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

Supplemental Experimental Procedures:

CONTACT FOR REAGENT AND RESOURCE SHARING (pts7h@virginia.edu, kep9v@virginia.edu)

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 CaCl2, 1mM MgSO4, 5mM HEPES (pH7.8), 2.5 mM NaHCO3, 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:CTGTGTCCTGGTAGTGCAGCTACTACAGATCCTCGCAGCG.

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 diH2O 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-α-tubulin (Dm1α, 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

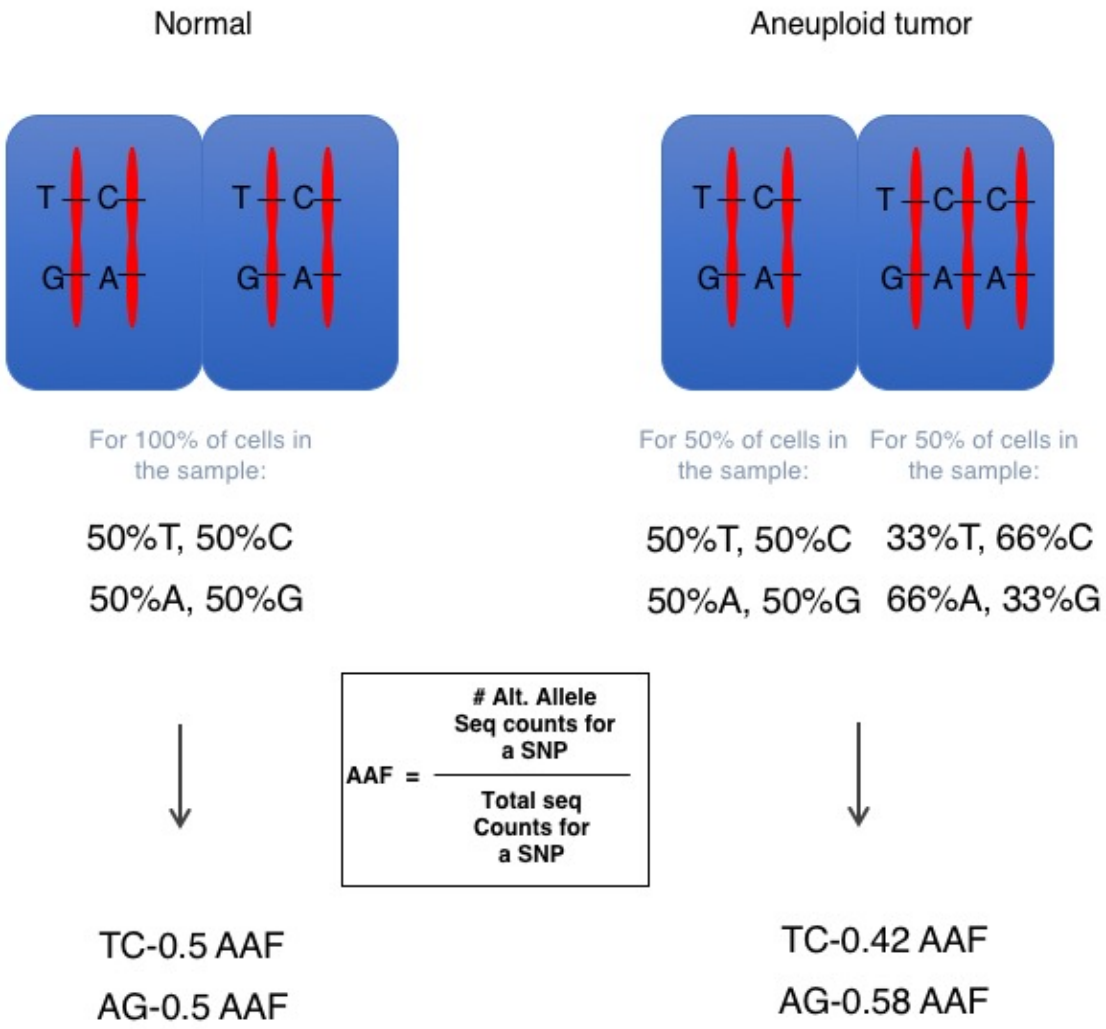
Co-association analysis was performed at the CBioPortal Web site by inputting the four genes and determining if there were any associations of either mutations or RNA expression patterns. Venn Diagram in figure 4D was generated by inputting the referenced gene sets into the Venn diagram maker at Bioinformations 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) for Matlab code to generate Allele Frequency Plots.

RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
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
Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
Dextran, tetramethylrhodamine (and fluorescein), 70,000 MW, Lysine Fixable	ThermoFisher	D1818, D1823
TO-PRO-3 Stain	ThermoFisher	T3605
Critical Commercial Assays		
mMessage mMachine	ThermoFisher	AM1344, AM1345
Deposited Data		
Experimental Models: Cell Lines		
Experimental Models: Organisms/Strains		
Xenopus laevis Embryos	Nasco	In house colony, Dr. R Keller laboratory
Recombinant DNA		
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	
Sequence-Based Reagents		
Primers for generating Control RNAs Full sequences in supplemental materials	Invitrogen, Fisher Scientific	
Xenopus p53 Morpholinos Full sequences in supplemental materials	Gene Tools, LLC.	Xp53-TrBl: GFP MO, Xp53-SpliceMO: GFP
Software and Algorithms		
VAAST version 1.0.4		
CBioPortal		
Bioinformations and Evolutionary genomics http://bioinformatics.psb.ugent.be/webtools/Venn/		
http://biovenn.nl		
Other		

Supplemental Figures and Legends.

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

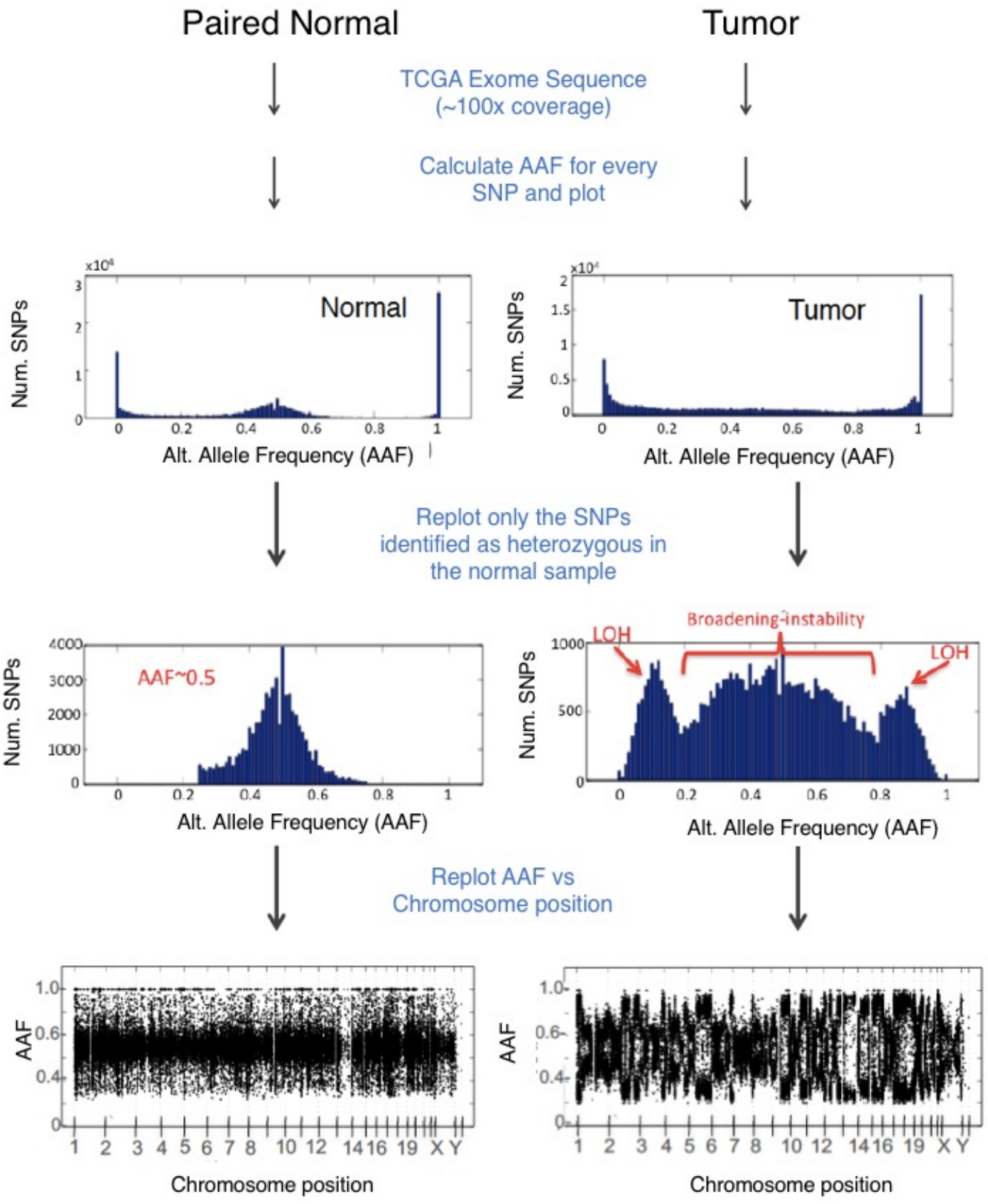

Paired Normal
Tumor
TCGA Exome Sequence
(~100x coverage)
Calculate AAF for every
SNP and plot
Num. SNPs
Normal
Alt. Allele Frequency (AAF)
Num. SNPs
Tumor
Alt. Allele Frequency (AAF)
Replot only the SNPs
identified as heterozygous in
the normal sample
Num. SNPs
AAF~0.5
Alt. Allele Frequency (AAF)
Num. SNPs
LOH
Broadening-instability
LOH
Alt. Allele Frequency (AAF)
Replot AAF vs
Chromosome position
AAF
1 2 3 4 5 6 7 8 10 12 14 16 19 X Y
Chromosome position
AAF
1 2 3 4 5 6 7 8 10 12 14 16 19 X Y
Chromosome position

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>=0.25 and AAF<=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

