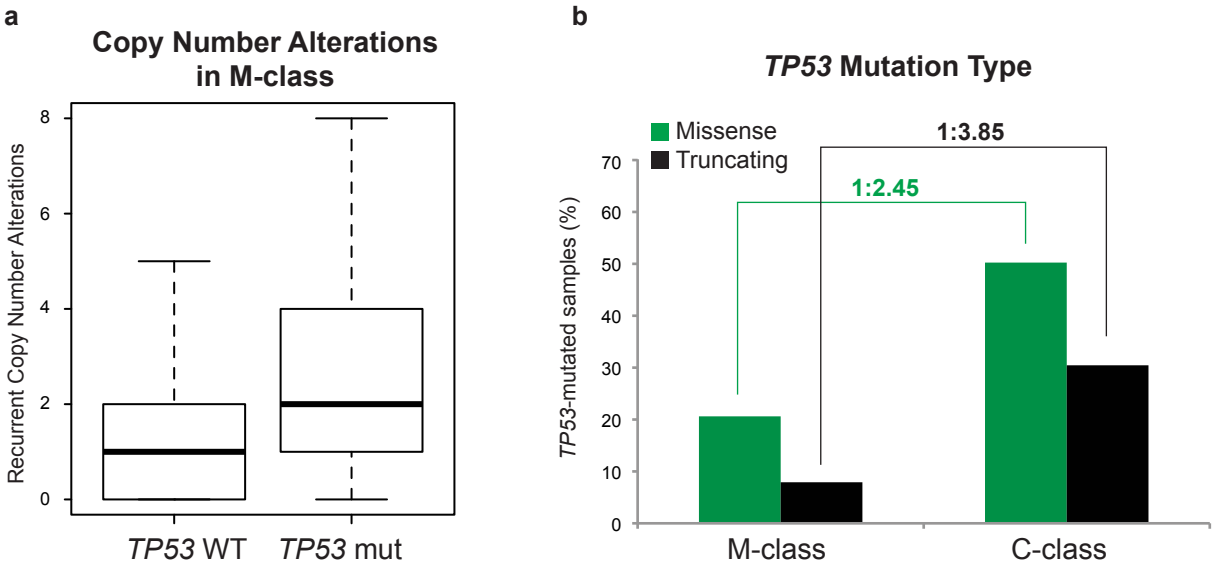
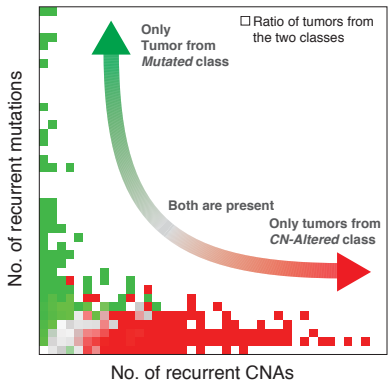
Supplementary Figure 1



Supplementary Figure 1: ***TP53*-mutated tumors a**, *TP53*-mutated tumors within the M-class have a higher number of recurrent copy number alterations (for all SFEs) than *TP53* wild type samples. **b**, While both missense (residue-changing) and truncating *TP53* mutations are more frequent in the C-class than in the M-class, this enrichment is significantly stronger for truncating mutations.

