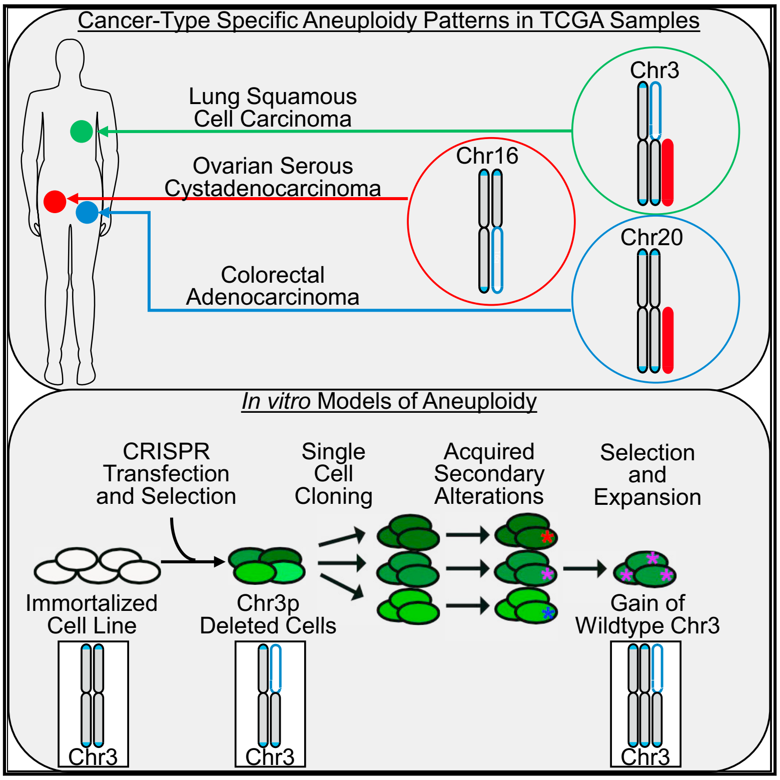
Cancer Cell

Genomic and Functional Approaches to Understanding Cancer Aneuploidy

Graphical Abstract



Highlights

- Aneuploidy, whole chromosome or chromosome-arm imbalance, occurs in 88% of cancers
- Aneuploidy correlates with cell-cycle genes and anti-correlates with immune levels
- Patterns of aneuploidy alterations are tumor-type specific
- Engineered chromosome 3p deletion does not promote proliferation in human lung cells

Authors

Alison M. Taylor, Juliann Shih,
Gavin Ha, ..., Andrew D. Cherniack,
Rameen Beroukhim,
Matthew Meyerson

Correspondence

matthew_meyerson@dfci.harvard.edu

In Brief

Analyzing >10,000 human cancers, Taylor et al. show that aneuploidy is correlated with somatic mutation rate, expression of proliferation genes, and decreased leukocyte infiltration. Loss of chromosome arm 3p is common in squamous cancers, but deletion of chromosome 3p reduces cell proliferation *in vitro*.

Taylor et al., 2018, Cancer Cell *33*, 676–689
April 9, 2018
https://doi.org/10.1016/j.ccell.2018.03.007

Cancer Cell
Article

Genomic and Functional Approaches to Understanding Cancer Aneuploidy

Alison M. Taylor,[1,2,3] Juliann Shih,[2] Gavin Ha,[1,2,3] Galen F. Gao,[2] Xiaoyang Zhang,[1,2,3] Ashton C. Berger,[2]
Steven E. Schumacher,[1,2] Chen Wang,[4,5] Hai Hu,[6] Jianfang Liu,[6] Alexander J. Lazar,[7]
The Cancer Genome Atlas Research Network, Andrew D. Cherniack,[1,2,3] Rameen Beroukhim,[1,2,3]
and Matthew Meyerson[1,2,8,9,*]
[1]Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA
[2]Cancer Program, Broad Institute, 415 Main Street, Cambridge, MA 02142, USA
[3]Department of Medicine, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA
[4]Department of Health Sciences Research, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905, USA
[5]Department of Obstetrics and Gynecology, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905, USA
[6]Chan Soon-Shiong Institute of Molecular Medicine at Windber, Windber, PA 15963, USA
[7]Departments of Pathology, Genomic Medicine, and Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 85, Houston, TX, USA
[8]Department of Pathology, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA
[9]Lead Contact
*Correspondence: matthew_meyerson@dfci.harvard.edu
https://doi.org/10.1016/j.ccell.2018.03.007

SUMMARY

Aneuploidy, whole chromosome or chromosome arm imbalance, is a near-universal characteristic of human cancers. In 10,522 cancer genomes from The Cancer Genome Atlas, aneuploidy was correlated with *TP53* mutation, somatic mutation rate, and expression of proliferation genes. Aneuploidy was anti-correlated with expression of immune signaling genes, due to decreased leukocyte infiltrates in high-aneuploidy samples. Chromosome arm-level alterations show cancer-specific patterns, including loss of chromosome arm 3p in squamous cancers. We applied genome engineering to delete 3p in lung cells, causing decreased proliferation rescued in part by chromosome 3 duplication. This study defines genomic and phenotypic correlates of cancer aneuploidy and provides an experimental approach to study chromosome arm aneuploidy.

INTRODUCTION

Aneuploidy, an unbalanced number of chromosomes, was first observed in tumor cells over 100 years ago (reviewed in Holland and Cleveland, 2009) and is a predominant cancer feature occurring in ~90% of solid tumors (Weaver and Cleveland, 2006). Even though it is among the oldest described cancer alterations, and even though genomics efforts have allowed high-throughput "karyotyping" of patient cancers, the role of aneuploidy in tumorigenesis is still a mystery.

Aneuploidy and focal copy-number alterations represent two classes of somatic copy-number alteration (SCNA) (Tang and Amon, 2013). Studies of genomic and phenotypic correlates of aneuploidy and SCNAs have examined copy number based on cytoband (Carter et al., 2006) or based on number of SCNAs, including broad and focal events (Davoli et al., 2017; Buccitelli et al., 2017). By these definitions, increased SCNA levels were reported to correlate with proliferation pathways (Carter et al., 2006; Davoli et al., 2017; Buccitelli et al., 2017) and to anti-correlate with immune signaling within individual tumor types (Davoli et al., 2017; Buccitelli et al., 2017).

We define aneuploidy as SCNAs of whole chromosomes and of chromosome arms. In cancer, aneuploidy affects more of the genome than any other somatic genetic alteration (Beroukhim et al., 2010; Zack et al., 2013; Mitelman, 2000). The most frequent recurrent arm alterations occur in over 30% of tumors, whereas the

Significance

Although its universality makes aneuploidy a salient feature of cancer genomes, the functional significance of cancer aneuploidy remains a mystery. Powerful technologies for genome analysis and genome engineering now make the study of aneuploidy more tractable. Here, we look across cancer to identify universal and cancer-type-specific characteristics of aneuploidy which further hint at its importance for the process of cancer pathogenesis. Furthermore, the ability to engineer aneuploid chromosome arms will now enable the development of more complex cellular models to generate and investigate the proliferative and survival impact of cancer aneuploidy.

most frequent recurrent focal copy-number alterations occur at a frequency below 0.15 (Beroukhim et al., 2010). Chromosome arm SCNAs are more common than whole chromosome SCNAs, occurring at higher frequencies in 12 of 16 cancer types analyzed (Beroukhim et al., 2010). Cancer subtypes are often characterized by tumor-specific patterns of chromosomal arm-level alterations (Ried et al., 2012), including squamous subtypes of lung, esophageal, and bladder tumors (Hoadley et al., 2014). Specific chromosome arm-level alterations have also defined groups of tumors that are responsive to particular therapies, such as low-grade gliomas with 1p/19q co-deletions that have been shown to be responsive to specific chemoradiotherapy regimens (Cairncross et al., 2013).

The functional effect of an individual aneuploidy is often thought to be due to the deletion of a tumor suppressor or overexpression of an oncogene (Liu et al., 2016). However, increasing evidence suggests that the effects of broad SCNAs result from the alteration of a combination of genes (Xue et al., 2012; Bonney et al., 2015). Recent SCNA models suggest that the phenotypic effect of chr_17p loss is due to more than *TP53* loss of heterozygosity (Liu et al., 2016). Chr_3p loss is an early event in lung squamous cell carcinoma observed in preneoplastic regions in the lung (Hung et al., 1995; Sundaresan et al., 1992), yet studies suggest that its contributions to tumor development are not the result of loss of one gene (Wistuba et al., 2000).

To study the role of aneuploidy in tumor development, models of whole chromosome aneuploidy have been developed. Trisomies and monosomies have been extensively modeled in yeast, where they slow proliferation and induce proteosomal stress (Torres et al., 2007; Sheltzer et al., 2011). Mouse cells with Robertsonian translocations can be used to model one trisomy at a time (Williams et al., 2008), or cells can be compared that differ only by single chromosomes added by microcell-mediated cellular transfer (Sheltzer et al., 2017; Stingele et al., 2012). These studies also show that whole chromosome aneuploidy leads to cellular senescence and decreased proliferation and transformation capabilities; however, karyotype evolution can lead to a rescue of proliferation rates (Sheltzer et al., 2017). There is one characterized model of targeted chromosome arm-level deletion, where chromosome arm 8p was deleted in mammary epithelial cells by targeting TALENs at either end of the chromosome arm and screening for recombination (Cai et al., 2016). In this model, 8p deletion does not lead to an increase in growth rate or tumorigenic potential.

Recent advances in targeting of endonucleases allow approaches to generating broad chromosomal alterations *in vitro*. CRISPR-Cas9 systems are particularly advantageous for their high efficiency and design tools available for targeting DNA throughout the genome (Mali et al., 2013; Hsu et al., 2013). CRISPR-targeting, when combined with an artificial telomere-containing plasmid, can be used to truncate a chromosome arm (Uno et al., 2017).

Here, we apply methods that define chromosome arm-level aneuploidy and a global cancer aneuploidy score to 10,522 tumors of 33 types in The Cancer Genome Atlas (TCGA), and develop and analyze an experimental cellular model of chromosome arm-level aneuploidy. By combining analysis of highly annotated cancer genomes and the experimental ability to manipulate chromosomes, we can advance our understanding of the effects of aneuploidy and specific chromosome arm-level alterations in cancer development and progression.

RESULTS

Generation of Aneuploidy Scores for 10,522 TCGA Cancers

To study features associated with aneuploidy, we first generated an aneuploidy score reflecting the total number of chromosome arms with arm-level copy-number alterations in a sample. When identifying these arm-level alterations, we were cognizant that two SCNAs that did not overlap could be misconstrued as a single event if we simply tallied arms with more than 50% of their length altered—for instance, if the SCNAs originated on opposite ends of the chromosome arm. Conversely, a simple length cutoff might exclude smaller SCNAs even if they cover most of the genomic region covered by larger SCNAs called as arm-level events (Figure S1A). We therefore clustered SCNAs on each arm based upon their locations and lengths (using a Gaussian mixture model) to identify events with likely similar consequences (Figure S1B and Method S1). Clusters in which the mean length of that SCNA was greater than 80% of the chromosome arm were considered positive for an arm alteration, samples whose SCNA length was less than 20% of a chromosome arm were considered negative, and clusters whose SCNAs were of intermediate length were not called.

We applied this approach to 10,522 samples spanning 33 cancer types from the TCGA pan-cancer dataset (Table S1). Somatic DNA copy-number was determined from Affymetrix SNP 6.0 array profiling of tumor samples. From SNP array data and mutational data, we used the ABSOLUTE algorithm (Carter et al., 2012) to generate segmented absolute copy-number and estimate sample purity, ploidy, and number of whole genome doublings. SCNAs, identified by ABSOLUTE to be clonal, were termed deviations from the euploid level of the sample: we considered a purely tetraploid cell (identified in 18 samples in our analysis) as having no arm-level SCNAs. (See the STAR Methods and Table S1 for additional information about expression and mutation data.) Using this arm-calling approach, we determined the arm-level SCNA status of more than 400,000 chromosome arms and more than 175,000 non-acrocentric whole chromosomes for the 10,522 cancer genomes (Table S2).

We then calculated an aneuploidy score that reflects the total burden, or number, of arm-level events in each sample. This aneuploidy score ranged from 0 to 39, the total number of human autosomal chromosome arms: p and q arms for chromosomes 1–12 and 16–20, together with q arms for the acrocentric chromosomes 13–15 and 21–22 (Figures 1A and S1C). Previous studies have demonstrated that there is less than a 2-fold change in frequency of arm alteration due to arm length (Beroukhim et al., 2010), and our aneuploidy score highly correlates with fraction of genome altered by aneuploidy (Spearman's rank correlation coefficient = 0.975, Figure S1D). For these reasons, we chose the sum of arms altered for subsequent analyses.

Samples that have undergone whole genome doubling, as defined by ABSOLUTE, have a higher degree of aneuploidy (Spearman's rank correlation coefficient = 0.55), suggesting that tumors with increased ploidy are more prone to aneuploidy events (Figure 1B). However, among samples of the same genome doubling status, increased aneuploidy is generally associated with decreasing ploidy, indicating that the

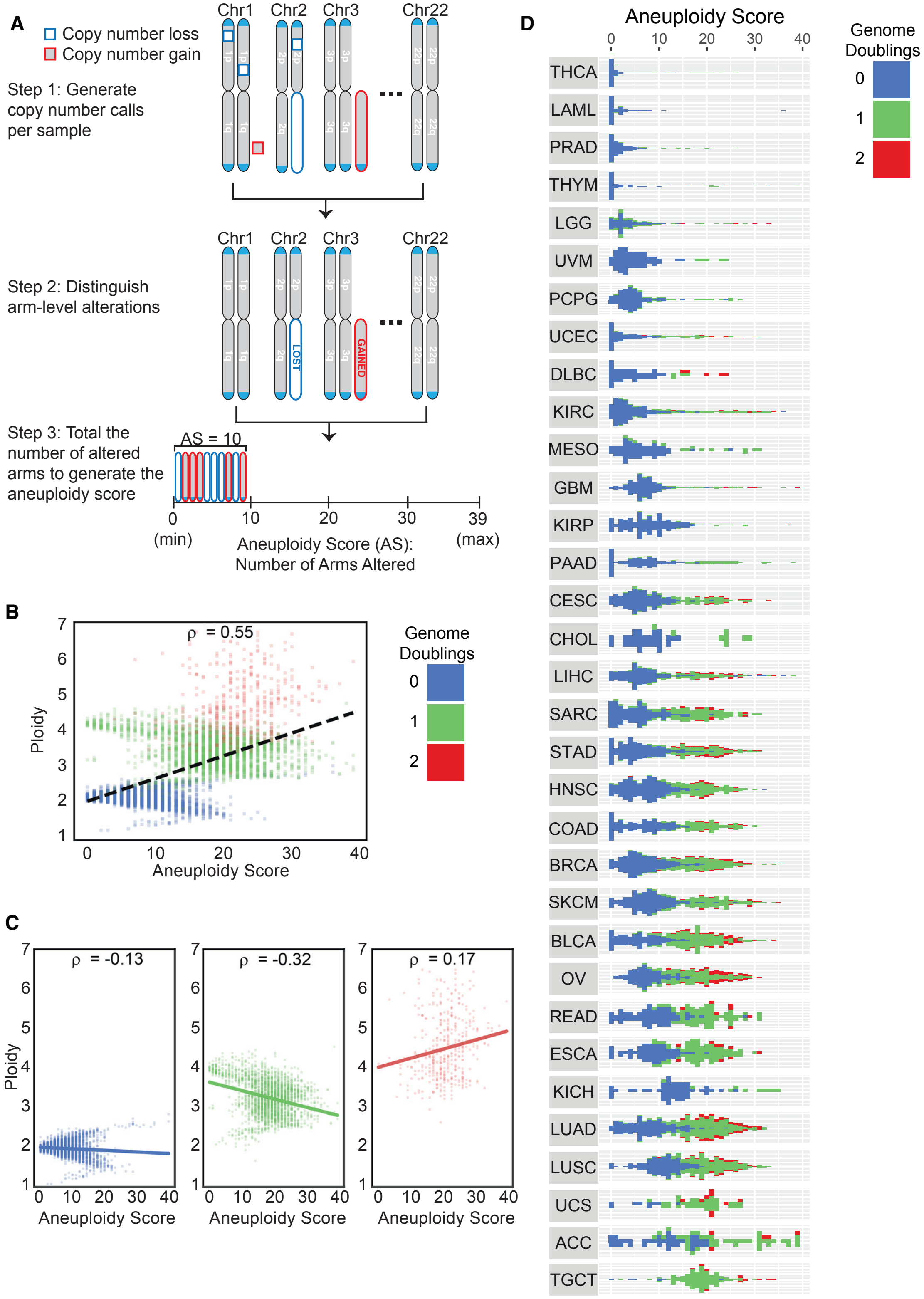


Figure 1. Aneuploidy Correlates with Ploidy, Genome Doubling Status, and Tumor Type

(A) Schematic of aneuploidy score. Step 1 is to generate copy-number calls per sample, including somatic copy-number alterations (SCNAs) of all sizes. Step 2 is to distinguish arm-level alterations within these SCNAs. Step 3 is to total the number of altered arms to generate the aneuploidy score.

