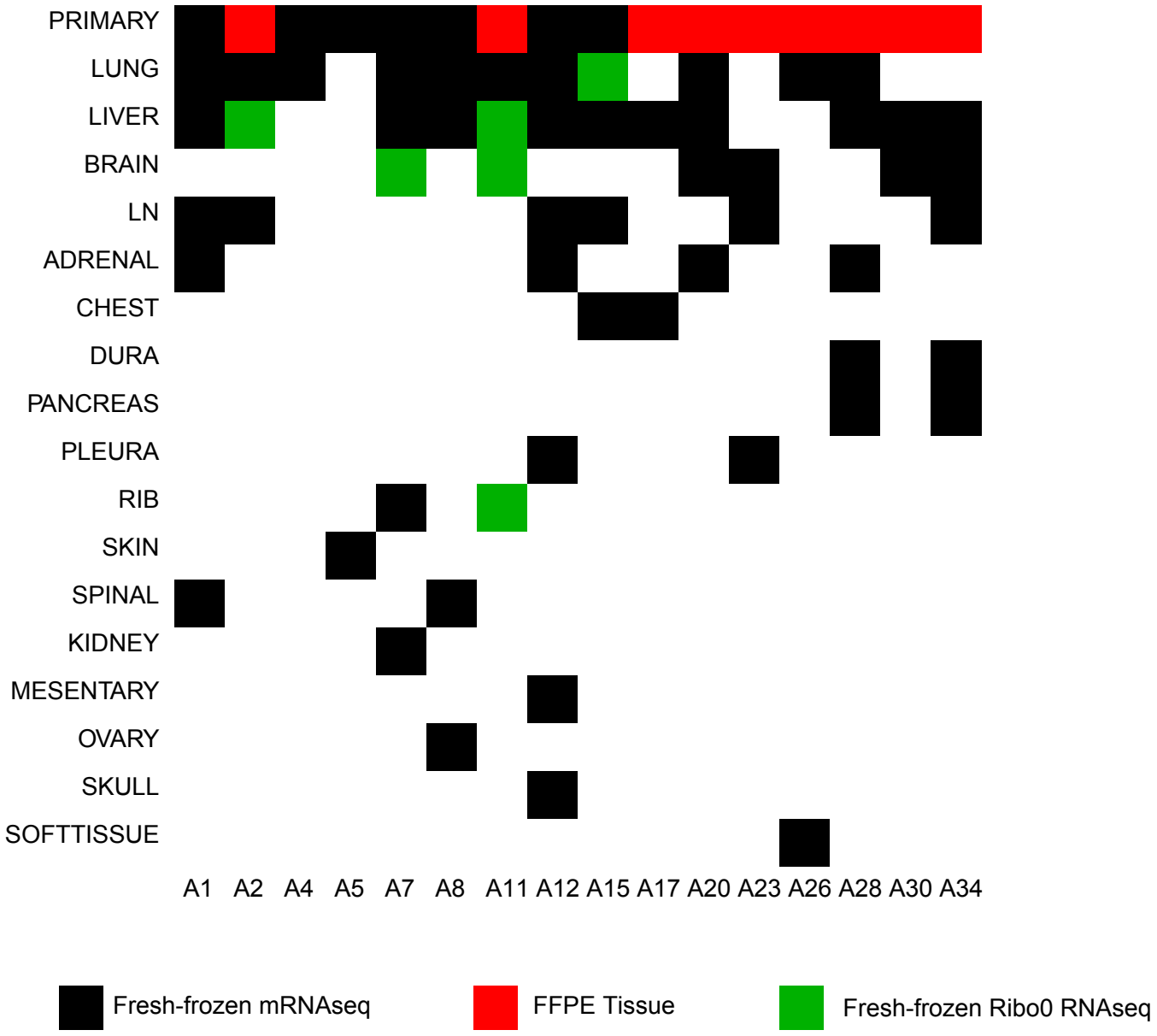
Supplemental Figure 1



Supplemental Figure 1. Distribution of tumor specimens for each patient. Diagrammatic view of the tumors from the 16 patients in the UNC Rapid Autopsy Program (RAP) that were sequenced with both RNA and DNA whole exome sequencing by site of disease (black = sequenced).