Nature Genetics: doi:10.1038/ng.2762

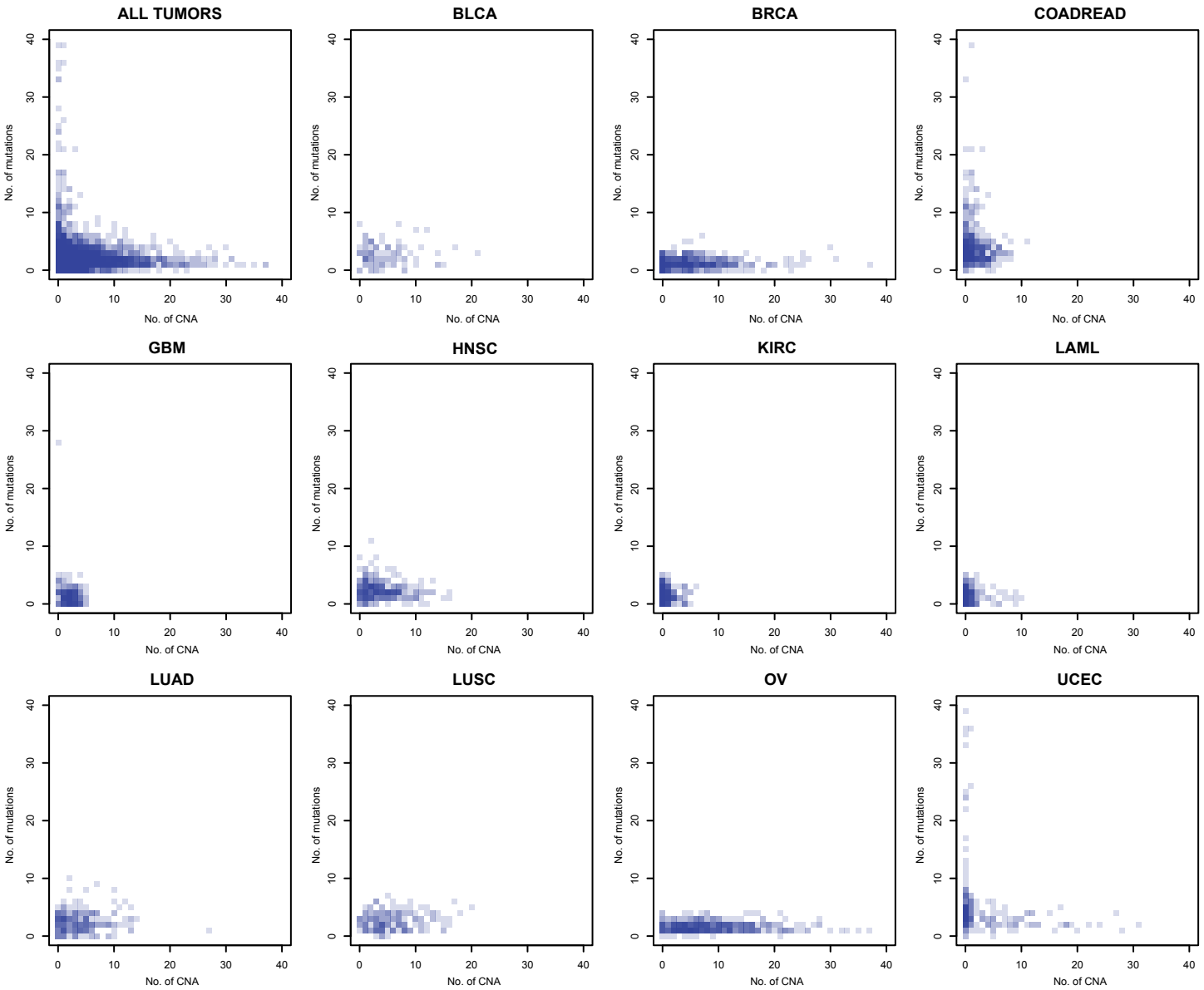
Supplementary Figure 2



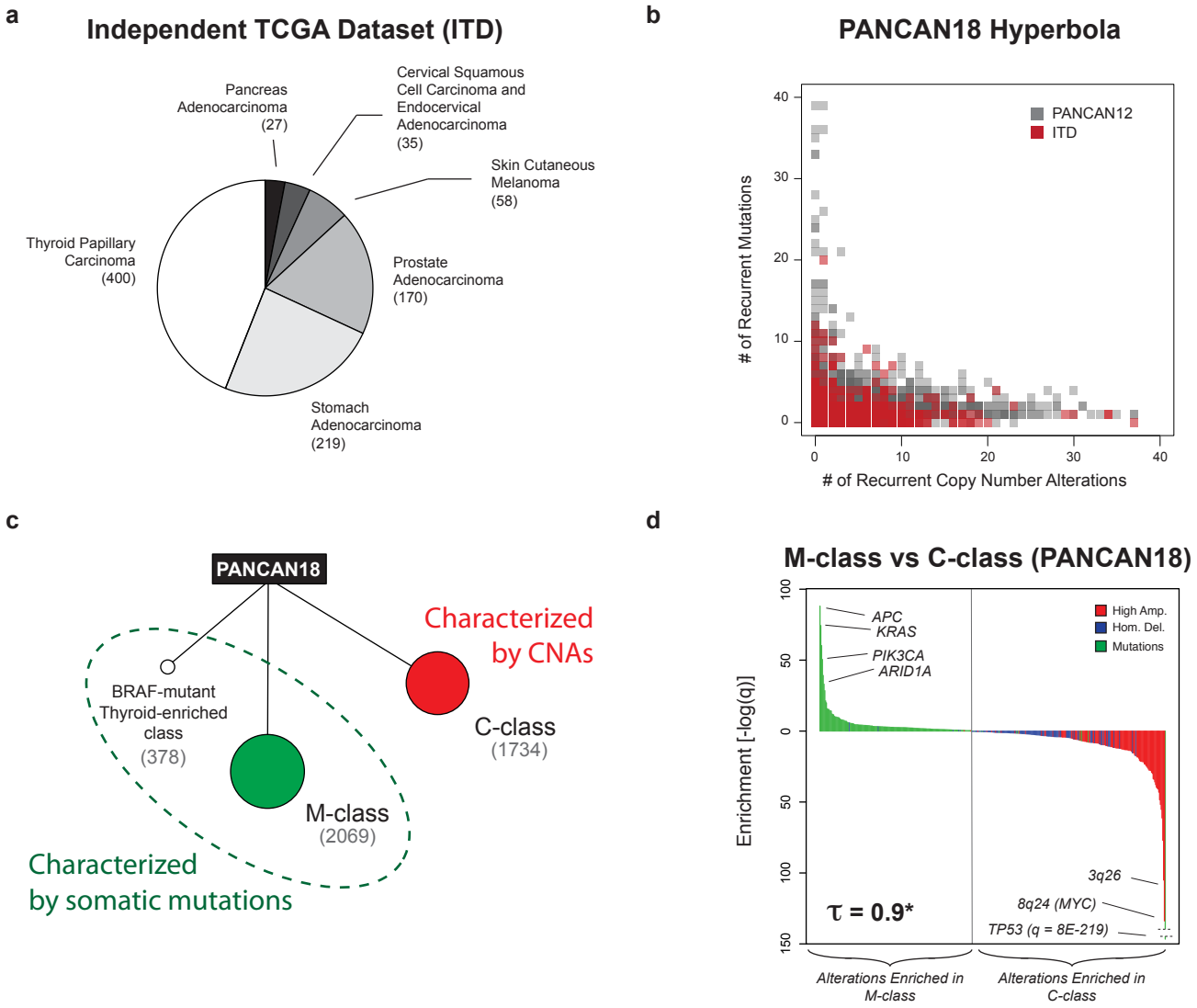
Supplementary Figure 2: Mutations versus copy number alterations. Tumors with more recurrent mutations than copy number alterations predominantly are in the M-class (increasing intensity of green), while tumors with more recurrent copy number alterations mostly are in the C-class (increasing intensity of red). Similar levels of both type of alterations are in tumors at the border between the two classes (grey).

