 bp, which is coded as -1, 0, or 1 for CN loss, neutral, or gain, respectively. Genomic regions were filtered by the CN alterations presented in at least 2 samples. Similarly, SNV data were also filtered by the presence of SNVs in at least 2 samples.

SNVs are coded as 0 or 1 for the absence for presence of mutations. Heatmap plots were generated using the ComplexHeatmap package.

### Sanger sequencing

Polymerase chain reaction (PCR) was performed using AmpliTaq Gold polymerase (Thermo Fisher Scientific) according to the manufacturer's instruction. PCR products were then purified by gel electrophoresis using the QIAquick Gel Extraction Kit (Qiagen). The Sanger reaction and electrophoresis for the forward and reverse sequences of each purified amplicon was performed by the NCI CCR Genomics Core. Sequences were manually assessed for single nucleotide variations using Geneious software. If the original sample was RNA, then cDNA was synthesized by reverse transcription using the iScript cDNA Synthesis Kit (Bio-Rad).

### Genomic Regions Enrichment of Annotations Tool (GREAT) analysis

BED files containing the recurrent regions of genomic gain or loss for each mouse non-FVB mouse strain were loaded into the GREAT tool website using the default settings for gene assignment. GREAT calculates statistics by associating genomic regions with nearby genes and applying the gene annotations to the regions.

### Ingenuity Pathway Analysis

RNA sequencing DESeq2 results of mammary tumors compared to lung metastases was downloaded from the 2019 Alzubi et.al article Additional File 10 [35]. Pathway analysis was performed using Ingenuity Pathway Analysis (IPA) (Qiagen). Differentially regulated gene lists from each paired comparison (HCI01, HCI02, WHIM2) were uploaded into IPA for Core Expression Analysis of expression data. The Ingenuity Knowledge Base was chosen as the reference set of genes, and both direct and indirect relationships were considered. No other analysis parameters were specified, and the default settings were selected.

### Cell culture

Mouse mammary carcinoma cell lines 6DT1, 4T1, and MET1 were a generous gift from Dr Lalage Wakefield (NCI, Bethesda, MD). All cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% Fetal Bovine Serum (FBS), 1% Penicillin and Streptomycin (P/S) (complete DMEM), and 1% L-Glutamine (Gibco), and maintained at 37°C with 5% CO<sub>2</sub>. Short interfering RNA (siRNA)-mediated knockdown and overexpression cells were cultured in the same conditions with an addition of 10 µg/ml puromycin and 5 µg/ml blasticidin, respectively.

### Plasmid constructs

Lentiviral pDEST Gateway Entry clones expressing *Kras* WT or *Kras* G12D with a C-terminal myc tag, and under the control of the *Pol2* promoter were obtained from the NCI Ras Initiative. An empty lentiviral pDEST Gateway Entry clone (Thermo Fisher Scientific) was used as the empty vector control.

### Virus transduction

1 × 10<sup>6</sup> 293T cells were plated in 6 cm dishes 24 hours prior to transfection in P/S-free 10% FBS DMEM media. Cells were transfected with 1 µg *Kras* or control expression plasmid and 1 µg of viral packaging plasmids (250 ng pMD2.G and 750 ng psPAX2) using 6 µl of X-

tremeGENE 9 transfection reagent (Roche). After 24 hours of transfection, media was refreshed with complete DMEM. The following day, virus-containing supernatant was passed through a 45- $\mu$ m filter to obtain viral particles, which were then transferred to 100,000 4T1 or MET1 cells. The viral media was removed and fresh complete DMEM was added 24 hours post-transduction. Finally, the cells were selected with 10  $\mu$ g/ml puromycin- or 5  $\mu$ g/ml blastidicin-containing complete DMEM beginning 48 hours after transduction.

### siRNA transfection

6DT1 cells were plated in in P/S-free 10% FBS DMEM media. 24 hours after plating, cells were transfected with AllStars Mouse Negative Control siRNA (Qiagen) or siKras (siKRAS\_234 as described by Yuan et al. [77]) using RNAiMAX (Invitrogen).

### *In vivo* metastasis experiments

Female virgin FVB/NJ or BALB/cJ mice were obtained from Jackson Laboratory, and athymic NCI Nu/Nu from NCI Frederick at 6–8 weeks of age. Two days prior to *in vivo* experiments, cells were plated at  $1 \times 10^6$  cells/condition into T-75 flasks (Corning) in non-selective DMEM. A total of 100,000 cells per mouse was injected into the fourth mammary fat pad of FVB/NJ (MET1 and 6DT1 cells), BALB/cJ (4T1 cells) mice. The mice were euthanized between 28–30 days post-injection. Primary tumors were resected, weighed, and lung metastases counted. Statistical significance was calculated with a Kruskal-Wallis test followed by Conover Inman test. All animal experiments were performed in compliance with the National Cancer Institute's Animal Care and Use Committee guidelines.