Supplemental Figure 2

Sample Collection

1. Tumor and normal (>4 cm from tumor) tissue collected within 6 hrs of death were snap-frozen in liquid nitrogen
2. Previous surgical tissue, if available, was acquired from Surgical Pathology archives (Table S2, Figure S1)
3. 1 FFPE H&E section from all tissues (fresh-frozen and FFPE) sequenced were reviewed by a pathologist to confirm presence/abscence of tumor

RNA Analyses

RNA Sequencing

1. RNA extracted with QIAGEN RNeasy
2. Libraries prepared Illumina TruSeq +/- Ribo0
3. Illumina HiSeq 2x50 bp paired-end reads

Bioinformatic Analyses

TCGA RNASeq Pipeline

1. Aligned reads with MapSplice
2. Quantified normalized reads with RSEM
3. Log2 of upper quantile normalized counts was used for all downstream analyses

Platform Normalization

1. Identical samples were sequenced on 3 platforms
   a) fresh-frozen polyA select
   b) fresh-frozen Total RNA + Ribo0
   c) FFPE Total RNA + Ribo0
2. Found the mean difference at each gene for
   a) fresh-frozen polyA vs. fresh-frozen Total
   b) fresh-frozen polyA vs. FFPE Total
3. Applied (a) to fresh-frozen Total RNA seq and (b) to FFPE Total RNAseq such that all RNAseq was scaled to fresh-frozen polyA

PAM50 Molecular Subtyping

1. Calculated centroid for fresh-frozen polyA using 50/50 ER+/- TCGA primaries

Signature Scores

1. Calculated mean gene expression of genes comprising ~470 publicly available gene signatures

DNA Analyses

DNA Sequencing

1. DNA extracted with QIAGEN DNeasy
2. Libraries prepared with Agilent SureSelect XT
3. Illumina HiSeq 2x100 bp paired-end reads

Bioinformatic Analyses

UNCSeq Pipeline (Zhao et al., 2015)

1. Aligned (BWA), sorted (biobambam), and realigned (ABRA)
2. Somatic Mutation Detection with Quality > 30 (STRELKA, Cadabra, UNCeqR)
3. Filter mutations if present at >$1x10^{-5}$ in ExAC and in less than 10 patients in COSMIC

Reinterrogation

1. All mutations from the metastases within each patient were recounted within both (a) all tumors from that patient and (b) all normals in the cohort
2. Mutations were kept as real if:
   a) at least 20 reads and 1 alt allele
   b) not in 2 Normals at 20% VAF

Somatic Copy Number Caller: SynthEx

1. For each tumor, SynthEx selects 3 normals with the most similar technical metrics
2. Log2 ratios of tumor vs the mean normal at 100 kB non-overlapping bins were calculated

Mutational Analyses

1. Phylogenetic relationship with phylip
2. Subclonal analyses with Gibbs-Dirchilet
3. Mutational signatures with *SomaticSignatures* R package