Supplementary Figure 3: Distribution of mutations and copy number changes across different tumor types. Samples from different tumor types and subtypes vary in the number of recurrent copy number alterations (X-axis) and number of recurrent mutations (Y-axis).

Supplementary Figure 3



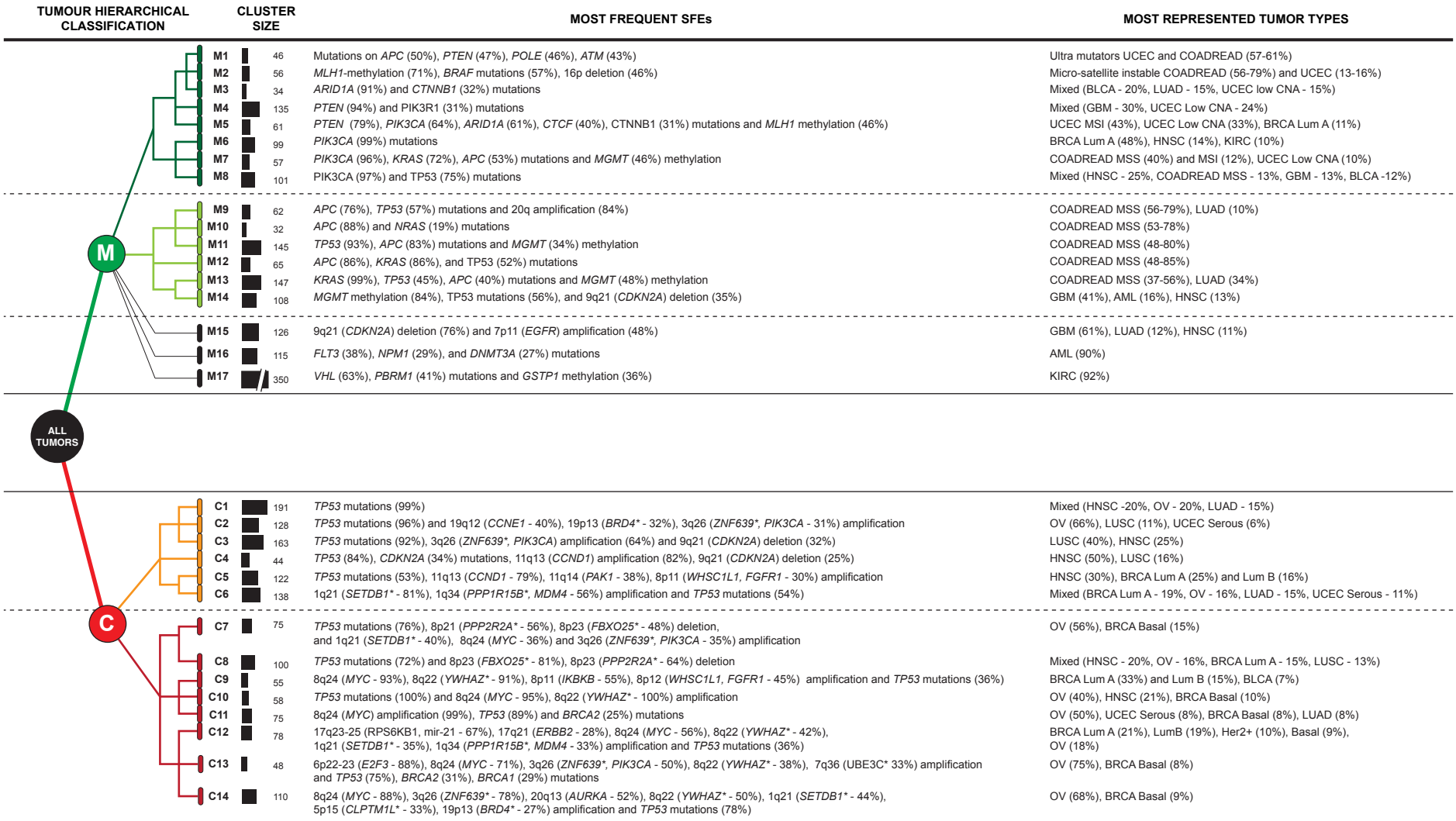
Supplementary Figure 4



Supplementary Figure 4: Robustness of classes and features - validation in enlarged dataset. a, An independent dataset consisting of 909 samples from 6 tumor types was integrated with our original dataset (PANCAN12). **b,** Samples from the resulting set of 18 tumor types (PANCAN18) confirmed the inverse relationships between copy number alterations and mutations, with the newly added samples following the same inverse trend (red squares) we originally observed (gray squares). **c,** A stratification of the PANCAN18 dataset identifies two main modules of similar size and a small subgroups enriched in BRAF-mutant thyroid samples. These thyroid tumors emerged separately as strongly characterized by BRAF V600E mutations and otherwise depleted of any other SFE. Nonetheless, these tumors do not violate the distinction between