### Western blotting

Protein lysate from  $1 \times 10^6$  cells were extracted on ice using Golden Lysis Buffer (10 mM Tris pH 8.0, 400 mM NaCl, 1% Triton X-100, 10% Glycerol+Complete protease inhibitor cocktail (Roche), phosphatase inhibitor (Sigma)). Protein concentration was measured using a BCA Protein Assay Kit (Pierce) and analyzed on the Versamax spectrophotometer at a wavelength of 560 nm. Appropriate volumes containing 20  $\mu$ g of protein lysates combined with NuPage LDS Sample Buffer and NuPage Reducing Agent (Invitrogen) were run on 4–12% (or otherwise indicated) NuPage Bis-Tris gels in MOPS buffer. Proteins were transferred onto a PVDF membrane (Millipore), blocked in 5% milk (TBST + dry milk) for 1 hour and incubated in the primary antibody (in 5% milk) overnight at 4°C. Membranes were washed with 0.05% TBST (TBS + 5% Tween) and secondary antibody incubations were done at room temperature for 1 hour. Proteins were visualized using Amersham ECL Prime Western Blotting Detection System and Amersham Hyperfilm ECL (GE Healthcare).

The following primary antibodies were used: mouse anti-Actin (1:10,000; Abcam), mouse anti-myc-tag (1:1000; Cell Signaling Technology), mouse anti-KRAS (1:1,000; Sigma). Secondary antibodies, including goat anti-rabbit (Santa Cruz) and goat-anti-mouse (GE Healthcare), were used at a concentration of 1:10,000.

### Statistics

Statistical significance between groups in *in vivo* assays was determined using the Mann-Whitney unpaired nonparametric test using Prism (version 5.03, GraphPad Software, La Jolla, CA). Statistical significance between samples in RT-qPCR analysis was determined by an unpaired t test, also using Prism.

## Supporting information

**S1 Fig. Sample collection and analysis workflow.** **A.** Schematic of sample collection for analysis. Preclinical mouse model development followed by sample collection from 14 FVB x MMTV-Her2 and 80 outcrossed MMTV-PyMT mice. Overlapping analyses performed on paired samples, numbers in Venn diagram represent number of animals. **B.** Schematic of analysis workflow. Next generation sequence data was analyzed for single nucleotide variants, copy number variants, and structural variants. The results were then filtered for those events enriched in metastatic tissue only.

(TIF)

**S2 Fig.** Overlap of three SNV calling algorithms, Strelka (purple), Mpileup (yellow), and Mutect2 (green), used with exome-seq data collected from primary tumor (PT) and lung metastases (LM) from 65 mice. **A.** SNVs called in PT tissue when compared to normal (strain-specific) gDNA. **B.** SNVs called in LM when compared to normal (strain-specific) gDNA. **C.** SNVs called in LM when compared to paired PT tissue using 0.3 allele frequency cutoff.

(TIF)

**S3 Fig. Sanger sequencing spectra showing the validation of SNVs in metastatic gDNA.** **A.** C (blue trace)-T (green trace) SNV within the *Kras* gene resulting in the G12D amino acid substitution, Y indicates ambiguity in calling T or C. **B.** G (yellow trace)-T (green trace) substitution within the *Shc1* gene resulting in the P561S amino acid substitution. K indicates ambiguity in calling T or G.

(TIF)

**S4 Fig.** Kaplan-Meier plots generated using METABRIC for 23 genes identified with metastasis-driver SNVs by exome-seq in mice that significantly stratify patient survival when altered in primary tumor tissue (blue = no CNV, red = CNV present).

(TIF)

**S5 Fig.** Oncoprint schema from the METABRIC human primary tumor dataset showing copy number variation rates of the **A.** 17 genes with recurrent SNVs and **B.** 147 singly mutated genes identified by exome-seq as putative metastasis-driver mutations (red = amplification, blue = deletion, green = SNV).

(TIF)

**S6 Fig. Venn diagram of the genes associated with CNV in each mouse strain.** Numbers represent the number of genes, and numbers in overlapping regions represent the number of common CNV-associated CNVs.

(TIF)

**S7 Fig.** **A.** Western blot showing expression of MYC-tagged KRAS and total KRAS in 4T1 and MET1 transduced with empty vector (EV), *Kras* wildtype (WT), and *Kras* G12D (G12D). **B.** Western blot showing knock down of KRAS in 6DT1 cells 24 hours after transfection siCtrl or si*Kras*.

(TIF)

**S8 Fig.** **A.** Dot plots showing normalized *Kras* and EMT gene transcript counts by RNA-seq of metastatic nodules from PyMT and Her2 animals with *Kras* wildtype (blue) or *Kras* mutations (red). **B.** Dot plots showing normalized EMT gene transcript counts by RNA-seq of 4T1 cells stably transduced with empty vector (purple), *Kras* wildtype (blue), or *Kras* G12D (yellow) expression vectors. 4.

