Supplemental Methods

GISTIC Copy-number Thresholds

To perform GISTIC analysis of the SCNA profiles, only those detected SCNAs whose intensity exceeds a given threshold are included in the analysis. All copy-numbers were normalized to a genome-wide average of two copies, assuming diploid cells. On this scale, histogram analyses suggested that copy-number changes of less than 0.2 often reflected artifact. Therefore, we only included amplifications and deletions resulting in inferred copy-numbers above 2.2 and below 1.8, respectively. SCNAs are also divided into those that are chromosome arm-level (defined as exceeding half the length of a chromosome arm) and focal (shorter than this). We considered as significant all events with False Discovery Rate q-values <0.01. We considered all genes within the 95% confidence interval for each peak region as candidate targets.

SCNA Quantification

To quantify the numbers of SCNAs in different sample sets, we relied upon the automated deconstruction of copy-number profiles into arm-level and focal SCNAs in GISTIC. We selected subsets of these SCNAs in two ways. In the first, we included all samples, and all events in these samples exceeding the copy-number thresholds for GISTIC (2.2 and 1.8 for amplifications, and deletions, respectively; see above). In the second, we included only those samples in which we could clearly detect arm-level SCNAs exceeding these samples, so as to exclude samples without extensive aneuploidy

(e.g. MSI colorectal cancers) and samples for which SCNAs could not be detected for technical reasons (e.g. samples with extensive admixture of contaminating normal DNA).

We also evaluated the number of more intense SCNAs (multi-copy focal amplifications and deletions) across different sample sets. We performed this analysis in two ways. In the first, we used fixed arbitrary event thresholds of >4 inferred copies for amplifications and <1.3 copies for deletions. In the second (reflected in Fig. S3), we selected SCNAs which surpassed a sample-specific threshold reflecting the most pronounced arm-level SCNA in that sample. This analysis also only included samples with detected arm-level SCNAs surpassing the GISTIC thresholds (2.2 and 1.8 for amplifications and deletions, respectively; see above). The intent of this analysis was again to correct for potential biases relating to samples without extensive aneuploidy or for which SCNAs could not be detected for technical reasons.

The significance of differences in event rates between cancer types was determined by permuting class labels.

Copy-number Quantitative Real-time PCR

Amplification of *FGFR2* was determined in triplicate by real-time quantitative PCR using validated *FGFR2* (Hs05182482_cn) and *LINE-1* (Hs01098704_cn) Taqman copy-number assays (Applied Biosystems, Carlsbad, CA) as described in previously (1), and using normal human genomic DNA (EMD Biosciences, Darmstadt, Germany) as a diploid control. Normalized copy-numbers of *FGFR2* were calculated using the comparative *C* method (2).

Microsatellite Instability Testing

MSI (microsatellite instability) analysis was performed using 10 microsatellite markers (D2S123, D5S346, D17S250, BAT25, BAT26, BAT40, D18S55, D18S56, D18S67 and D18S487). Cases were determined to harbor evidence of MSI if they showed MSI-high or MSI-low status (3), defined as the presence of instability in ≥ 30% and in 10% to 29% of the markers, respectively. MSS (microsatellite stability) was defined as no unstable marker.

Correlations Between Events

Correlations and exclusivity between these events was determined by comparing the observed rate of their co-occurrence to the distribution of rates of co-occurrence after permuting the sample labels associated with each event. Lineage was maintained in these permutations. P-values were determined from the results of 200,000 permutations and false discovery rate q-values were generated using the method of Benjamini and Hochberg (4) to account for multiple hypotheses.

Unsupervised hierarchical clustering was performed using a correlation metric and inferred copy-numbers at the GISTIC peak locations from the combined analysis across all three tumor types.

Cell Lines and Culture Conditions

The human esophageal adenocarcinoma cell line OE33 was purchased from the Sigma (St. Louis, MO) and maintained in RPMI 1640. A549, a non-small cell lung cancer cell line, was purchased from American Type Culture Collection (ATCC,

Manassas, VA) and grown in RPMI 1640. All cultures were supplemented with 10% fetal bovine serum, 1 mM penicillin/streptomycin, and 2 mM L-glutamine. Cell lines have been authenticated by SNP arrays within the last year.

Lentiviral Infection

RUNX1 and *GFP* were cloned into the pLEX-puro lentiviral vector. Viral production and infections were performed as described previously (5). Protein expression was confirmed via immunoblotting with antibodies to RUNX1 (Cell Signaling; Danvers, MA) with Vinculin (Sigma) as a loading control.

Anchorage-Independent Growth Assays

OE33 and A549 cells stably expressing *RUNX1* or *GFP* were plated in a top layer of growth medium with 0.33% Noble agar at a density of 10,000 cells per well and plated onto a bottom layer of medium with 0.5% agar in six-well plates in triplicate. Colonies were counted at 2–4 weeks based on growth rates. Images were acquired at 6.3X magnification using Magnifire software by inverted microscopy (Olympus SZX9). ImageJ software (http://rsb.info.nih.gov/ij/) was used to quantify colony number.

Supplemental Methods References

1. Soh J, Okumura N, Lockwood WW, Yamamoto H, Shigematsu H, Zhang W, et al. Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. PLoS One. 2009;4:e7464.

2. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25:402-8.

3. Ogino S, Kawasaki T, Kirkner GJ, Yamaji T, Loda M, Fuchs CS. Loss of nuclear p27 (CDKN1B/KIP1) in colorectal cancer is correlated with microsatellite instability and CIMP. Mod Pathol. 2007;20:15-22.

4. Benjamini Y, Hochberg A. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B (Methodological). 1995;57:289-300.

5. The RNAi Consortium. 2011 [cited; Available from: http://www.broadinstitute.org/rnai/trc