

Washington University School of Medicine

Digital Commons@Becker

---

Open Access Publications

---

2018

## Identification of drivers of aneuploidy in breast tumors

Katherine Pfister  
*University of Virginia*

Justyna L. Pipka  
*University of Virginia*

Colby Chiang  
*Washington University School of Medicine in St. Louis*

Yunxian Liu  
*University of Virginia*

Royden A. Clark  
*University of Virginia*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open\\_access\\_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

**Please let us know how this document benefits you.**

---

### Recommended Citation

Pfister, Katherine; Pipka, Justyna L.; Chiang, Colby; Liu, Yunxian; Clark, Royden A.; Keller, Ray; Skoglund, Paul; Guertin, Michael J.; Hall, Ira M.; and Stukenberg, P. Todd, "Identification of drivers of aneuploidy in breast tumors." *Cell Reports*. 23, 9. 2758-2769. (2018).  
[https://digitalcommons.wustl.edu/open\\_access\\_pubs/6887](https://digitalcommons.wustl.edu/open_access_pubs/6887)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact [vanam@wustl.edu](mailto:vanam@wustl.edu).

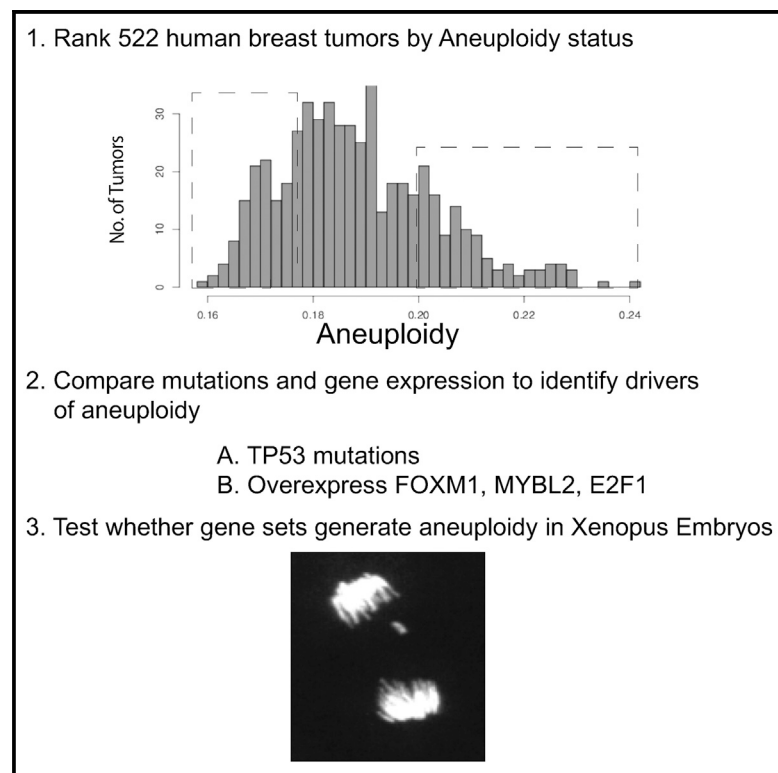
---

## Authors

Katherine Pfister, Justyna L. Pipka, Colby Chiang, Yunxian Liu, Royden A. Clark, Ray Keller, Paul Skoglund, Michael J. Guertin, Ira M. Hall, and P. Todd Stukenberg

## Identification of Drivers of Aneuploidy in Breast Tumors

### Graphical Abstract



### Authors

Katherine Pfister, Justyna L. Pipka, Colby Chiang, ..., Michael J. Guertin, Ira M. Hall, P. Todd Stukenberg

### Correspondence

pts7h@virginia.edu

### In Brief

Pfister et al. analyzed TCGA sequence data to identify drivers of aneuploidy in breast tumors. TP53 is mutated in most aneuploid tumors, and a large number of genes that control mitosis are overexpressed. The oncogenes E2F1, MYBL2, and FOXM1 that regulate mitotic transcription drive the overexpression of mitotic proteins to lower the fidelity of chromosome segregation.

### Highlights

- The authors analyze TCGA sequence data to identify highly aneuploid breast tumors
- Mitotic transcriptional regulators and target genes are overexpressed in aneuploid tumors
- Mutations in TP53 co-associate with the overexpression of mitotic transcription factors
- Overexpression of MYBL2, E2F1, and FOXM1 immediately reduces mitotic fidelity in tissues



# Identification of Drivers of Aneuploidy in Breast Tumors

Katherine Pfister,<sup>1,2</sup> Justyna L. Pipka,<sup>1</sup> Colby Chiang,<sup>1,3</sup> Yunxian Liu,<sup>1</sup> Royden A. Clark,<sup>1</sup> Ray Keller,<sup>2</sup> Paul Skoglund,<sup>2</sup> Michael J. Guertin,<sup>1</sup> Ira M. Hall,<sup>1,3,4</sup> and P. Todd Stukenberg<sup>1,5,\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA

<sup>2</sup>Department of Biology, University of Virginia, Charlottesville, VA 22908, USA

<sup>3</sup>McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA

<sup>4</sup>Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA

<sup>5</sup>Lead Contact

\*Correspondence: [pts7h@virginia.edu](mailto:pts7h@virginia.edu)

<https://doi.org/10.1016/j.celrep.2018.04.102>

## SUMMARY

Although aneuploidy is found in the majority of tumors, the degree of aneuploidy varies widely. It is unclear how cancer cells become aneuploid or how highly aneuploid tumors are different from those of more normal ploidy. We developed a simple computational method that measures the degree of aneuploidy or structural rearrangements of large chromosome regions of 522 human breast tumors from The Cancer Genome Atlas (TCGA). Highly aneuploid tumors overexpress activators of mitotic transcription and the genes encoding proteins that segregate chromosomes. Overexpression of three mitotic transcriptional regulators, E2F1, MYBL2, and FOXM1, is sufficient to increase the rate of lagging anaphase chromosomes in a non-transformed vertebrate tissue, demonstrating that this event can initiate aneuploidy. Highly aneuploid human breast tumors are also enriched in TP53 mutations. TP53 mutations co-associate with the overexpression of mitotic transcriptional activators, suggesting that these events work together to provide fitness to breast tumors.

## INTRODUCTION

Aneuploidy is a hallmark of cancer and is high in breast tumors. Current estimates suggest that 90% of solid tumors contain whole-chromosome gains or losses (Danielsen et al., 2016; Roylance et al., 2011). The amount of aneuploidy between tumors and within a tumor is highly variable, suggesting that in addition to current breast cancer classifications, the degree of aneuploidy could be used to tailor treatments. However, classification of tumors requires either simple assays to measure aneuploidy in tumors or a deeper understanding of the mechanisms that drive aneuploidy.

The fidelity of mitosis is critical in maintaining a normal karyotype. Tumor cells sometimes lower the fidelity of mitosis by unknown mechanisms to generate chromosomal instability (CIN). It is estimated that ~45% of breast tumors develop CIN, while the rest remain karyotypically stable (Lingle et al., 2002;

Roylance et al., 2011). CIN could drive tumor evolution by driving loss of heterozygosity (LOH) of tumor suppressors or by generating imbalances or structural changes that overexpress oncogenes. How tumor cells develop CIN is an important unanswered question. Mutations in mitotic proteins can drive aneuploidy in experimental systems; however, the major mitotic regulators are rarely mutated in cancer (Bakhoun and Compton, 2012; Bakhoun et al., 2009; Barber et al., 2008; Cahill et al., 1998, 1999). There are examples implicating known tumor suppressors and oncogenes in driving CIN; for example, the overexpression of cyclin D in mouse mammary tissue generates a CIN phenotype and late-onset tumors. In addition, the Mad2 spindle checkpoint protein is regulated by E2F and Rb1 and contributes to their potent oncogenic roles (Casimiro and Pestell, 2012; Schwartzman et al., 2011; van Deursen, 2007).

Mutations in the TP53 tumor suppressor are associated with aneuploid tumors (Clark et al., 1992; Kihana et al., 1992), but there is a debate about how it prevents aneuploidy. Current models suggest that mutating TP53 does not lower the fidelity of mitosis; however, cells that arrest in mitosis for an extended period can trigger tp53 responses that senesce cells (Thompson and Compton, 2010; Uetake and Sluder, 2010). In addition, the missegregation of chromosomes can also trigger a tp53 response, although there is a current debate regarding whether this is an immediate response or requires multiple cell divisions (Li et al., 2010; Santaguida et al., 2017; Soto et al., 2017; Thompson and Compton, 2010). Together, these data suggest that loss of tp53 function does not cause aneuploidy, but it is critical to allow aneuploid cells to remain proliferative.

Transcription factors that regulate the transcription of genes required for mitosis can be oncogenes or tumor suppressors (Fischer et al., 2016). It is unclear how they provide a selective advantage to tumor cells and drive tumor progression. The MuvB transcription complex binds cell cycle promoters and associates with MYBL2 and FOXM1 in S-phase/G2 to activate the promoters of a number of genes required for mitosis (Sadasivam et al., 2012). Both MYBL2 and FOXM1 are oncogenes that are overexpressed in a subset of breast tumors (Laoukili et al., 2007). E2F1 drives the transcription of cell-cycle genes in G1 phase, and it plays critical roles in the cell's ability to initiate proliferation, which is a well-established role in tumor development. However, E2F1 also binds the promoters of many mitotic genes that are transcribed in G2, and it is unclear whether this also





contributes to its oncogenic potential (Cao et al., 2011). Interestingly, loss of the tumor suppressor Rb1, which inhibits the E2F1 transcription factor, is associated with aneuploidy (Manning and Dyson, 2012). Meta-analysis of TP53 downregulated genes suggest that MYBL2, FOXM1, and E2F1 cell-cycle transcriptional regulators are important targets of the tp53 response (Fischer et al., 2016).

Measuring CIN requires knowing cell-to-cell variability in chromosome number and structure within a tumor or an assessment of the rate at which chromosomal changes occur. Therefore, it is more feasible to measure aneuploidy within tumors. Chromosomal changes can be measured by multigene fluorescent *in situ* hybridization (FISH), comparative genomic hybridization (CGH), and counting of anaphase lagging chromosomes, but these are difficult to perform on biopsies (Bakhoun and Compton, 2012; McGranahan et al., 2012). Therefore, neither CIN nor aneuploidy is routinely tested in the clinic. An algorithm to measure the consequences of CIN (changes in copy number of whole or parts of chromosomes of dominant cell populations within the tumor) was developed by comparing the overexpression of adjacent sets of genes along chromosomes in transcriptional array data (Carter et al., 2006). Carter et al. coined the term “functional aneuploidy” (FA) to distinguish the direct measurement of aneuploidy from CIN, which measures a rate of missegregation. Several cell-cycle regulators were found in CIN tumors, and they identified a CIN signature based on the top 70 genes that were overexpressed in CIN tumors (CIN70) (Carter et al., 2006). High expression of the top 25 of these genes indicates poor prognosis in 12 studies and six different tumor types (Carter et al., 2006). This report supported a number of other studies where CIN was correlated with poor prognosis by either FISH analysis or CGH (Chandhok and Pellman, 2009). There is an opportunity to learn more about how CIN develops by segregating tumors by aneuploidy status in large well-annotated tumor databases such as The Cancer Genome Atlas (TCGA) (2012).

We developed a computational method for measuring aneuploidy and large structural changes of chromosomes from exome sequence data based on the heterogeneity of allele frequencies. This allowed us to segregate breast tumors from the TCGA datasets based on their amount of aneuploidy. We compared gene expression profiles from tumors with high and low FA to identify the mechanisms that generate highly aneuploid tumors. Our data confirm that there is a striking correlation between mutations in TP53 and FA in breast tumors (Pati et al., 2004; Sigurdsson et al., 2000). Although the genes encoding the proteins that segregate chromosomes were rarely mutated, they dominate the proteins that are specifically overexpressed in highly aneuploid tumors. In addition, regulators of mitotic transcription, MYBL2, E2F1, and FOXM1, were overexpressed in highly aneuploid breast tumors. We expressed these oncogenes in *Xenopus* embryos and followed the first divisions that are driven by transcription to show the sufficiency of the overexpression of these three oncogenes to lower the fidelity of mitosis in non-transformed vertebrate epithelial tissues. Our analysis of human breast tumors suggests a two-hit model for generating aneuploidy or CIN (although it cannot predict which hit comes first). First, the overexpression of the E2F1, FOXM1, and MYBL2 oncogenes lowers the fidelity of chromosome segrega-

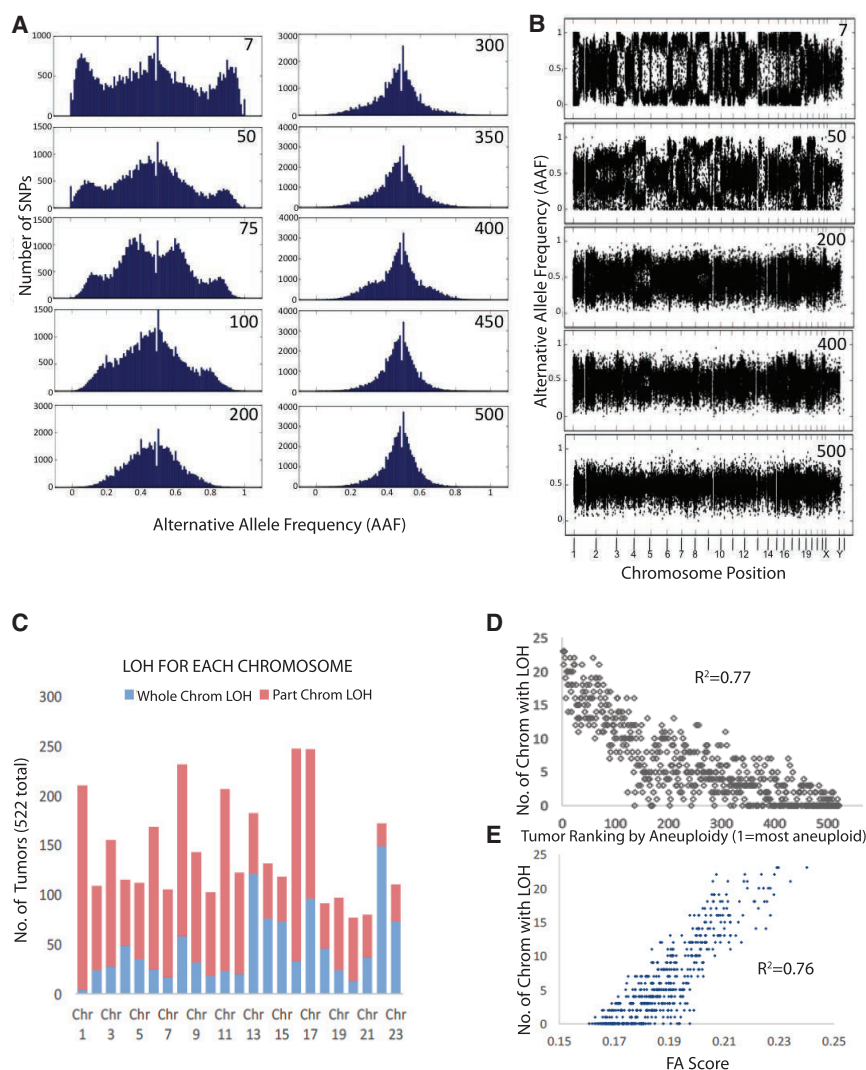
tion. Second, the loss of TP53 function enables cells that missegregate chromosomes to remain proliferative.