(legend continued on next page)

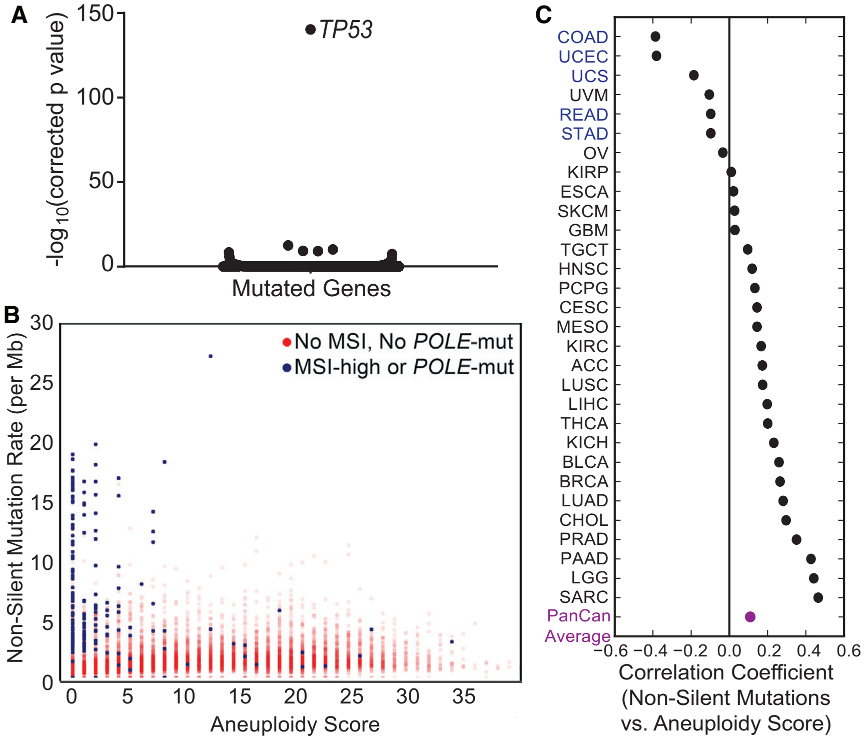


Figure 2. Aneuploidy Score Correlates with *TP53* Mutations and Overall Mutation Rate

(A) The y axis is $-\log_{10}$ Bonferroni corrected p value for linear model coefficient of aneuploidy score. Dots represent every mutated gene.

(B) The x axis is aneuploidy score. The y axis is rate of non-silent mutations per megabase (square root). Blue samples have been called as microsatellite instability (MSI)-high or *POLE* mutated, whereas red samples do not have these features or have not been called.

(C) Spearman correlation coefficients for aneuploidy score and mutation rate across TCGA tumor types, arranged from smallest to largest value. Tumor types in blue have MSI-high or *POLE* mutated samples. Average of correlation coefficients across cancer types is in purple.

See also Figure S2 and Tables S3 and S4.

arm-level events that contribute to aneuploidy are more often absolute losses than gains (Figure 1C).

Across the entire sample panel, 88% of cancers had at least some detectable aneuploidy (mean aneuploidy score of 10.0); however, this rate varied substantially across cancer types (SD between cancer type aneuploidy means = 4.8; Figure 1D). For example, only 26% of thyroid carcinomas have any chromosome arm-level alteration (mean aneuploidy score of 0.87), and less than half of acute myeloid leukemias and thymomas have these alterations (mean aneuploidy scores of 1.6 and 3.8, respectively). In contrast, virtually all glioblastomas (99%, mean aneuploidy score of 8.2), uterine carcinosarcomas (96%, mean aneuploidy score of 17.2), and testicular germ cell tumors (99%, mean aneuploidy score of 18.7) have at least one aneuploidy event (Figure 1D).

Relationship between Cancer Mutation Frequencies and Aneuploidy

A variety of studies have assessed the relationship between chromosomal copy-number alteration and somatic mutation (Ciriello et al., 2013; Zack et al., 2013; Davoli et al., 2017). We applied the aneuploidy scores of the full pan-cancer cohort of TCGA cases to examine this relationship. To assess whether gene mutations are associated with aneuploidy, we used a multivariable linear regression model that accounts for cancer type and the number of mutations per sample.

In this analysis, *TP53* was an outlier, with the highest coefficient in the linear model (toward enrichment of mutations among aneuploid samples) and the highest statistical significance (Figure 2A; Table S3), consistent with previous studies (Ciriello et al., 2013; Zack et al., 2013; Davoli et al., 2017). We also detected significant associations between aneuploidy and mutation rate for 34 other genes. However, the correlation coefficients for these genes were always negative (toward enrichment of mutations in low aneuploidy samples) and never exceeded a magnitude of 0.027, whereas the coefficient for *TP53* was positive 0.13.

Next, we assessed the relationship between somatic mutation rate and aneuploidy. It has been previously reported that there is an inverse relationship between the frequencies of recurrent copy-number alterations and of recurrent somatic mutations in cancer (Ciriello et al., 2013; Figure S2A). Although we do also observe that cancers with very high mutation rates have low aneuploidy scores (Figure 2B), these high mutation rate/low aneuploidy cancers largely exhibit high levels of microsatellite instability (MSI) or *POLE* mutation, primarily in colon adenocarcinoma and endometrial cancers (Figure 2B).

In contrast, when these hypermutated tumors are excluded, we observed a positive correlation between mutation frequency

---

(B) Each tumor sample is organized by genome doubling status (blue, not doubled; green, one genome doubling; red, two or more genome doublings). The x axis is aneuploidy score, the sum of the number of altered chromosome arms. The y axis is ploidy as determined by ABSOLUTE. Spearman's rank correlation coefficient = 0.55.

(C) The x axis is aneuploidy score, the sum of the number of altered chromosome arms. The y axis is ploidy as determined by ABSOLUTE. Samples are separated by whole genome doubling status: samples without genome doubling (left, blue; Spearman's rank correlation coefficient = −0.13), samples with one genome doubling (middle, green; Spearman's rank correlation coefficient = −0.32), and samples with two or more genome doublings (right, red; Spearman's rank correlation coefficient = 0.17).

(D) Each tumor sample is organized by tumor type and genome doubling status (blue, not doubled; green, one genome doubling; red, two or more genome doublings). Samples are organized by tumor type, and ranked from least to most aneuploid samples within a tumor type. The x axis is aneuploidy score, the y axis is sample number.

See also Figure S1 and Tables S1 and S2.

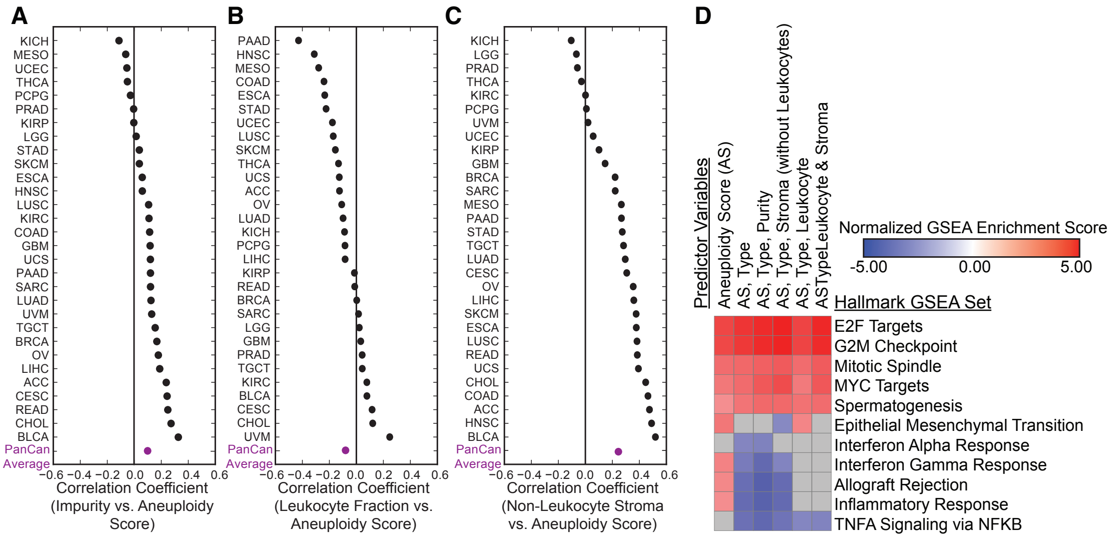


Figure 3. Aneuploidy Score Negatively Correlates with Immune Infiltrate, Which Contributes to Decreased Expression of Immune Genes

(A) Spearman correlation coefficients for aneuploidy score and impurity across TCGA tumor types, arranged from smallest to largest value. Average of correlation coefficients across tumor types is in purple.

(B) Spearman correlation coefficients for aneuploidy score and leukocyte fraction across TCGA tumor types, arranged from smallest to largest value. Average of correlation coefficients across tumor types is in purple.

(C) Spearman correlation coefficients for aneuploidy score and non-leukocyte stroma across TCGA tumor types, arranged from smallest to largest value. Average of correlation coefficients across tumor types is in purple.

(D) Heatmap of normalized enrichment scores for Hallmark gene sets in GSEA (gene set enrichment analysis), with FWER (family-wise error rate) p value < 0.01. Gray, not significant or not enriched. Predictor variables describe variables included in linear regression model for gene expression. Pathways are those identified from genes with significant coefficients in linear regression analysis.

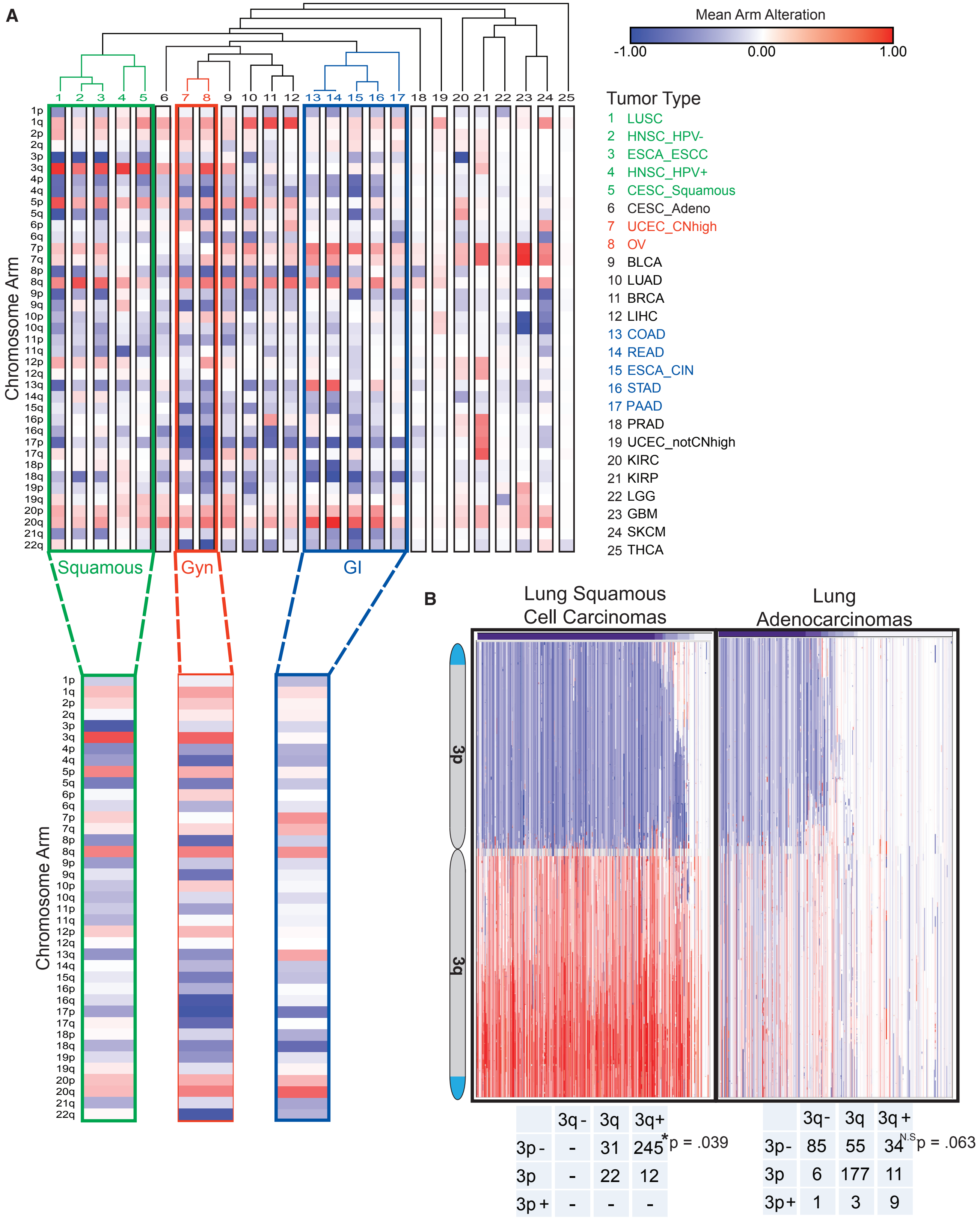
See also Figure S2 and Tables S4, S5, and S6.

and aneuploidy score (Spearman's rank correlation coefficient = 0.38), as well as between recurrent SCNAs and recurrent mutations (Spearman's rank correlation coefficient = 0.34). The positive correlation between mutation and aneuploidy was found in most cancer types with the notable exceptions of colorectal carcinoma (COAD and READ) and uterine carcinoma (UCEC and UCS), as well as uveal melanoma and stomach adenocarcinoma (Figure 2C; Table S4). All of these exceptions other than uveal melanoma had some cases of MSI or *POLE* mutation.

Relationship between Immune Infiltrates and Aneuploidy

Another important question about the cancer genome is the relationship between genome alterations and the immune response. Tumor mutational burden is known to be associated with response to immune checkpoint inhibition in cancer (Topalian et al., 2016; Rizvi et al., 2015), while aneuploidy has been reported to be associated with decreased immune infiltrate across many tumor types (Davoli et al., 2017). To assess the relationship between aneuploidy and immune infiltrates, we performed analyses of gene expression as a function of aneuploidy scores. The contribution of cells such as fibroblasts, leukocytes, endothelial cells, and other cell types is a confounding factor in the analysis of gene expression in primary tumors, so we also controlled for these features in our analysis. We measured immune and stromal cell populations using computational analyses rather than pathology-based estimates, as sections of the tumor used for histology were different from those used for molecular profiling. For estimating purity, we used ABSOLUTE, an established copy-number pipeline used in most TCGA studies that has been benchmarked histologically (Carter et al., 2012; Zack et al., 2013). For the leukocyte fraction, estimates based on methylation (Thorsson et al., 2018) and expression (Aran et al., 2017) correlate with a Spearman's rank correlation coefficient of 0.706 (Figure S2B). Namely, we first identified features reflecting cellular composition that associated with aneuploidy, and included these in our regression model tying gene expression to aneuploidy.

We first found that cancer impurity, as measured by ABSOLUTE (Carter et al., 2012), positively correlated with aneuploidy; in other words, cancers with high aneuploidy were associated with a higher fraction of non-cancerous cells (Figure 3A; Table S4). However, the measurement of leukocyte fractions based on methylation signatures, as described previously (Carter et al., 2012; Thorsson et al., 2018), showed a negative correlation with aneuploidy (Figure 3B; Table S4). Most individual tumor types showed a negative correlation between aneuploidy and leukocyte fraction, which was strongest in pancreatic adenocarcinoma and head and neck squamous cell carcinoma (HNSC) (Spearman's rank correlation coefficients = −0.428 and −0.312, respectively) (Figure 3B; Table S4). This observation is consistent with previous reports that aneuploidy is associated with decreased levels of cancer immune infiltrate (Davoli et al., 2017). Paradoxically, pan-cancer analysis shows a slight positive correlation between aneuploidy score and leukocyte fraction (Spearman's rank



	3q -	3q	3q+
3p -	-	31	245
3p	-	22	12
3p +	-	-	-

*p = .039

	3q-	3q	3q +
3p-	85	55	34
3p	6	177	11
3p+	1	3	9

N.S p = .063

(legend on next page)

correlation = 0.0568, Figure S2C), but this is driven by differences between tumor types as there is a negative correlation within most tumor types (Figure S2D). Combining the aneuploidy and leukocyte fraction results, the observed positive correlation of aneuploidy with impurity could be related to a non-leukocyte cellular fraction that is associated with aneuploidy (Figure 3C; Table S4).

Another way to assess immune infiltrates is by gene expression rather than methylation data. Using RNA sequencing data from 9,670 TCGA cancers for which there was also aneuploidy data (Table S1; STAR Methods), we generated a linear regression model relating the expression of each gene (measured by RNA-Seq by Expectation Maximization [RSEM] values of RNA sequencing data) in each sample as a function of aneuploidy score, across all cancer types (Table S5). A ranked list of significant aneuploidy score coefficients for each gene was analyzed by gene set enrichment (GSEA) with the MSigDB hallmark gene sets.