tumors primarily characterized by copy number alterations (C-class) and tumors characterized by mutations (M-class), of which thyroid samples are a subset. **d,** The ranked list of SFEs, with the first of the list being the most enriched in the M-class and the last the most enriched in C-class, is almost perfectly concordant with the same ranked list obtained in the PANCAN12 dataset (Kendall correlation coefficient tau = 0.9).

Nature Genetics: doi:10.1038/ng.2762

Supplementary Figure 5

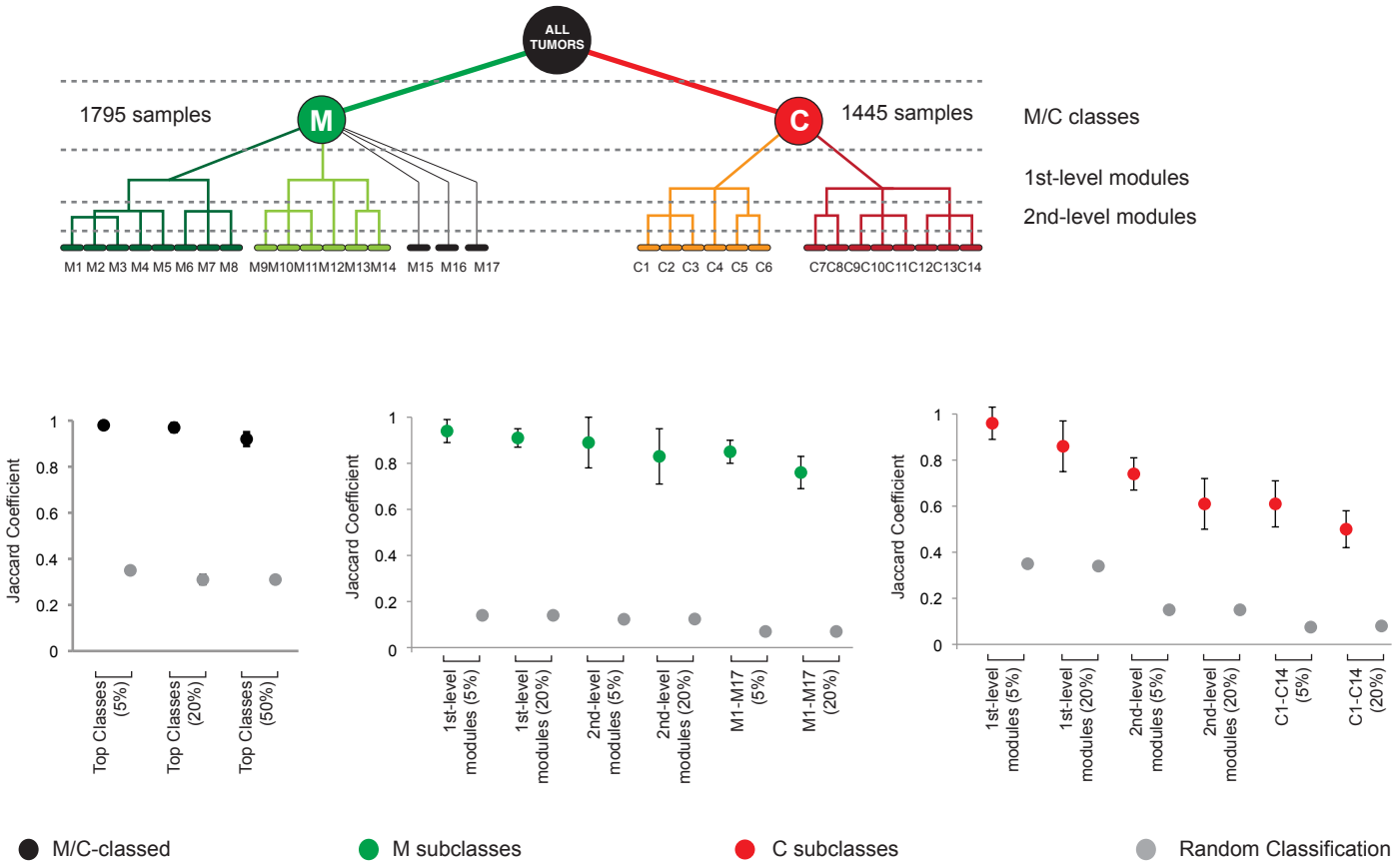


* Most differentially expressed gene in the region when compared to diploid samples

Supplementary Figure 5: Hierarchical stratification. The hierarchical stratification of tumors based on selected functional events identifies 31 oncogenic signature sub-classes (dendrogram on the left). For each sub-class, we report the most frequent SFEs and the most represented tumor types. For each recurrent copy number alteration, we report known oncogenes or tumor suppressors in the region (if present). For large regions spanning multiple genes, we also report the gene with the most significant differential mRNA expression between altered and diploid cases.