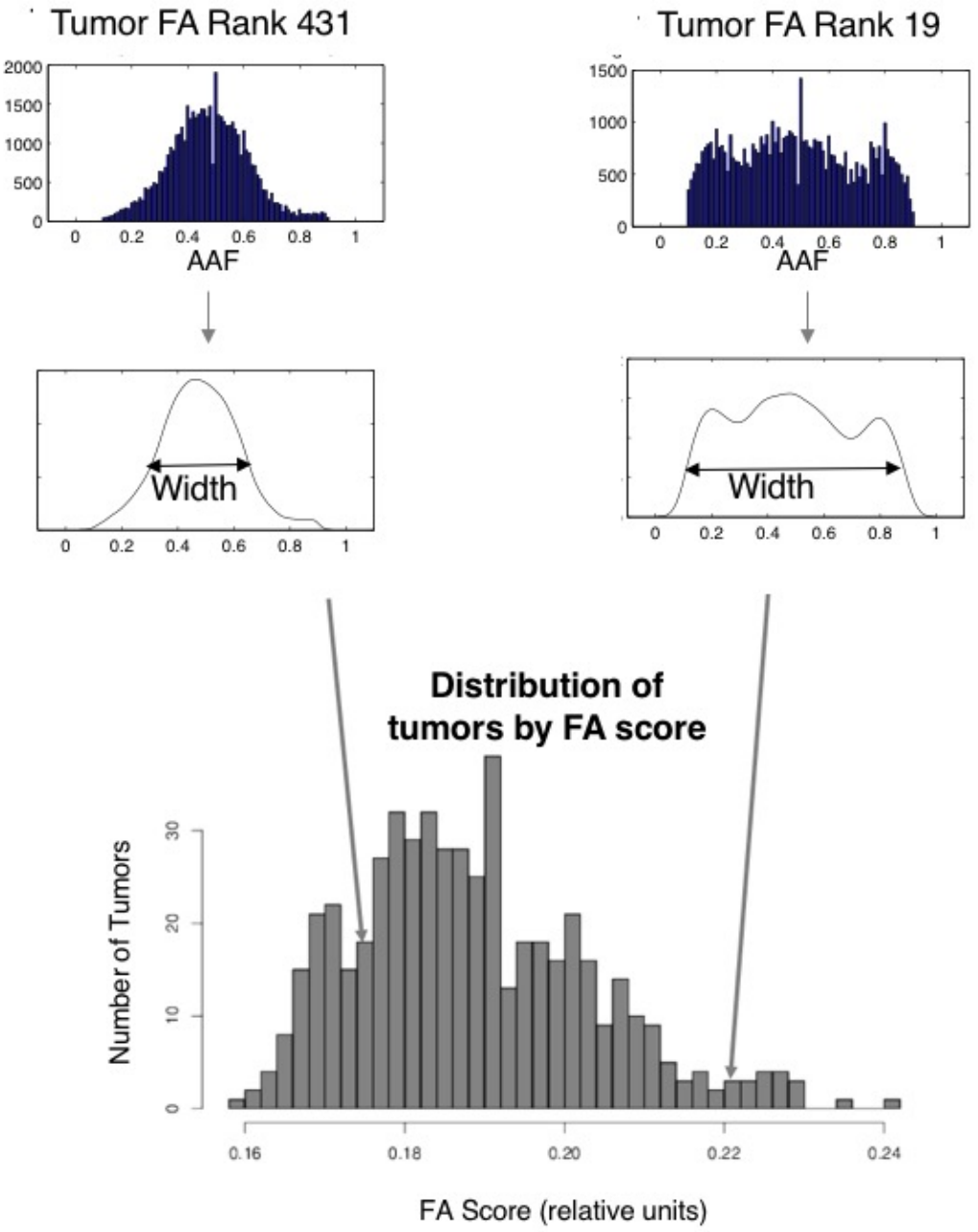

Tumor FA Rank 431
AAF
Width
Tumor FA Rank 19
AAF
Width
Distribution of tumors by FA score
Number of Tumors
FA Score (relative units)