In the case of univariate linear regression, we observed a statistically significant (family-wise error rate [FWER] p value < 0.01) enrichment for proliferation pathways (such as E2F targets, mitotic spindle, and G2M checkpoint) and immune pathways (including interferon gamma response, allograft rejection, and immune response) (Figure 3D, column 1; Table S6). At first glance, this result appears contradictory to previously published studies (Davoli et al., 2017; Buccitelli et al., 2017). However, when we added tumor type as an additional variable in the linear regression model, immune gene sets were negatively associated with aneuploidy scores (Figure 3D, column 2; Table S6), consistent with other studies that assessed correlations within specific tumor types (Davoli et al., 2017; Buccitelli et al., 2017). When calculating an immune gene set expression score per sample, we also observed a pan-cancer-positive correlation which is driven by tumor type, consistent with the linear regression model (Figures S2E and S2F).

To determine whether the purity and immune infiltrate factors contribute to the aneuploidy-correlated expression patterns, we added them to our linear model. Adding purity or non-leukocyte stroma as variables did not affect the enrichment of immune signatures (Figure 3D, columns 3 and 4; Table S6), but addition of leukocyte fraction as a variable resulted in the loss of the immune signature enrichment (Figure 3D, column 5; Table S6). These results suggest that the enrichment of the immune expression signature was due to the fraction of leukocyte infiltrate present in that sample. In contrast, pro-proliferative and cell cycle pathways were significantly positively correlated with aneuploidy score, regardless of other predictors in the model (Figure 3D and Table S6).

Cancer-type-specific Patterns of Chromosome Arm-Level Aneuploidies

Subsequently, we assessed the rates of whole chromosome and chromosome arm alteration across human cancer types. Arm or whole chromosome alterations occurred in 88% of cancer samples. Not including acrocentric chromosomes, whole chromosome alterations occurred in 66% of samples and chromosome arm-level alterations occurred in 78% of samples (Table S2). Individual arms and chromosomes were altered at different frequencies (Table S7); 8p and 17p were the most frequently deleted (at 33% and 35%, respectively), 8q was the most commonly gained (33% of samples), and 2p and 2q the least commonly altered (total 18% and 16%, respectively). Chromosome arms 6p, 12q, 17q, and 19q were gained and lost in equal percentages (difference between gain and loss frequency < 0.03), but others were predominantly gained (1q, 7p, 8q, and 20q) or predominantly lost (3p and 17p).

We observed that every cancer type harbors a unique pattern of aneuploidy, with different arms or whole chromosomes altered at different frequencies (Table S7). To compare different cancer types and molecular subtypes, we performed hierarchical clustering of mean arm-level calls (Figure 4A; Table S8), previously performed across 3,000 TCGA samples (Beroukhim et al., 2010). Within an individual cancer type, molecular subtypes often clustered together, such as glioblastoma subtypes and testicular germ cell tumor subtypes (Figure S3A). Consistent with previous analyses, endometrial cancer was a notable exception, as copy-number-high "serous like tumors" cluster separately from other endometrial subtypes (Figure 4A and Cancer Genome Atlas Research Network et al., 2013). Testicular germ cell tumors were characterized by chromosome 12p gain among other aneuploidy events (Taylor-Weiner et al., 2016). Glioblastomas without IDH mutations are characterized by chromosome 7 gain and chromosome 10 loss, consistent with previous studies (Brennan et al., 2013). Low-grade gliomas cluster near glioblastomas, but are characterized by distinct alterations of chromosome 1p loss and chromosome 19q gain correlating with IDH mutation (Cancer Genome Atlas Research Network et al., 2015), a pattern shown to be therapeutically relevant (Cairncross et al., 2013).

Several cross-tumor clusters emerged from our analysis (Figure 4A and Table S8). Gastrointestinal tumors (colorectal, non-squamous esophageal, stomach, and pancreatic) clustered together with co-occurring gains of 8q, 13q, and chromosome 20, regardless of the status of microsatellite or chromosome instability. We also observed a cluster of some gynecological tumors (ovarian cancer and endometrial cancers with high copy-number alterations) and a second cluster of epithelial tumors characterized by 1q gain (lung adenocarcinoma, breast cancers, and liver hepatocellular carcinoma). Neural lineage cancers (low-grade glioma, glioblastoma, and melanoma) formed a cluster neighboring additional mesoderm-derived tumors (endometrial cancers with few copy-number alterations, renal clear cell carcinoma, and renal papillary cell carcinoma), characterized by recurrent chromosome 7 gain and fewer aneuploidy alterations.

Figure 4. Patterns of Arm-Level Alterations Cluster by Tumor Site, Tissue of Origin, and Squamous Morphology

(A) Matrix of mean arm-level alteration within each tumor type/subtype. Hierarchical clustering of tumor type by Pearson's method.

(B) Integrated genomics viewer (IGV) plots of chromosome 3 copy-number alterations in lung squamous cell carcinoma or lung adenocarcinoma. Blue, loss; red, gain. Numbers for lung squamous and lung adenocarcinoma samples that had arm calls for both 3p and 3q. The p values represent chi-square test for enrichment of co-occurrence of chromosome 3p loss and chromosome 3q gain.

See also Figure S3 and Tables S7 and S8.

Previous clustering efforts by multiple data types across 12 tumor types showed that squamous cancers from different tissues of origin (lung, esophagus, and bladder) clustered together (Hoadley et al., 2014). Our analyses also separated out squamous cancers based on aneuploidy data alone, suggesting that broad SCNAs are a major determinant of the squamous cluster. A dominant feature of the squamous cancer cluster was chromosome arm 3p loss and chromosome arm 3q gain, which is present in cervical squamous cell carcinomas and HPV-positive HNSCs, and strongest in esophageal squamous cell carcinomas, lung squamous cell carcinoma, and HPV-negative HNSCs.

Chromosome 3 alterations are a known feature of lung squamous cell carcinoma (Bass et al., 2009; Hagerstrand et al., 2013; Zabarovsky et al., 2002), with chromosome 3p loss present in preneoplastic lesions in the lung (Hung et al., 1995; Sundaresan et al., 1992). In our lung squamous cell carcinoma dataset, chromosome 3p was deleted in almost 80% of tumors and chromosome arm 3q was gained in over 60% of tumors (Figure 4B). The co-occurrence of chr_3p loss and chr_3q gain was significantly more frequent than would be expected by chance (Figure 4B; chi-square test $p = 0.0386$). Alterations in the reverse direction, chr_3p gain and chr_3q loss, were not observed in our dataset. In contrast, in lung adenocarcinoma, less than half of tumors had 3p loss and only 13% had 3q gain (Campbell et al., 2016), and these did not significantly co-occur ($p = 0.0626$). Chr_3p gains occurred, though rarely, and chr_3q loss occurs at higher rates. As expected, deletion of 3p correlated with a reduction of 3p gene expression across TCGA samples (Figure S3B). Based on the linear modeling of gene expression, chr_3p alteration negatively correlated with hallmark sets of cell cycle (E2F targets/G2M checkpoint; Figure S3C, FWER p value < 0.01), epithelial mesenchymal transition, interferon gamma response, and tumor necrosis factor alpha signaling (Figures S3D and S3E, FWER p value < 0.01). These pathways were downregulated when 3p is gained and upregulated when 3p is lost.

Genome Engineering Approach to Delete Chromosome Arm 3p *In Vitro*

Chromosome arm-level aneuploidies occur in almost 80% of cancers, yet have been rarely modeled in human cells (Cai et al., 2016; Uno et al., 2017). Given the continuing scientific mystery regarding the function of cancer aneuploidy, and given that deletion of chromosome 3p occurs in almost 80% of squamous cell lung cancers, we wanted to further understand the effect of chr_3p deletion in lung epithelial cells. We developed a recombination directed approach to remove the 3p chromosome arm. Using the CRISPR-Cas9 system, we generated double-strand breaks (DSBs) centromeric to all genes on the 3p arm. We modified a plasmid containing an artificial telomere and puromycin selection cassette (Uno et al., 2017) by addition of 1 kb DNA of sequence homologous to the region centromeric of the chr_3p DSB (Figure 5A). The CRISPR plasmid and telomere-containing plasmid were co-transfected into cells, and the recombination event was selected for with puromycin treatment. We observed successful recombination verified by Sanger sequencing, in a human immortalized lung epithelial cell line (AALE, immortalized by SV40 large T antigen) (Figure S4A). To test for chr_3p hemizygous deletion, we performed qPCR on genomic DNA to compare chr_3p and chr_3q levels. Single-cell clones that were positive for recombination were positive for chr_3p hemizygous deletion by qPCR (Figure 5B), as well as whole genome sequencing (Figures 5C and S4B) and karyotyping (Figure 5D).

Chromosome Arm 3p Deleted Cells Evolve over Time

We evaluated the growth of cells with chromosome 3p truncation (Figure 6A). At the first round of characterization, ten passages post-single-cell cloning, chromosome 3p deleted cells proliferated more slowly than their non-deleted siblings (Figure 6B, p value < 0.05). The chr_3p hemizygously deleted cells did not undergo increased apoptosis as measured by propidium iodide staining (Figure S4C). However, we observed more cells in G1 of the cell cycle by propidium iodide staining in fixed cells, indicative of cell cycle arrest (Figure S4D, p value <0.001). AALE parental cells do not form colonies in soft agar (Lundberg et al., 2002), and 3p deleted cells were also negative in this assay for anchorage independent growth (data not shown). RNA sequencing of deleted clones and their non-deleted siblings confirmed downregulation of chr_3p genes in *cis*, statistically significant for 64% of genes (false discovery rate [FDR] < 0.05) (Figure S4E). *STAC* and *ROBO1* were the most downregulated 3p genes, decreased by more than 15-fold, and chr_3p genes *UBA7* and *LMCD1* were upregulated more than 2-fold. Chr_3p hemizygously deleted cells also had upregulation of interferon and immune response pathways by GSEA (FWER < 0.01) (Figure 6C), consistent with the finding that 3p copy-number was anti-correlated with these pathways in the pan-cancer analysis (Figures S3D and S3E).

After an additional 4–5 passages (approximately 10 population doublings) and one round of freeze-thaw, 3p deleted cells no longer proliferated more slowly (Figure 6D). Fifty-three percent of chr_3p genes were still significantly downregulated (FDR $<$ 0.05) (Figure S4F), with *STAC* and *ROBO1* decreased more than 13-fold. Interestingly, *UBA7*, an ubiquitin enzyme on 3p that targets interferon gene *ISG15*, was no longer upregulated. By GSEA, interferon pathways were still upregulated (Figure 6C), though enrichment scores were lower and other immune signatures were no longer significant. Later stage chr_3p deleted cells had downregulation of genes involved in the epithelial mesenchymal transition and angiogenesis (GSEA FWER < 0.01), but the implications of this remain obscure. We also observed subclones with duplication of the remaining full copy of chromosome 3 in two of the three deleted clones (Figures 6E and S4G). In these subclones with chromosome 3 duplications, 3q is gained and 3p is no longer lost. We conclude that selection of advantageous alterations or expression changes allow chr_3p deleted cells to overcome the negative growth effects of aneuploidy in this model system (Figure 6F). In this section, we have demonstrated that it is possible to model aneuploidy. We have not observed a neutral or positive affect on aneuploidy on cell proliferation. Results are summarized in Table 1.

DISCUSSION

Here, we present parallel computational and experimental approaches that provide insight into the largely unexplored role of aneuploidy in cancer. By calculating aneuploidy level in

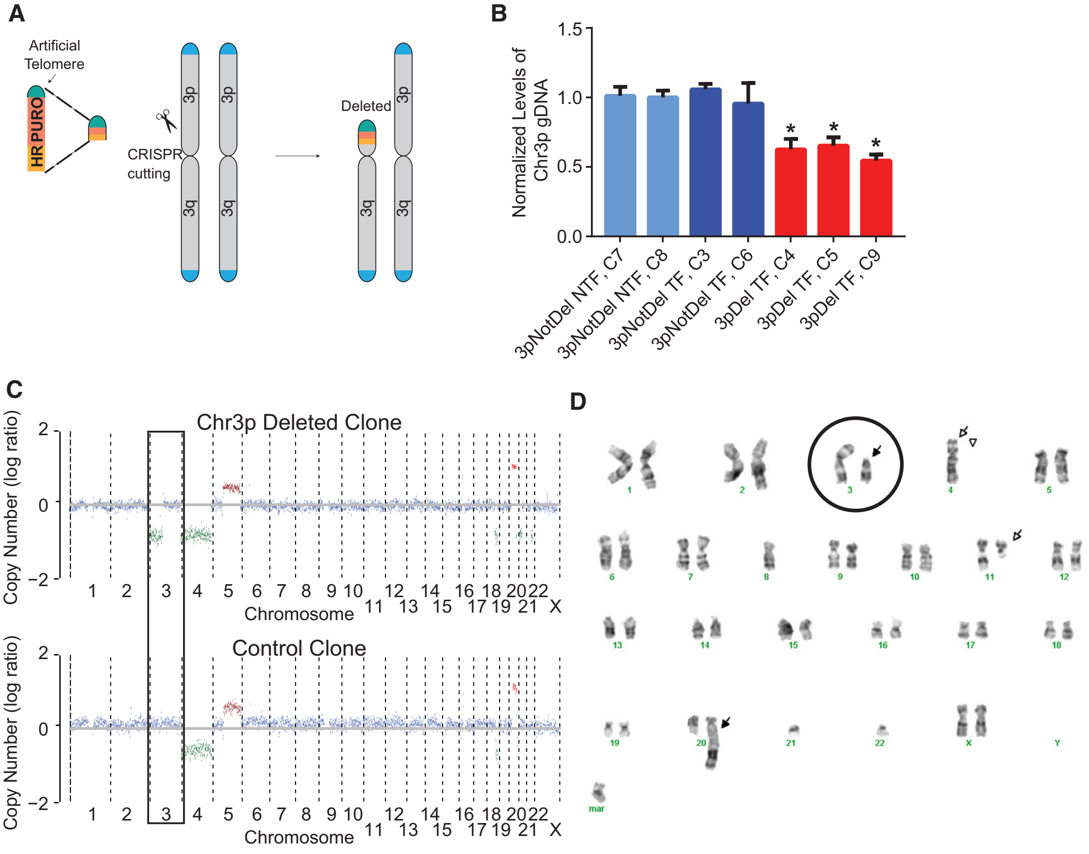


Figure 5. CRISPR-Based Approach can Delete a Chromosome Arm in Human Immortalized Cells

(A) Schematic of CRISPR and recombination-based approach to delete chr_3p *in vitro*. A linearized plasmid containing homologous DNA, a puromycin selection marker, and an artificial telomere is co-transfected with a CRISPR-Cas9 construct to target DNA sequence adjacent to the centromere. Upon transfection, a double-strand break is produced and repaired by homologous directed recombination, removing a chromosome arm and replacing it with an artificial telomere.

(B) qPCR measuring chr_3p gDNA normalized to chr_3q gDNA in single-cell clones. Bars represent means with error bars ± SD. Light blue, single-cell clones that were not transfected (NTF) and did not have chr_3p deletion; dark blue, single-cell clones that were transfected (TF) but did not have chr_3p deletion; and red, single-cell clones that were TF and have chr_3p deletion. *p value <0.05.

(C) Whole genome sequencing output of HMMCopy, for one chr_3p deleted cell clone (top) and a non-deleted control clone (bottom).

(D) Karyotype of one chr_3p deleted cell clone. Chromosome 3 is circled, arrows point to chromosomal abnormalities, and arrowheads point to missing chromosomes. Mar, marker chromosome.

See also Figure S4.

10,522 tumors, we uncovered the correlation of aneuploidy with *TP53* mutations, overall somatic mutation rate, and proliferative signatures, and an inverse correlation with leukocyte fraction. Our analysis revealed expression changes in cell cycle and immune hallmarks associated with individual chromosome arm-level alterations, independent of aneuploidy level. We also observe tissue-specific patterns of aneuploidy, and squamous tumors of different tissue origins clustering together. Using CRISPR technology, we modeled one of the chromosome arm alterations observed in squamous tumors (chr_3p deletion) in human immortalized lung epithelial cells. A decrease in cellular proliferation rate is associated with chr_3p deletion, a phenotype that is reversed by gain of chromosome 3.

Examining Pan-cancer Aneuploidy Analyses

In this study, we defined aneuploidy to include chromosome and arm-level SCNAs, but not smaller SCNAs. In some cases, computationally derived definitions of aneuploidy and analyses of SCNAs have included both broad and focal SCNAs (Davoli et al., 2017; Buccitelli et al., 2017). Broad and focal events occur by different mechanisms and have different effects; broad events affect a large number of genes by one or few copies, while focal events affect fewer genes with higher amplitudes of gain or with homozygous deletion (Beroukhim et al., 2007). Therefore we chose to restrict our analysis to the broad events, to harmonize the definition of aneuploidy more with the mechanism of generation of the structural and transcriptional consequences.