## RESULTS

### Measuring the Amount of Aneuploidy in Tumors from the TCGA by Calculating the Variance of Allelic Frequency Ratios

We developed a computational method to measure the amount of aneuploidy in tumors. The method measures changes in copy number of both whole chromosomes and large fragments of chromosomes, which we refer to as FA (“functional aneuploidy”), as originally termed in the CIN70 work (Carter et al., 2006). Aneuploidy changes alternate allele frequencies (AAFs), which can be measured using the exome sequence data available for TCGA samples. AAF in a tumor is a function of the chromosome number in each cell and the percentage of cells that are aneuploid (Figure S1). We developed a simple pipeline to visualize the FA within a single human breast tumor using publicly available data from the TCGA consortium (Figure S2). We identified all heterozygous SNPs for each patient as defined by allele frequencies between 0.3 and 0.7 in the non-transformed exome sequence reads. For each heterozygous SNP, we calculate the AAF in the tumor by dividing the number of sequence reads that identify the alternate allele relative to the total number of reads aligning to that genomic region. We plot histograms of the number of SNPs per AAF for each patient’s tumor (Figure 1A; Data S1). Tumors with minimal FA retained a single peak in the histogram of heterozygous alleles that was normally distributed between AAFs of 0.3 and 0.7, as is seen in all matched samples. In contrast, the range of distributions broadens to include AAFs between 0.1 and 0.9 in tumors whose heterogeneity is driven by events that encompass large fragments or whole chromosomes, since these change the AAF for thousands of SNPs (Figures 1A, S1, and S2). A trimodal AAF distribution was found in many of the tumor plots, even though they lack alleles that were homozygous in the normal matched sample. These trimodal distributions arise from LOH events that convert SNP ratios that were heterozygous and close to 0.5 in the matched normal tissue to ratios that appear near homozygous in the tumor.

The width of these AAF histograms correlates to the amount of FA in each tumor. We measured the width of the peaks for 522 breast tumors of various grades and types from the TCGA dataset by calculating the SD of the allele frequency distribution of all AAFs between 0.1 and 0.9 (Figure S3; Table S1). We decided on this simple approach because it is robust and assumption-free. We used these SDs to rank the tumors, assigning the tumor with the widest peaks (largest SD) as number 1, while the tumor with the narrowest peak (smallest SD) as number 522. Examples of 10 tumors and their relative ranks are shown in Figure 1A. Note that all the tumors ranked below 200 are qualitatively broader than plots generated from matched normal sequence data (compare Figure 1A to the corresponding plots of the normal samples in Figure S4), while most of the tumors ranked above 300 were similar to plots generated from non-tumor sequence datasets (Figure S4). Each of the tumors ranked below 200 has its own characteristic shape and extent of AAF peak broadening, which represents the amount of tumor heterogeneity. There is a



**Figure 1. Measuring the Aneuploidy in Tumors by the Broadening of Allelic Frequency Ratios Calculated from Sequencing Data**

(A) 10 examples of FA histograms, each portraying the allelic frequencies of the heterozygous SNPs of human breast tumors. The number in the top right corner of each plot represents the FA ranking based on the broadening of the allele frequency peaks, with 1 being the widest peak and 510 the narrowest peak. Note that the lower the number, the broader the central peak of allelic frequencies. (B) Replots of 5 of the tumors from (A) showing the allele frequency ratios along chromosome positions corresponding to coordinates in the GRCh37 genome (so each chromosome runs p arm to q arm). Note that the LOH events that separate the central peaks into two often span large segments of chromosomes or whole chromosomes, which is consistent with chromosome missegregation driving the LOH event.

(C) Plot of the frequency of whole (blue) or partial (red) chromosome LOH for each chromosome.

(D and E) The number of chromosomes with an LOH event in breast tumors correlates with the ranking (D) and FA score (E) of the tumors.  $R^2$  values were calculated by fitting to a second-order polynomial curve in Excel. Our method of LOH quantification is summarized in Figure S5.

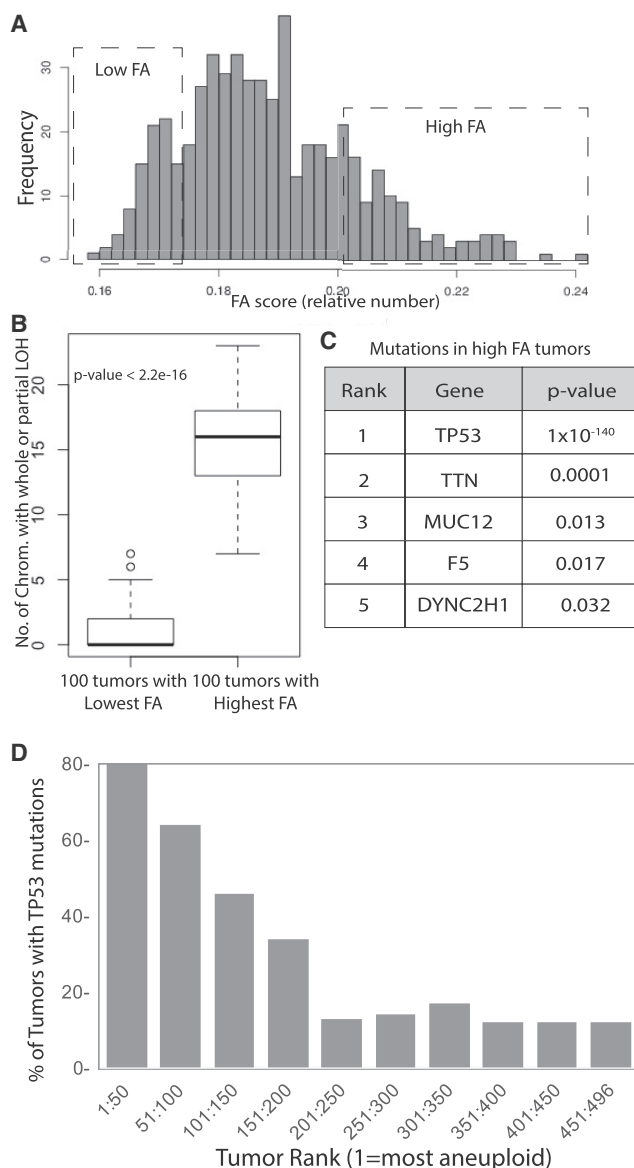
continuum of AAF distributions (Figure S3). The histograms of the ~115 highest FA scored tumors have both broad central peaks as well as outer peaks from LOH (Data S1). Between ~115 and ~226, there are broad central peaks only. In the tumors ranked from 227 to 320, the peaks transition between noticeable broadening and peaks that appear similar to the matched normal samples. The final 200 tumors (321–522) resemble the matched control plots, suggesting that they have low amounts of FA.

For each tumor, we also generated complementary plots that show the chromosome position of each SNPs AAF score (Figures 1B and S2; Data S2). LOH events generate areas with few SNPs in the 0.5 AAF regions, with corresponding dense areas of SNPs scoring closer to zero and one. The AAF frequencies of most SNPs are contiguous along whole chromosomes or large chromosome regions, confirming that aneuploidy and not random point mutations drives the redistribution of AAF frequencies in tumor samples (Figure 1B). We quantified the number of chromosomes with a region of LOH that spanned a large region of a chromosome (scoring method is outlined in Figure S5A; see also Figure 1C). Our

analysis confirmed the high percentages of LOH in chromosomes 11, 16, and 17 found in breast tumors (Lindblom et al., 1993; Nagayama and Watatani, 1993), which validates our analysis (Figure 1C). We also found large amounts of LOH in chromosomes 1 and 8. They were then injected with either control RNA or the RNA encoding the three transcription factors (FOXM1, E2F1, and MYBL2) in two of the four cells of Stage 3 embryos (Figures 5E and 5F). The tumors ranked lower than ~200 showed a stronger correlation of chromosomes with LOH events than the other 322 tumors, which is consistent with about 40% of breast tumors having strong CIN. Whole-chromosome events that are expected from missegregation of chromosomes during mitosis were more strongly represented in the tumors ranked 1–200 ( $R^2 = 0.66$ ; Figure S5B), while they were rare in the tumors ranked from 300 to 522. LOH along large regions of chromosomes, but not spanning the entire chromosome, had similar distributions but were not as tightly linked to our ranking ( $R^2 = 0.53$ ; Figure S5C). Together, these analyses confirm that our method can segregate tumors by the amount of aneuploidy of the cells within the tumor and also that missegregation of whole chromosomes contributes to the FA of high-ranking tumors.

### Identification of the Mechanisms Generating FA in Breast Tumors

We used the Variant Annotation, Analysis and Search Tool (VAAS) (Yandell et al., 2011) to identify the genes that had higher



**Figure 2. Identification of Genes Mutated in Aneuploid Tumors**

(A) Histogram showing the distribution of FA scores of the 522 analyzed human breast tumors. The boxes highlight the distributions of the 100 highest and lowest scoring tumors.

(B) The number of chromosomes with LOH events was compared in the 100 highest and lowest aneuploid tumors to demonstrate that the analysis stratifies tumors by aneuploid status. The p value was generated by Welch two-sample t test in R.

(C) All genes significantly mutated in the high aneuploid tumor sets were identified by comparing the sequence data from the 100 highest and lowest ranked tumors using the VAAST program. The p values were calculated by VAAST.

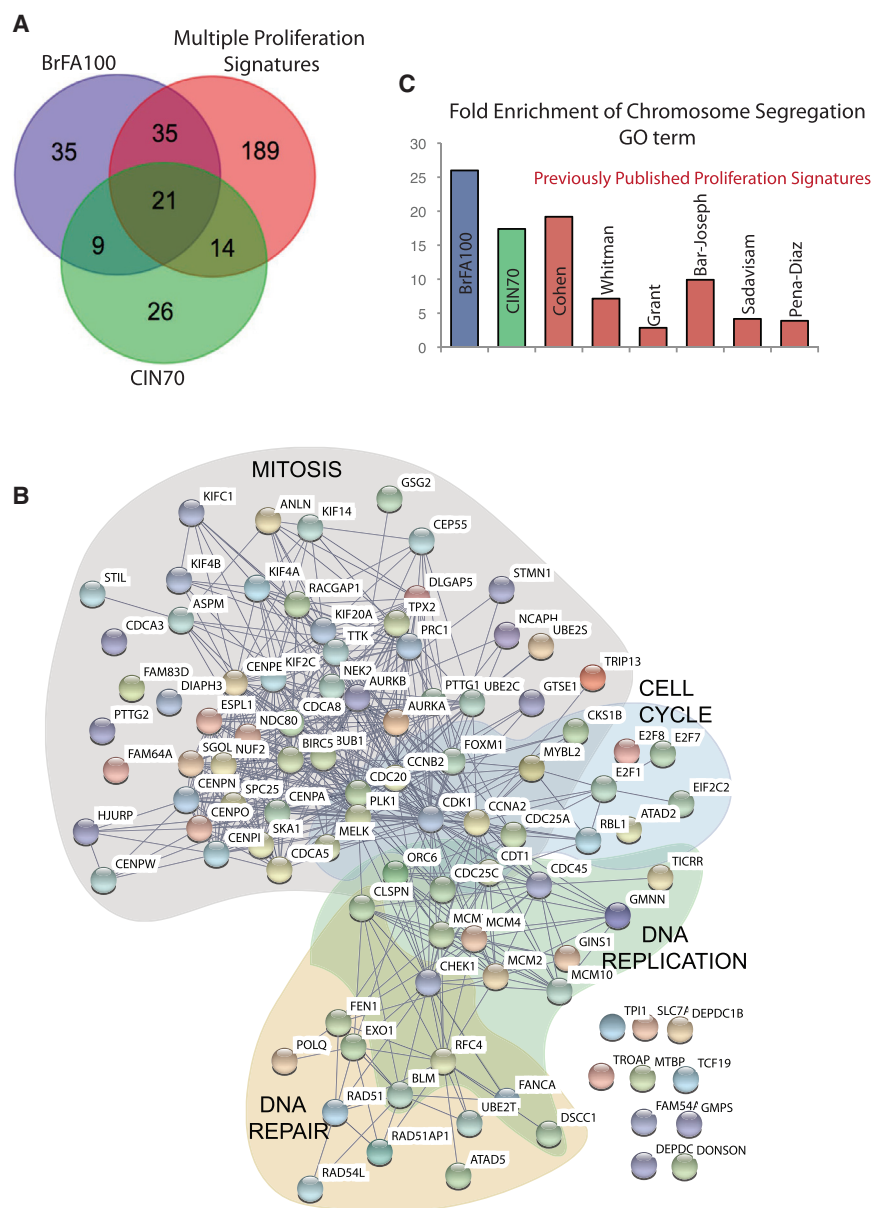
(D) p53 mutations are correlated with functional aneuploidy in the 250 top ranked tumors.

numbers of mutations in the 100 tumors with the highest FA score compared to the 100 with the lowest FA score (dotted boxes in Figure 2A). The average number of chromosomes with an LOH event (whole or partial) in the two groups were 15.6 and 0.97, respectively, demonstrating that we are segregating tumors by

FA status (Figure 2B). Mutations in TP53 were highly enriched in the high FA tumors ( $p = 1.5 \times 10^{-140}$ ; Figure 2C). TP53 mutations have previously been connected to aneuploidy in tumors, validating our method (Li et al., 2010; Thompson and Compton, 2010). The strong association is consistent with a growing body of literature arguing that tp53 arrests non-transformed cells in G1 following a prolonged mitotic arrest or the missegregation of a chromosome (Hayashi et al., 2012; Jeon and Lee, 2013; Krzywicka-Racka and Sluder, 2011; Uetake and Sluder, 2010). Forty of the 50 tumors with the most FA contained TP53 mutations. Interestingly, the percentage of tumors with TP53 mutations decreased as a function of FA rank for the 250 tumors with the most FA (linear-by-linear association test, test statistic = 30.21842;  $p = 3.86 \times 10^{-8}$ ), and then there was poorer correlation in the following 246 tumors that are likely to have low FA status (test statistic = 1.222109;  $p = 2.689 \times 10^{-1}$ ) (Figure 2D) (Agresti, 2007). There were no other genes strongly associated with FA, although mutations in four other genes, including spindle regulator cytoplasmic dynein (heavy chain), were statistically significant. However, the relevance of these genes is unclear, as they tend to be large genes and the enrichment in FA tumors was not very strong. Activating mutations in PIK3CA ( $p = 3 \times 10^{-3}$ ), followed by mutations in MAP3K1 ( $p = 8 \times 10^{-3}$ ), were enriched in the low-FA tumors, although this is probably driven by the predominance of these mutations in luminal tumors that were enriched in the low-FA tumor group (Figure 6). Thus, the only gene whose mutation was strongly enriched in high-FA tumors was TP53.