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

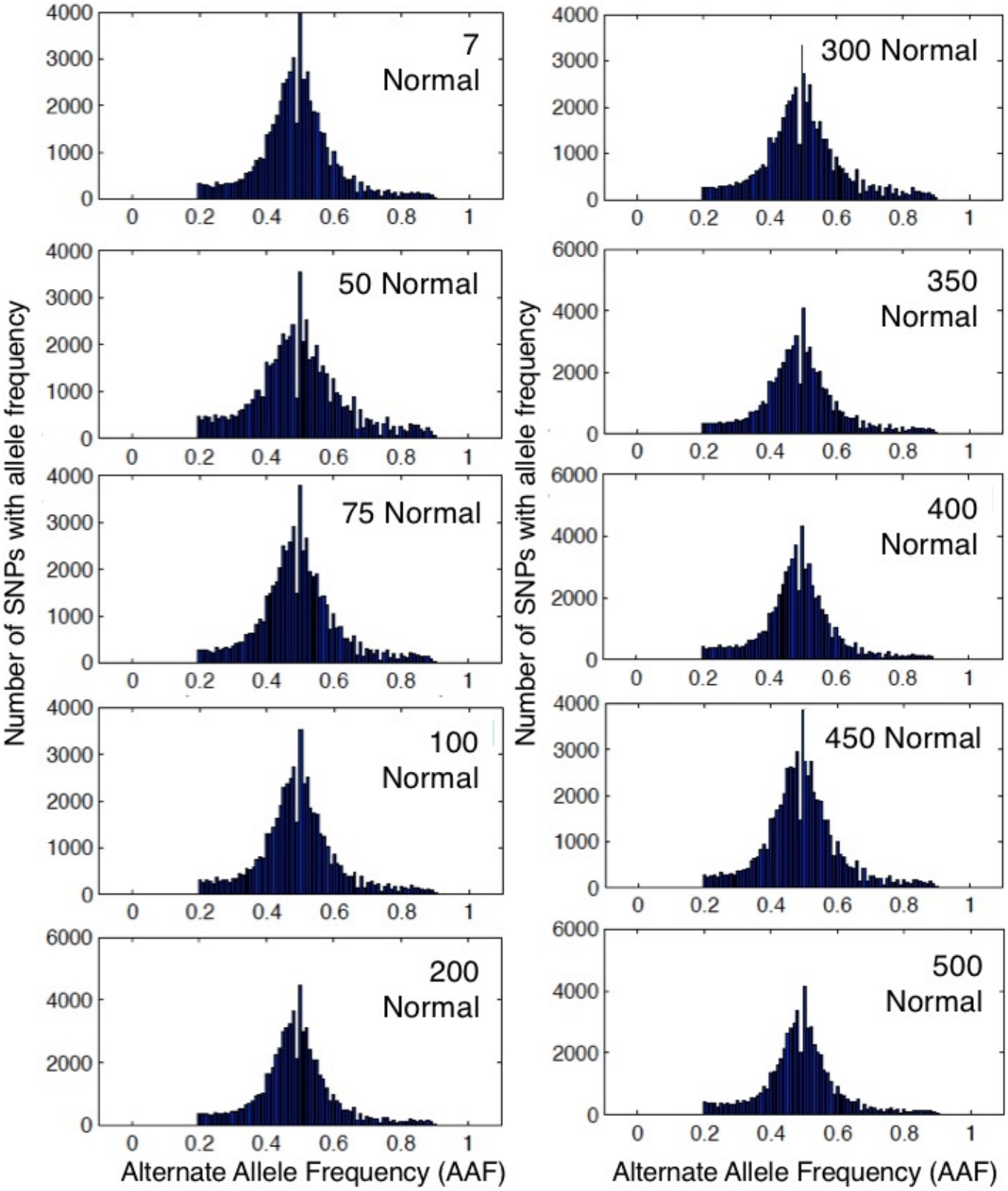

7 Normal
300 Normal
50 Normal
350 Normal
75 Normal
400 Normal
100 Normal
450 Normal
200 Normal
500 Normal
Number of SNPs with allele frequency
Alternate Allele Frequency (AAF)