Nature Genetics: doi:10.1038/ng.2762

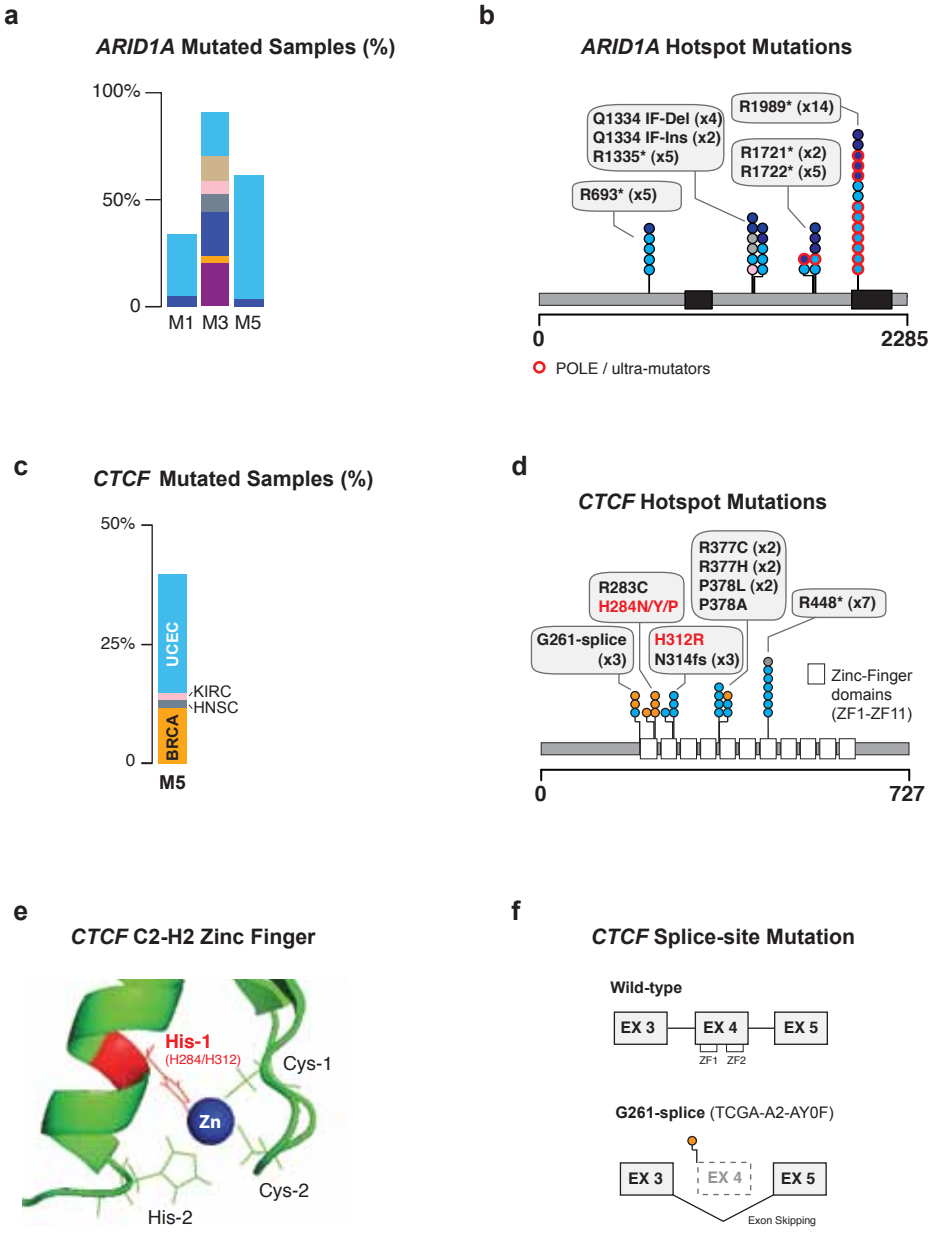
Supplementary Figure 6



Supplementary Figure 6: Robustness of cluster assignments. Subclasses at all levels of the hierarchical stratification were separately tested for robustness to random sample removal. The top classes M and C (black dots) show almost perfect reproducibility upon removal of 5%, 20%, and 50% of the samples. Similarly, the M subclasses (green dots) maintain a constantly high robustness descending the classification. C subclasses (red dots) at the bottom the classification are the most affected by sample removals, despite a robustness significantly higher than expected (gray dots). The non-focal nature of copy number alterations and the difficulty in identifying the corresponding functional targets are likely to impact the definition of these groups, weakening their robustness.