We also compared the RNA expression data to identify the genes that were overexpressed in FA tumors. All genes were ranked from 1 being the most overexpressed in the high-FA group to ~18,000 being the most overexpressed in the low-FA group. We refer to the Breast FA100 gene set (BrFA100; Table S2) as the 100 genes the were the most overexpressed in the high-FA tumors. The BrFA100 contained 30 genes that were also in the CIN70 list generated by independent methodology in multiple tumor types (Figure 3A; Table S3). Gene ontology (GO) analysis demonstrated that the BrFA100 is highly enriched in genes involved in the cell cycle, as the top two GO terms were cell cycle ( $p = 5.3 \times 10^{-81}$ ) and cell-cycle phase ( $p = 1.3 \times 10^{-77}$ ). “Cell cycle” is a broad GO term set that consists of 1,355 genes, so we asked which stage of the cell cycle was best represented. Twenty of the top 30 GO terms represented mitotic segregation of chromosomes, suggesting that genes controlling mitosis were driving this gene signature (Table S4). We ran the BrFA100 gene sets through the STRING database, which uncovered a highly noded network of 58 critical mitotic regulators (Figure 3B), including nine mitotic kinases (AURKA, AURKB, TTK/MPS1, GSG2/Haspin, CHEK1, NEK2, BUB1, CDK1, and MELK) and 13 kinetochore proteins (Table S2). There was also a subset of cell-cycle regulators, but these were dominated by the proteins that control the entry into mitosis and contained proteins such as Cyclin B2/CCNB2, Cyclin A2/CCNA2, CDK1, CKS1B, CDC25A, and CDC25C. Finally, there were 17 genes associated with DNA replication (DNA Rep), and 14 in DNA repair. These data corroborate the CIN70 study (Carter et al., 2006) and greatly extend the idea that aneuploidy (and, likely, CIN) is associated with overexpression of mitotic proteins.





**Figure 3. Regulators of Mitosis Are Overexpressed in Aneuploid Breast Tumors**

The 100 most overexpressed genes in the high-FA-ranking tumors (BrFA100) were identified by comparing RNA expression data of the 100 high-FA and low-FA tumors.

(A) The overlap of the BrFA100 genes, the CIN70 list, and genes present in 3–6 previously published proliferation signatures (Multiple Proliferation Signatures).

(B) STRING diagrams (<http://string-db.org>) of the BrFA100 list show a highly noded grouping of mitotic regulators, mitotic cell-cycle genes, and DNA replication and repair proteins. The top 10 GO terms of the BrFA100 list shows a strong enrichment of cell-cycle genes, which is driven by a high M-phase gene enrichment (Table S4).

(C) Specifically, we plotted the relative fold enrichment of genes in the chromosome segregation GO term in the BrFA100, CIN70, and six different proliferation signatures. See also Table S4.

our 100 proteins, and 6 of these control the cell cycle (MYBL2, FOXM1, E2F1, RBL1/p107, E2F7, and E2F8). Specifically, the oncogenes MYBL2, FOXM1, and E2F1 (ranked 9th, 10th, and 38th) were of interest due to their roles as master transcriptional activators of genes required for mitosis (Cao et al., 2011; Grant et al., 2013; Sadasivam and DeCaprio, 2013; Sadasivam et al., 2012). Combining the target genes of the DREAM, MMB, and FoxM1/MuvB transcriptional complexes accounts for 92 of the 100 genes in the BrFA100 (Figure 4A) (Fischer et al., 2016). Overexpression of FOXM1, MYBL2, and E2F1 mRNA ( $z > 2$ ) correlated with FA scores (Figure 4B). Overexpression of these three genes was correlated with FA status in all four breast tumor subtypes (Figure S5D).

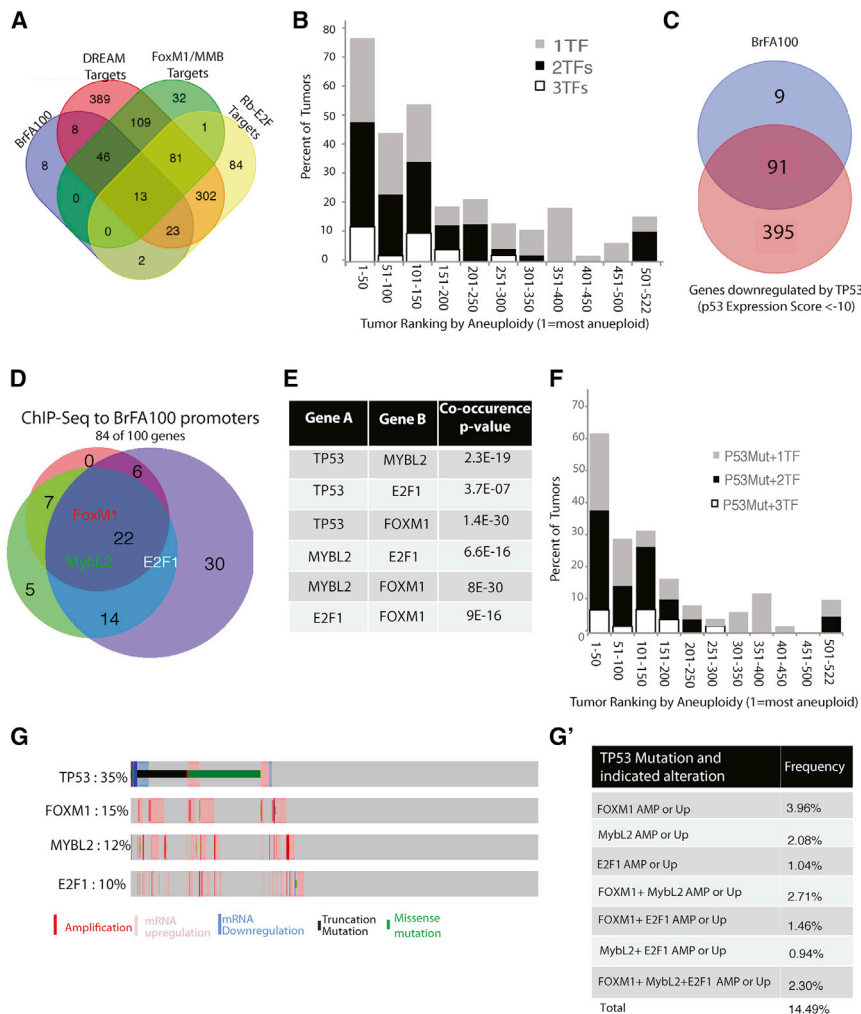
There was also strong overlap of the BrFA100 with genes identified in a meta-analysis as downregulated by TP53 (Figure 4C). The drivers of this gene signature were shown to be DREAM, FOXM1, and MMB. The BrFA100 contained 91 of the top 486 genes that were overexpressed after multiple treatments that downregulated tp53 function.

We mined publically available chromatin immunoprecipitation followed by sequencing (ChIP-seq) datasets to examine whether MYBL2, FOXM1, and E2F1 bind the promoters of the genes in the BrFA100 lists (Cao et al., 2011; Grant et al., 2013; Sadasivam et al., 2012). At least one of the three transcription factors binds to the promoter of 82 of the 100 genes, and 49 of the BrFA100 genes are bound by at least two of the transcription factors (Figures 4D and S6B; Table S3). Twenty-two of the genes bound all three transcription factors, including the three most overexpressed proteins TPX2, Plk1, and Aurora B/AURKB. For this

Fifty-six of the BrFA100 genes are found in at least 3 of 6 proliferation signatures (Figures 3A and S6A) (Cohen et al., 2013; Fischer et al., 2016). We found that most proliferation signatures contained about 10-fold enrichment of mitotic regulators, while the CIN70 contained ~18-fold enrichment and the BrFA100 contained over 25-fold enrichment (Figure 3C). We conclude that aneuploid tumors overexpress a large number of genes associated with proliferation, but there is stronger enrichment for genes involved in mitosis.

### Mitotic Transcriptional Regulators Drive the Overexpression Patterns Seen in High-FA Tumors

To determine the mechanism of overexpressing the mitotic genes, we asked whether any transcription factors were overexpressed in high-FA tumors. There were 8 transcription factors in



**Figure 4. Mutations in TP53 and Overexpression of E2F1, MYBL2, and FOXM1 Are Highly Associated in Breast Tumors**

(A) The overlap of BrFA100 and target genes of mitotic transcriptional regulation complexes DREAM, FoxM1-MuvB/MMB, and Rb-E2F.

(B) The percentage of tumors in each group of 50 (ranked by aneuploid status) with 1, 2, or 3 of the transcription factors MYBL2, FOXM1, and E2F1.

(C) Venn diagram of the overlap of the BrFA100 with the top 400 genes downregulated by TP53 (p53 expression score of less than  $-10$ , as listed in Fischer et al., 2016).

(D) Venn diagram to show the overlap of ChIP-seq datasets for E2F1, MYBL2, and FOXM1 with the BrFA100 list. Gene lists are shown in Table S3.

(E) Association p values of TP53, E2F1, MYBL2, and FOXM1 as individual pairs. p values were obtained through Fisher exact tests with Benjamini-Hochberg multiple test corrections.

(F) The percentage of tumors in each group of 50 that have a TP53 mutation and 1, 2, or 3 overexpressed transcription factors.

(G) Association of TP53, MYBL2, E2F1, and FOXM1 in 960 human breast tumors of the TCGA. Plots were generated at the cBioPortal ([www.cbioportal.org](http://www.cbioportal.org)).

(G') The percentage of the 960 tumors with a TP53 mutation and either an amplification (AMP) of the gene as defined by a positive GISTIC score or an upregulation (Up) of the mRNA as defined by a Z score  $> 2$ . TF, transcription factor.

See also Table S3.

analysis, we used a subset of the FOXM1 ChIP-seq data that was listed as the 270 “high-confidence” binding regions; interestingly, they contained 35 of the BrFA100 genes (exact hypergeometric probability,  $p < 9 \times 10^{-18}$ ). These data suggest that the overexpression of the three transcription factors could drive the overexpression of mitotic proteins that we find overexpressed in functionally aneuploid breast tumors. Interestingly, E2F1 and FOXM1 were members of the 16 genes that were not bound by any of the transcription factors, while MYBL2 was only in the E2F1 ChIP-seq set. This suggests that the mechanisms that overexpress these oncogenes are independent of the events that cause the expression of the bulk of the BrFA100 list.

We conclude that the BrFA100 signature can be explained by the overexpression of the mitotic transcriptional regulators MYBL2, FOXM1, and E2F1 and that this correlation is stronger (91/100 genes) than general proliferation signatures (56/100).

### A Two-Event Model for Aneuploidy in Breast Tumors

To determine whether TP53 works in conjunction with the mitotic transcriptional regulators to generate FA, we asked whether the

tumors that contained mutated TP53 also overexpressed E2F1, FOXM1, and MYBL2. The co-association of every combination of TP53 mutations and the overexpression of MYBL2, FOXM1, or E2F1 are highly significant and provided in Figure 4E. This association suggests that these genes work together to provide an advantage to tumor progression.

Overall, our bioinformatics suggest a two-event model for the generation of aneuploidy in breast tumors. First, the overexpression of the oncogenes MYBL2, E2F1, and FOXM1 is the causative event of the chromosome missegregation, perhaps because this event leads to the transcriptional overexpression of a large number of mitotic regulators. Second, loss of tp53 is required, which, we postulate, allows the cells that missegregate chromosomes to remain proliferative as suggested by the literature (Hayashi et al., 2012; Jeon and Lee, 2013; Krzywicka-Racka and Sluder, 2011; Santaguida et al., 2017; Uetake and Sluder, 2010). To estimate the number of aneuploid tumors that could be explained by this two-hit model, we ranked the percentages of tumors that had both TP53 mutations and overexpressed the transcription factors according to our FA scores (Figure 4F). Sixty percent of the 50 top tumors could be explained by this two-hit model and the co-occurrence of TP53 mutations, and overexpression of the mitotic transcription factors correlated strongly with FA score. In Figures 4G and 4G', we visualize the

co-occurrence in over 900 breast tumors. This analysis also shows the co-occurrence of TP53 mutations and overexpression of one of the three transcription factors in 14.5% of all breast tumors (Figure 4G'). It makes the point that these events may have oncogenic potential beyond generating aneuploidy, since there is a small set of tumors that either have TP53 mutated without overexpressing mitotic transcriptional regulators or that overexpress either MYBL2, FOXM1, or E2F1 without mutating TP53. Overall, our data suggest that 39.6% of the 100 most aneuploid breast tumors become aneuploid because they lose tp53 function and overexpress one of the transcriptional regulators of mitosis ( $Z > 2$ ). If we increase the stringency to requiring overexpression of 2 of the 3 transcription factors, then approximately 25.4% of highly aneuploid tumors would be explained by the model.

### Overexpression of MYBL2, E2F1, and FOXM1 Generates Lagging Anaphase Chromosomes in Non-transformed Vertebrate Epithelial Tissues

We developed a system to connect the results of our bioinformatics approach to missegregation of chromatids in mitosis. Our goal was to visualize the first mitosis after overexpression of FoxM1, E2F1, and MybL2 in a non-transformed vertebrate tissue. To do this, we used early embryos of the frog *Xenopus laevis*. RNA encoding transcription factors can be injected into embryos and translated into proteins; however, they are not active during the first 12 division cycles that lack transcription (Newport and Kirschner, 1982). After this time (~8 hr post-fertilization), the embryo undergoes the mid-blastula transition (MBT) when transcription begins. Therefore, the immediate action of transcription factors can be measured by visualizing the mitoses of divisions directly after MBT (stages 8 and 9).

We injected mRNA encoding the human MYBL2, E2F1, and FOXM1 transcription factors into two-cell *Xenopus* embryos. For controls, we injected mRNAs with a stop codon inserted after 10 amino acids of each transcription factor. Western blot analysis of post-MBT embryo lysates confirmed the expression of hMYBL2, hE2F1, and hFOXM1 in the embryos injected with the functional RNAs (Figure 5A). The epithelial tissue above the blastocoel (animal cap) can be easily dissected from the embryos without affecting normal cell divisions, providing an opportunity to collect relatively flat tissue from multiple embryos. The animal cap tissue from FOXM1, MYBL2, and E2F1 (triple)-overexpressing or control RNA-expressing embryos was fixed, stained with the DNA stain (TOPRO-3), and imaged on a confocal microscope to visualize the chromosomes (representative images are shown in Figure 5B). Individual chromatids between segregating anaphase chromosome masses (lagging chromatid) could be detected in 25% of all anaphase events in triple-overexpressing embryos, compared to less than 5% in singly injected control RNA (30 pg) and 8% in embryos injected with three times the amount of control RNA (100 pg) (Figure 5C). In embryos where RNA for only one of the transcription factors was injected, there was an increase in the number of lagging chromatids over controls, but in each case, these values were also significantly lower than the triple-overexpressing clutchmates. To further characterize these mistakes in mitosis, we also examined micronuclei in fixed animal cap tissue. Anaphase

lagging chromosomes often become micronuclei in the subsequent interphase, because nuclear envelope formation happens before the lagging chromatid joins the mass of segregating chromatids (Zhang et al., 2013). Approximately 3.5% of triple-overexpressing cells had visible micronuclei, while the single-overexpressing embryos had micronuclei in 1.5%–2% of cells. Again, the triple-injected embryos displayed a significantly higher percentage of micronuclei than both controls or singly overexpressing clutchmates (Figure 5D).