(TIF)

**S1 Table. PyMT and Her2 exome-seq high probability metastasis-specific SNVs Sheets: 1.** Instances of Her2 metastasis-specific (met. spec.) SNVs, 2. Instances of PyMT met. spec. SNVs, 3. Singly mutated genes, 4. Recurrently mutated genes. Abbreviations: Chr (Chromosome number), Position (mm10 genomic position of mutated SNV), type (mutation type: synonymous, nonsynonymous, or stop gain), alt.fraction (allele fraction within the metastatic tissue), Transcript (NCBI accession number for isoform), Exon (exon harbouring SNV within designated transcript), Codon (codon harbouring SNV within designated transcript), Nuc sub (nucleotide position within designated transcript and substitution), AA sub (amino acid position within gene isoform and resulting substitution)  
(XLSX)

**S2 Table. Sequencing validation and overlap Sheets: 1.** Sanger sequencing (seq.) summary, 2. All seq. summary. Abbreviations: Y (yes), N (no), “/” (and), E (exome seq), R (RNA-seq), W (whole genome seq)  
(XLSX)

**S3 Table. PyMT regions of CNV in PT and metastatic tissue compared to normal.** Number of CNV events observed in PT and metastases compared to normal tissue. This table stratifies CNVs by mouse chromosome number and mouse strain. Also listed is the number of animals used in this study per stain, as well as the number of PT or metastatic samples collected from that strain total. Blue cells represent deletion events termed “loss,” and red cells represent amplification events termed “gain.”  
(XLSX)

**S4 Table. PyMT regions of CNV and associated genes specific to MOLF/Ei metastatic tissue.** Sheet1: “Strain” loss/gain associated (assoc.) genes, 2. “Strain” loss/gain assoc. pathways. Abbreviations: Name (gene symbol), ID (Term identifier from GREAT ontology), Rank (ordinal rank of the p-value compared to the p-values of other annotations), Raw p-value (uncorrected p-value from the binomial test over genomic regions), FDR q-Value (False discovery rate q-value), Fold Enrichment (fold enrichment of number of genomic regions in the test set with the annotation), Observed Region Hits (actual number of genomic regions in the test set with the annotation), Region Set Coverage (the fraction of all genomic regions in the test set that lie in the regulatory domain of a gene with the annotation, Sheet 2: “Strain” recurrent regions of loss or gain. Abbreviations: chr (chromosome), start (position of amplification or deletion start), end (position of amplification or del end), overlap.region (length in bp of overlap in recurrent region of amplification or deletion), freq. (number of individual animals with overlapping region of amplification or deletion), s1 / s2 (region identified in individual animals 1 and 2).  
(XLSX)

**S5 Table. PyMT regions of CNV and associated genes specific to CAST/Ei metastatic tissue.** Sheet1: “Strain” loss/gain associated (assoc.) genes, 2. “Strain” loss/gain assoc. pathways. Abbreviations: Name (gene symbol), ID (Term identifier from GREAT ontology), Rank (ordinal rank of the p-value compared to the p-values of other annotations), Raw p-value (uncorrected p-value from the binomial test over genomic regions), FDR q-Value (False discovery rate q-value), Fold Enrichment (fold enrichment of number of genomic regions in the test set with the annotation), Observed Region Hits (actual number of genomic regions in the test set with the annotation), Region Set Coverage (the fraction of all genomic regions in the test set that lie in the regulatory domain of a gene with the annotation, Sheet 2: “Strain” recurrent regions of loss or gain. Abbreviations: chr (chromosome), start (position of amplification or

deletion start), end (position of amplification or del end), overlap.region (length in bp of overlap in recurrent region of amplification or deletion), freq. (number of individual animals with overlapping region of amplification or deletion), s1 / s2 (region identified in individual animals 1 and 2).

(XLSX)

**S6 Table. PyMT regions of CNV and associated genes specific to C57BL10/nJ metastatic tissue.** Sheet1: “Strain” loss/gain associated (assoc.) genes, 2. “Strain” loss/gain assoc. pathways. Abbreviations: Name (gene symbol), ID (Term identifier from GREAT ontology), Rank (ordinal rank of the p-value compared to the p-values of other annotations), Raw p-value (uncorrected p-value from the binomial test over genomic regions), FDR q-Value (False discovery rate q-value), Fold Enrichment (fold enrichment of number of genomic regions in the test set with the annotation), Observed Region Hits (actual number of genomic regions in the test set with the annotation), Region Set Coverage (the fraction of all genomic regions in the test set that lie in the regulatory domain of a gene with the annotation, Sheet 2: “Strain” recurrent regions of loss or gain. Abbreviations: chr (chromosome), start (position of amplification or deletion start), end (position of amplification or del end), overlap.region (length in bp of overlap in recurrent region of amplification or deletion), freq. (number of individual animals with overlapping region of amplification or deletion), s1 / s2 (region identified in individual animals 1 and 2).