Aneuploidy level varies greatly across tumor type, impacting the analyses of aneuploidy correlations with mutation rate and immune signatures and potentially leading to conflicting results in the literature (Ciriello et al., 2013; Davoli et al., 2017; Buccitelli et al., 2017). By controlling for tumor type, as well as confounders such as sample purity and cellular composition, our analysis reconciles discrepancies in the literature. Regarding

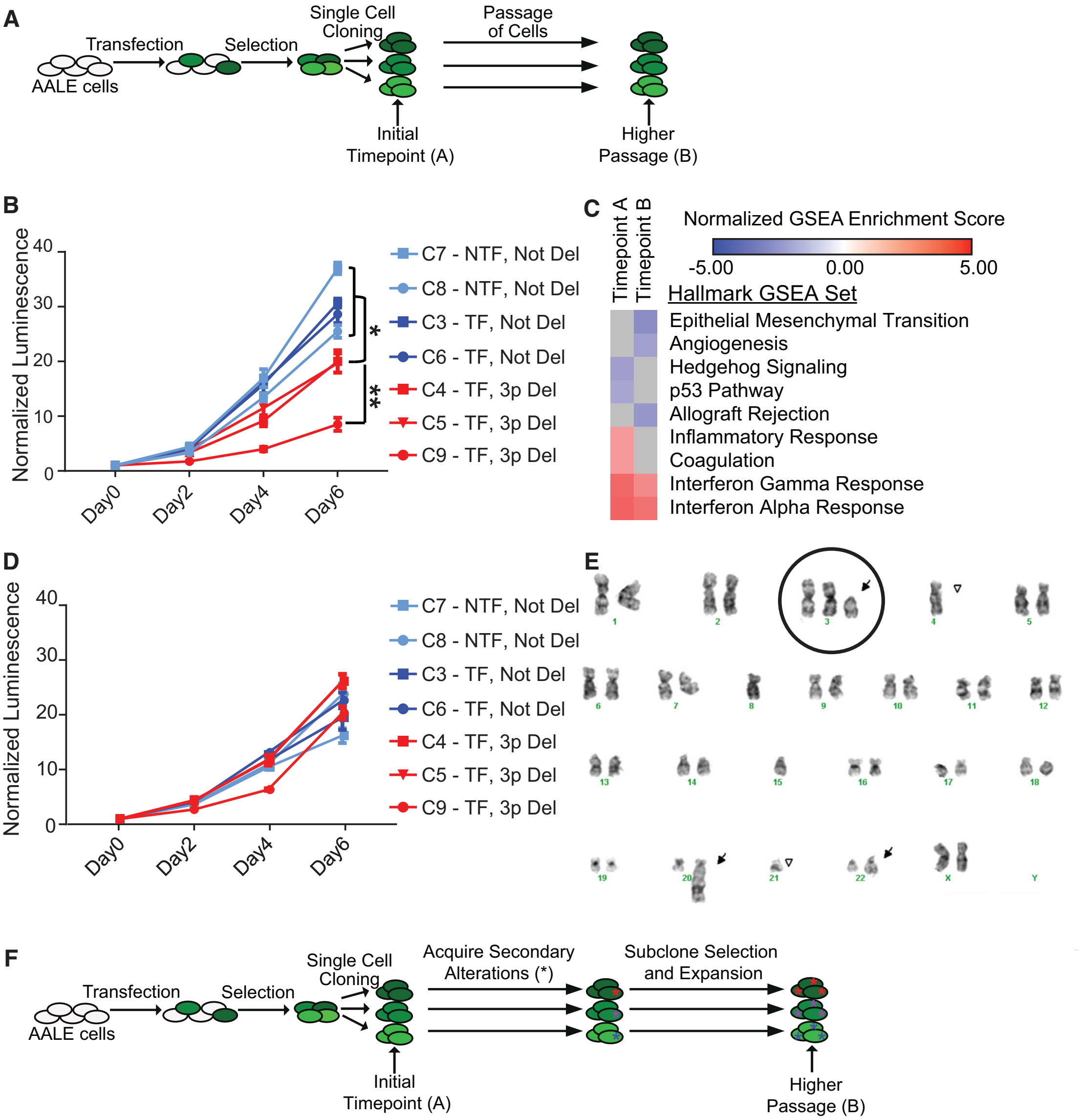


Figure 6. Chromosome 3p Lung Cells Initially have Slower Proliferation, but Normalize over Time

(A) AALE cells were transfected, and 1 day later cells were selected with puromycin for cells that had incorporated transfected DNA. Cells were single-cell cloned to isolate chr_3p deleted clones, and assayed before and after extensive passaging.

(B) Proliferation curves were generated using CellTiter-Glo over 6 days (x axis). The y axis is relative luminescence units, normalized to day 0. Data plotted are means with error bars ± SD. *p value <0.01, **p value <0.001. Cells were from time point (A).

(C) Heatmap of normalized enrichment scores for Hallmark gene sets in GSEA.

(D) Proliferation curves were generated using CellTiter-Glo over 6 days (x axis). The y axis is relative luminescence units, normalized to day 0. Data plotted are means with error bars ± SD. Cells were from the higher passage population.

(E) Karyotype from one of the chromosome 3p deleted clones. Chromosome 3 is circled, arrows point to chromosomal abnormalities, and arrowheads point to missing chromosomes.

(F) During passaging, cells have gained additional changes (*) to adapt to chr_3p deletion.

See also Figure S4.

mutation rate and aneuploidy correlations across cancers, mutation rate is inversely proportional to aneuploidy, but this anti-correlation is driven by tumors with high MSI (mostly colon adenocarcinoma and endometrial cancers). Once MSI tumors are removed, mutation and aneuploidy are positively correlated (consistent with Davoli et al., 2017 and Buccitelli et al., 2017).

When controlling for tumor type, our results are consistent with a negative correlation between aneuploidy and leukocyte infiltrate, which explains a decrease in expression of immune signatures. Interestingly, some studies have shown that abnormal karyotypes or hyperploidy can trigger an immune response (Santaguida et al., 2017; Senovilla et al., 2012). This result suggests the possibility that aneuploid tumor cells may have to

Table 1. Summary of Results

Aneuploidy and Ploidy

- Across 10,522 pan-cancer samples, the correlation coefficient between ploidy and aneuploidy score is 0.55.
  - Across 6,800 samples without genome doubling, the correlation coefficient is -0.13.
  - Across 3,242 samples with one genome doubling, the correlation coefficient is -0.32.
  - Across 480 samples with two or more genome doublings, the correlation coefficient is 0.17.

Aneuploidy and Mutations

- Across 9,702 samples (with copy number and mutation data) and corrected by tumor type and mutation rate, TP53 mutation is enriched in samples with higher aneuploidy, with a coefficient of 0.126 and a Bonferroni corrected p value of $5.58*10^{-141}$.
- Across 9,766 pan-cancer samples in 30 tumor types (with copy number, mutation rate, and leukocyte fraction data), the correlation coefficient between non-silent mutation rate and aneuploidy score is 0.336.
  - Across 191 samples with high microsatellite instability or *POLE* mutations, the correlation coefficient between non-silent mutation rate and aneuploidy score is -0.239.
  - Across 9,575 samples without high microsatellite instability or *POLE* mutations, the correlation coefficient between non-silent mutation rate and aneuploidy score is 0.379.

Aneuploidy and Microenvironment Correlations

- Across 9,766 pan-cancer samples, the average tumor-type correlation coefficient between impurity and aneuploidy score is 0.089.
- Across 9,766 pan-cancer samples, the average tumor-type correlation coefficient between leukocyte fraction and aneuploidy score is -0.077.
- Across 9,766 pan-cancer samples, the average tumor-type correlation coefficient between non-leukocyte stroma and aneuploidy score is 0.241.

Chromosome Arm-Alterations and Cross-Tumor Clusters

- Clustering 25 tumor subtypes revealed a group of gastrointestinal tumors, including colon adenocarcinoma, rectal adenocarcinoma, esophageal adenocarcinoma, stomach adenocarcinoma, and pancreatic adenocarcinoma. The Pearson's correlation coefficients range from 0.37 to 0.99.
- Clustering revealed a group of gynecological tumors, including endometrial carcinoma with high copy number levels and ovarian cancers. The Pearson's correlation coefficient is 0.90.
- Clustering revealed a group of squamous cell carcinomas, including lung squamous cell carcinoma, head and neck squamous cell carcinoma, esophageal squamous carcinoma, and cervical squamous carcinoma. The Pearson's correlation coefficients range from 0.58 to 0.94.

Phenotypes of Chromosome 3p Deletion

- Chromosome 3p deletion and chromosome 3q gain are significantly enriched in lung squamous cell carcinoma, with a p value of 0.039.
- Chromosome 3p deletion does not induce apoptosis.
- Chromosome 3p deletion induces slowed cell cycle (more cells in G_1), with a p value of 0.0006.
- Chromosome 3p deletion initially induces slower cellular proliferation (p value < 0.0017), rescued over time.
- Chromosome 3 duplication occurs in chromosome 3p deleted cells.

overcome or evade this ploidy-related immune response for tumors to progress.

When controlling for tumor type, aneuploidy level, and cellular composition of each sample, we identified immune expression signatures that can independently correlate with individual arm-level copy-number alterations. Deletions of some arms, including 3p, 8p, 13q, or 17p, are positively correlated with immune signatures, whereas deletions of other arms, including 4q, 5q, and 14q, are anti-correlated with immune signatures. The distinct correlations of immune signatures with deletion of different arms also suggest that immune signature changes might be due to specific genes or regions within each affected arm rather than to overall aneuploidy.

Genome Engineering to Model an Individual Cancer Aneuploidy

We have developed a method to generate chromosome arm deletions by initiating centromeric DNA breaks with CRISPR-Cas9 to promote replacement of an arm with a selection cassette and an artificial telomere. The method described here can be applied to deletion of any chromosome arm. In the case of chr_3p deletion, differential expression of chr_3p genes and interferon response genes is consistent with expression changes across TCGA samples, supporting the utility of both our computational and experimental approaches.

Consistent with single-chromosome aneuploidy models in human cells (Stingele et al., 2012; Sheltzer et al., 2017), chromosome arm-level deletion of human chr_3p has an anti-proliferative effect that cells can overcome with time. Even though p53 is inhibited in the lung epithelial cells, as they were immortalized with SV40 large T antigen (Lundberg et al., 2002), chr_3p deletion initially caused a decrease in cell proliferation. We initially predicted that deletion of chr_3p would not be as anti-proliferative as other aneuploidies, as it is thought to be an early event in lung tumorigenesis (Hung et al., 1995; Wistuba et al., 2000). However, in our cells, selective pressures have likely induced some secondary changes in the chr_3p deleted cells to promote proliferation.

One previously observed mechanism of selection is the evolution of cell line karyotypes over time (Sheltzer et al., 2017). In two of the three chr_3p deleted clones, we observed subclones with a duplicated wild-type copy of chromosome 3, changing an individual cell from chr_3p deleted to chr_3q gained. Interestingly, chr_3q gain is also a defining feature of squamous cell carcinoma, and contains oncogenes such as *SOX2*, *PIK3CA*, and *TERC*. In the future, we plan to expand our chromosome deletion model to different chromosome arms (Cai et al., 2016; Uno et al., 2017), and thereby to interrogate how specific aneuploidies can induce distinct expression patterns, phenotypes, and selection mechanisms.

Open Questions in Aneuploidy: Tissue Specificity and Selection

The tissue specificity we observe in patterns of aneuploidy could be due to the specific transcriptional effects of different aneuploidies. Our work here begins to analyze expression changes induced by each arm alteration while controlling for overall aneuploidy level and other confounders. By focusing our analysis within tumor types and comparing between them, we can uncover whether specific aneuploidies have different transcriptional consequences in different tissues. Determining whether transcriptional consequences have distinct selective advantages and disadvantages in different tissues will require further integration of *in vitro* models and computational analysis. In addition, specific aneuploidies may trigger unique dependencies; future directions involve interrogating existing *in vitro* CRISPR and small hairpin RNA cell line data to identify potential therapeutic targets that can be validated in the *in vitro* models.

The major question about cancer aneuploidy remains open: is aneuploidy positively selected for in cancer? We favor the view that aneuploidy is positively selected for, as it is a universal and tissue-consistent feature of epithelial cancer. However, the experimental data on cancer aneuploidy do not currently support this model. There are a few possible explanations for this contradiction. In the case of oncogenic mutations, overexpression in model systems often leads to oncogene-induced senescence (Larsson, 2011); individual aneuploidies may elicit a similar response. In addition, we have not yet identified the exact environment for positive selection of aneuploidy; p53/RB inhibition and *TERT* activation may not be sufficient to allow selection of aneuploidies. Future work will involve combining models of diverse aneuploidy events, including chromosome arm gains and losses, along with different mutations and different cellular contexts. It will be increasingly important to incorporate models of stromal interaction into prospective studies of aneuploidy and its selection, whether by advanced cellular modeling, single-cell sequencing, or use of animal models. In particular, CRISPR manipulation in human lung organoids will be a useful approach to assess interactions of different lung cell types with aneuploidy alterations, as well as potential stromal cell interactions in a 3D tissue model (Barkauskas et al., 2017). Continued integration of computational and experimental approaches will be required to understand how an aneuploidy alteration, affecting hundreds of genes simultaneously, results in an aneuploidy phenotype and contributes to tumor development or, alternatively, is an unselected event.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ Pan-Cancer Computational Analyses
  - ○ Cell Line Analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, eight tables, and one method and can be found with this article online at https://doi.org/10.1016/j.ccell.2018.03.007.

ACKNOWLEDGMENTS

All cytogenomics analysis was performed by the CytoGenomics Core at Brigham and Women's Hospital. A plasmid containing an artificial telomere was kindly shared by Narumi Uno and Mitsuo Oshimura (Uno et al., 2017). A.M.T. was funded by grants from Uniting Against Lung Cancer and the Lung Cancer Research Foundation and is an American Cancer Society postdoctoral fellow (grant number 128194-PF-16-001-01-CCG). M.M. acknowledges support from the National Cancer Institute (1R35CA197568) and the Norman R. Seaman Endowment fund. M.M. is an American Cancer Society Research Professor (grant number 127046-RP-14-230-01-TBG). Funding sources for The Cancer Genome Atlas Research Network include: U54 HG003273 (PI Richard A. Gibbs), U54 HG003067 (PIs Stacey Gabriel and Eric S. Lander), U54 HG003079 (PI Richard K. Wilson), U24 CA143799 (PI Terence Paul Speed and Paul T. Spellman), U24 CA143835 (PI Ilya Shmulevich), U24 CA143840 (PIs Marc Ladanyi and Chris Sander), U24 CA143843 (PIs Richard A. Gibbs and David Andrew Wheeler), U24 CA143845 (PIs Lynda Chin and Gad Getz), U24 CA143848 (PI David N. Hayes and Charles M. Perou), U24 CA143858 (PIs Joshua Stuart, Christopher Benz, and David H. Haussler), U24 CA143866 (PI Marco Antonio Marra), U24 CA143867 (PI Stacey Gabriel and Matthew L. Meyerson), U24 CA143882 (PIs Stephen B. Baylin and Peter W. Laird), U24 CA143883 (PIs Gordon B. Mills, John N. Weinstein, Rehan Akbani, and W.K. Alfred Yung), U24 CA144025 (PI Raju S. Kucherlapati), and P30 CA016672 (PI Gordon B. Mills).

AUTHOR CONTRIBUTIONS

A.M.T., J.S., G.H., G.F.G., A.C.B., S.E.S., H.H., and J.L. performed computational analyses, A.M.T. and X.Z. performed *in vitro* experiments and sequencing, TCGA Research Network provided genomic data of patient samples, A.D.C., R.B., and M.M. supervised the work, C.W. and A.J.L. provided feedback and advice on analyses, A.M.T. and M.M. wrote the manuscript.

DECLARATION OF INTERESTS

Michael Seiler, Peter G. Smith, Ping Zhu, Silvia Buonamici, and Lihua Yu are employees of H3 Biomedicine, Inc. Parts of this work are the subject of a patent application: WO2017040526 titled "Splice variants associated with neomorphic sf3b1 mutants." Shouyoung Peng, Anant A. Agrawal, James Palacino, and Teng Teng are employees of H3 Biomedicine, Inc. Andrew D. Cherniack, Ashton C. Berger, and Galen F. Gao receive research support from Bayer Pharmaceuticals. Gordon B. Mills serves on the External Scientific Review Board of Astrazeneca. Anil Sood is on the Scientific Advisory Board for Kiyatec and is a shareholder in BioPath. Jonathan S. Serody receives funding from Merck, Inc. Kyle R. Covington is an employee of Castle Biosciences, Inc. Preethi H. Gunaratne is founder, CSO, and shareholder of NextmiRNA Therapeutics. Christina Yau is a part-time employee/consultant at NantOmics. Franz X. Schaub is an employee and shareholder of SEngine Precision Medicine, Inc. Carla Grandori is an employee, founder, and shareholder of SEngine Precision Medicine, Inc. Robert N. Eisenman is a member of the Scientific Advisory Boards and shareholder of Shenogen Pharma and Kronos Bio. Daniel J. Weisenberger is a consultant for Zymo Research Corporation. Joshua M. Stuart is the founder of Five3 Genomics and shareholder of NantOmics. Marc T. Goodman receives research support from Merck, Inc. Andrew J. Gentles is a consultant for Cibermed. Charles M. Perou is an equity stock holder, consultant, and Board of Directors member of BioClassifier and GeneCentric

Diagnostics and is also listed as an inventor on patent applications on the Breast PAM50 and Lung Cancer Subtyping assays. Matthew Meyerson receives research support from Bayer Pharmaceuticals; is an equity holder in, consultant for, and Scientific Advisory Board chair for OrigiMed; and is an inventor of a patent for EGFR mutation diagnosis in lung cancer, licensed to LabCorp. Eduard Porta-Pardo is an inventor of a patent for domainXplorer. Han Liang is a shareholder and scientific advisor of Precision Scientific and Eagle Nebula. Da Yang is an inventor on a pending patent application describing the use of antisense oligonucleotides against specific lncRNA sequence as diagnostic and therapeutic tools. Yonghong Xiao was an employee and shareholder of TESARO, Inc. Bin Feng is an employee and shareholder of TESARO, Inc. Carter Van Waes received research funding for the study of IAP inhibitor ASTX660 through a Cooperative Agreement between NIDCD, NIH, and Astex Pharmaceuticals. Raunaq Malhotra is an employee and shareholder of Seven Bridges, Inc. Peter W. Laird serves on the Scientific Advisory Board for AnchorDx. Joel Tepper is a consultant at EMD Serono. Kenneth Wang serves on the Advisory Board for Boston Scientific, Microtech, and Olympus. Andrea Califano is a founder, shareholder, and advisory board member of DarwinHealth, Inc. and a shareholder and advisory board member of Tempus, Inc. Toni K. Choueiri serves as needed on advisory boards for Bristol-Myers Squibb, Merck, and Roche. Lawrence Kwong receives research support from Array BioPharma. Sharon E. Plon is a member of the Scientific Advisory Board for Baylor Genetics Laboratory. Beth Y. Karlan serves on the Advisory Board of Invitae.