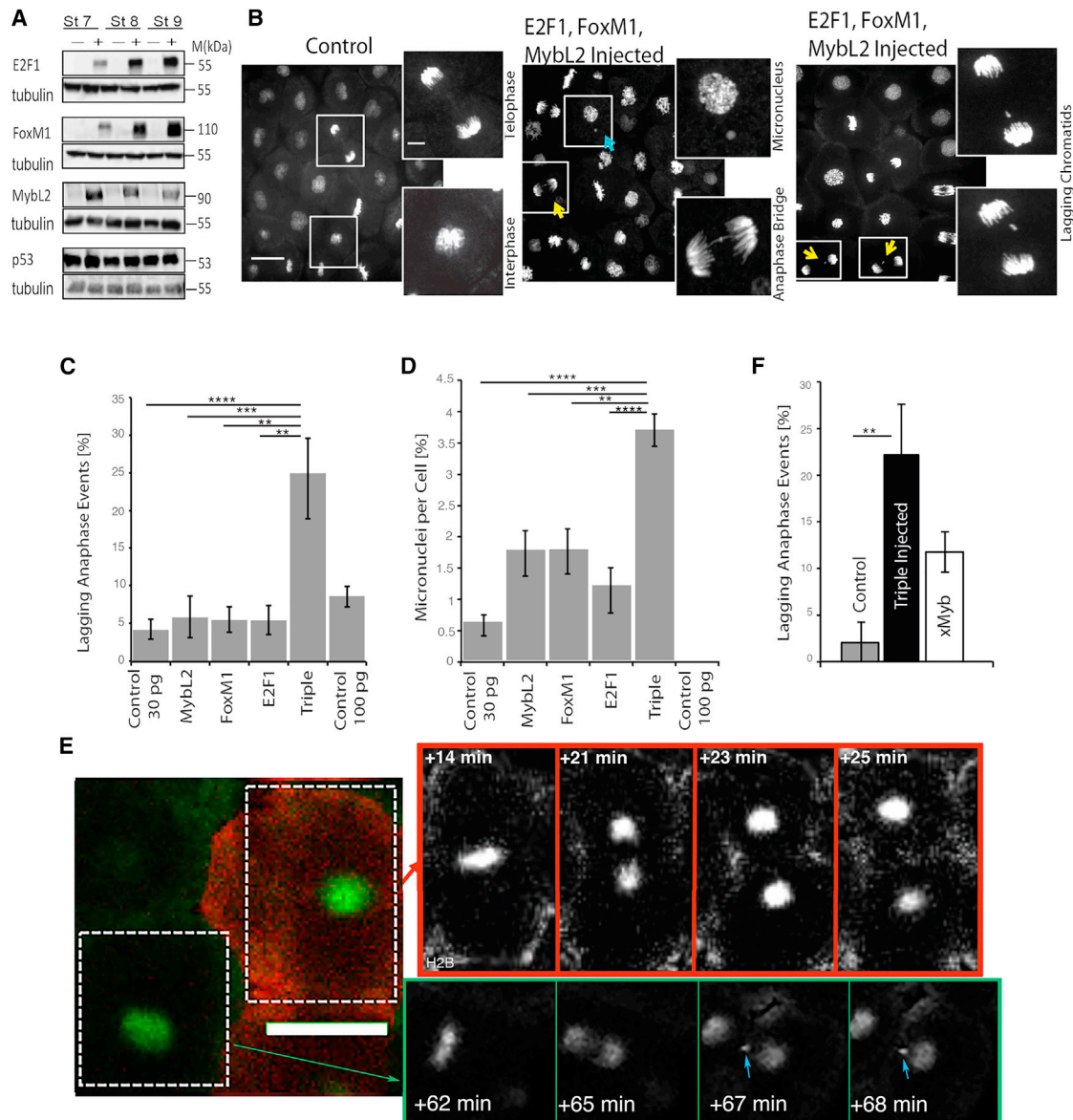
To confirm that the lagging chromatids and micronuclei were caused by missegregation of chromosomes, we performed time-lapse confocal microscopy of animal caps directly after MBT. Moreover, we designed an experiment where the controls and triple-injected cells could be visualized in the same animal cap to control for frog-to-frog variability. In these experiments, embryos were injected at the one-cell stage with RNA for a GFP-tagged histone H2B to follow mitotic events. They were then injected with either control RNA or the RNA encoding the three transcription factors (FOXM1, E2F1, and MYBL2) in two of the four cells of stage 3 embryos (Figures 5E and 5F). Fluorescent dextran was used to label the cells that were either control or triply overexpressing, and there was no experimental variability if the dextran was injected into control or triple-overexpressing cells (data not shown). After MBT, this treatment will generate areas of the animal cap that contain both control and triply overexpressing cells that can be distinguished by the fluorescent dextran. More than 20 videos of animal caps from separate *in vitro* fertilizations were made and analyzed. Again, we saw a significantly higher level of abnormal divisions in the triply overexpressing half of animal caps, compared to the internal control (Figure 5E). Embryos overexpressing only one of the transcription factors—in this case, the *Xenopus* version of MybL2—showed an intermediate value of lagging chromatids during mitosis, similar to the trend seen in fixed cap analysis (Figure 5F). We conclude that overexpression of three oncogenic transcription factors in human breast tumors was sufficient to generate aneuploidy in a nontransformed vertebrate epithelial tissue.

We also used the *Xenopus* system to determine whether preventing new translation of tp53 lowered chromosome segregation. However, this analysis is complicated by the fact that tp53 protein is maternally loaded, and the protein is required for early embryonic development (Wallingford et al., 1997). We created two Morpholino oligonucleotides specific to *Xenopus* TP53. While injection of either Morpholino had little effect on protein levels at stage 9, we saw a significant decrease in protein levels by later stages (Figure S6C). We find that there is no increase in the number of lagging chromosomes or micronuclei upon injection of TP53 Morpholino at stage 9 (Figure S6D). These data are consistent with a large body of literature that suggests TP53 mutations do not directly lower the fidelity of mitosis (Santaguida et al., 2017; Soto et al., 2017; Thompson and Compton, 2010; Uetake and Sluder, 2010).

### Segregation of Human Breast Tumors by FA Status

Tumors of all four subtypes of breast cancer (basal-like, luminal A, luminal B, and Her2 enriched) were represented in the 522 tumors analyzed by our computational method. All four of the tumor subtypes contained both high- and low-FA-score tumors





**Figure 5. Overexpression of hE2F1, hFOXM1, and hMYBL2 Is Sufficient to Generate CIN Phenotypes in *Xenopus* Embryos**

(A) 2-cell-stage embryos were injected with either RNA containing stop codon after 33 nt (-) or functional hE2F1, hFOXM1, and hMYBL2 (+), as detected by western blot.

(B) Representative images of TOPRO-stained normally dividing animal caps and two of the most common CIN phenotypes seen in triple-overexpressing embryos. Yellow arrows indicate a lagging chromatid; blue arrow indicates a micronucleus.

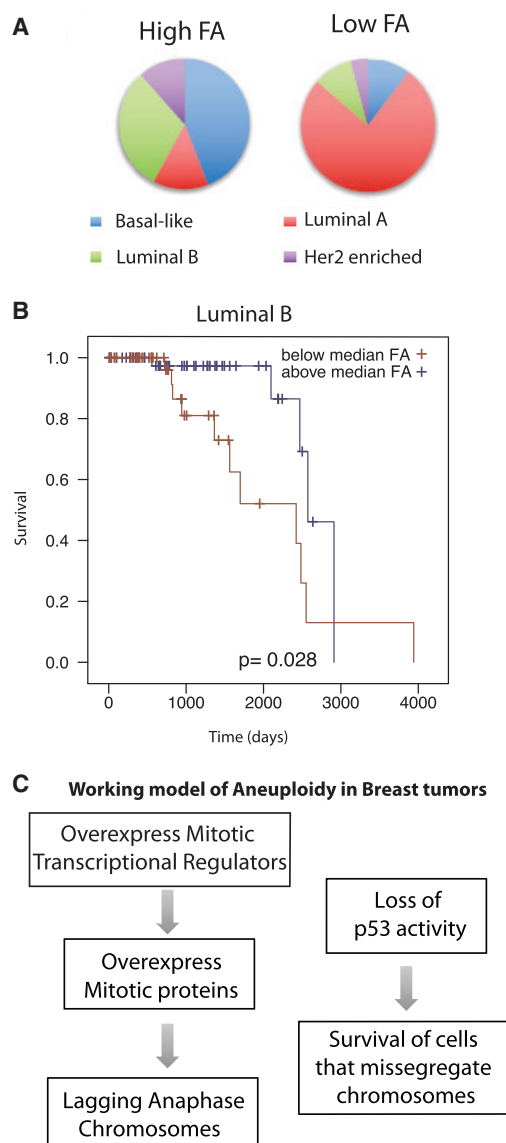
(C and D) Quantification of lagging chromatids (C) and micronuclei (D) in control, triply overexpressing, or singly injected embryos through fixed-animal cap analysis.

(E) Representative time-lapse series of an animal cap expressing H2B:GFP with normally dividing control cells (co-injected with Ruby-Dextran) or CIN-like phenotypes seen in neighboring triple-overexpressing cells. Blue arrows indicate abnormal divisions seen as lagging chromatids and micronuclei. Time points chosen to show anaphase events.

(F) Quantification of lagging chromatid events as seen in time-lapse videos of control embryos, triply overexpressing embryos, and overexpression of only xMYBL2. Full supplemental videos are available upon request. Scale bars represent 40  $\mu$ m in all images. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; \*\*\*\* $p < 0.0001$ , one-way ANOVA and Bonferroni post-test statistics. Error bars represent  $\pm$ SEM. 8 hpf, 8 hr post-fertilization.

(Figure 6A). Over one third of the high FA tumors were of the Basal-like subtype, compared to the low-FA group, which only contained 8 basal tumors. Luminal A tumors dominated the low FA group. These data suggest that aneuploidy could be used to further stratify tumors beyond current diagnostics.

It has been previously suggested that CIN is a predictor of prognosis. We generated Kaplan-Meier plots to compare the 200 highest and lowest FA-scoring tumors for all four tumor types. We did not have enough statistical power to predict diagnosis for most tumor subtypes. However, women with luminal B tumors fared



**Figure 6. Characterization of the Tumors Scored as High and Low FA**  
(A) Tumor subtype distribution of the 100 tumors scored as the highest FA and lowest FA.  
(B) Kaplan-Meier curve demonstrating that FA status indicates good prognosis for the luminal B subtype of tumors.  
(C) Our two-hit model for the generation and propagation of functional aneuploidy; note that we do not indicate which event takes place first.

better if they had high FA scores (Figure 6B). The concept that patients with high levels of CIN have improved prognoses is consistent with a number of previous studies (Birkbak et al., 2011; Roylance et al., 2011; Zhang et al., 2016). We conclude that our algorithm to measure FA has prognostic value for breast tumors.

## DISCUSSION

We have used a bioinformatic approach to segregate breast tumors in the TCGA that are highly aneuploid (both whole-

chromosome loss and loss of large fragments of chromosomes) from breast tumors of more normal ploidy. TP53 was the only protein whose loss of function was strongly associated with the high-FA tumors. In addition, our analysis argues that the overexpression of oncogenes E2F1, FOXM1, and MYBL2, which we find overexpressed specifically in high-FA tumors, is sufficient to lower the fidelity of chromosome segregation in vertebrate epithelial tissues. The frequency of anaphase lagging chromosomes was consistent with those predicted for optimal tumor progression by CIN (Bakhoum and Compton, 2012). Not only are E2F1, FOXM1, and MYBL2 oncogenes, but they also have a role in non-transformed tissues as master controllers of the mitotic transcriptional program. In fact, E2F1, FOXM1, and MYBL2 bind the promoters of a large percentage of the mitotic regulators that we also found overexpressed in high-FA breast tumors.

## Bioinformatic Methods to Measure FA

There is an unmet need for a simple reliable measurement of aneuploidy or CIN status for tumors (McGranahan et al., 2012; Zasadil et al., 2016). Our data suggest that a combination of TP53 mutation and overexpression of MYBL2, FOXM1, and E2F1 may be diagnostically relevant. In addition, as tumor sequencing becomes more prevalent, our bioinformatic method provides the basis for a simple diagnostic test to measure FA in a human tumor.

Our FA algorithm is an improvement in several ways on the CIN70 algorithm, which has been useful and is the only available technology to measure FA from genomic data (Carter et al., 2006). First, our scoring is based on a direct measurement of genome-wide tumor variation, not simply on correlations of gene expression data. Second, the increase in publically available datasets allowed us to focus on mechanisms for breast tumors where the CIN70 work required the mixing of tumor types. Our methodology should be generalizable to other tumor types. Third, we generated not only a score for approximating CIN but also two plots that provide great insight into the subtleties and complexities of heterogeneity within the tumor. Fourth, it is compatible with next-generation sequencing, which will be the future of personalized medicine.

## A Model for Generation of Aneuploidy in Breast Tumors

Our data support a two-event model for generation of FA (Figure 6C). First, TP53 is a CIN suppressor, and our data support previous findings that the loss of tp53 function is strongly associated with aneuploidy in tumors (Li et al., 2010; Thompson and Compton, 2010). Second, overexpression of mitotic transcriptional regulators drive the missegregation of chromosomes in mitosis. While it is unlikely that this is the only mechanism to generate CIN, we calculate that this two-step mechanism may act in about 40% of the most aneuploid breast tumors and 15% of all breast tumors.

Our data are consistent with a growing body of literature reporting that an important role of tp53 is to trigger senescence or cell-death pathways after the missegregation of chromosomes (Hinchcliffe et al., 2016; Lambrus et al., 2015; Orth et al., 2012; Uetake and Sluder, 2010), and we suggest that aneuploidy or CIN usually develops after this pathway is lost. The strong co-association of TP53 mutations with the



overexpression of MMB/FOXM1 mitotic regulators has also been previously shown; in fact, MMB/FOXM1 drives the transcriptional responses that are found after tp53 loss. Our data suggest that the evolutionary advantage of this co-association would be the generation of genomic instability through CIN, although it is formally possible that the MMB/FOXM1 mitotic regulators provide a function that allows tumors to survive the loss of tp53 without causing aneuploidy. To distinguish between these models, we tested whether overexpression of MYBL2, FOXM1, and E2F1 generated CIN phenotypes. In untransformed *Xenopus* embryonic tissue, mitotic transcriptional regulator overexpression was sufficient to generate anaphase lagging chromatids at rates comparable to those of CIN in tumors, providing a strong causal link between mitotic transcriptional regulation and CIN.