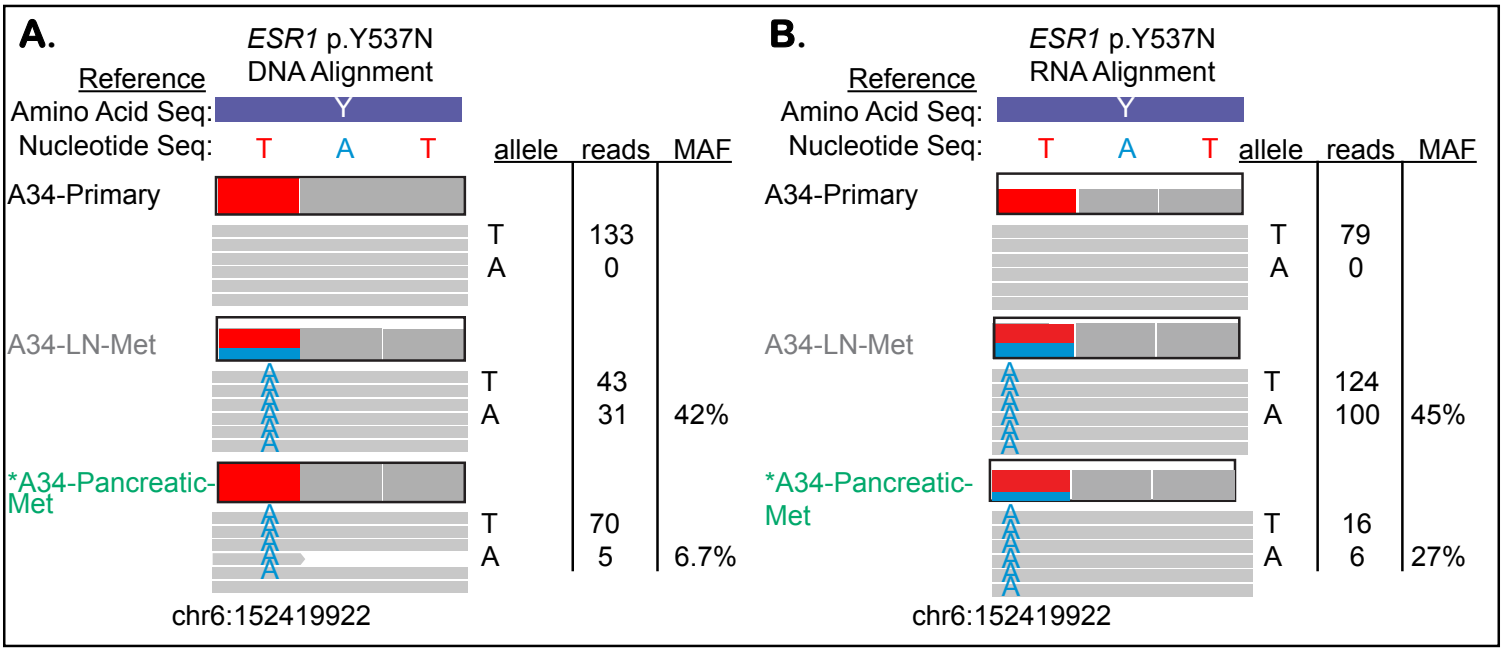
Integrated Analyses

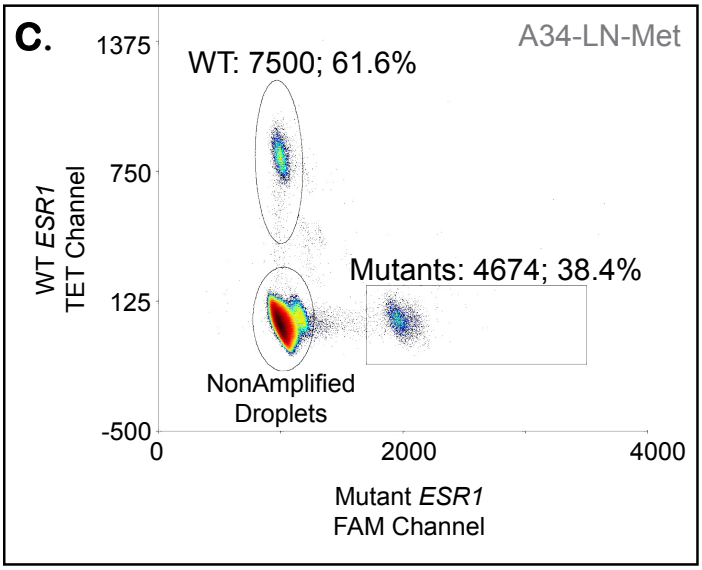
DawnRank - computational prediction of genetic drivers