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

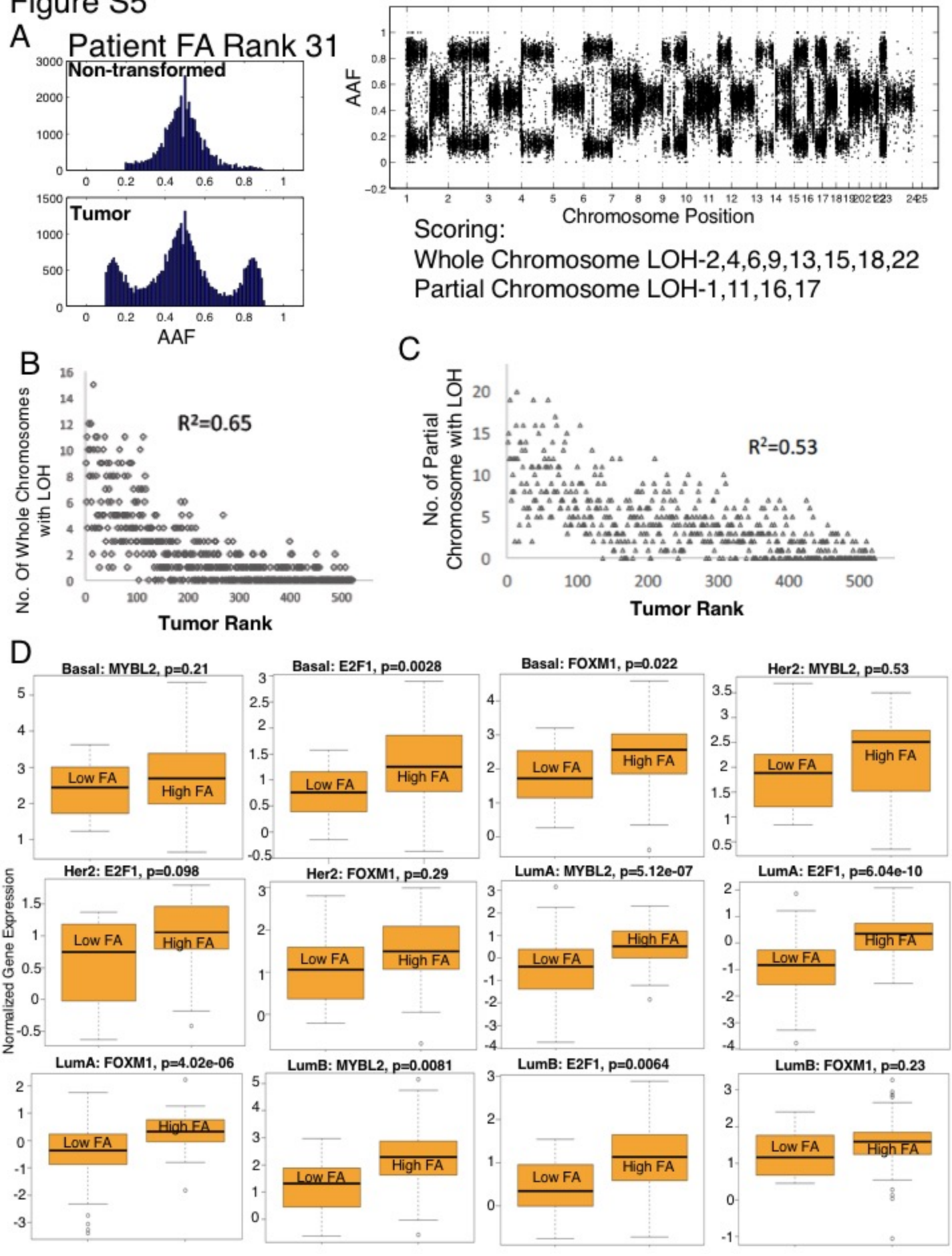

A
Patient FA Rank 31
Non-transformed
Tumor
AAF
AAF
Chromosome Position
Scoring:
Whole Chromosome LOH-2,4,6,9,13,15,18,22
Partial Chromosome LOH-1,11,16,17
B
No. Of Whole Chromosomes with LOH
R²=0.65
Tumor Rank
C
No. of Partial Chromosome with LOH
R²=0.53
Tumor Rank
D
Normalized Gene Expression
Basal: MYBL2, p=0.21
Basal: E2F1, p=0.0028
Basal: FOXM1, p=0.022
Her2: MYBL2, p=0.53
Her2: E2F1, p=0.098
Her2: FOXM1, p=0.29
LumA: MYBL2, p=5.12e-07
LumA: E2F1, p=6.04e-10
LumA: FOXM1, p=4.02e-06
LumB: MYBL2, p=0.0081
LumB: E2F1, p=0.0064
LumB: FOXM1, p=0.23
Low FA
High FA