We don't know how E2F1, FOXM1, and MYBL2 overexpression lowers the fidelity of mitosis, but the simplest model is that they drive the overexpression of a large number of mitotic regulators, which lowers the robustness of mitotic pathways. For example, the proteins involved in the resolution and prevention of merotelic kinetochore microtubule attachments (Aurora B, Borealin, Survivin, Bub1, MPS1, Plk1, Sgo1, and Ndc80) are all regulated by E2F1 and FOXM1/MMB and overexpressed in human CIN breast tumors, which could be the underlying cause of lagging anaphase chromatids. It is also consistent with recent studies that FOXM1 is a central regulator of kinetochore gene transcription and the fact that centromere gene expression can predict patient outcomes (Banerjee et al., 2014; Laoukili et al., 2005; Thiru et al., 2014; Zhang et al., 2016).

We suggest that the *Xenopus* animal cap system is an ideal system to assay the development and the consequences of FA because it combines (1) very low background missegregation rates with (2) outstanding observation of mitotic events by microscopy and (3) the employment of a non-transformed vertebrate tissue that would become the skin and neural tissues of a frog, if not manipulated; (4) the ability to rapidly manipulate these events by loss- and gain-of-function methodologies; and (5) the ability to quickly replace loss-of-function experiments with mutants.

## EXPERIMENTAL PROCEDURES

Further details of all procedures can be found in the [Supplemental Experimental Procedures](#).

### *Xenopus* Embryo Injection and Analysis

Capped RNA was diluted to 25 pg/nL at 2.5-nL injections, with a final concentration of 30–100 pg per embryo. Animal caps were cut and visualized at stages 8–9. Embryos were injected with transcription factor (TF)-specific or control RNAs (dual stop codons 10 amino acids downstream of the translational start site); for live-cell imaging, embryos were coinjected with nuclear label-H2B:GFP in PCS2+ vectors as described earlier in the text.

### Scoring of FA from Exome Sequencing Data

The FA score was calculated from tumor exome sequencing data by measuring the variant allele frequency (also known as “allele balance”) distribution at heterozygous germline SNPs, where heterozygous germline SNPs are defined by variant calls and genotypes in the matched normal data. The variant allele frequency (VAF) of each SNP is calculated by dividing the number of reads containing the alternate SNP allele by the total number of

reads that align to that base in the reference genome. In high-FA tumors, numerous LOH and CNA events will cause VAF deviations at large numbers of SNPs across the entire genome, resulting in markedly broader VAF distributions that are obvious from genome-wide VAF histograms (as in [Figure 1](#)). The FA score is derived from the SD of VAF measurements in the tumor exome sequencing data across all heterozygous germline SNPs.

### Quantification and Statistical Analysis

All statistical tests are specified in the text or figure legends, and details can be found in the [Supplemental Experimental Procedures](#).

## DATA AND SOFTWARE AVAILABILITY

Contact Ira Hall ([ihall@wustl.edu](mailto:ihall@wustl.edu)) for MATLAB code to generate allele frequency plots.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, four tables, and two data files and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.04.102>.

## ACKNOWLEDGMENTS

This work was initiated by a pilot project from the University of Virginia Cancer Center (P30 CA044579). K.P. was supported by grants DR03437, R01 GM099108, R01 GM118798, T32 CA009109, and T32 GM008136. P.T.S. was supported by grant HD069352 from the NICHD and grants R01 GM118798 and DR03437 from the Charlottesville Women's Four Miler. We acknowledge the Keck Center for Cellular Imaging for the usage of the Zeiss 780 confocal microscope (OD016446). An NIH New Innovator Award (DP2OD006493-01) and a Burroughs Wellcome Fund Career Award funded I.M.H., C.C., and R.A.C. Full supplemental videos are available upon request.

## AUTHOR CONTRIBUTIONS

P.T.S. oversaw all aspects of the project; performed the analysis in [Figures 1C–1E, 2, 3, 4, and 6A](#); and wrote the paper with K.P. I.M.H., R.A.C., and C.C. performed the bioinformatic analysis for the original CIN screen and generated the plots in [Figure 1](#), [Data S1](#), and [Data S2](#). They also performed the analysis to identify the genes mutated and overexpressed in the CIN tumor set. K.P. and J.L.P. performed the *Xenopus* embryo experiments for [Figure 5](#). M.J.G. and Y.L. generated [Figures 6B and S5D](#). P.S. and R.K. provided funds to support K.P.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 11, 2017

Revised: January 4, 2018

Accepted: April 25, 2018

Published: May 29, 2018

## REFERENCES

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, Second Edition (Wiley-Interscience Publications).
- Bakhom, S.F., and Compton, D.A. (2012). Chromosomal instability and cancer: a complex relationship with therapeutic potential. *J. Clin. Invest.* 122, 1138–1143.
- Bakhom, S.F., Genovese, G., and Compton, D.A. (2009). Deviant kinetochore microtubule dynamics underlie chromosomal instability. *Curr. Biol.* 19, 1937–1942.
- Banerjee, B., Kestner, C.A., and Stukenberg, P.T. (2014). EB1 enables spindle microtubules to regulate centromeric recruitment of Aurora B. *J. Cell Biol.* 204, 947–963.

- Barber, T.D., McManus, K., Yuen, K.W., Reis, M., Parmigiani, G., Shen, D., Barrett, I., Nouhi, Y., Spencer, F., Markowitz, S., et al. (2008). Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proc. Natl. Acad. Sci. USA* **105**, 3443–3448.
- Birkbak, N.J., Eklund, A.C., Li, Q., McClelland, S.E., Endesfelder, D., Tan, P., Tan, I.B., Richardson, A.L., Szallasi, Z., and Swanton, C. (2011). Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res.* **71**, 3447–3452.
- Cahill, D.P., Lengauer, C., Yu, J., Riggins, G.J., Willson, J.K., Markowitz, S.D., Kinzler, K.W., and Vogelstein, B. (1998). Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303.
- Cahill, D.P., da Costa, L.T., Carson-Walter, E.B., Kinzler, K.W., Vogelstein, B., and Lengauer, C. (1999). Characterization of MAD2B and other mitotic spindle checkpoint genes. *Genomics* **58**, 181–187.
- Cancer Genome Atlas, N.; Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70.
- Cao, A.R., Rabinovich, R., Xu, M., Xu, X., Jin, V.X., and Farnham, P.J. (2011). Genome-wide analysis of transcription factor E2F1 mutant proteins reveals that N- and C-terminal protein interaction domains do not participate in targeting E2F1 to the human genome. *J. Biol. Chem.* **286**, 11985–11996.
- Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N., and Szallasi, Z. (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat. Genet.* **38**, 1043–1048.
- Casimiro, M.C., and Pestell, R.G. (2012). Cyclin d1 induces chromosomal instability. *Oncotarget* **3**, 224–225.
- Chandhok, N.S., and Pellman, D. (2009). A little CIN may cost a lot: revisiting aneuploidy and cancer. *Curr. Opin. Genet. Dev.* **19**, 74–81.
- Clark, J.S., George, W.D., and Campbell, A.M. (1992). Dual colour flow cytometry of p53 and c-erbB-2 expression related to DNA aneuploidy in primary and metastatic breast cancer. *Cancer Lett.* **66**, 193–200.
- Cohen, M., Vecsler, M., Liberzon, A., Noach, M., Zlotorynski, E., and Tzur, A. (2013). Unbiased transcriptome signature of in vivo cell proliferation reveals pro- and antiproliferative gene networks. *Cell Cycle* **12**, 2992–3000.
- Danielsen, H.E., Pradhan, M., and Novelli, M. (2016). Revisiting tumour aneuploidy - the place of ploidy assessment in the molecular era. *Nat. Rev. Clin. Oncol.* **13**, 291–304.
- Fischer, M., Grossmann, P., Padi, M., and DeCaprio, J.A. (2016). Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res.* **44**, 6070–6086.
- Grant, G.D., Brooks, L., 3rd, Zhang, X., Mahoney, J.M., Martynov, V., Wood, T.A., Sherlock, G., Cheng, C., and Whitfield, M.L. (2013). Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Mol. Biol. Cell* **24**, 3634–3650.
- Hayashi, M.T., Cesare, A.J., Fitzpatrick, J.A., Lazzerini-Denchi, E., and Karlseder, J. (2012). A telomere-dependent DNA damage checkpoint induced by prolonged mitotic arrest. *Nat. Struct. Mol. Biol.* **19**, 387–394.
- Hinchcliffe, E.H., Day, C.A., Karanjeet, K.B., Fadness, S., Langfald, A., Vaughan, K.T., and Dong, Z. (2016). Chromosome missegregation during anaphase triggers p53 cell cycle arrest through histone H3.3 Ser31 phosphorylation. *Nat. Cell Biol.* **18**, 668–675.
- Jeon, H.Y., and Lee, H. (2013). Depletion of Aurora-A in zebrafish causes growth retardation due to mitotic delay and p53-dependent cell death. *FEBS J.* **280**, 1518–1530.
- Kihana, T., Tsuda, H., Teshima, S., Okada, S., Matsuura, S., and Hirohashi, S. (1992). High incidence of p53 gene mutation in human ovarian cancer and its association with nuclear accumulation of p53 protein and tumor DNA aneuploidy. *Jpn. J. Cancer Res.* **83**, 978–984.
- Krzywicka-Racka, A., and Sluder, G. (2011). Repeated cleavage failure does not establish centrosome amplification in untransformed human cells. *J. Cell Biol.* **194**, 199–207.
- Lambrus, B.G., Uetake, Y., Clutario, K.M., Daggubati, V., Snyder, M., Sluder, G., and Holland, A.J. (2015). p53 protects against genome instability following centriole duplication failure. *J. Cell Biol.* **210**, 63–77.
- Laoukili, J., Kooistra, M.R., Brás, A., Kauw, J., Kerkhoven, R.M., Morrison, A., Clevers, H., and Medema, R.H. (2005). FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nat. Cell Biol.* **7**, 126–136.
- Laoukili, J., Stahl, M., and Medema, R.H. (2007). FoxM1: at the crossroads of ageing and cancer. *Biochim. Biophys. Acta* **1775**, 92–102.
- Li, M., Fang, X., Baker, D.J., Guo, L., Gao, X., Wei, Z., Han, S., van Deursen, J.M., and Zhang, P. (2010). The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. *Proc. Natl. Acad. Sci. USA* **107**, 14188–14193.
- Lindblom, A., Rotstein, S., Skoog, L., Nordenskjöld, M., and Larsson, C. (1993). Deletions on chromosome 16 in primary familial breast carcinomas are associated with development of distant metastases. *Cancer Res.* **53**, 3707–3711.
- Lingle, W.L., Barrett, S.L., Negron, V.C., D'Assoro, A.B., Boeneman, K., Liu, W., Whitehead, C.M., Reynolds, C., and Salisbury, J.L. (2002). Centrosome amplification drives chromosomal instability in breast tumor development. *Proc. Natl. Acad. Sci. USA* **99**, 1978–1983.
- Manning, A.L., and Dyson, N.J. (2012). RB: mitotic implications of a tumour suppressor. *Nat. Rev. Cancer* **12**, 220–226.
- McGranahan, N., Burrell, R.A., Endesfelder, D., Novelli, M.R., and Swanton, C. (2012). Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep.* **13**, 528–538.
- Nagayama, K., and Watatani, M. (1993). Analysis of genetic alterations related to the development and progression of breast carcinoma. *Jpn. J. Cancer Res.* **84**, 1159–1164.
- Newport, J., and Kirschner, M. (1982). A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* **30**, 687–696.
- Orth, J.D., Loewer, A., Lahav, G., and Mitchison, T.J. (2012). Prolonged mitotic arrest triggers partial activation of apoptosis, resulting in DNA damage and p53 induction. *Mol. Biol. Cell* **23**, 567–576.
- Pati, D., Haddad, B.R., Haegeler, A., Thompson, H., Kittrell, F.S., Shepard, A., Montagna, C., Zhang, N., Ge, G., Otta, S.K., et al. (2004). Hormone-induced chromosomal instability in p53-null mammary epithelium. *Cancer Res.* **64**, 5608–5616.
- Roylance, R., Endesfelder, D., Gorman, P., Burrell, R.A., Sander, J., Tomlinson, I., Hanby, A.M., Speirs, V., Richardson, A.L., Birkbak, N.J., et al. (2011). Relationship of extreme chromosomal instability with long-term survival in a retrospective analysis of primary breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **20**, 2183–2194.
- Sadasivam, S., and DeCaprio, J.A. (2013). The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat. Rev. Cancer* **13**, 585–595.
- Sadasivam, S., Duan, S., and DeCaprio, J.A. (2012). The MuvB complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression. *Genes Dev.* **26**, 474–489.
- Santaguida, S., Richardson, A., Iyer, D.R., M'Saad, O., Zasadil, L., Knouse, K.A., Wong, Y.L., Rhind, N., Desai, A., and Amon, A. (2017). Chromosome mis-segregation generates cell-cycle-arrested cells with complex karyotypes that are eliminated by the immune system. *Dev. Cell* **41**, 638–651.e5.
- Schvartzman, J.M., Duijff, P.H., Sotillo, R., Coker, C., and Benezra, R. (2011). Mad2 is a critical mediator of the chromosome instability observed upon Rb and p53 pathway inhibition. *Cancer Cell* **19**, 701–714.
- Sigurdsson, S., Bödvarsdóttir, S.K., Ananthawat-Jonsson, K., Steinarsdóttir, M., Jonasson, J.G., Ogmundsdóttir, H.M., and Eyfjörð, J.E. (2000). p53 abnormality and chromosomal instability in the same breast tumor cells. *Cancer Genet. Cytogenet.* **121**, 150–155.
- Soto, M., Raaijmakers, J.A., Bakker, B., Spierings, D.C.J., Lansdorp, P.M., Foijer, F., and Medema, R.H. (2017). p53 prohibits propagation of chromosome segregation errors that produce structural aneuploidies. *Cell Rep.* **19**, 2423–2431.