Nature Genetics: doi:10.1038/ng.2762

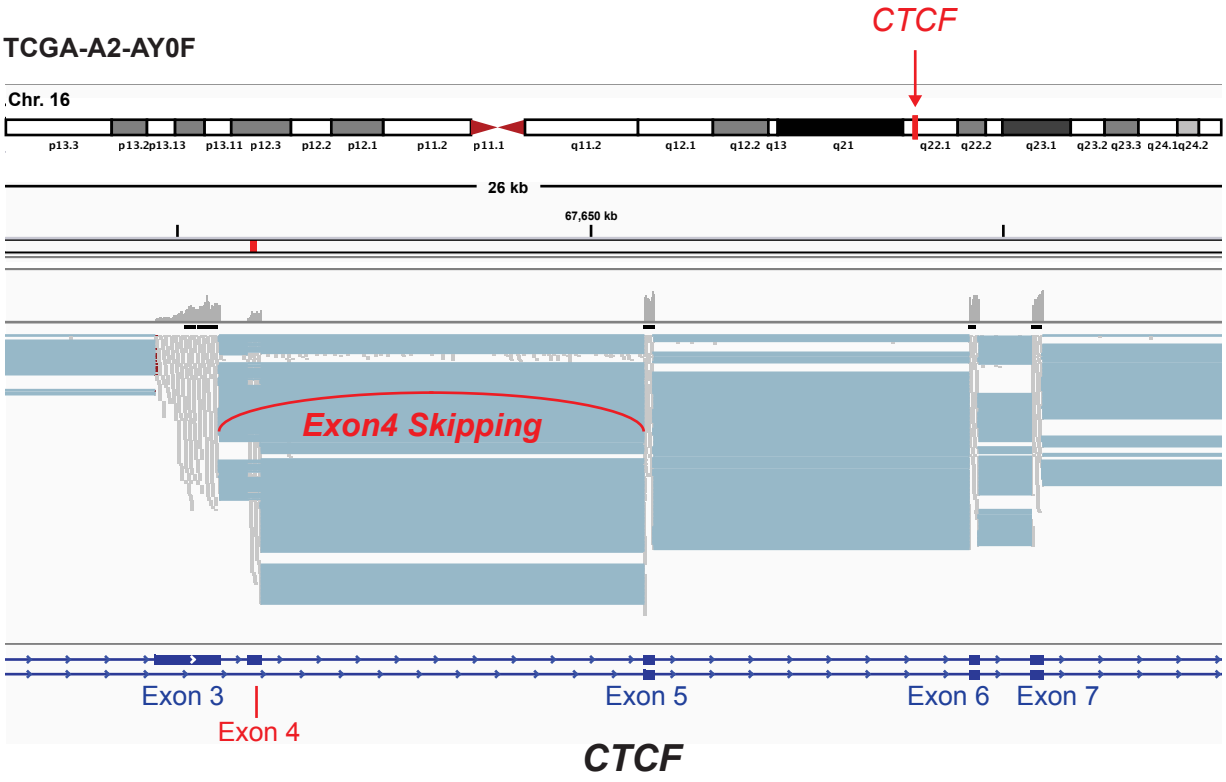
Supplementary Figure 7



Supplementary Figure 7: Recurrent mutations identified by cross-cancer analysis. a, *ARID1A* mutations are observed in sub-classes M1, M3, and M5, which include diverse tumor subtypes (colored stacked bars). **b,** The frequency of mutations of *ARID1A* across all tumors indicates hotspot regions (colored according to the tumor type of the mutated samples). **c,** *CTCF* mutations are characteristic of sub-class M5 and **d,** hotspot somatic mutations target different zinc-finger domains (white boxes). **e,** hotspot mutations frequently occur at the zinc-binding histidine. **f,** This splice-site mutation causes an in-frame exon-skipping event. The skipped exon codes for zinc-finger 1 and 2.

Nature Genetics: doi:10.1038/ng.2762

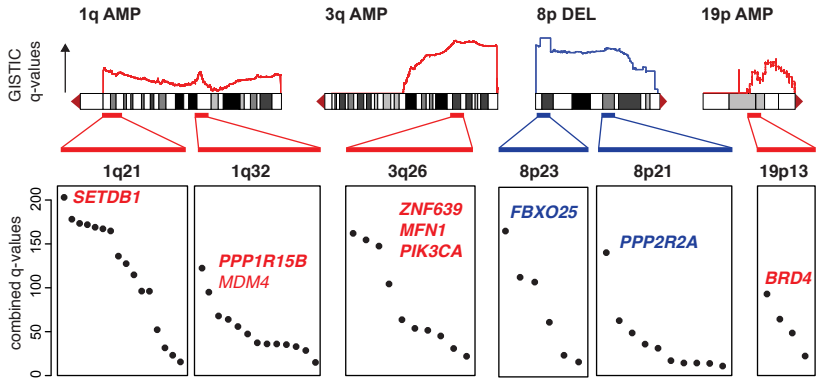
Supplementary Figure 8



Supplementary Figure 8: Splice-site mutation in *CTCF*. Splice-site mutation on G261 on *CTCF* (located on 16q22) caused the skipping of Exon 4 in the TCGA-A2-AY0F sample. Multiple RNA-seq reads (grey thick lines, one sample per row) span sequences starting from Exon 3 to Exon 5, therefore skipping Exon 4 (endpoints of reads spanning more than one exon are connected by blue lines).