(XLSX)

**S7 Table. PyMT regions of CNV and associated genes specific to C57BL6/nJ metastatic tissue.** Sheet1: “Strain” loss/gain associated (assoc.) genes, 2. “Strain” loss/gain assoc. pathways. Abbreviations: Name (gene symbol), ID (Term identifier from GREAT ontology), Rank (ordinal rank of the p-value compared to the p-values of other annotations), Raw p-value (uncorrected p-value from the binomial test over genomic regions), FDR q-Value (False discovery rate q-value), Fold Enrichment (fold enrichment of number of genomic regions in the test set with the annotation), Observed Region Hits (actual number of genomic regions in the test set with the annotation), Region Set Coverage (the fraction of all genomic regions in the test set that lie in the regulatory domain of a gene with the annotation, Sheet 2: “Strain” recurrent regions of loss or gain. Abbreviations: chr (chromosome), start (position of amplification or deletion start), end (position of amplification or del end), overlap.region (length in bp of overlap in recurrent region of amplification or deletion), freq. (number of individual animals with overlapping region of amplification or deletion), s1 / s2 (region identified in individual animals 1 and 2).

(XLSX)

**S8 Table. PyMT strain summary for CNVs associated genes specific to metastatic tissue.**

Genes associated with metastasis-specific CNVs and found in more than one mouse strain.

Table associated with [S5 Fig](#).

(XLSX)

**S9 Table. Fusion events for primary tumor and metastatic tissue compared to normal.** This table shows the data generated from BreakDancer algorithm analysis of 10X WGS from 20 matched pairs of primary and metastatic lesions from FVB/NJ PyMT x MOLF/EiJ and FVB/NJ PyMT x CAST/EiJ animals. We show here the metastasis specific gene fusion data used as input data for [Fig 3B](#). We also show here primary tumor tissue and metastatic tissue gene fusion events versus normal strain specific tissue. Sheet1: Met. spec gene fusion events circos plot input. Sheet2: PT v Norm gene fusion events. Sheet 3: Met v Norm gene fusion event. Abbreviations: Sample (mouse ID and tissue type), Chr1 (chromosome where gene 1 is located), Pos1

(position within Chr1 where fusion event break point is located), Chr2 (chromosome where gene 2 is located), Pos2 (position within Chr2 where fusion event break point is located), Type (intra-chromosome events with different strands (ITX) inversion (INV), deletion (DEL), num\_ - Reads (number of reads).

(XLSX)

**S10 Table. Pathway analysis of CHI01 PDX mammary gland tumor vs. lung met RNA-seq comparative analysis.** Pathway analysis of RNA sequencing DESeq2 results of PDX mammary gland tumors compared to spontaneous lung metastases downloaded from the 2019 Alzubi et. al article Additional File 10 [35]. Sheet1: Pathway analysis. Sheet2: Upstream regulator analysis (XLSX)

**S11 Table. Pathway analysis of CHI02 PDX mammary gland tumor vs. lung met RNA-seq comparative analysis.** Pathway analysis of RNA sequencing DESeq2 results of PDX mammary gland tumors compared to spontaneous lung metastases downloaded from the 2019 Alzubi et. al article Additional File 10 [35]. Sheet1: Pathway analysis. Sheet2: Upstream regulator analysis (XLSX)

**S12 Table. Pathway analysis of WHIM2 PDX mammary gland tumor vs. lung met RNA-seq comparative analysis.** Pathway analysis of RNA sequencing DESeq2 results of PDX mammary gland tumors compared to spontaneous lung metastases downloaded from the 2019 Alzubi et. al article Additional File 10 [35]. Sheet1: Pathway analysis. Sheet2: Upstream regulator analysis (XLSX)

## Acknowledgments

This work utilized the computational resources of the NIH HPC Biowulf cluster. (<http://hpc.nih.gov>).

## Author Contributions

**Conceptualization:** Christina Ross, Karol Szczepanek, Kent Hunter.

**Data curation:** Christina Ross, Maxwell Lee, Howard Yang, Kent Hunter.

**Formal analysis:** Christina Ross, Maxwell Lee, Howard Yang, Kent Hunter.

**Investigation:** Christina Ross, Karol Szczepanek.

**Methodology:** Christina Ross, Karol Szczepanek, Kent Hunter.

**Resources:** Kent Hunter.

**Supervision:** Kent Hunter.

**Validation:** Christina Ross, Tinghu Qiu, Jack D. Sanford.

**Visualization:** Christina Ross.

**Writing – original draft:** Christina Ross.

**Writing – review & editing:** Kent Hunter.

## References

1. Sonnenblick A, Pondé N, Piccart M. Metastatic breast cancer: The Odyssey of personalization. *Mol Oncol.* 2016; 10: 1147–1159. <https://doi.org/10.1016/j.molonc.2016.07.002> PMID: 27430154