- Thiru, P., Kern, D.M., McKinley, K.L., Monda, J.K., Rago, F., Su, K.C., Tsinman, T., Yarar, D., Bell, G.W., and Cheeseman, I.M. (2014). Kinetochore genes are coordinately up-regulated in human tumors as part of a FoxM1-related cell division program. *Mol. Biol. Cell* 25, 1983–1994.
- Thompson, S.L., and Compton, D.A. (2010). Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. *J. Cell Biol.* 188, 369–381.
- Uetake, Y., and Sluder, G. (2010). Prolonged prometaphase blocks daughter cell proliferation despite normal completion of mitosis. *Curr. Biol.* 20, 1666–1671.
- van Deursen, J.M. (2007). Rb loss causes cancer by driving mitosis mad. *Cancer Cell* 11, 1–3.
- Wallingford, J.B., Seufert, D.W., Virta, V.C., and Vize, P.D. (1997). p53 activity is essential for normal development in *Xenopus*. *Curr. Biol.* 7, 747–757.
- Yandell, M., Huff, C., Hu, H., Singleton, M., Moore, B., Xing, J., Jorde, L.B., and Reese, M.G. (2011). A probabilistic disease-gene finder for personal genomes. *Genome Res.* 21, 1529–1542.
- Zasadil, L.M., Britigan, E.M., Ryan, S.D., Kaur, C., Guckenberger, D.J., Beebe, D.J., Moser, A.R., and Weaver, B.A. (2016). High rates of chromosome missegregation suppress tumor progression but do not inhibit tumor initiation. *Mol. Biol. Cell* 27, 1981–1989.
- Zhang, C.Z., Leibowitz, M.L., and Pellman, D. (2013). Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev.* 27, 2513–2530.
- Zhang, W., Mao, J.H., Zhu, W., Jain, A.K., Liu, K., Brown, J.B., and Karpen, G.H. (2016). Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. *Nat. Commun.* 7, 12619.

**Cell Reports, Volume 23**

## **Supplemental Information**

### **Identification of Drivers of Aneuploidy in Breast Tumors**

**Katherine Pfister, Justyna L. Pipka, Colby Chiang, Yunxian Liu, Royden A. Clark, Ray Keller, Paul Skoglund, Michael J. Guertin, Ira M. Hall, and P. Todd Stukenberg**

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### ***Xenopus Embryo preparation and injection***

Embryos were obtained from *X. laevis* females which were injected with 800 U of human chronic gonadotropin into the dorsal lymph sac 16h before use. Eggs were laid into 1/3x MBS (88 mM NaCl, 1 mM KCl, 0.7 mM CaCl<sub>2</sub>, 1mM MgSO<sub>4</sub>, 5mM HEPES (pH7.8), 2.5 mM NaHCO<sub>3</sub>, pH 7.8) and fertilized by adding macerated testes. At 20 min after fertilization, embryos were dejellied in 2% cysteine (in 1/3 MBS, pH 8.0) and rinsed several times with 1/3x MBS. Completely dejellied embryos were maintained in 1/3x MBS at RT until microinjection. Injections were performed in 5% Ficoll solution in 1x MBS at one-cell to four-cell stages depending on the experiment. Half an hour after the last injection, embryos were returned to 1/3xMBS. DNA plasmids containing human E2F1 (HA-E2F-1 wtpRcCMV) and MybL2 (pCDNA3) were purchased from Addgene. The FOXM1 clone was obtained from the human ORFeome and cloned into pCSF107mT through Gateway cloning. The control RNAs were generated by inserting a stop codon 33 nucleotides after the start site of transcription for each of the plasmids of E2F1, MybL2, and FOXM1. The primers used to generate these clones are as follows:

E2F1 5'GCGGCCCATGATAGCCGGCGCTGGAG,

R:CTCCAGCGCCGGCTATCATGGGCCGC

FoxM1: 5'CGTCGGCCACTGATTTAGTAAAGACGGAGGCTGC,

R:GCAGCCTCCGTCTTTACTAAATCAGTGGCCGACG

MybL2: 5'CGCTGCGAGGATCTGTAGTAGCTGCACTACCAGGACACAG

R:CTGTGTCCTGGTAGTGACGCTACTACAGATCCTCGCAGCG.

Capped RNA was generated using mMESSAGE mMACHINE SP6 or T7 kit (AM1344, AM1345; Thermo Fischer Scientific). A dilution to 25pg/nl at 2.5nl injection, with a final concentration of 30-100pg per embryo, gave phenotypes in the triply-over expressing embryos and no discernable phenotype in the control RNA injections, therefore this concentration was used in all experiments. For live cell imaging embryos were coinjected at 2-cell stage with a membrane label-GAP43:RFP and nuclear label-H2B:GFP in PCS2+ vectors. As directed by Gene Tools LLC, the Morpholino stock was made at 1mM in diH<sub>2</sub>O so that when injected into the embryo there was a final concentration of 5-10uM in the embryo, a concentration that has been validated through multiple other Morpholino studies and shows no off-target toxicities or effects. The sequences are as follows:

Translation Blocking: 5' CCATGCCGGTCTCAGAGGAAGGTTC 3'

Splice Blocking: 5' GGGACTCACCGTGCAGGTAACAGAC 3'

## METHOD DETAILS

### ***Scoring of FA from exome sequencing data***

The FA score was calculated from tumor exome sequencing data by measuring the variant allele frequency (also known as “allele balance”) distribution at heterozygous germline single nucleotide polymorphisms (SNPs), where heterozygous germline SNPs are defined by variant calls and genotypes in the matched normal data. The variant allele frequency (VAF) of each SNP is calculated by dividing the number of reads containing the alternate SNP allele by the total number of reads that align to that base in the reference genome. Heterozygous germline SNPs are expected to be found at a VAF of approximately 0.5, with minor deviations occurring due to stochastic sampling of chromosomes during exome sequencing. Loss of heterozygosity (LOH) events and copy number alterations (CNA) cause VAF deviations at heterozygous SNPs by altering the copy number of one SNP allele relative to the other SNP allele. For example, an LOH event that is fixed among tumor cells will result in heterozygous SNPs within the altered genomic segment to have VAF values of 0 or 1 (depending on which allele is lost), with subclonal alterations causing intermediate values. In high-FA tumors, numerous LOH and CNA events will cause VAF deviations at large numbers of SNPs across the entire genome, resulting in markedly broader VAF distributions that are obvious from genome-wide VAF histograms (as in Figure 1). The FA score is derived from the standard deviation of VAF measurements in the tumor exome sequencing data across all heterozygous germline SNPs.

### ***Immunofluorescence and Live Imaging***

Fixed cell immunofluorescence microscopy was performed on St. 8.5-9 embryos (approximately 8-9 hours post fertilization) that were fixed overnight at -20C in Dent's solution (80% methanol, 20% DMSO). They were postfixed in methanol for 9h or overnight at -20C. Embryos were then hydrated and dissected in Tris-buffered saline (150mM NaCl, 50mM Tris; pH 7.4) with 1% SDS detergent (TBS and Tween 20 [TBST]). The dissected caps were subjected to a DNA stain (TOPRO 3, 1:20000 dilution) for 30- 60 min at RT, and dehydrated in Methanol overnight. Pre-imaging, fixed caps were cleared in Murray's solution (2:1 Benzyl Benzoate, Benzyl Alcohol) 15- 30 min, mounted on slides (Histomount, Life Technologies). Images were taken by Axiovert 200 microscope (Carl Zeiss) with PerkinElmer-RS spinning disk confocal system illuminated by a krypton/argon laser, using a 40x or 60x oil immersion objective (NA 1.4; Carl Zeiss), with images acquired by an electron multiplying charge coupled device camera (Hamamatsu C9100) using Velocity software. For live imaging experiments, capped RNA was generated using an SP6 mMessage mMachine kit (Ambion). *Xenopus laevis* embryos were injected while suspended in 5% Ficoll solution at 2-cell stage with 50-150pg of RNA. To visualize mitotic events embryos were co-injected membrane label (GAP43:RFP) and nuclear label (H2B:GFP). Triple mutants were additionally co-injected with 30pg of each of the transcription factors overexpressed, whereas the controls were injected with an equal concentration of control RNA. At NF stage 8, embryos were devitellined and the animal cap portion was removed by microsurgery and sandwiched between coverslips before imaging. A time-lapse movie of the layer of cells that make up the blastocoel roof was made on a Zeiss 780 Confocal Microscope with the 25x objective and a framing rate of 30 s. We thank Robert Rotzin for help with the live imaging acquisition. To quantify chromosome instability phenotypes in animal caps, we assessed every anaphase event that could be clearly visualized and scored it as having an anaphase lagging chromosome (a chromosome clearly left at the area of the metaphase plate without significant stretching of chromatin toward either pole), a multipolar anaphase (chromosomes segregating in more than 2 directions), anaphase bridging (chromatin stretched between the segregating anaphase masses) or normal anaphase segregation (controls n=187 in 18 movies, triple injection n=436 in 12 movies). Movies included anaphase events in stages 8-9.5. Statistical significance was measured using a one-way ANOVA with a Bonferonni post-test.

### ***Embryos lysates and immunoblotting***

Embryos lysates for western blotting were collected at stages 7, 8, 9 for both control and triply-overexpressing embryos as described in [S5]. Blots were stained with a 1:500 dilution of anti- E2F1 (clones KH20 &KH95, mixed mouse monoclonal IgGs, 05-379; Millipore), anti-FoxM1 (G- 5: sc-376471, Santa Cruz Biotechnology, Inc.), anti- MybL2 (phospho T487, ab76009; Abcam), p53 X-77, MA1-12549, Thermo Fisher Scientific), and anti- $\alpha$ -tubulin (Dm1 $\alpha$ , 1:2000 dilution) used as a loading control. Secondary antibodies conjugated to HRP at a 1:10,000 dilution were visualized using Amersham ECL Prime Western Blotting Detection Reagent, (GE Healthcare) and imaged with Chemi Doc Bio-Rad system.

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

### ***Identification of significantly mutated genes in high- vs. low-FA tumors.***

For mutation burden analysis, we used the official somatic mutation calls from the Level 2 Illumina exome sequencing VCF files in the TCGA project database, which are derived from precisely the same exome sequencing datasets that we used to measure FA. We then used the VAAST software package (version 1.0.4) [S1] to identify genes that were significantly mutated in the 100 tumors with the highest FA scores relative to the 100 tumors with the lowest FA scores using the following command line parameters: "-m lrt--codon\_bias --gp 10000 -r 0.001". We then performed the reverse enrichment experiment to identify genes that were significantly mutated in the 100 tumors with the lowest FA scores relative to the 100 tumors with the highest FA scores using the same parameters. To test whether the number of TP53 mutations was linearly correlated with aneuploidy ranking, we performed a linear-by-linear association test using R. We performed separate tests for the most aneuploid (Rank 1-250) and least aneuploid (Rank 251-522) tumors binned into groups of 50.

### ***Identification of genes that are overexpressed in high-FA tumors (the BrFA100 gene list)***

To identify genes whose expression was correlated with FA, we calculated the Pearson correlation coefficient between the highest and lowest FA scored tumors described above and gene-level RNA expression values obtained from TCGA. This correlation-based approach is identical to that previously used to identify the CIN70

gene list [S2] although we note that the prior study used a different CIN scoring method and RNA expression array platform.

## DATA AND SOFTWARE AVAILABILITY

### ***Bioinformatic analysis***

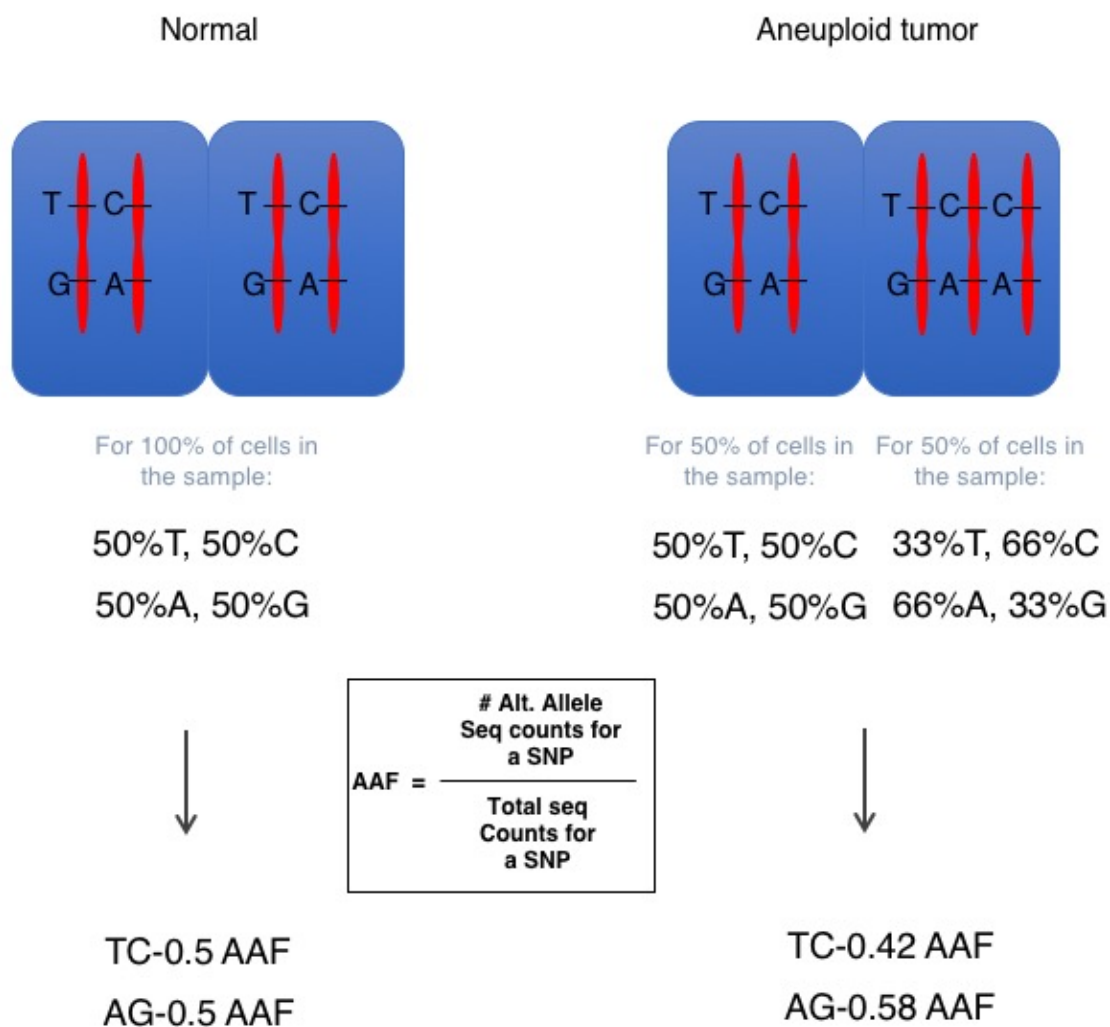
Co-association analysis was performed at the CBioPortal Web site by inputting the four genes and determining if there were any associations of either mutations or RNA expression patterns. Venn Diagram in figure 4D was generated by inputting the referenced gene sets into the Venn diagram maker at Bioinformatics and Evolutionary genomics at Ghent (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). GO analysis was performed at <http://www.geneontology.org/page/go-enrichment-analysis> [S3]. String network diagram was produced using the String Webportal tool (<http://string-db.org>) limiting interactions to actions with the highest confidence ratios[S4]. To generate the masks that highlight the genes involved in mitosis, cell cycle and DNA replication and began with G0 terms but also performed manual curations based on pubmed searches using the gene and associated processes. The masks were generated by hand to include these manually curated lists. Contact Ira Hall ([ihall@wustl.edu](mailto:ihall@wustl.edu)) for Matlab code to generate Allele Frequency Plots.

## RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Anti-E2F1 (clones KH20 & KH95)	Milipore	05-379
Anti-MybL2 phospho T487	Abcam	Ab76009
Anti-FoxM1	Santa Cruz	Sc-376471
Anti-alpha-tubulin	Abcam	Ab7291
Anti-p53	ThermoFisher	MA1-12549
<b>Biological Samples</b>		
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Dextran, tetramethylrhodamine (and fluorescein), 70,000 MW, Lysine Fixable	ThermoFisher	D1818, D1823
TO-PRO-3 Stain	ThermoFisher	T3605
<b>Critical Commercial Assays</b>		
mMessage mMachine	ThermoFisher	AM1344, AM1345
<b>Deposited Data</b>		
<b>Experimental Models: Cell Lines</b>		
<b>Experimental Models: Organisms/Strains</b>		
<i>Xenopus laevis</i> Embryos	Nasco	In house colony, Dr. R Keller laboratory
<b>Recombinant DNA</b>		
pCS2H2B-GFP/pCS2H2B-RFP	Stukenberg Lab, developed at Wallingford Lab, University of Texas, Austin	
pCS2GAP43-RFP/pCS2GAP43-GFP	Ray Keller Lab	
pCDNA3-MybL2, HA-E2F-1-wt-pRcCMV	Addgene	
pCSF107mT-FoxM1	humanORFeome, XenopusORFeome	
<b>Sequence-Based Reagents</b>		
Primers for generating Control RNAs Full sequences in supplemental materials	Invitrogen, Fisher Scientific	
<i>Xenopus</i> p53 Morpholinos Full sequences in supplemental materials	Gene Tools, LLC.	Xp53-TrBl: GFP MO, Xp53-SpliceMO: GFP
<b>Software and Algorithms</b>		
VAAST version 1.0.4		
CBioPortal		
Bioinformatics and Evolutionary genomics <a href="http://bioinformatics.psb.ugent.be/webtools/Venn/">http://bioinformatics.psb.ugent.be/webtools/Venn/</a> <a href="http://biovenn.nl">http://biovenn.nl</a>		
<b>Other</b>		



Figure S1



\*Note that all of the SNPs along the entire chromosome will have similar AAF spreading

Figure S1. Why aneuploidy changes the AAF of heterozygous alleles (Related to Figure 1). We calculate the AAF of a SNP by the number of next gen sequence reads of the alternate allele divided by the number of total reads for that locus. All heterozygous SNPs will generate an AAF around 0.5 in a normal sample. However, in the theoretical case where there is an extra chromosome in about half the cells in the tumor then the heterozygous SNPs on that chromosome will both increase (if they are the alternate allele) or decrease if they are the allele on the reference genome. The AAF is dependent upon both the amount of aneuploidy (in this case 3 chromosomes) and the percentage of the cells in the tumor that have that aneuploidy (50% of the cells). In this theoretical case, the AAF for all the heterozygous SNPs on the extra chromosome will be either 0.42 or 0.58.

Figure S2

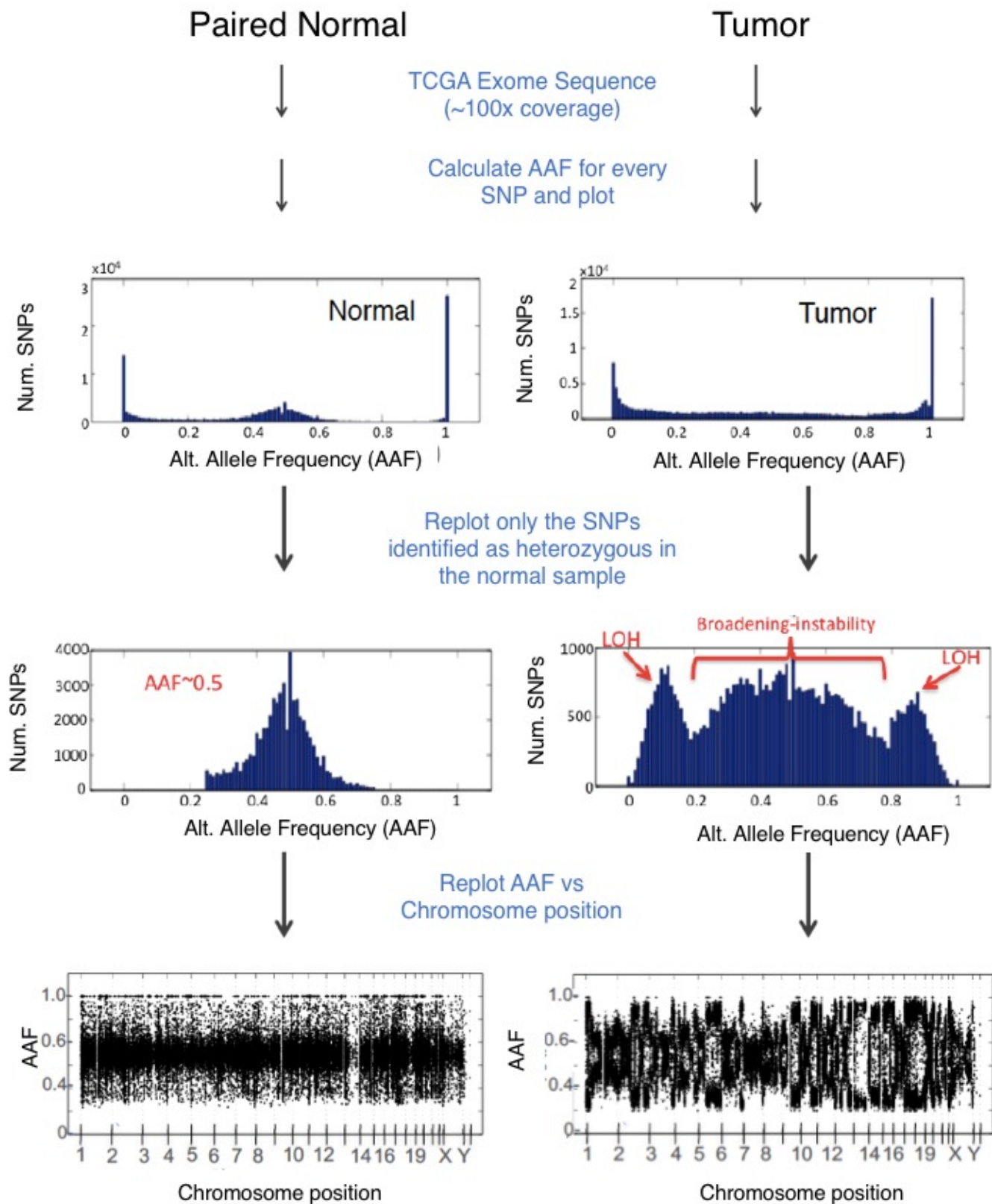


Figure S2. The pipeline used to measure aneuploidy in tumors (Related to Figure 1). Germline heterozygous SNPs are identified from breast cancer TCGA exome data based on their presence in paired normal samples. After calculation of alternate allele frequencies (AAFs; top plots), heterozygous SNPs are defined as those with an  $AAF \geq 0.25$  and  $AAF \leq 0.75$  in the normal sample (middle plots). When one generates histograms quantifying the number of initially heterozygous SNPs with various AAF the distribution in the tumor samples then one can detect aneuploidy and tumor heterogeneity by two different mechanisms. First the central peak around AAF broadens. Second, if there is LOH of chromosomes in a large percentage of the cells then all of the SNPs now generate AAF peaks at 0 or 1, which generate peaks that are outside the central peak. We also generate a second plot for each tumor (bottom plots). In these plots the AAF is on the Y-axis and Chromosome position is on the X-axis and each SNP is given a single dot. The tumor used in this example was scored as having the 14th most functional aneuploidy of 522 tumors.

Figure S3

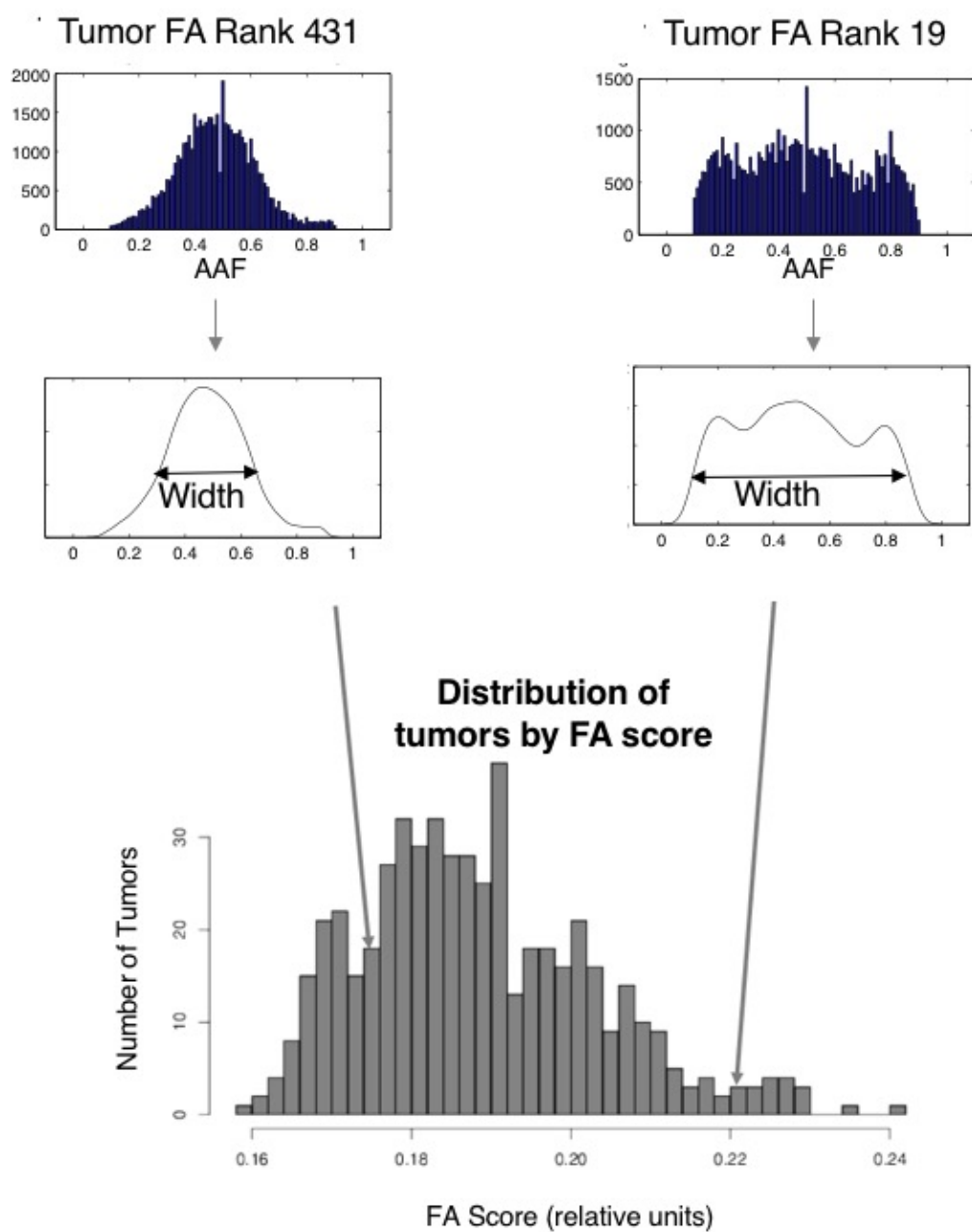


Figure S3. Method of scoring FA according to the AAF plots (Related to Figures 1,2). We generated line graphs that represent the shape of the associated AAF plot and then calculated a standard deviation of the associated curve, as visualized by the width of the peaks.

Figure S4

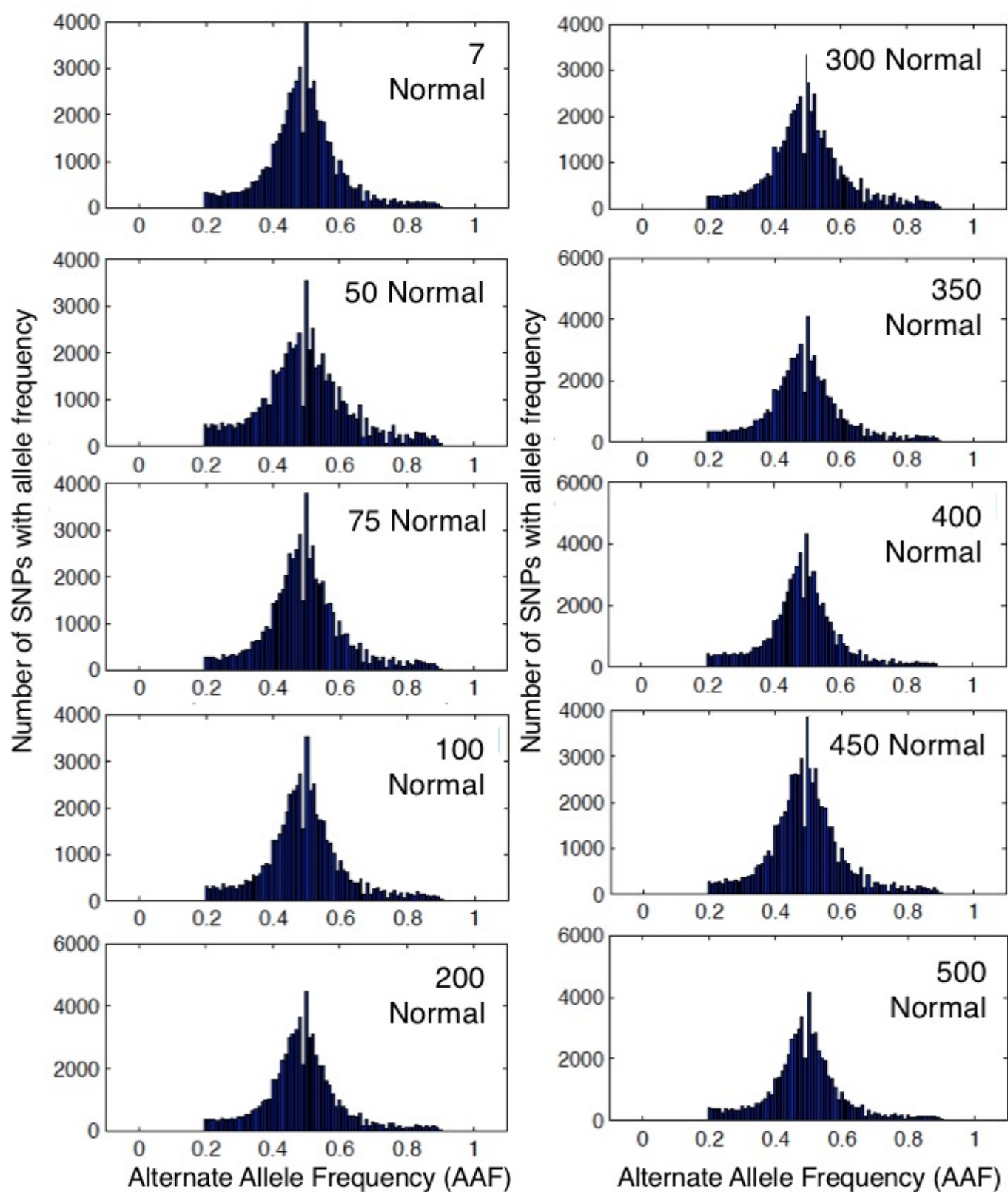
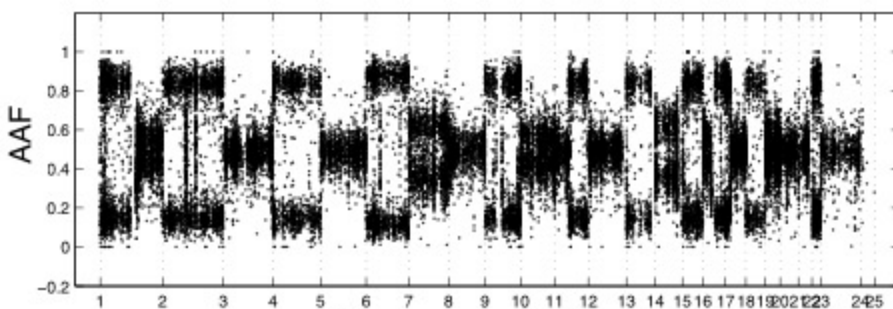
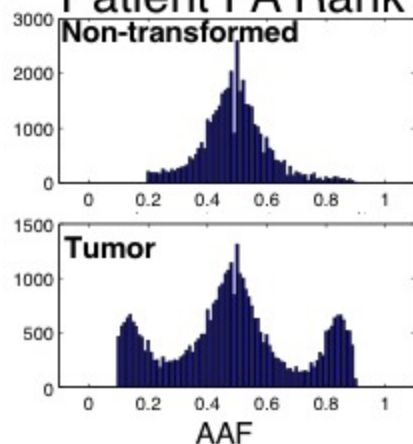