Received: December 15, 2017
Revised: February 14, 2018
Accepted: March 6, 2018
Published: April 2, 2018

REFERENCES

Adalsteinsson, V.A., Ha, G., Freeman, S.S., Choudhury, A.D., Stover, D.G., Parsons, H.A., Gydush, G., Reed, S.C., Rotem, D., Rhoades, J., et al. (2017). Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat. Commun. *8*, 1324.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

Barkauskas, C.E., Chung, M.I., Fioret, B., Gao, X., Katsura, H., and Hogan, B.L. (2017). Lung organoids: current uses and future promise. Development *144*, 986–997.

Bass, A.J., Watanabe, H., Mermel, C.H., Yu, S., Perner, S., Verhaak, R.G., Kim, S.Y., Wardwell, L., Tamayo, P., Gat-Viks, I., et al. (2009). SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat. Genet. *41*, 1238–1242.

Beroukhim, R., Getz, G., Nghiemphu, L., Barretina, J., Hsueh, T., Linhart, D., Vivanco, I., Lee, J.C., Huang, J.H., Alexander, S., et al. (2007). Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc. Natl. Acad. Sci. USA *104*, 20007–20012.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Bonney, M.E., Moriya, H., and Amon, A. (2015). Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. Genes Dev. *29*, 898–903.

Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462–477.

Buccitelli, C., Salgueiro, L., Rowald, K., Sotillo, R., Mardin, B.R., and Korbel, J.O. (2017). Pan-cancer analysis distinguishes transcriptional changes of aneuploidy from proliferation. Genome Res. *27*, 501–511.

Cai, Y., Crowther, J., Pastor, T., Abbasi Asbagh, L., Baietti, M.F., De Troyer, M., Vazquez, I., Talebi, A., Renzi, F., Dehairs, J., et al. (2016). Loss of chromosome 8p governs tumor progression and drug response by altering lipid metabolism. Cancer Cell *29*, 751–766.

Cairncross, G., Wang, M., Shaw, E., Jenkins, R., Brachman, D., Buckner, J., Fink, K., Souhami, L., Laperriere, N., Curran, W., et al. (2013). Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402. J. Clin. Oncol. *31*, 337–343.

Campbell, J.D., Alexandrov, A., Kim, J., Wala, J., Berger, A.H., Pedamallu, C.S., Shukla, S.A., Guo, G., Brooks, A.N., Murray, B.A., et al. (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat. Genet. *48*, 607–616.

Cancer Genome Atlas Research Network, Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., et al. (2013). Integrated genomic characterization of endometrial carcinoma. Nature *497*, 67–73.

Cancer Genome Atlas Research Network, Brat, D.J., Verhaak, R.G., Aldape, K.D., Yung, W.K., Salama, S.R., Cooper, L.A., Rheinbay, E., Miller, C.R., Vitucci, M., et al. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N. Engl. J. Med. *372*, 2481–2498.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421.

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.

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., and Sander, C. (2013). Emerging landscape of oncogenic signatures across human cancers. Nat. Genet. *45*, 1127–1133.

Davoli, T., Uno, H., Wooten, E.C., and Elledge, S.J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science *355*, https://doi.org/10.1126/science.aaf8399.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al. (2018). Scalable open scienceapproach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Systems *6*, 271–281.

Hagerstrand, D., Tong, A., Schumacher, S.E., Ilic, N., Shen, R.R., Cheung, H.W., Vazquez, F., Shrestha, Y., Kim, S.Y., Giacomelli, A.O., et al. (2013). Systematic interrogation of 3q26 identifies TLOC1 and SKIL as cancer drivers. Cancer Discov. *3*, 1044–1057.

Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D.M., Niu, B., McLellan, M.D., Uzunangelov, V., et al. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell *158*, 929–944.

Holland, A.J., and Cleveland, D.W. (2009). Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. Nat. Rev. Mol. Cell Biol. *10*, 478–487.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hung, J., Kishimoto, Y., Sugio, K., Virmani, A., McIntire, D.D., Minna, J.D., and Gazdar, A.F. (1995). Allele-specific chromosome 3p deletions occur at an early stage in the pathogenesis of lung carcinoma. JAMA *273*, 558–563.

Lai, D., Ha, G., and Shah, S. (2016). HMMcopy: Copy Number Prediction with Correction for GC and Mappability Bias for HTS Data. R Package Version 1.18.0. https://rdrr.io/bioc/HMMcopy/.

Larsson, L.G. (2011). Oncogene- and tumor suppressor gene-mediated suppression of cellular senescence. Semin. Cancer Biol. *21*, 367–376.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Liu, Y., Chen, C., Xu, Z., Scuoppo, C., Rillahan, C.D., Gao, J., Spitzer, B., Bosbach, B., Kastenhuber, E.R., Baslan, T., et al. (2016). Deletions linked to

TP53 loss drive cancer through p53-independent mechanisms. Nature *531*, 471–475.

Lundberg, A.S., Randell, S.H., Stewart, S.A., Elenbaas, B., Hartwell, K.A., Brooks, M.W., Fleming, M.D., Olsen, J.C., Miller, S.W., Weinberg, R.A., et al. (2002). Immortalization and transformation of primary human airway epithelial cells by gene transfer. Oncogene *21*, 4577–4586.

Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. Nat. Methods *10*, 957–963.

Mitelman, F. (2000). Recurrent chromosome aberrations in cancer. Mutat. Res. *462*, 247–253.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Ried, T., Hu, Y., Difilippantonio, M.J., Ghadimi, B.M., Grade, M., and Camps, J. (2012). The consequences of chromosomal aneuploidy on the transcriptome of cancer cells. Biochim. Biophys. Acta *1819*, 784–793.

Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science *348*, 124–128.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

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.

Senovilla, L., Vitale, I., Martins, I., Tailler, M., Pailleret, C., Michaud, M., Galluzzi, L., Adjemian, S., Kepp, O., Niso-Santano, M., et al. (2012). An immunosurveillance mechanism controls cancer cell ploidy. Science *337*, 1678–1684.

Sheltzer, J.M., Blank, H.M., Pfau, S.J., Tange, Y., George, B.M., Humpton, T.J., Brito, I.L., Hiraoka, Y., Niwa, O., and Amon, A. (2011). Aneuploidy drives genomic instability in yeast. Science *333*, 1026–1030.

Sheltzer, J.M., Ko, J.H., Replogle, J.M., Habibe Burgos, N.C., Chung, E.S., Meehl, C.M., Sayles, N.M., Passerini, V., Storchova, Z., and Amon, A. (2017). Single-chromosome gains commonly function as tumor suppressors. Cancer Cell *31*, 240–255.

Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. Mol. Syst. Biol. *8*, 608.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Sundaresan, V., Ganly, P., Hasleton, P., Rudd, R., Sinha, G., Bleehen, N.M., and Rabbitts, P. (1992). p53 and chromosome 3 abnormalities, characteristic of malignant lung tumours, are detectable in preinvasive lesions of the bronchus. Oncogene *7*, 1989–1997.

Tang, Y.C., and Amon, A. (2013). Gene copy-number alterations: a cost-benefit analysis. Cell *152*, 394–405.

Taylor-Weiner, A., Zack, T., O'Donnell, E., Guerriero, J.L., Bernard, B., Reddy, A., Han, G.C., AlDubayan, S., Amin-Mansour, A., Schumacher, S.E., et al. (2016). Genomic evolution and chemoresistance in germ-cell tumours. Nature *540*, 114–118.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.-H., Porta Pardo, E., Gao, G., Eddy, J.A., Plaisier, C.L., et al. (2018). The iImmune landscape of cancer. Immunity, https://doi.org/10.1016/j.immuni.2018.03.023.

Topalian, S.L., Taube, J.M., Anders, R.A., and Pardoll, D.M. (2016). Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. Nat. Rev. Cancer *16*, 275–287.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science *317*, 916–924.

Uno, N., Hiramatsu, K., Uno, K., Komoto, S., Kazuki, Y., and Oshimura, M. (2017). CRISPR/Cas9-induced transgene insertion and telomere-associated truncation of a single human chromosome for chromosome engineering in CHO and A9 cells. Sci. Rep. *7*, 12739.

Weaver, B.A., and Cleveland, D.W. (2006). Does aneuploidy cause cancer? Curr. Opin. Cell Biol. *18*, 658–667.

Williams, B.R., Prabhu, V.R., Hunter, K.E., Glazier, C.M., Whittaker, C.A., Housman, D.E., and Amon, A. (2008). Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. Science *322*, 703–709.

Wistuba, I.I., Behrens, C., Virmani, A.K., Mele, G., Milchgrub, S., Girard, L., Fondon, J.W., 3rd, Garner, H.R., McKay, B., Latif, F., et al. (2000). High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. Cancer Res. *60*, 1949–1960.

Xue, W., Kitzing, T., Roessler, S., Zuber, J., Krasnitz, A., Schultz, N., Revill, K., Weissmueller, S., Rappaport, A.R., Simon, J., et al. (2012). A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions. Proc. Natl. Acad. Sci. USA *109*, 8212–8217.

Zabarovsky, E.R., Lerman, M.I., and Minna, J.D. (2002). Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. Oncogene *21*, 6915–6935.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhang, C.Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat. Genet. *45*, 1134–1140.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
Nextera DNA Sample Preparation Kit	Illumina	Catalog: FC-121-1030
NEBNext Ultra Directional RNA Library Prep Kit	New England Biolabs	Catalog: E7420S
QIAamp DNA Mini Kit	Qiagen	Catalog: 51304
Qiagen RNeasy Mini Kit	Qiagen	Catalog: 74104
CellTiter-Glo Luminescence Viability Assay	VWR	Catalog: PAG7572
Deposited Data		
Raw data files for RNA-sequencing of *in vitro* AALE clones	This paper	https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP133935
Raw data files for DNA-sequencing of *in vitro* AALE clones	This paper	https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP133935
Experimental Models: Cell Lines		
Immortalized Lung Epithelial Cells (AALE)	Lundberg et al., 2002	#N/A
Oligonucleotides		
Chr_3p CRISPR (5’ TGATCAGTCAGGTAAGGATG 3’)	This paper	#N/A
Recombination PCR Forward (5’ CTACCCGCTTCCATTGCTCA 3’)	This paper	#N/A
Recombination PCR Reverse (5’ TTGGTTGAGCAGTTGGACAT 3’)	This paper	#N/A
Chr_3p qPCR Forward (5’ ACAATCCAAACTAGCATGCACA 3’)	This paper	#N/A
Chr_3p qPCR Reverse (5’ AGCGTTAGAGGGAGGGGAG 3’)	This paper	#N/A
Chr_3q qPCR Forward (5’ CGTGTCCGGGGTAGATCTTG 3’)	This paper	#N/A
Chr_3p qPCR Reverse (5’ GCTTACATCCTCGGGCAGAA 3’)	This paper	#N/A
Software and Algorithms		
ABSOLUTE	Carter et al., 2012	Version 1.5
Python Package SciKit-Learn	Pedregosa et al., 2011	Version 0.16.1
Python Package scipy stats	www.scipy.org	Version 0.19.0
edgeR	Robinson et al., 2010	
STAR	Dobin et al., 2013	
RSEM	Li and Dewey, 2011	
GSEA	Subramanian et al., 2005	Version 3.0
HMMCopy	Lai et al., 2016	
IchorCNA	Adalsteinsson et al., 2017	
Morpheus	https://software.broadinstitute.org/morpheus/index.html	

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Matthew Meyerson (matthew_meyerson@dfci.harvard.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All *in vitro* experiments were performed in XX immortalized lung epithelial cells (AALE cells), derived at Dana-Farber and immortalized by SV40 large-T antigen (Lundberg et al., 2002). Cells were maintained at 37 degrees Celsius and 5% CO_2 in Lonza small airway growth medium (CC-3118).

METHOD DETAILS

Pan-Cancer Computational Analyses

Datasets

For 10,522 TCGA samples, somatic DNA copy number was determined from Affymetrix SNP 6.0 arrays. For 9,670 of these 10,522 samples, RSEM (RNA-seq by Experimentation Maximization) expression values were determined from Illumina mRNA sequencing data. For 9,756 of these 10,522 samples, mutations were called from whole-exome DNA sequencing data (Ellrot et al., 2018). (See also Table S1.)

Calculating Arm-Level Events and Aneuploidy Score

Using the ABSOLUTE algorithm (Carter et al., 2012) for each sample, we determined the likeliest ploidy and absolute total copy number of each segment in the genome. Each segment was designated as amplified, deleted, or neutral based on whether its copy number was greater than, smaller than, or equal to the sample's ploidy (rounded to the nearest integer) respectively. For amplifications and deletions separately (collectively "alterations"), segments were joined until either the entire chromosome was considered altered or more than 20% of the genomic length between the start and ends were not altered in the same direction; e.g. >20% deleted or neutral for joining amplification segments. For every combination of arm/chromosome and direction of alteration within each TCGA tumor type, the start coordinates, end coordinates, and percentage length of the longest joined segment were clustered across samples using a Gaussian Mixture Model (Pedregosa et al., 2011, Python package SciKit-Learn). The optimal clustering solution between 2-9 clusters inclusive was chosen based on the lowest BIC (Bayesian information criterion). Tumors in clusters whose mean fraction altered in either specific direction was >=80% were considered "aneuploid." Tumors altered <20% (in both directions) were considered "non-aneuploid," and others were designated "other." Each arm was assigned -1 if lost, +1 if gained, 0 if non-aneuploid, and "NA" if other. Aneuploidy score (number of altered arms) for each tumor is calculated as the sum total of altered arms, for a range of 0 (no arms) to 39 (all arms – long and short arms for each non-acrocentric chromosome, and only long arms for chromosomes 13, 14, 15, 21, and 22).

Other Scores. Fraction of genome altered by aneuploidy was determined by multiplying each arm altered by its length, and dividing by the length of the genome. Recurrent SCNA scores are the sum of recurrent SCNAs as identified by GISTIC2.0 (output to be available at NCI Genomic Data Commons). Recurrent mutation scores are the sum of recurrent mutations, defined in the TCGA MC3 manuscript (Ellrott et al., 2018).

Spearman Correlations

Spearman correlation coefficients were calculated using the spearmanr function in the stats package of scipy-0.19.0, which was run using Python-3.5.4 or using cor.test in R (method = "spearman"), which was run using R version 3.2.3.

Linear Modeling

Linear modeling was performed in R version 3.2.3, using lm. The equations used:

$$Expression_{(gene\ X)} \sim \beta_1 {}^{*}predictor\ variable + \beta_2 {}^{*}aneuploidy\ score$$

$$Mutation_{(gene\ X)} \sim \beta_1 {}^{*}predictor\ variable + \beta_2 {}^{*}aneuploidy\ score$$

$$Expression_{(gene\ X)} \sim \beta_1 {}^{*}chromosome\ arm + \beta_2 {}^{*}aneuploidy\ score$$

P values for each coefficient were calculated by the lm function in R.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was performed using Broad GSEA v.3.0 (Subramanian et al., 2005). Ranked lists of genes and coefficients were entered, and enrichment was assessed using Hallmark and Positional Gene Sets.

Clustering

Hierarchical clustering of tissue-means of arm-level alterations was performed using 1-Pearson's in Morpheus (https://software.broadinstitute.org/morpheus/index.html).

Cell Line Analyses

Karyotype Analysis

Karyotype analysis was performed by the Brigham and Women's Cytogenomics Core.

RNA Sequencing and Analysis

RNA was isolated from cells using the Qiagen RNAeasy kit (Qiagen 74104), with DNAse treatment. PolyA isolation and sequencing library preparation was performed using the NEBNext Ultra Directional RNA Library Prep Kit. Each set of samples were pooled and sequenced in one lane of HiSeq2500 RR, 100 basepair paired end. Sequencing reads were aligned using STAR (Dobin et al., 2013) and expression level was quantified by RSEM (Li and Dewey, 2011). Differential expression analysis was performed using the edgeR pipeline (Robinson et al., 2010). Ranked lists of genes with fold change were entered into GSEA as described above.