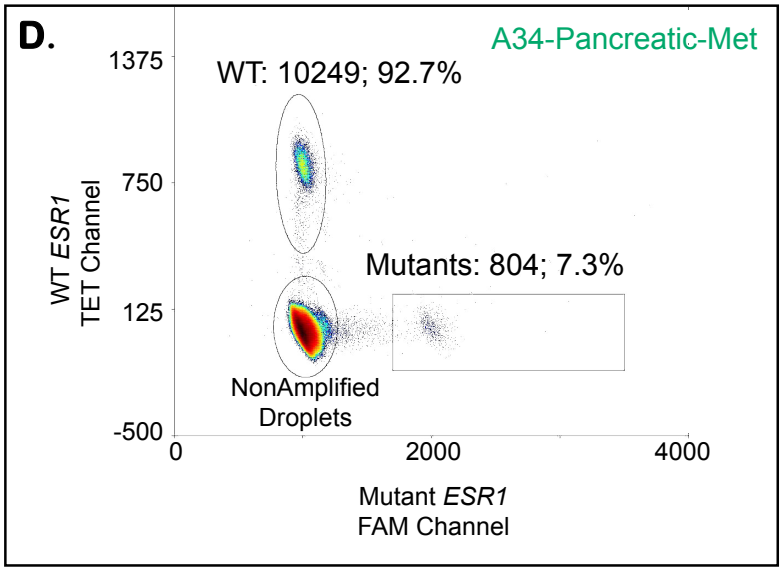
Expression of DNA Mutations in the RNA

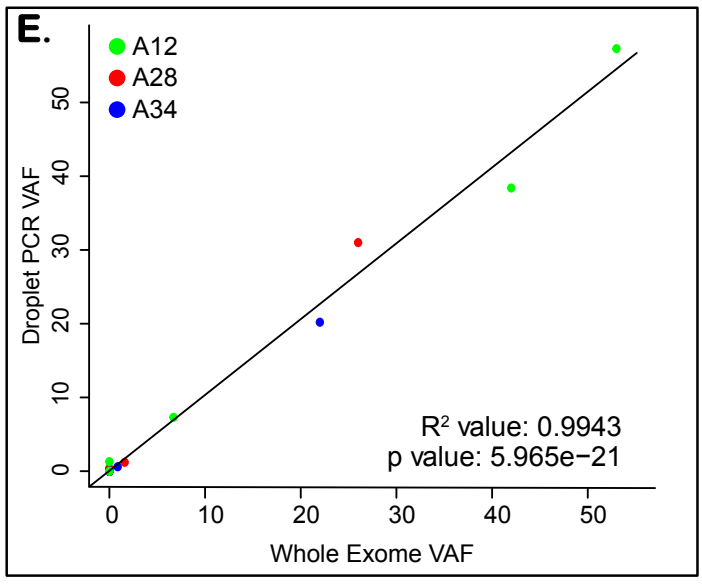
Supplemental Figure 2. Overview of methods for the UNC Rapid Autopsy Program.