2. Female Breast Cancer—Cancer Stat Facts [Internet]. [cited 11 Apr 2018]. Available: <https://seer.cancer.gov/statfacts/html/breast.html>
3. Mattson J, Huovinen R. [Treatment of disseminated breast cancer]. *Duodecim*. 2015; 131: 1033–1040.
4. Rachdi H, Mokrani A, Batti R, Ayadi M, Chraiet N, Mezlini A. Target therapy for metastatic breast cancer. *Tunis Med*. 2018; 96: 465–471.
5. Steeg PS. Targeting metastasis. *Nat Rev Cancer*. 2016; 16: 201–218. <https://doi.org/10.1038/nrc.2016.25> PMID: 27009393
6. Lu J, Steeg PS, Price JE, Krishnamurthy S, Mani SA, Reuben J, et al. Breast cancer metastasis: challenges and opportunities. *Cancer Res*. 2009; 69: 4951–4953. <https://doi.org/10.1158/0008-5472.CAN-09-0099> PMID: 19470768
7. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976; 194: 23–28. <https://doi.org/10.1126/science.959840> PMID: 959840
8. Hunter KW, Amin R, Deasy S, Ha N-H, Wakefield L. Genetic insights into the morass of metastatic heterogeneity. *Nat Rev Cancer*. 2018; 18: 211–223. <https://doi.org/10.1038/nrc.2017.126> PMID: 29422598
9. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol*. 2018; 15: 81–94. <https://doi.org/10.1038/nrclinonc.2017.166> PMID: 29115304
10. Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, et al. Toward understanding and exploiting tumor heterogeneity. *Nat Med*. 2015; 21: 846–853. <https://doi.org/10.1038/nm.3915> PMID: 26248267
11. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature*. 2013; 501: 328–337. <https://doi.org/10.1038/nature12624> PMID: 24048065
12. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490: 61–70. <https://doi.org/10.1038/nature11412> PMID: 23000897
13. Faraji F, Hu Y, Yang HH, Lee MP, Winkler GS, Hafner M, et al. Post-transcriptional Control of Tumor Cell Autonomous Metastatic Potential by CCR4-NOT Deadenylase CNOT7. *PLoS Genet*. 2016; 12: e1005820. <https://doi.org/10.1371/journal.pgen.1005820> PMID: 26807845
14. Ha N-H, Long J, Cai Q, Shu XO, Hunter KW. The Circadian Rhythm Gene *Arntl2* Is a Metastasis Susceptibility Gene for Estrogen Receptor-Negative Breast Cancer. *PLoS Genet*. 2016; 12: e1006267. <https://doi.org/10.1371/journal.pgen.1006267> PMID: 27656887
15. Gilkes DM, Semenza GL, Wirtz D. Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nat Rev Cancer*. 2014; 14: 430–439. <https://doi.org/10.1038/nrc3726> PMID: 24827502
16. Chaffer CL, San Juan BP, Lim E, Weinberg RA. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev*. 2016; 35: 645–654. <https://doi.org/10.1007/s10555-016-9648-7> PMID: 27878502
17. Patel SA, Vanharanta S. Epigenetic determinants of metastasis. *Mol Oncol*. 2017; 11: 79–96. <https://doi.org/10.1016/j.molonc.2016.09.008> PMID: 27756687
18. Hosseini H, Obradović MMS, Hoffmann M, Harper KL, Sosa MS, Werner-Klein M, et al. Early dissemination seeds metastasis in breast cancer. *Nature*. 2016; 540: 552–558. <https://doi.org/10.1038/nature20785> PMID: 27974799
19. Linde N, Casanova-Acebes M, Sosa MS, Mortha A, Rahman A, Farias E, et al. Macrophages orchestrate breast cancer early dissemination and metastasis. *Nat Commun*. 2018; 9: 21. <https://doi.org/10.1038/s41467-017-02481-5> PMID: 29295986
20. Harper KL, Sosa MS, Entenberg D, Hosseini H, Cheung JF, Nobre R, et al. Mechanism of early dissemination and metastasis in Her2+ mammary cancer. *Nature*. 2016; 540: 588–592. <https://doi.org/10.1038/nature20609> PMID: 27974798
21. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR, et al. Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol*. 2013; 14: R125. <https://doi.org/10.1186/gb-2013-14-11-r125> PMID: 24220145
22. Guy CT, Cardiff RD, Muller WJ. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol*. 1992; 12: 954–961. <https://doi.org/10.1128/MCB.12.3.954> PMID: 1312220
23. Goddard KAB, Weinmann S, Richert-Boe K, Chen C, Bulkley J, Wax C. HER2 evaluation and its impact on breast cancer treatment decisions. *Public Health Genomics*. 2012; 15: 1–10. <https://doi.org/10.1159/000325746> PMID: 21540562
24. Yaziji H, Goldstein LC, Barry TS, Werling R, Hwang H, Ellis GK, et al. HER-2 testing in breast cancer using parallel tissue-based methods. *JAMA*. 2004; 291: 1972–1977. <https://doi.org/10.1001/jama.291.16.1972> PMID: 15113815