DNA Sequencing and Analysis

DNA was isolated from cells using the QiaAmp Mini DNA kit. Sequencing library preparation was performed using the Nextera DNA Sample Preparation Kit. Samples were pooled and sequenced by miSeq, 300 basepair paired end. Copy number profiles were generated by HMMCopy (Lai et al., 2016). Subclonal analysis was performed using IchorCNA (Adalsteinsson et al., 2017).

Cell Maintenance, Transfection and Proliferation Assays

Transfections were performed using Fugene-6 (Promega E2691) following manufacturer's instructions at a 3:1 ratio. 24 hours after transfection, cells were selected with 2 μg/mL puromycin through several passages and a round of single cell cloning. For single cell cloning, individual cells were plated in each well of a 96-well plate using cell sorting, with each well containing 100μL of 50% conditioned small airway growth media (SAGM). After single cell cloning and confirmation of telomere recombination, puromycin was no longer added in the media. For cell proliferation assays, 1500 cells per well in a 96-well plate were plated in 100μL of media. Plates were collected at time of plating and days 2, 4, and 6. 50μL of CellTiter-Glo Reagent was added, and plates were incubated for 20 minutes before luminescence readings.

Flow Cytometry

For apoptosis analysis, adherent cells were washed, trypsinized, collected, and stained in propidium iodide solution. Propidium iodide levels were measured on the BD LSRii.

For cell cycle analysis, adherent cells were washed, trypsinized, collected, and fixed in 10% formalin buffered solution and 70% ethanol. After fixing, cells were stained with propidium iodide solution for thirty minutes before analysis on the BD LSRii. Cells were assigned to G1, S, and G2 stages of the cell cycle by identifying the G1 peak and 2x G2 peak.

PCR and qPCR

PCR was performed using the Sigma Aldrich AccuTaq enzyme (D8045) with primers listed under Key Resources Table. Quantitative PCR (qPCR) was performed using the Power SYBR Green PCR mastermix (ThermoFisher 4367659), with the primers listed above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantitative and statistical methods are noted above.

DATA AND SOFTWARE AVAILABILITY

The raw TCGA data, processed data and clinical data can be found at the legacy archive of the GDC (https://portal.gdc.cancer.gov/legacy-archive/search/f) and the PancanAtlas publication page (https://gdc.cancer.gov/about-data/publications/pancanatlas). The mutation data can be found here (https://gdc.cancer.gov/about-data/publications/mc3-2017). TCGA data can also be explored through the Broad Institute FireBrowse portal (http://gdac.broadinstitute.org) and the Memorial Sloan Kettering Cancer Center cBioPortal (http://www.cbioportal.org). For the *in vitro* assays, RNA-sequencing BAM files are available at the Sequence Read Archive as Study PRJNA436953 (SRP133935), Run Numbers SRR6806553-SRR6806554 and SRR6806559-SRR6806570. Related DNA-sequencing BAM files are available at the Sequence Read Archive as Study PRJNA436953 (SRP133935), Run Numbers SRR6806543-SRR6806552 and SRR6806555-SRR6806558. Details for software availability are in the Key Resources Table.

Cancer Cell, Volume *33*

Supplemental Information

Genomic and Functional Approaches to Understanding Cancer Aneuploidy

Alison M. Taylor, Juliann Shih, Gavin Ha, Galen F. Gao, Xiaoyang Zhang, Ashton C. Berger, Steven E. Schumacher, Chen Wang, Hai Hu, Jianfang Liu, Alexander J. Lazar, The Cancer Genome Atlas Research Network, Andrew D. Cherniack, Rameen Beroukhim, and Matthew Meyerson

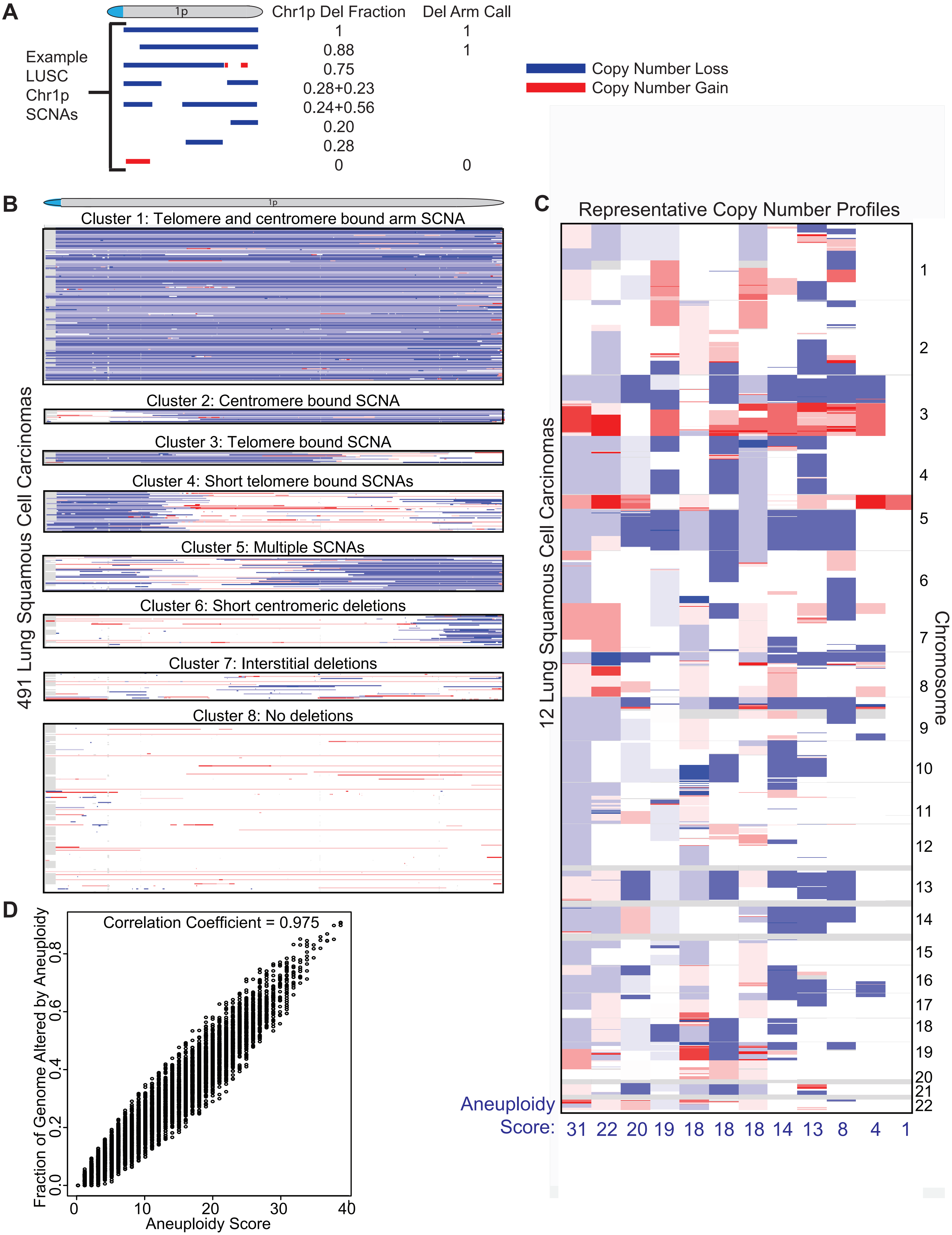

A
1p
Chr1p Del Fraction
Del Arm Call
Example
LUSC
Chr1p
SCNAs
1
1
0.88
1
0.75
0.28+0.23
0.24+0.56
0.20
0.28
0
0
Copy Number Loss
Copy Number Gain
B
1p
Cluster 1: Telomere and centromere bound arm SCNA
Cluster 2: Centromere bound SCNA
Cluster 3: Telomere bound SCNA
Cluster 4: Short telomere bound SCNAs
Cluster 5: Multiple SCNAs
Cluster 6: Short centromeric deletions
Cluster 7: Interstitial deletions
Cluster 8: No deletions
491 Lung Squamous Cell Carcinomas
C
Representative Copy Number Profiles
12 Lung Squamous Cell Carcinomas
Chromosome
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Aneuploidy
Score:
31
22
20
19
18
18
18
14
13
8
4
1
D
Correlation Coefficient = 0.975
Fraction of Genome Altered by Aneuploidy
0.0
0.2
0.4
0.6
0.8
0
10
20
30
40
Aneuploidy Score

Figure S1, related to Figure 1. Methods to call chromosome arm alterations and aneuploidy score.

(A) Schematic depicting 8 examples of chromosome 1p SCNAs in lung squamous cell carcinoma (LUSC). The fraction of arm deleted and final arm-level SCNA call are listed for each set of alterations on the arm. Blue = copy number loss, red = copy number gain.

(B) 491 lung squamous cell carcinomas are separated by cluster of chromosome 1p deletion status. Clusters 1 and 2 are called as arm-level deletions.

(C) 12 representative lung squamous cell carcinomas are depicted by absolute copy number, with aneuploidy score listed at the bottom.

(D) 10,522 cancers are plotted by aneuploidy score (x-axis) and fraction of genome altered by aneuploidy (y-axis), with Spearman's ranked correlation coefficient = 0.975.

Tumor Code	**Tumor Type**	**Samples with Aneuploidy Scores**	**Samples with Aneuploidy and Mutation Data**	**Samples with Aneuploidy and Expression Data**
ACC	Adrenocortical carcinoma	89	89	76
BLCA	Bladder urothelial carcinoma	402	401	397
BRCA	Breast invasive carcinoma	1048	977	1037
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma	296	280	291
CHOL	Cholangiocarcinoma	36	36	36
COAD	Colon adenocarcinoma	433	388	429
DLBC	Diffuse large b-cell lymphoma	47	37	47
ESCA	Esophageal carcinoma	163	163	161
GBM	Glioblastoma multiforme	569	381	144
HNSC	Head and neck squamous cell carcinoma	512	493	502
KICH	Kidney chromophobe	65	65	65
KIRC	Kidney renal clear cell carcinoma	483	345	481
KIRP	Kidney renal papillary cell carcinoma	280	273	280
LAML	Acute myeloid leukemia	124	116	109
LGG	Lower grade glioma	517	514	507
LIHC	Liver hepatocellular carcinoma	362	350	355
LUAD	Lung adenocarcinoma	503	499	497
LUSC	Lung squamous cell carcinoma	482	468	479
MESO	Mesothelioma	81	80	81
OV	Ovarian serous cystadenocarcinoma	552	393	287
PAAD	Pancreatic adenocarcinoma	166	162	158
PCPG	Pheochromocytoma and paraganglioma	164	164	160
PRAD	Prostate adenocarcinoma	472	472	471
READ	Rectum adenocarcinoma	155	137	154
SARC	Sarcoma	248	230	242
SKCM	Skin cutaneous melanoma	461	458	460
STAD	Stomach adenocarcinoma	427	423	401
TGCT	Testicular germ cell tumors	155	150	149
THCA	Thyroid carcinoma	471	465	461
THYM	Thymoma	106	106	103
UCEC	Uterine corpus endometrial carcinoma	517	505	514
UCS	Uterine carcinosarcoma	56	56	56
UVM	Uveal melanoma	80	80	80
	Total	**10522**	**9756**	**9670**

Table S1, related to Figure 1. TCGA sample information.

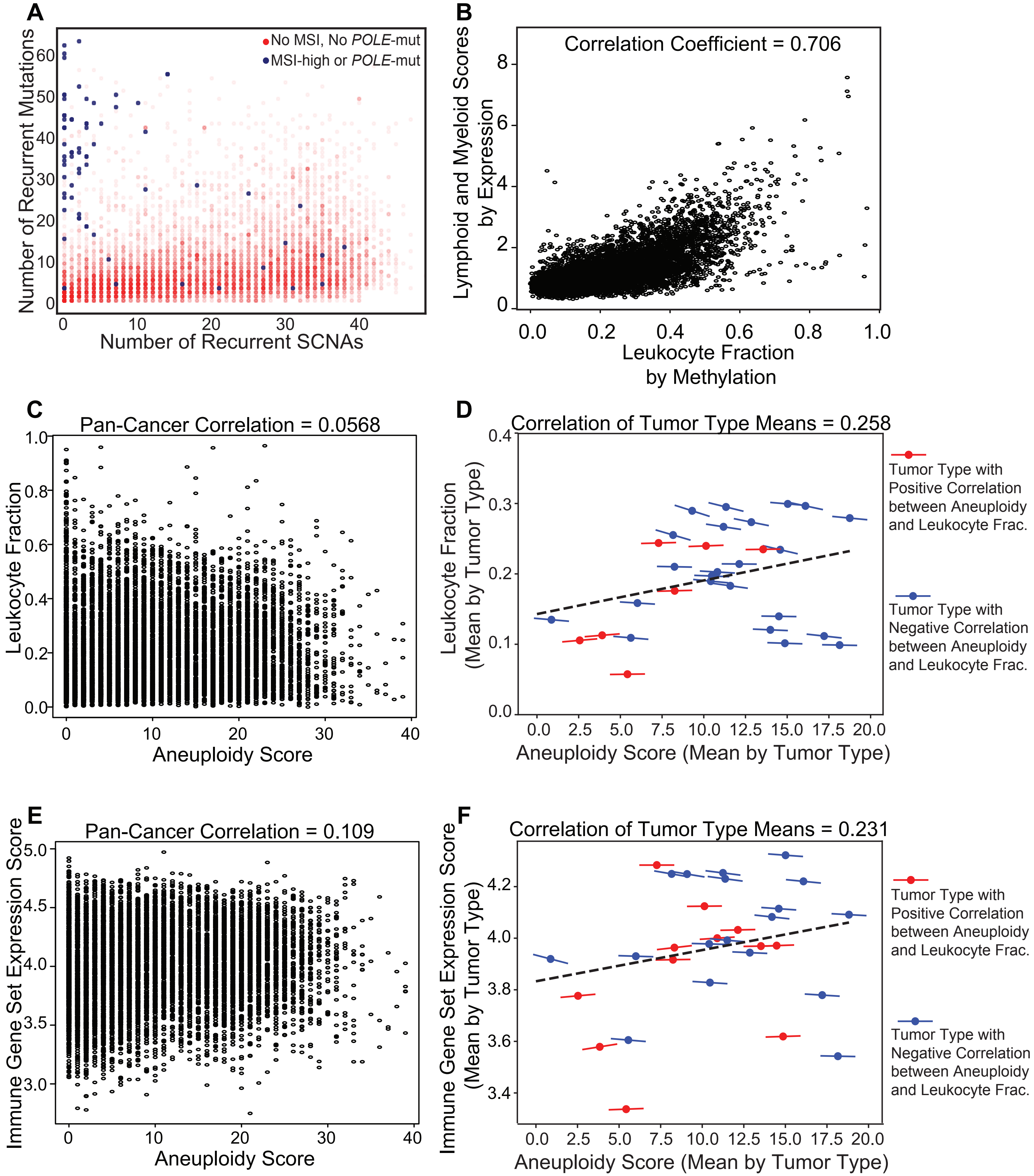

A
No MSI, No POLE-mut
MSI-high or POLE-mut
Number of Recurrent Mutations
0
10
20
30
40
50
60
Number of Recurrent SCNAs
0
10
20
30
40
B
Correlation Coefficient = 0.706
Lymphoid and Myeloid Scores by Expression
0
2
4
6
8
Leukocyte Fraction by Methylation
0.0
0.2
0.4
0.6
0.8
1.0
C
Pan-Cancer Correlation = 0.0568
Leukocyte Fraction
0.0
0.2
0.4
0.6
0.8
1.0
Aneuploidy Score
0
10
20
30
40
D
Correlation of Tumor Type Means = 0.258
Leukocyte Fraction (Mean by Tumor Type)
0.0
0.1
0.2
0.3
0.4
Aneuploidy Score (Mean by Tumor Type)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Tumor Type with Positive Correlation between Aneuploidy and Leukocyte Frac.
Tumor Type with Negative Correlation between Aneuploidy and Leukocyte Frac.
E
Pan-Cancer Correlation = 0.109
Immune Gene Set Expression Score
3.0
3.5
4.0
4.5
5.0
Aneuploidy Score
0
10
20
30
40
F
Correlation of Tumor Type Means = 0.231
Immune Gene Set Expression Score (Mean by Tumor Type)
3.4
3.6
3.8
4.0
4.2
Aneuploidy Score (Mean by Tumor Type)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Tumor Type with Positive Correlation between Aneuploidy and Leukocyte Frac.
Tumor Type with Negative Correlation between Aneuploidy and Leukocyte Frac.

Figure S2, related to Figures 2 and 3. Correlations between aneuploidy, immune populations, and gene expression.

A. X-axis is number of recurrent somatic copy number alterations (SCNAs), and y-axis is number of recurrent mutations. Blue samples have been called as microsatellite instability (MSI)-high or *POLE* mutated, whereas red samples do not have these features or have not been called.