Supplemental Figure 3





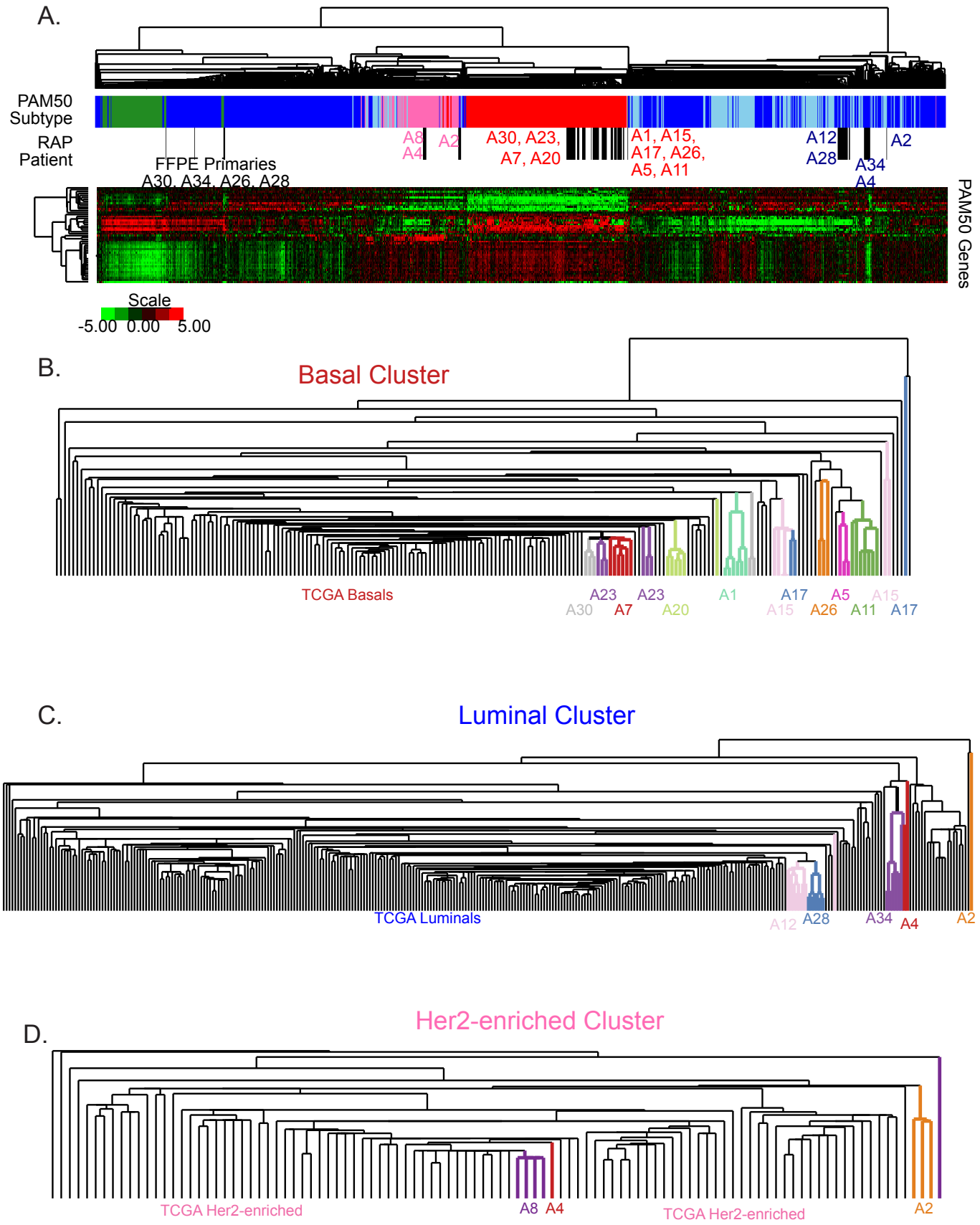




Supplemental Figure 3. Computational reinterrogation example and confirmation with Droplet PCR of mutant *ESR1*. The T to A mutation at chr6:152419922 was initially identified (gray text) in **A**. A34-LN-Met but not the primary. Upon re-interrogation (*green text), the mutation was also identified in A34-Pancreatic-Met and counted as mutant. **B**. *ESR1* mutation chr6-152419922-T-A was also observed in the RNA alignments from RNASeq for both the original call in A34-LN-Met and A34-Pancreatic-Met.