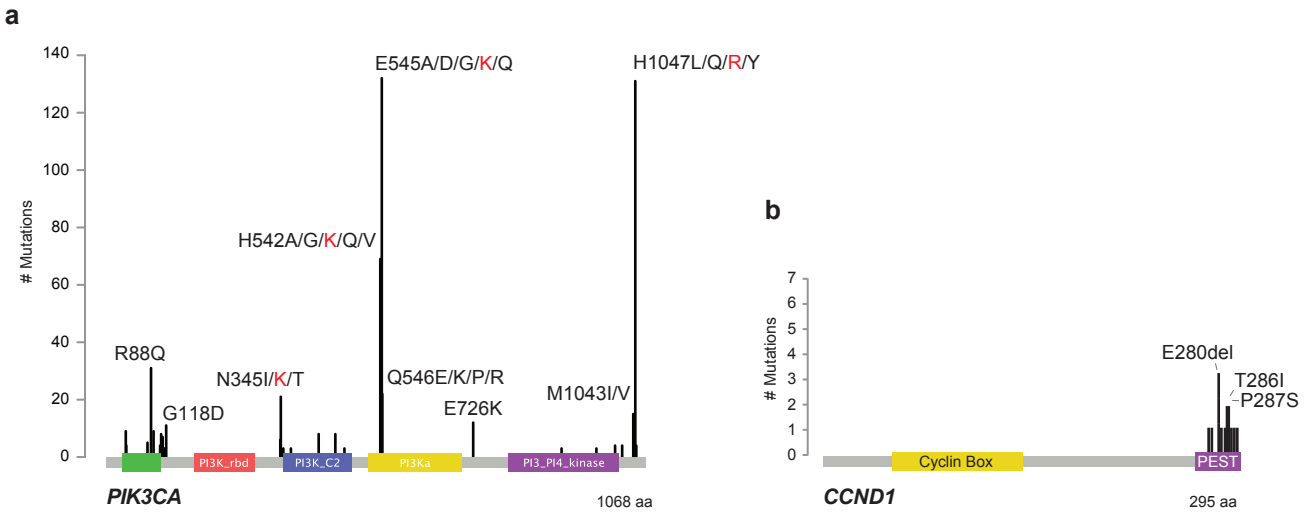
Nature Genetics: doi:10.1038/ng.2762

Supplementary Figure 9



Supplementary Figure 9: Significant loci within chromosomal events. Recurrent gains and losses on chromosomes 1q, 3q, 8p, and 19p span multiple genes. The Gistic algorithm identified the most significant loci in these regions (peaks of q-value histograms). We tested each gene in these peaks for differential mRNA expression concordant with the copy number change (over-expressed if amplified, down-regulated if deleted, relative to diploid cases). Combined q-values [-log(q)] serve to determine the most likely functional targets (labeled by corresponding gene symbol).