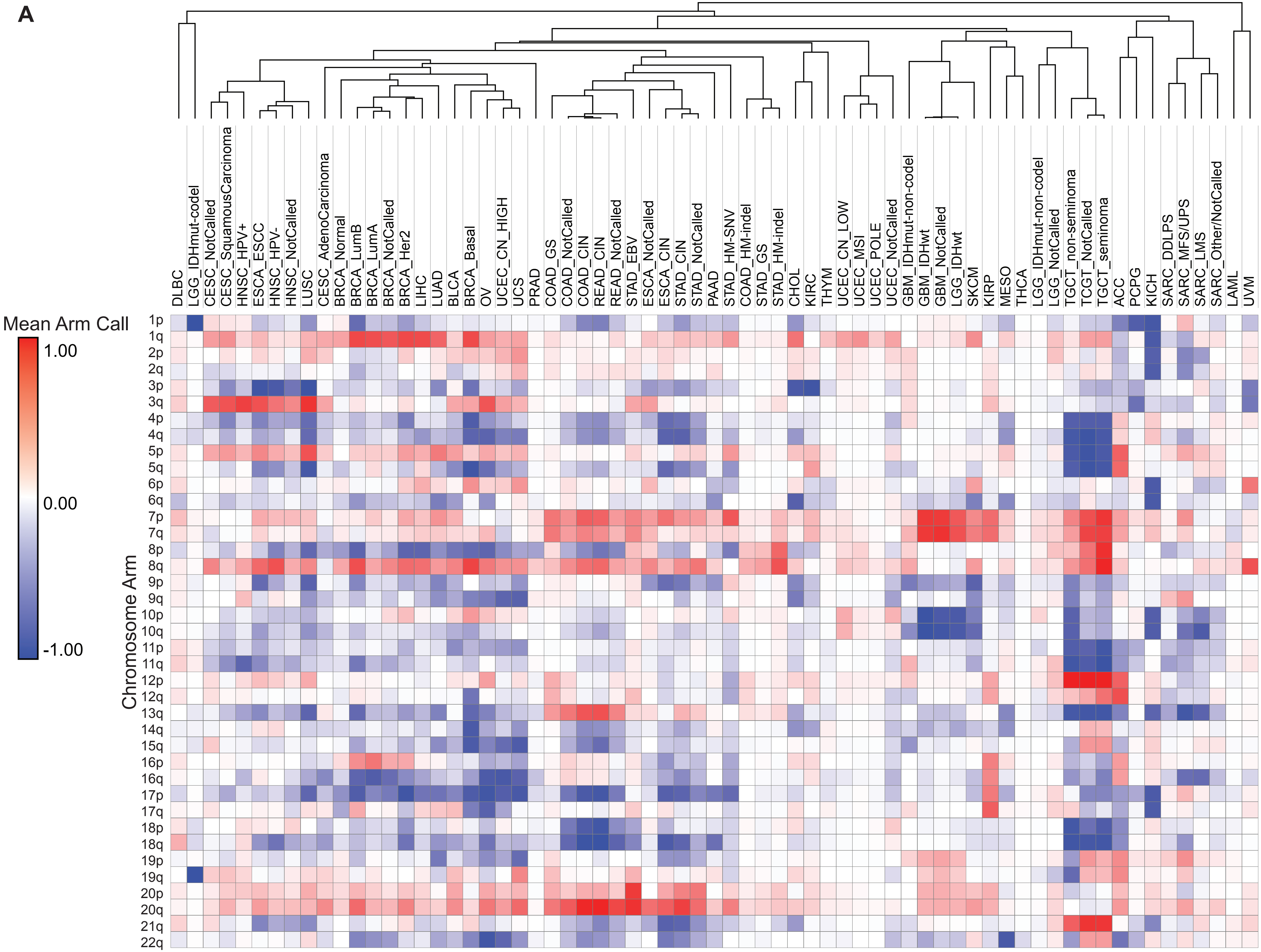
B. X-axis is leukocyte fraction by methylation (TCGA pan-cancer atlas immune paper, submitted), and y-axis is a sum of lymphoid and myeloid xCell scores based on expression data (Aran et al., 2017). Spearman's rank correlation coefficient = 0.706.

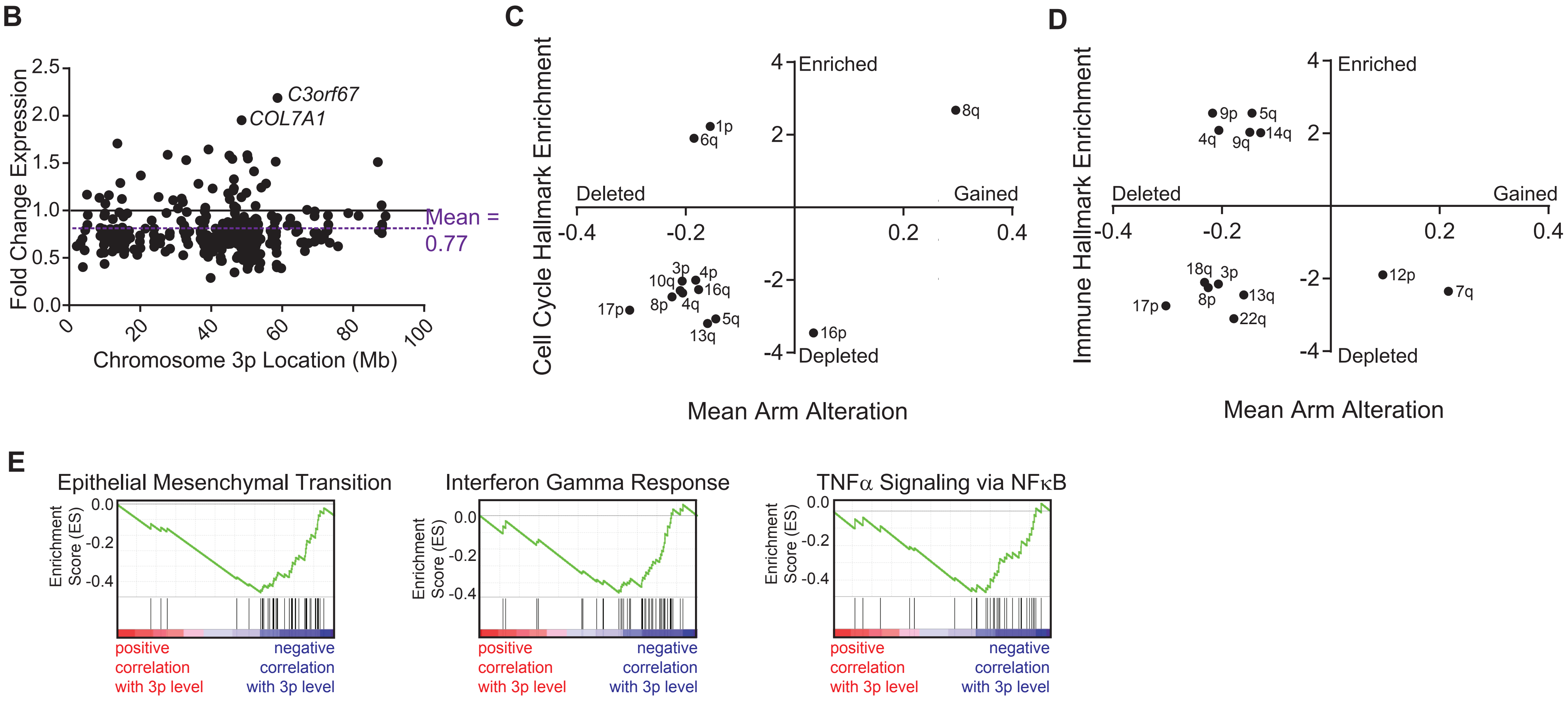
C. Samples from 30 cancer types are plotted by aneuploidy score (x-axis) and leukocyte fraction (from methylation data, y-axis). The pan-cancer Spearman's rank correlation coefficient = 0.0568.

D. X-axis is the mean aneuploidy score (for tumor type), y-axis is the mean leukocyte fraction. Colored dots each represent a tumor type, with the extended line reproducing the correlation slope within that tumor type. Positive correlations are colored red, negative correlations are colored blue. Dashed line is the correlation of tumor type means for aneuploidy score and leukocyte fraction, with Spearman's rank correlation coefficient = 0.258.

E. Samples from 30 cancer types are plotted by aneuploidy score (x-axis) and immune gene set expression score (y-axis). The pan-cancer Spearman's ranked correlation coefficient = 0.109.

F. X-axis is the mean aneuploidy score (for tumor type), y-axis is the mean immune gene set expression score. Colored dots each represent a tumor type, with the extended line reproducing the correlation slope within that tumor type. Positive correlations are colored red, negative correlations are colored blue. Dashed line is the correlation of tumor type means for aneuploidy score and immune expression, with Spearman's rank correlation coefficient = 0.231.


A
Mean Arm Call
1.00
0.00
-1.00
Chromosome Arm
DLBC
LGG_IDHmut-codel
CESC_NotCalled
CESC_SquamousCarcinoma
HNSC_HPV+
ESCA_ESCC
HNSC_HPV-
HNSC_NotCalled
LUSC
CESC_AdenoCarcinoma
BRCA_Normal
BRCA_LumB
BRCA_LumA
BRCA_NotCalled
BRCA_Her2
LIHC
LUAD
BLCA
BRCA_Basal
OV
UCEC_CN_HIGH
UCS
PRAD
COAD_GS
COAD_NotCalled
COAD_CIN
READ_CIN
READ_NotCalled
STAD_EBV
ESCA_NotCalled
ESCA_CIN
STAD_CIN
STAD_NotCalled
PAAD
STAD_HM-SNV
COAD_HM-indel
STAD_GS
STAD_HM-indel
CHOL
KIRC
THYM
UCEC_CN_LOW
UCEC_MSI
UCEC_POLE
UCEC_NotCalled
GBM_IDHmut-non-codel
GBM_IDHwt
GBM_NotCalled
LGG_IDHwt
SKCM
KIRP
MESO
THCA
LGG_IDHmut-non-codel
LGG_NotCalled
TGCT_non-seminoma
TCGT_NotCalled
TGCT_seminoma
ACC
PCPG
KICH
SARC_DDLPS
SARC_MFS/UPS
SARC_LMS
SARC_Other/NotCalled
LAML
UVM
1p
1q
2p
2q
3p
3q
4p
4q
5p
5q
6p
6q
7p
7q
8p
8q
9p
9q
10p
10q
11p
11q
12p
12q
13q
14q
15q
16p
16q
17p
17q
18p
18q
19p
19q
20p
20q
21q
22q



B
Fold Change Expression
2.5
2.0
1.5
1.0
0.5
0.0
C3orf67
COL7A1
Mean = 0.77
0
20
40
60
80
100
Chromosome 3p Location (Mb)
C
Cell Cycle Hallmark Enrichment
4
2
-2
-4
Enriched
Depleted
Deleted
Gained
-0.4
-0.2
0.2
0.4
1p
6q
8q
3p
4p
10q
16q
17p
8p
4q
5q
13q
16p
Mean Arm Alteration
D
Immune Hallmark Enrichment
4
2
-2
-4
Enriched
Depleted
Deleted
Gained
-0.4
-0.2
0.2
0.4
9p
5q
4q
9q
14q
18q
3p
8p
13q
17p
22q
12p
7q
Mean Arm Alteration
E
Epithelial Mesenchymal Transition
Interferon Gamma Response
TNFα Signaling via NFκB
Enrichment Score (ES)
0.0
-0.2
-0.4
positive correlation with 3p level
negative correlation with 3p level

Figure S3, related to Figure 4. Patterns of arm-level alterations cluster by tumor site, tissue of origin, and squamous morphology, and individual chromosome arm alterations affect gene expression in *cis* and in *trans*.

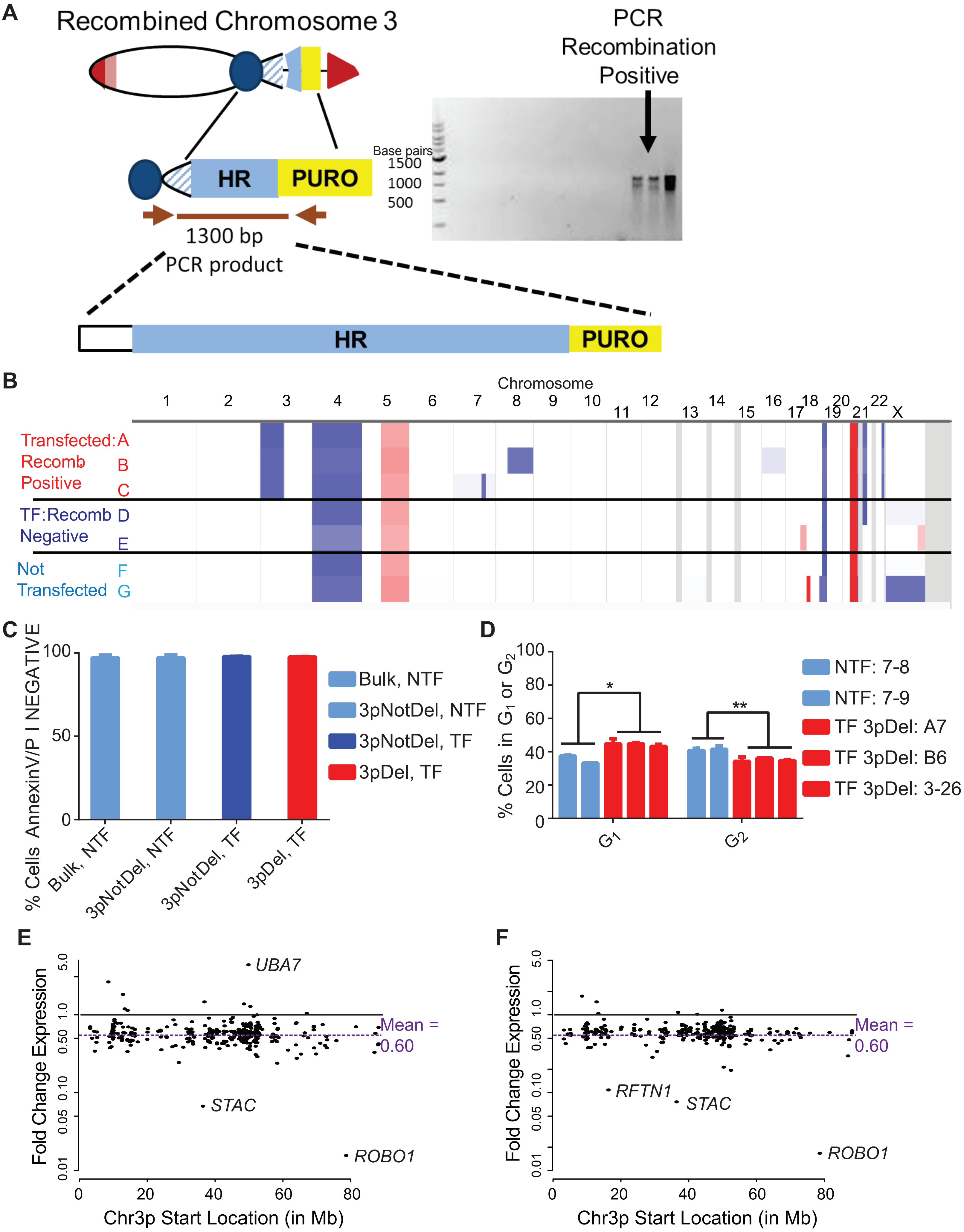
(A) Matrix of mean arm-level alteration within each tumor type and/or molecular subtype. Hierarchical clustering of tumor types by Pearson's method. Additional abbreviations: codel = co-deletion of chromosome 1p and chromosome 19q, ESCC = esophageal squamous cell carcinoma, Lum = Luminal, CN = copy number, CIN = chromosome instability, HM = hypermutated, SNV = single nucleotide variant, GS = genomically stable, DDLPS = dedifferentiated liposarcoma, MFS = myxofibrosarcoma, UPS = undifferentiated pleomorphic sarcoma, and LMS = leiomyosarcoma.

(B) Fold change for chromosome 3p genes comparing TCGA cancers with and without 3p deletion. Mean fold change = 0.77 (in purple).

(C) Mean arm alteration vs. average normalized enrichment score for proliferation and cell cycle pathways (includes E2F Transition, G2M Checkpoints, and Mitotic Spindle). Only statistically significant pathways are included (FWER p value < 0.01), so arms without these signatures are not included.

(D) Mean arm alteration vs. average normalized enrichment score for immune pathways (includes Interferon Response pathways and TNF Signaling). Only statistically significant pathways are included (FWER p value < 0.01), so arms without these signatures are not included.

(E) Gene set enrichment analysis was performed for genes with significant chromosome 3p coefficients in the pan-cancer dataset. Plots are for top significant negatively enriched hallmark datasets.