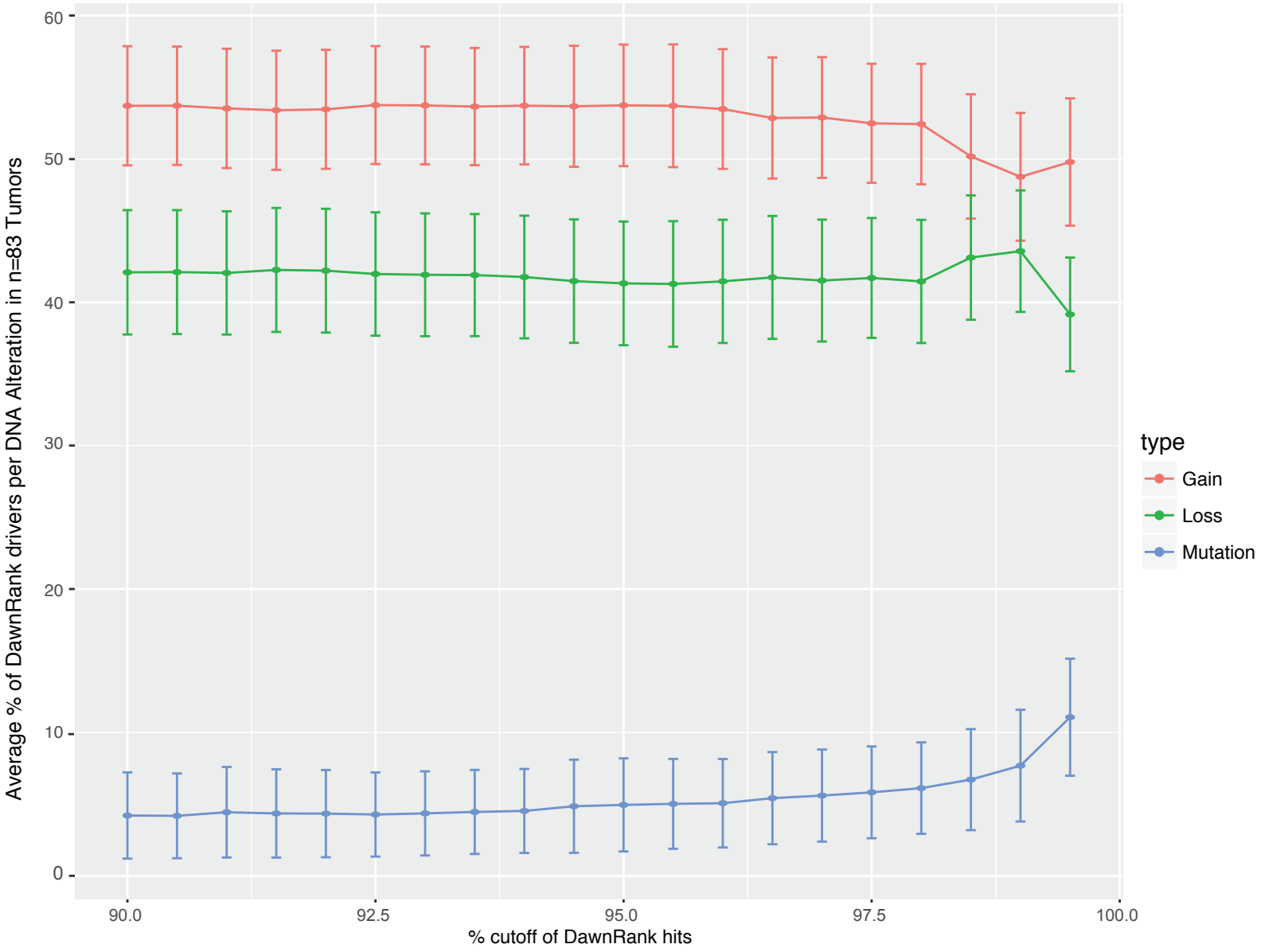
Droplet PCR of **C**. A34-LN-Met and **D**. A34-Pancreatic-Met confirms the presence of mutant *ESR1*. **E**. Droplet PCR in all 3 patients exhibiting *ESR1* mutation validates *ESR1* mutations down to 1/114 read accuracy.