25. Millis SZ, Ikeda S, Reddy S, Gatalica Z, Kurzrock R. Landscape of Phosphatidylinositol-3-Kinase Pathway Alterations Across 19 784 Diverse Solid Tumors. *JAMA Oncol.* 2016; 2: 1565–1573. <https://doi.org/10.1001/jamaoncol.2016.0891> PMID: 27388585
26. Liu Y, Patel L, Mills GB, Lu KH, Sood AK, Ding L, et al. Clinical significance of CTNNB1 mutation and Wnt pathway activation in endometrioid endometrial carcinoma. *J Natl Cancer Inst.* 2014; 106. <https://doi.org/10.1093/jnci/dju245> PMID: 25214561
27. Faraji F, Pang Y, Walker RC, Nieves Borges R, Yang L, Hunter KW. Cadm1 is a metastasis susceptibility gene that suppresses metastasis by modifying tumor interaction with the cell-mediated immunity. *PLoS Genet.* 2012; 8: e1002926. <https://doi.org/10.1371/journal.pgen.1002926> PMID: 23028344
28. Lancaster M, Rouse J, Hunter KW. Modifiers of mammary tumor progression and metastasis on mouse chromosomes 7, 9, and 17. *Mamm Genome.* 2005; 16: 120–126.
29. Le Voyer T, Rouse J, Lu Z, Lifsted T, Williams M, Hunter KW. Three loci modify growth of a transgene-induced mammary tumor: suppression of proliferation associated with decreased microvessel density. *Genomics.* 2001; 74: 253–261. <https://doi.org/10.1006/geno.2001.6562> PMID: 11414753
30. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010; 28: 495–501. <https://doi.org/10.1038/nbt.1630> PMID: 20436461
31. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature.* 2016; 534: 47–54. <https://doi.org/10.1038/nature17676> PMID: 27135926
32. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol.* 2011; 7: e1001138. <https://doi.org/10.1371/journal.pcbi.1001138> PMID: 21625565
33. Fan X, Abbott TE, Larson D, Chen K. BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping. *Curr Protoc Bioinformatics.* 2014; 45: 15.6.1–11. <https://doi.org/10.1002/0471250953.bi1506s45> PMID: 25152801
34. Yang Y, Yang HH, Hu Y, Watson PH, Liu H, Geiger TR, et al. Immunocompetent mouse allograft models for development of therapies to target breast cancer metastasis. *Oncotarget.* 2017; 8: 30621–30643. <https://doi.org/10.18632/oncotarget.15695> PMID: 28430642
35. Alzubi MA, Turner TH, Olex AL, Sohal SS, Tobin NP, Recio SG, et al. Separation of breast cancer and organ microenvironment transcriptomes in metastases. *Breast Cancer Res.* 2019; 21: 36. <https://doi.org/10.1186/s13058-019-1123-2> PMID: 30841919
36. Shao DD, Xue W, Krall EB, Bhutkar A, Piccioni F, Wang X, et al. KRAS and YAP1 converge to regulate EMT and tumor survival. *Cell.* 2014; 158: 171–184. <https://doi.org/10.1016/j.cell.2014.06.004> PMID: 24954536
37. Kim R-K, Suh Y, Yoo K-C, Cui Y-H, Kim H, Kim M-J, et al. Activation of KRAS promotes the mesenchymal features of basal-type breast cancer. *Exp Mol Med.* 2015; 47: e137. <https://doi.org/10.1038/emmm.2014.99> PMID: 25633745
38. Hollern DP, Swiatnicki MR, Andrechek ER. Histological subtypes of mouse mammary tumors reveal conserved relationships to human cancers. *PLoS Genet.* 2018; 14: e1007135. <https://doi.org/10.1371/journal.pgen.1007135> PMID: 29346386
39. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol.* 2019; 20: 69–84. <https://doi.org/10.1038/s41580-018-0080-4> PMID: 30459476
40. Hoefnagel LDC, van der Groep P, van de Vijver MJ, Boers JE, Wesseling P, Wesseling J, et al. Discordance in ER $\alpha$ , PR and HER2 receptor status across different distant breast cancer metastases within the same patient. *Ann Oncol.* 2013; 24: 3017–3023. <https://doi.org/10.1093/annonc/mdt390> PMID: 24114857
41. Bell R, Barraclough R, Vasieva O. Gene Expression Meta-Analysis of Potential Metastatic Breast Cancer Markers. *Curr Mol Med.* 2017; 17: 200–210. <https://doi.org/10.2174/1566524017666170807144946> PMID: 28782484
42. Krøigård AB, Larsen MJ, Thomassen M, Kruse TA. Molecular concordance between primary breast cancer and matched metastases. *Breast J.* 2016; 22: 420–430. <https://doi.org/10.1111/tbj.12596> PMID: 27089067
43. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature.* 2010; 464: 999–1005. <https://doi.org/10.1038/nature08989> PMID: 20393555