A
Recombined Chromosome 3
HR
PURO
1300 bp
PCR product
HR
PURO
PCR
Recombination
Positive
Base pairs
1500
1000
500
B
Chromosome
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Transfected: A
Recomb B
Positive C
TF:Recomb D
Negative E
Not F
Transfected G
C
% Cells AnnexinV/P I NEGATIVE
100
50
0
Bulk, NTF
3pNotDel, NTF
3pNotDel, TF
3pDel, TF
Bulk, NTF
3pNotDel, NTF
3pNotDel, TF
3pDel, TF
D
% Cells in G1 or G2
100
80
60
40
20
0
*
**
G1
G2
NTF: 7-8
NTF: 7-9
TF 3pDel: A7
TF 3pDel: B6
TF 3pDel: 3-26
E
Fold Change Expression
5.0
1.0
0.50
0.10
0.05
0.01
UBA7
STAC
ROBO1
Mean = 0.60
0
20
40
60
80
Chr3p Start Location (in Mb)
F
Fold Change Expression
5.0
1.0
0.50
0.10
0.05
0.01
RFTN1
STAC
ROBO1
Mean = 0.60
0
20
40
60
80
Chr3p Start Location (in Mb)

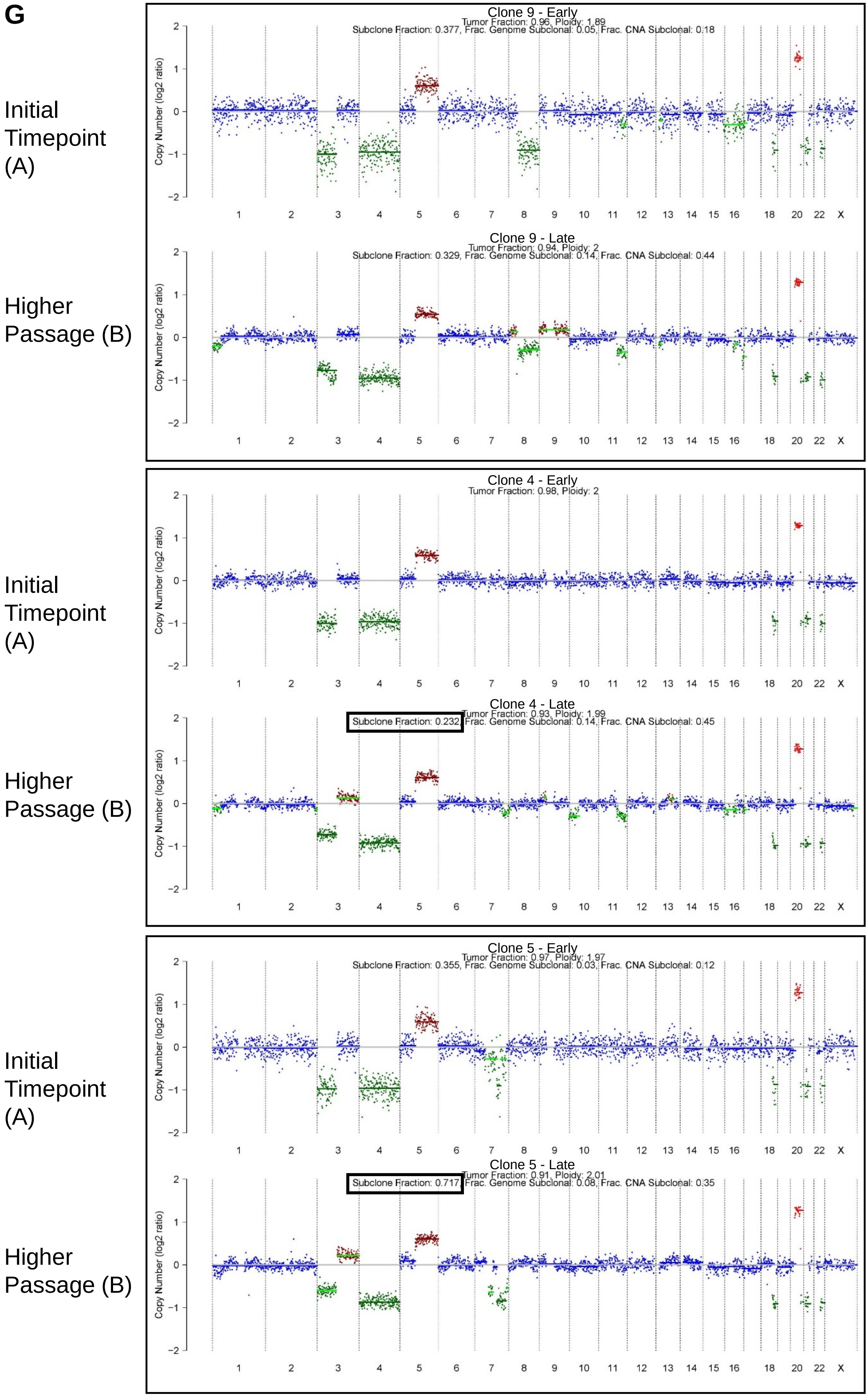

G
Initial Timepoint (A)
Higher Passage (B)
Initial Timepoint (A)
Higher Passage (B)
Initial Timepoint (A)
Higher Passage (B)
Clone 9 - Early
Tumor Fraction: 0.96, Ploidy: 1.89
Subclone Fraction: 0.377, Frac. Genome Subclonal: 0.05, Frac. CNA Subclonal: 0.18
Clone 9 - Late
Tumor Fraction: 0.94, Ploidy: 2
Subclone Fraction: 0.329, Frac. Genome Subclonal: 0.14, Frac. CNA Subclonal: 0.44
Clone 4 - Early
Tumor Fraction: 0.98, Ploidy: 2
Clone 4 - Late
Tumor Fraction: 0.93, Ploidy: 1.99
Subclone Fraction: 0.232, Frac. Genome Subclonal: 0.14, Frac. CNA Subclonal: 0.45
Clone 5 - Early
Tumor Fraction: 0.97, Ploidy: 1.97
Subclone Fraction: 0.355, Frac. Genome Subclonal: 0.03, Frac. CNA Subclonal: 0.12
Clone 5 - Late
Tumor Fraction: 0.91, Ploidy: 2.01
Subclone Fraction: 0.717, Frac. Genome Subclonal: 0.08, Frac. CNA Subclonal: 0.35
Copy Number (log2 ratio)
2
1
0
−1
−2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
18
20
22
X

Figure S4, related to Figures 5 and 6. Chromosome 3p deletion in immortalized lung epithelial cells.

(A) PCR-based approach to measure recombination in transfected cells. One primer binds in the endogenous centromeric DNA, and the other binds in the puromycin cassette from the recombination plasmid. Three single cell clones with 3p deletion are positive for the PCR product.

(B) Whole genome sequencing of transfected and non-transfected clones confirms chromosome 3p deletion for three clones. Single cell clones are rows A through G.

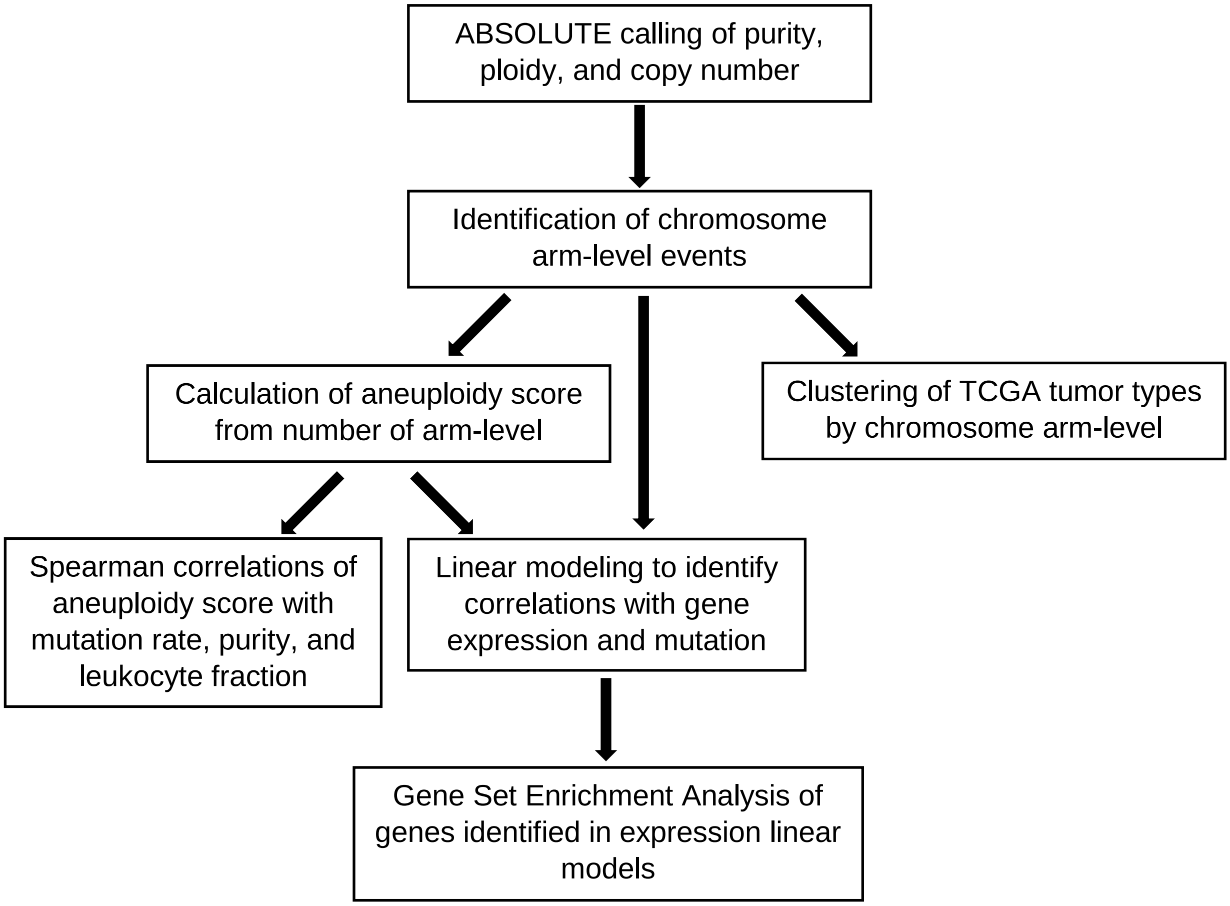
(C) Apoptosis profiling by measuring propidium iodide uptake in non-fixed cells. NTF = not transfected, TF = transfected. Bars are mean percentages of cells with error bars representing +/- standard deviation. Light blue are non-transfected clones, dark blue are transfected clones without 3p deletion, and red are clones with chromosome 3p deletion.

(D) Cell cycle profiling by measuring percentage of cells in each stage by propidium iodide level. Bars are mean percentages of cells with error bars representing +/- standard deviation. * = p value < 0.05, ** = p value < 0.01. NTF = not transfected, TF = transfected. Light blue are non-transfected clones and red are clones with chromosome 3p deletion.

(E) Fold change of chromosome 3p genes comparing non-deleted clones and chromosome 3p deleted clones (Timepoint A). Mean fold change = 0.60 (in purple).

(F) Fold change of chromosome 3p genes comparing non-deleted clones and chromosome 3p deleted clones (higher passage cells). Mean fold change = 0.60 (in purple).

(G) Three sets of sequences from deleted clones at the time of initial characterization (A) to the time of higher passage (B). Top set does not have a subclone with chromosome 3 gain, whereas the bottom two sets both have evidence of subclones with chromosome 3 gain.



ABSOLUTE calling of purity, ploidy, and copy number segments

The ABSOLUTE algorithm (Carter et al., 2012) was applied to 10,522 TCGA cancer samples, with an output of an absolute .seg file and a table calling ploidy, whole genome doubling, and purity per sample. All ABSOLUTE data will be available at the NCI Genome Data Commons.

Example .seg file:

Sample	Chromosome	Start	End	Num_Probes	Segment_Mean
SampleID1	1	3218923	17937997	4306	2
SampleID1	1	17954486	1.5E+08	31617	1

Example table file:

sample	call status	purity	ploidy	Genome doublings	Coverage for 80% power	Cancer DNA fraction	Subclonal genome fraction	solution
SampleID1	called	0.9	2	0	9	0.9	0.02	new
SampleID2	called	0.89	1.3	0	6	0.84	0.16	new

Identification of chromosome arm-level events

Analysis for identification of chromosome arm-level events was performed in Python.

```
##Python 2.7##
```

```
import numpy as np
import scipy.stats as stats
Import pandas as pd
from sklearn import mixture #scikit-learn v0.16.1#

##for each combination of arm/chromosome, alteration (amp/del), and tumor type##
X = data.as_matrix() #matrix containing [fraction altered, start coordinate, end
        coordinate] for each tumor; X.shape = (n,3)#
n = min(range(1,min(X.shape[0]+1,9)),key=lambda x:mixture.GMM(n_components=
        x,covariance_type=covtype).fit(X).bic(X)) #cluster solution with the most
        optimal BIC#
g=mixture.GMM(n_components=n,covariance_type=covtype)
g.fit(X)

cluster_assignment = {} #assign (cluster, probability) to each tumor based on max
                        likelihood#
for i in zip(data.index, g.predict_proba(X)):
        cluster_assignment[i[0]]=(np.argmax(i[1]),max(i[1]))

means = {} #mean fraction, start, and end per cluster#
for k in range(0,g.means_.shape[0]):
        means['k'+str(k)] = (g.means_[k,0],g.means_[k,1],g.means_[k,2])

data['GMM_cluster']=data.index.map(lambda x:'k'+str(cluster_assignment[x][0]))
data['clusterprob']=data.index.map(lambda x: cluster_assignment[x][1])
data['fractionmean']=data['GMM_cluster'].map(lambda x:means[x][0])
data['startmean']=data['GMM_cluster'].map(lambda x:means[x][1])
data['endmean']=data['GMM_cluster'].map(lambda x:means[x][2])
data['aneu']=data.apply(lambda x: 1 if x.loc['fractionmean'] >= 0.8 else 0 if
        x.loc['fraction'] <= 0.2 else np.nan, axis=1) #aneuploid if cluster fraction mean
        >=0.8, non-aneuploid if individual fraction <=0.2, otherwise NA#
```

Calculation of aneuploidy score from number of arm-level events

Calculation was performed in Python or Excel by summing the total number of altered arms - long and short arms for each non-acrocentric chromosome, and long arms for chromosomes 13, 14, 15, 21, and 22.

Spearman correlations of aneuploidy score with mutation rate, purity, and leukocyte fraction

Spearman correlation coefficients were calculated in Python or R.

```
##Python 3.6##
From scipy.stats important spearmanr
spearmanr(<list of x>, <list of y>)
Output: rho and p

##R version 3.2.3##
Cor.test(<list of x>, <list of y>, method = “spearman”)
```

Output: rho and p

Inputs:

Mutation rate: Non-silent mutation rate per megabase, as well as microsatellite instability level and *POLE* status, were provided by the DNA Damage Response TCGA PanCanAtlas paper (submitted) for Spearman correlations.

Stromal and immune infiltrates: Purity/impurity was measured using ABSOLUTE. Leukocyte fraction was estimated based on methylation data by the TCGA Immune subgroup (paper submitted). Non-leukocyte stroma was calculated by subtracting leukocyte fraction from impurity fraction. Expression based lymphoid and myeloid scores were from derived from xCell (Aran et al., 2017), a sum of all individual lymphoid and myeloid scores from the xCell analysis (performed in Excel). To calculate immune gene set expression scores, normalized RSEM expression values for genes from 4 hallmark gene sets (Interferon Alpha Response, Interferon Gamma Response, Allograft Rejection, and Immune Response) were summed and divided by 1000 (performed in Excel). All correlations between aneuploidy score and purity/immune measures were Spearman rank correlations.

Linear modeling to identify correlations with gene expression and mutation

All linear modeling was performed in R version 3.2.3, using lm.

```
##R version 3.2.3##
i = GeneX
form <- as.formula(paste0(i, " ~ ", "Type + AS"))
lmFit <- lm(form, data = armsandexpscores)
output <- c(i, lmFit$coefficients, summary(lmFit)$coefficients[,4])    #summary
                                                                         outputs p values
```

Equations used:

$\mathit{Expression}_{(gene\ X)} \sim \beta_1 * \mathit{predictor\ variable} + \beta_2 * \mathit{aneuploidy\ score}$

Predictor variables: tumor type, purity, and/or leukocyte fraction

Expression input: RSEM from Illumina mRNA sequencing data

Output: β_2 coefficients for aneuploidy score for each gene, p values for every coefficient

$\mathit{Mutation}_{(gene\ X)} \sim \beta_1 * \mathit{predictor\ variable} + \beta_2 * \mathit{aneuploidy\ score}$

Predictor variables: tumor type, mutation rate

Mutation input: Mutation calls from whole-exome sequencing data

Output: β_2 coefficients for aneuploidy score for each gene, p values for every coefficient

$\mathit{Expression}_{(gene\ X)} \sim \beta_1 * \mathit{chromosome\ arm} + \beta_2 * \mathit{aneuploidy\ score}$

Expression input: RSEM from Illumina mRNA sequencing data

Output: β_1 coefficients for each chromosome arm for each gene, p values for every coefficient

Gene Set Enrichment Analysis of genes identified in expression linear models

Gene set enrichment analysis (GSEA) was performed using Broad GSEA v.3.0 (Subramanian et al., 2005).

1) Gene lists were filtered for genes with coefficients having a Bonferroni p value of less than 0.01.

2) Coefficients for significant genes were entered as a ranked list for “GSEA-preranked”. Enrichment was assessed using Hallmark and Positional Gene Sets.
3) Enriched gene sets were considered significant if all presented gene sets have a family-wise error rate less than 0.01.

Clustering of TCGA tumor types by chromosome arm-level events

Within each tumor subtype, mean of arm-level event was calculated for each arm in Excel. Patterns were clustered using Morpheus (https://software.broadinstitute.org/morpheus/index.html), by 1-Pearson’s correlations.

Method S1, related to Figures 1, 2, 3, and 4. Pan-cancer computational analyses.

Additional information for computational analyses, including flowchart and code.