Supplemental Figure 4

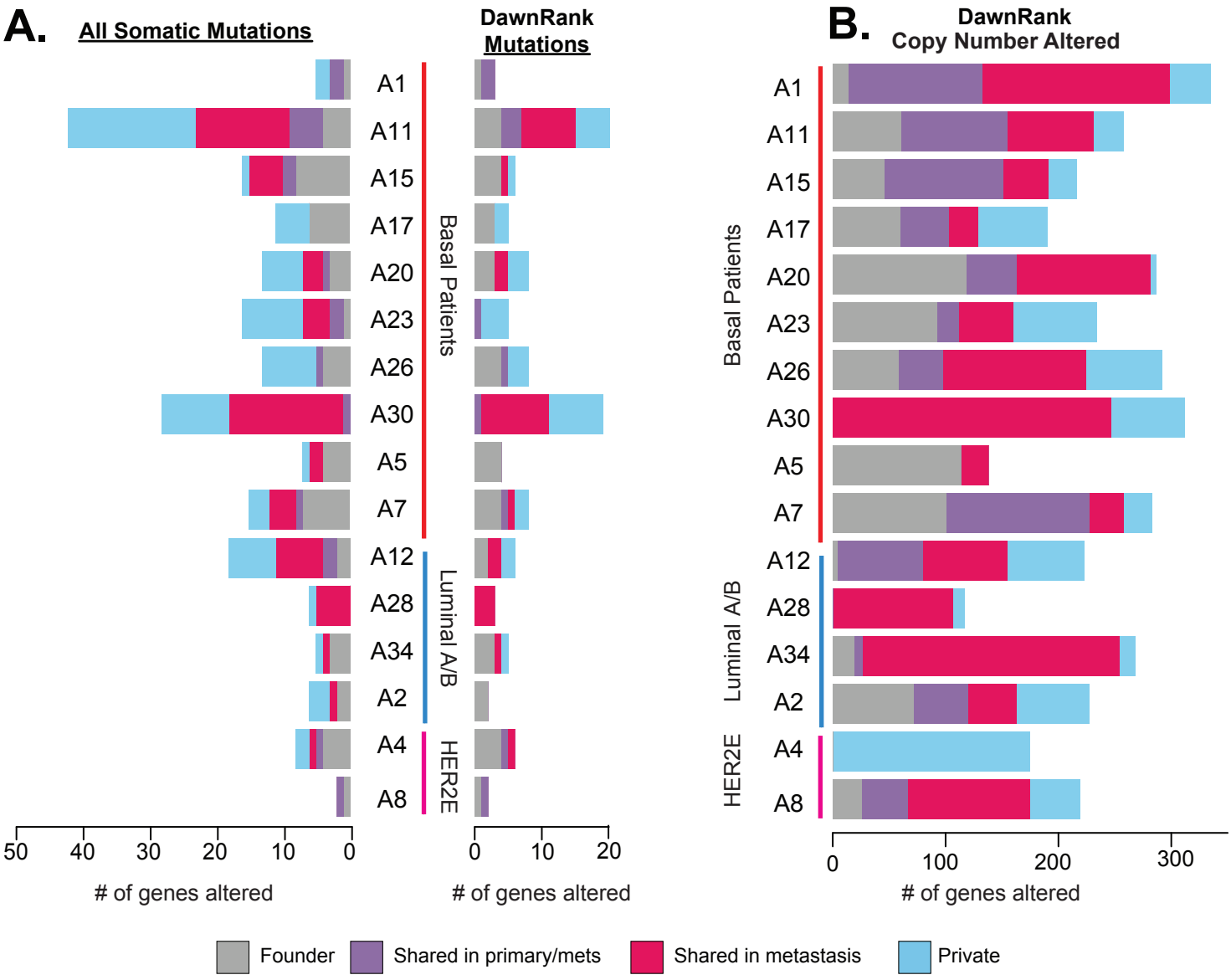


Supplemental Figure 4. Hierarchical clustering of TCGA breast cancer with the RAP dataset. A. Supervised hierarchical clustering using the PAM50 gene set with TCGA primary breast cancers and RAP primaries and metastses. PAM50 subtype and positioning of RAP tumor sshown in the second row of the color bar. Zoomed in view of the dendrogram of each subtype showing the location of tumors from each RAP patient for B. basal-like, C. luminal, and D. HER2-enriched sample associated clusters.