Figure S4. AAF histograms of the Normal (non-transformed) samples (Related to Figure 1). Here we show the matched normal samples from the patients whose tumors are shown in Figure 1A. The ranking of each tumor is shown as the number in the top left corner of each histogram.



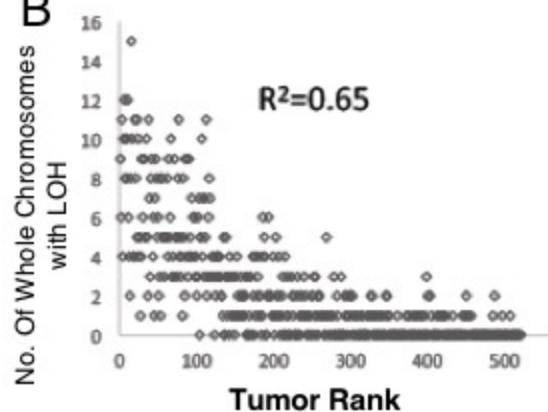
Figure S5

**A** Patient FA Rank 31

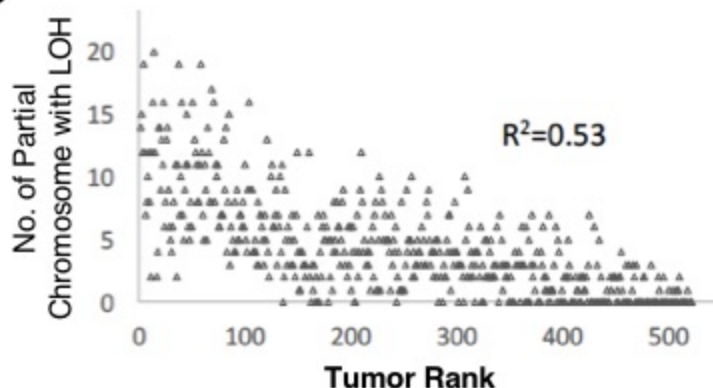


Scoring:  
Whole Chromosome LOH-2,4,6,9,13,15,18,22  
Partial Chromosome LOH-1,11,16,17

**B**



**C**



**D**

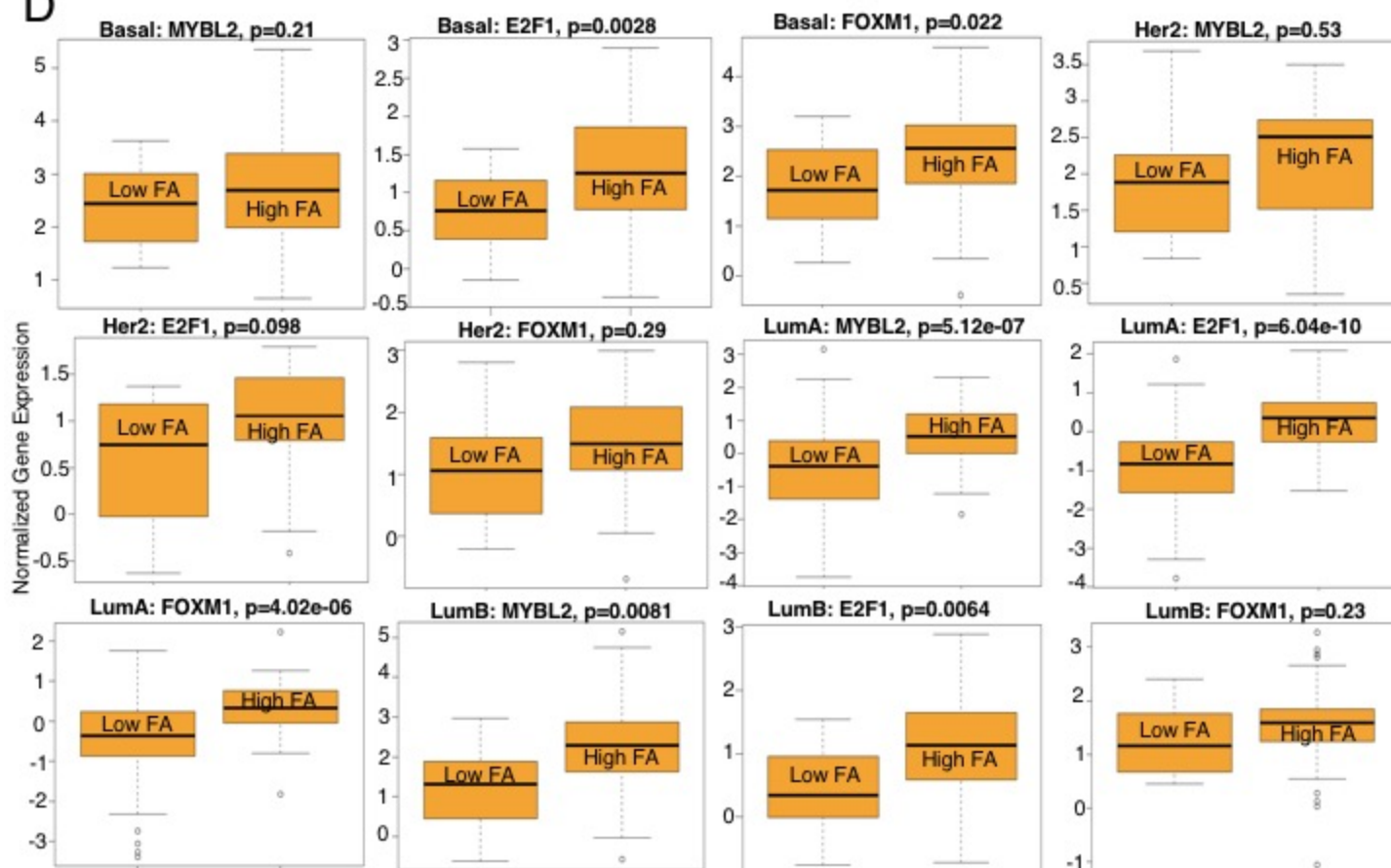


Figure S5. Method of distinguishing between whole and partial LOH events (Related to Figure 1 and 4). A) We manually visualized the chromosome position vs. AAF plots. We scored every chromosome of every tumor by determining if there were two major peaks maxima below AAF 0.25 and greater than 0.75 that spanned along a chromosome. These AAF ratios were chosen to rule triploidization events that generated peaks at 0.33 and 0.67. Note that it is possible that we miscall a chromosome LOH event if there are more than 4 times the number of one parental chromosome over its homolog. Each chromosome was scored as: 1) entire chromosome that had a split peak across all positions for a chromosome was scored as “Whole Chromosome LOH”, if we could find splitting of some regions of chromosome but others with allele frequencies between 0.25 and 0.75 it was scored as “Partial Chromosome LOH”, if we could not find any splitting of peaks along a contiguous region of a chromosome the tumor was scored as “No Chromosome LOH”. The number of chromosomes with an LOH events comprising a whole chromosome (B) in breast tumors correlates with the FA score (similar to Fig 1D). R2 value was generated by fitting the points to a linear regression in Excel. C) Plot to show that partial chromosome events correlated with tumor ranking although this correlation was lower than either the total number of LOH events or the whole chromosome events. R2 value was generated by fitting the points to a second order polynomial curve in Excel. D) TCGA RNA-seq gene expression data from primary solid tumor sample of breast cancer patients for MYBL2, E2F1, and FOXM1 was compared for the 200 highest and 200 lowest FA scoring tumors. We stratified the data by different subtypes of breast cancer patients, including Basal, HER2+, Luminal A, and Luminal B and performed a T-test between high FA and low FA TCGA breast cancer patients, and report the p-value.

Figure S6

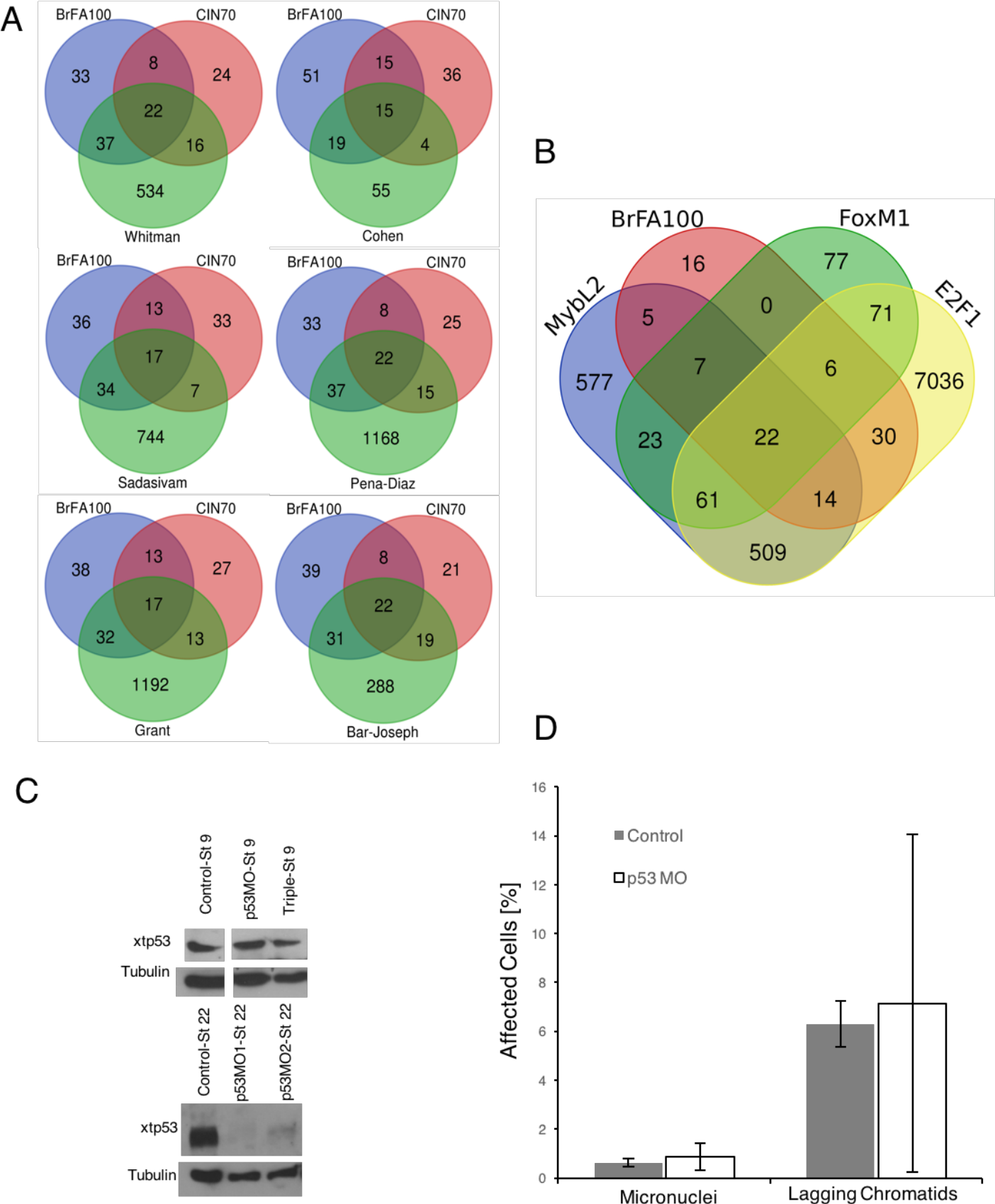


Figure S6. BrFA100 is significantly from proliferation signatures (Related to Figure 4) and xtp53 knockdown alone is not responsible for *Xenopus* phenotypes (Related to Figure 5). The overlap of the BrFA100, CIN70 and each of 6 different proliferation signatures are shown through Venn Diagrams (A). Gene lists are available in Table S3. (B) A different visual representation of the overlap of the BrFA100 and the ChIP-Seq data sets for E2F1, FoxM1, and MybL2. (C) Western Blot analysis of p53 Morpholino injected *Xenopus* embryos at Stage 9 and Stage 22 shows that a significant decrease in tp53 protein level is not seen until much later than when most of our in vivo experiments take place. This is why we do not see an increase in the number of micronuclei or lagging chromosomes in p53MO injected embryos (D) (n=30 for controls, n=15 for p53MO experiments).

### Supplemental References

- S1. Yandell, M., Huff, C., Hu, H., Singleton, M., Moore, B., Xing, J., Jorde, L.B., and Reese, M.G. (2011). A probabilistic disease-gene finder for personal genomes. *Genome Res* 21, 1529-1542.
- S2. Carter, S.L., Eklund, A.C., Kohane, I.S., Harris, L.N., and Szallasi, Z. (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 38, 1043-1048.
- S3. Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 8, 1551-1566.
- S4. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43, D447-452.
- S5. Stukenberg, P.T., Lustig, K.D., McGarry, T.J., King, R.W., Kuang, J., and Kirschner, M.W. (1997). Systematic identification of mitotic phosphoproteins. *Curr Biol* 7, 338-348