44. Vakiani E, Janakiraman M, Shen R, Sinha R, Zeng Z, Shia J, et al. Comparative genomic analysis of primary versus metastatic colorectal carcinomas. *J Clin Oncol*. 2012; 30: 2956–2962. <https://doi.org/10.1200/JCO.2011.38.2994> PMID: 22665543
45. Lin EY, Jones JG, Li P, Zhu L, Whitney KD, Muller WJ, et al. Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am J Pathol*. 2003; 163: 2113–2126. [https://doi.org/10.1016/S0002-9440\(10\)63568-7](https://doi.org/10.1016/S0002-9440(10)63568-7) PMID: 14578209
46. Kuukasjärvi T, Karhu R, Tanner M, Kähkönen M, Schäffer A, Nupponen N, et al. Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer Res*. 1997; 57: 1597–1604.
47. De Mattos-Arruda L, Sammut S-J, Ross EM, Bashford-Rogers R, Greenstein E, Markus H, et al. The genomic and immune landscapes of lethal metastatic breast cancer. *Cell Rep*. 2019; 27: 2690–2708. e10. <https://doi.org/10.1016/j.celrep.2019.04.098> PMID: 31141692
48. Echeverria GV, Powell E, Seth S, Ge Z, Carugo A, Bristow C, et al. High-resolution clonal mapping of multi-organ metastasis in triple negative breast cancer. *Nat Commun*. 2018; 9: 5079. <https://doi.org/10.1038/s41467-018-07406-4> PMID: 30498242
49. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell*. 2018; 34: 427–438.e6. <https://doi.org/10.1016/j.ccell.2018.08.008> PMID: 30205045
50. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell*. 2017; 32: 169–184.e7. <https://doi.org/10.1016/j.ccell.2017.07.005> PMID: 28810143
51. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet*. 2019; 51: 1450–1458. <https://doi.org/10.1038/s41588-019-0507-7> PMID: 31570896
52. Bertucci F, Ng CKY, Patsouris A, Droin N, Piscuoglio S, Carbuccia N, et al. Genomic characterization of metastatic breast cancers. *Nature*. 2019; 569: 560–564. <https://doi.org/10.1038/s41586-019-1056-z> PMID: 31118521
53. Nayar U, Cohen O, Kapstad C, Cuoco MS, Waks AG, Wander SA, et al. Acquired HER2 mutations in ER+ metastatic breast cancer confer resistance to estrogen receptor-directed therapies. *Nat Genet*. 2019; 51: 207–216. <https://doi.org/10.1038/s41588-018-0287-5> PMID: 30531871
54. Pereira AAL, Rego JFM, Morris V, Overman MJ, Eng C, Garrett CR, et al. Association between KRAS mutation and lung metastasis in advanced colorectal cancer. *Br J Cancer*. 2015; 112: 424–428. <https://doi.org/10.1038/bjc.2014.619> PMID: 25535726
55. Jancík S, Drábek J, Radziach D, Hajdúch M. Clinical relevance of KRAS in human cancers. *J Biomed Biotechnol*. 2010; 2010: 150960. <https://doi.org/10.1155/2010/150960> PMID: 20617134
56. Lièvre A, Bachet J-B, Le Corre D, Boige V, Landi B, Emile J-F, et al. KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res*. 2006; 66: 3992–3995. <https://doi.org/10.1158/0008-5472.CAN-06-0191> PMID: 16618717
57. Sánchez-Muñoz A, Gallego E, de Luque V, Pérez-Rivas LG, Vicioso L, Ribelles N, et al. Lack of evidence for KRAS oncogenic mutations in triple-negative breast cancer. *BMC Cancer*. 2010; 10: 136. <https://doi.org/10.1186/1471-2407-10-136> PMID: 20385028
58. Wills MKB, Jones N. Teaching an old dogma new tricks: twenty years of Shc adaptor signalling. *Biochem J*. 2012; 447: 1–16. <https://doi.org/10.1042/BJ20120769> PMID: 22970934
59. D’Cruz CM, Gunther EJ, Boxer RB, Hartman JL, Sintasath L, Moody SE, et al. c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous Kras2 mutations. *Nat Med*. 2001; 7: 235–239. <https://doi.org/10.1038/84691> PMID: 11175856
60. Campbell KM, O’Leary KA, Rugowski DE, Mulligan WA, Barnell EK, Skidmore ZL, et al. A spontaneous aggressive era+ mammary tumor model is driven by kras activation. *Cell Rep*. 2019; 28: 1526–1537.e4. <https://doi.org/10.1016/j.celrep.2019.06.098> PMID: 31390566
61. Liu H, Murphy CJ, Karreth FA, Ermdal KB, White FM, Elemento O, et al. Identifying and Targeting Sporadic Oncogenic Genetic Aberrations in Mouse Models of Triple-Negative Breast Cancer. *Cancer Discov*. 2018; 8: 354–369. <https://doi.org/10.1158/2159-8290.CD-17-0679> PMID: 29203461
62. Jeong W-J, Ro EJ, Choi K-Y. Interaction between Wnt/ $\beta$ -catenin and RAS-ERK pathways and an anti-cancer strategy via degradations of  $\beta$ -catenin and RAS by targeting the Wnt/ $\beta$ -catenin pathway. *npj Precision Onc*. 2018; 2: 5. <https://doi.org/10.1038/s41698-018-0049-y> PMID: 29872723
63. Malek JA, Mery E, Mahmoud YA, Al-Azwani EK, Roger L, Huang R, et al. Copy number variation analysis of matched ovarian primary tumors and peritoneal metastasis. *PLoS One*. 2011; 6: e28561. <https://doi.org/10.1371/journal.pone.0028561> PMID: 22194851