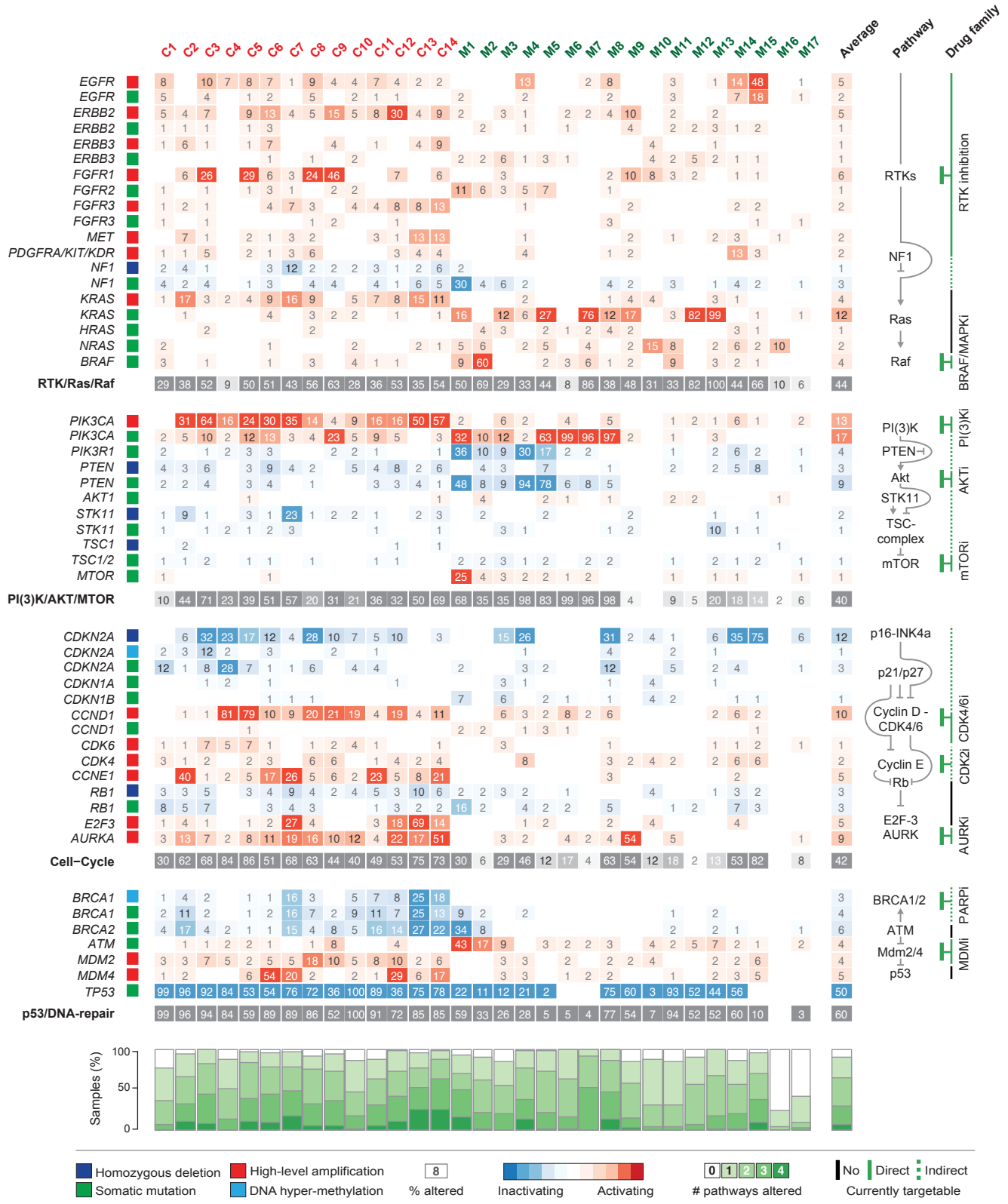
Supplementary Figure 10



Supplementary Figure 10: Hotspot mutations identified by cross-cancer analysis. a, PIK3CA hotspot mutations. **b,** Hotspot mutations in *CCND1* cluster in the PEST domain of the protein, preventing degradation and stabilizing the protein. Mutations are indicated by vertical bars (bar height proportional to the number of mutations in the specified residue, summed over all affected samples in all tumor types).

Nature Genetics: doi:10.1038/ng.2762

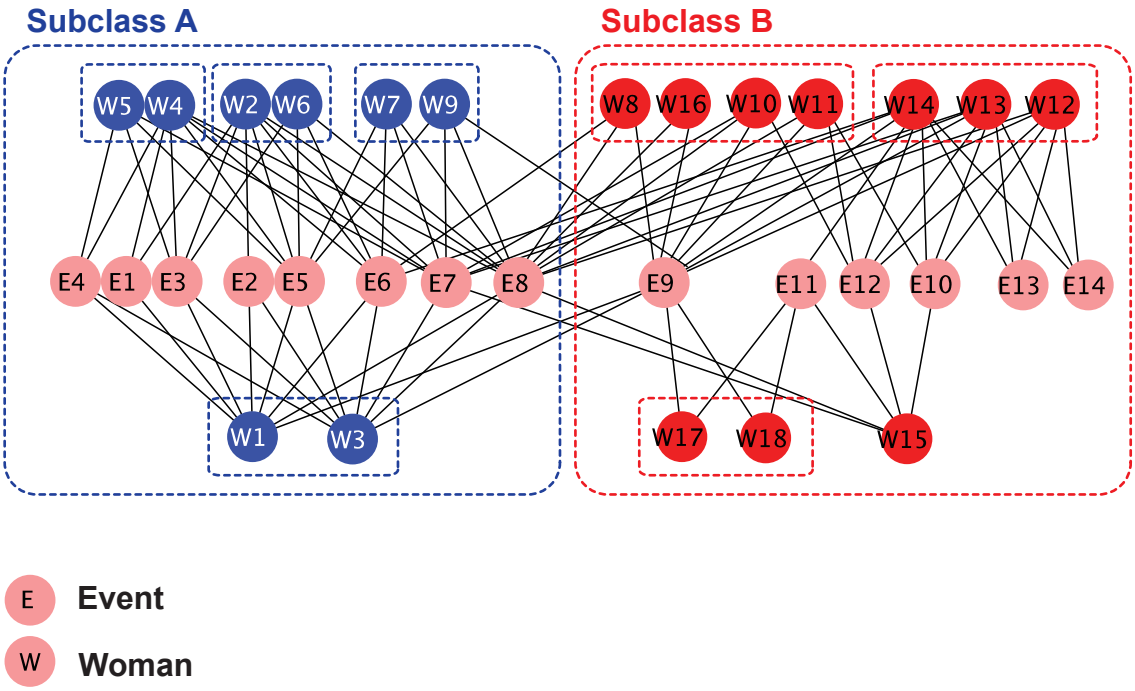
Supplementary Figure 11



Supplementary Figure 11: Map of functional and actionable alterations across 31 tumor subclasses. Four major oncogenic pathways (RTK/Ras/Raf, PI(3)K/Akt/mTOR, Rb, and DNA-repair/p53 signaling – simplified in the "Pathway" column) are altered by multiple different mechanisms (arranged vertically) across tumor types and subtypes (from left to right). Alterations to at least one of these pathways are observed in almost all samples of almost all tumor types (stacked barplots at the bottom). Several of these alterations are directly or indirectly therapeutically actionable ("Drug Class" column).

Nature Genetics: doi:10.1038/ng.2762

Supplementary Figure 12



Supplementary Figure 12: Exercise to test algorithm: network subclasses identified in the Southern Women Club Attendance network. We identified two main classes (blue and red) and within each main class we identified smaller groups of women, who attended the same set of events. Grouping by dotted lines. These results are consistent with the ones in the original publications (see Methods).