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

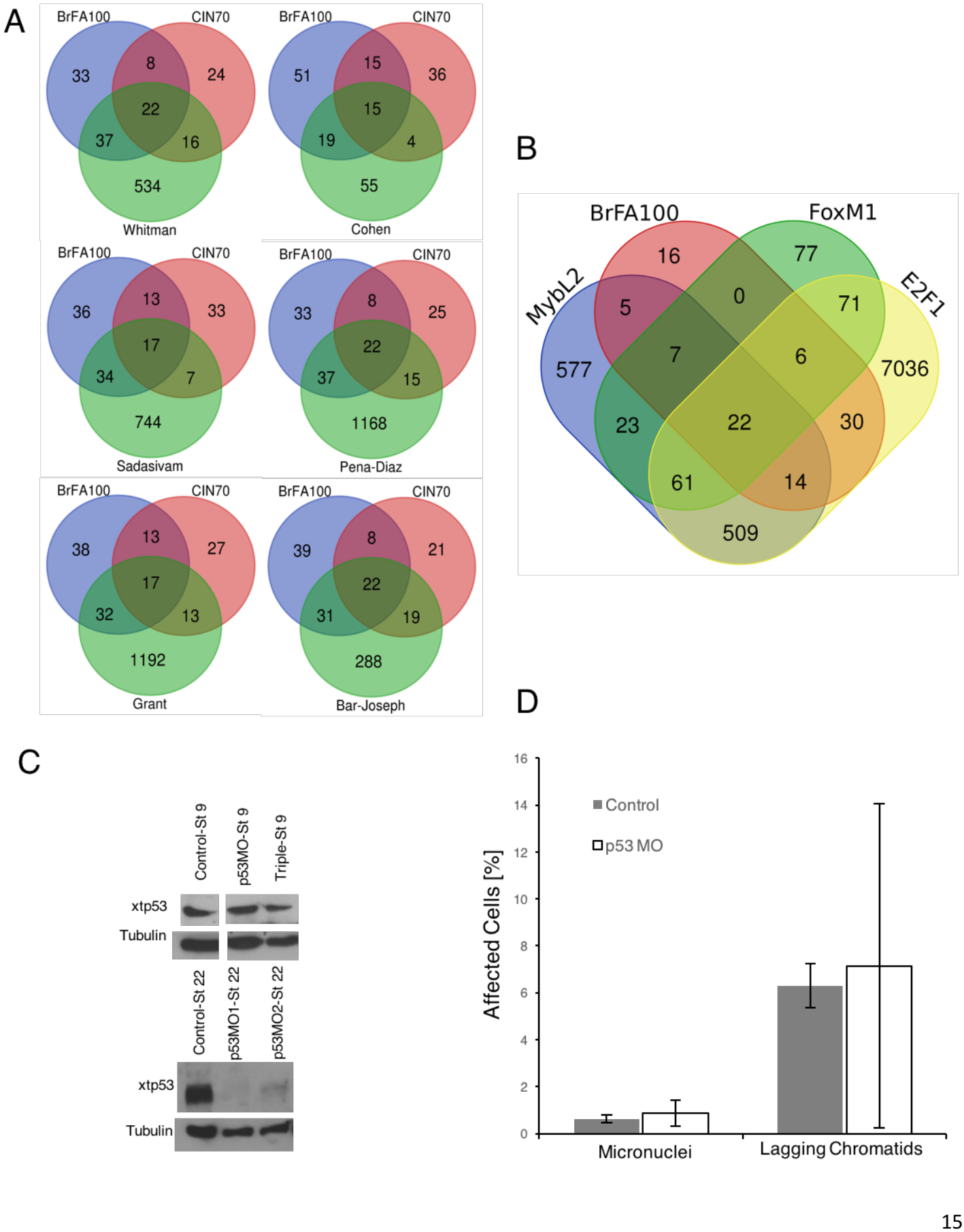

A
BrFA100
CIN70
33
8
24
22
37
16
534
Whitman
51
15
36
15
19
4
55
Cohen
36
13
33
17
34
7
744
Sadasivam
33
8
25
22
37
15
1168
Pena-Diaz
38
13
27
17
32
13
1192
Grant
39
8
21
22
31
19
288
Bar-Joseph
B
BrFA100
FoxM1
MybL2
E2F1
16
77
5
0
71
577
7
6
7036
23
22
30
61
14
509
C
Control-St 9
p53MO-St 9
Triple-St 9
xtp53
Tubulin
Control-St 22
p53MO1-St 22
p53MO2-St 22
xtp53
Tubulin
D
Affected Cells [%]
16
14
12
10
8
6
4
2
0
Control
p53 MO
Micronuclei
Lagging Chromatids

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