64. Behring M, Shrestha S, Manne U, Cui X, Gonzalez-Reymundez A, Grueneberg A, et al. Integrated landscape of copy number variation and RNA expression associated with nodal metastasis in invasive ductal breast carcinoma. *Oncotarget*. 2018; 9: 36836–36848. <https://doi.org/10.18632/oncotarget.26386> PMID: 30627325
65. Kutilin DS, Leyman IA, Lazutin YN, Chubaryan AV, Anistratov PA, Stateshny ON, et al. Genes copy number variation in tumor cells of patients with metastatic and non-metastatic lung adenocarcinoma. *J Clin Oncol*. 2019; 37: e14502–e14502. [https://doi.org/10.1200/JCO.2019.37.15\\_suppl.e14502](https://doi.org/10.1200/JCO.2019.37.15_suppl.e14502)
66. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA*. 2010; 107: 16910–16915. <https://doi.org/10.1073/pnas.1009843107> PMID: 20837533
67. El Gammal AT, Brüchmann M, Zustin J, Isbarn H, Hellwinkel OJC, Köllermann J, et al. Chromosome 8p deletions and 8q gains are associated with tumor progression and poor prognosis in prostate cancer. *Clin Cancer Res*. 2010; 16: 56–64. <https://doi.org/10.1158/1078-0432.CCR-09-1423> PMID: 20028754
68. Sato K, Qian J, Slezak JM, Lieber MM, Bostwick DG, Bergstralh EJ, et al. Clinical Significance of Alterations of Chromosome 8 in High-Grade, Advanced, Nonmetastatic Prostate Carcinoma. *JNCI Journal of the National Cancer Institute*. 1999; 91: 1574–1580. <https://doi.org/10.1093/jnci/91.18.1574> PMID: 10491435
69. Kang JU. Chromosome 8q as the most frequent target for amplification in early gastric carcinoma. *Oncol Lett*. 2014; 7: 1139–1143. <https://doi.org/10.3892/ol.2014.1849> PMID: 24944681
70. Beuten J, Gelfond JAL, Martinez-Fierro ML, Weldon KS, Crandall AC, Rojas-Martinez A, et al. Association of chromosome 8q variants with prostate cancer risk in Caucasian and Hispanic men. *Carcinogenesis*. 2009; 30: 1372–1379. <https://doi.org/10.1093/carcin/bgp148> PMID: 19528667
71. Bilal E, Vassallo K, Toppmeyer D, Barnard N, Rye IH, Almendro V, et al. Amplified loci on chromosomes 8 and 17 predict early relapse in ER-positive breast cancers. *PLoS One*. 2012; 7: e38575. <https://doi.org/10.1371/journal.pone.0038575> PMID: 22719901
72. Yong ZWE, Zaini ZM, Kallarakkal TG, Karen-Ng LP, Rahman ZAA, Ismail SM, et al. Genetic alterations of chromosome 8 genes in oral cancer. *Sci Rep*. 2014; 4: 6073. <https://doi.org/10.1038/srep06073> PMID: 25123227
73. Weiss L. Dynamic aspects of cancer cell populations in metastasis. *Am J Pathol*. 1979; 97: 601–608.
74. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002; 347: 1999–2009. <https://doi.org/10.1056/NEJMoa021967> PMID: 12490681
75. Baudot A, Real FX, Izarzugaza JMG, Valencia A. From cancer genomes to cancer models: bridging the gaps. *EMBO Rep*. 2009; 10: 359–366. <https://doi.org/10.1038/embor.2009.46> PMID: 19305388
76. Fugmann RA, Anderson JC, Stolfi RL, Martin DS. Comparison of adjuvant chemotherapeutic activity against primary and metastatic spontaneous murine tumors. *Cancer Res*. 1977; 37: 496–500.
77. Yuan TL, Fellmann C, Lee C-S, Ritchie CD, Thapar V, Lee LC, et al. Development of siRNA payloads to target KRAS-mutant cancer. *Cancer Discov*. 2014; 4: 1182–1197. <https://doi.org/10.1158/2159-8290.CD-13-0900> PMID: 25100204