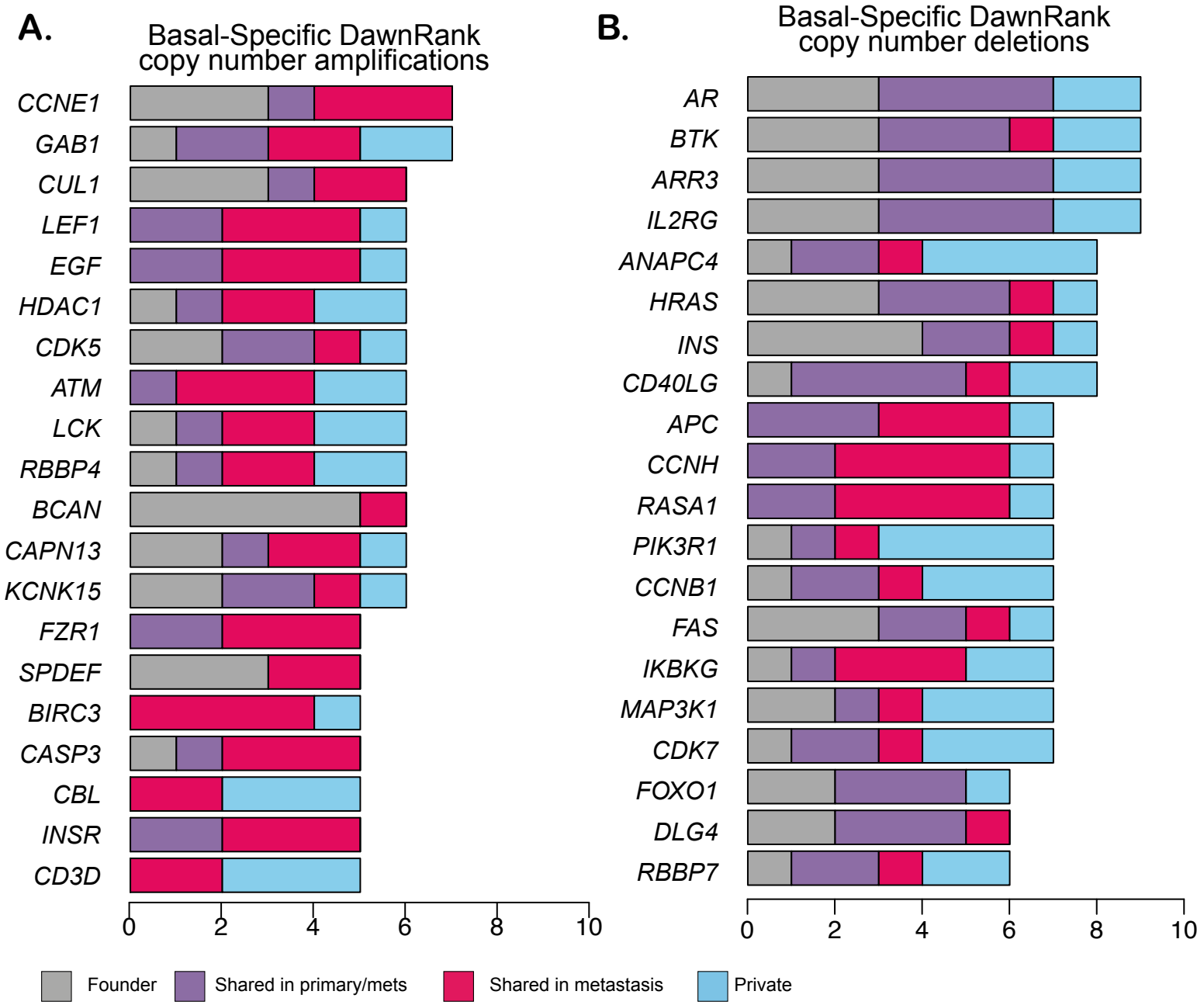
Supplemental Figure 5



Supplemental Figure 5. Varying the percent cutoff of DawnRank hits and its effect on type of DawnRank hit. Stepping from 90.0% cutoff (p = 0.10) to 99.5% cutoff (p = 0.005), the average percent of DawnRank hits with standard deviation from DNA copy number gains (red), DNA copy number losses (green), and DNA mutations (blue) from n=83 tumors.

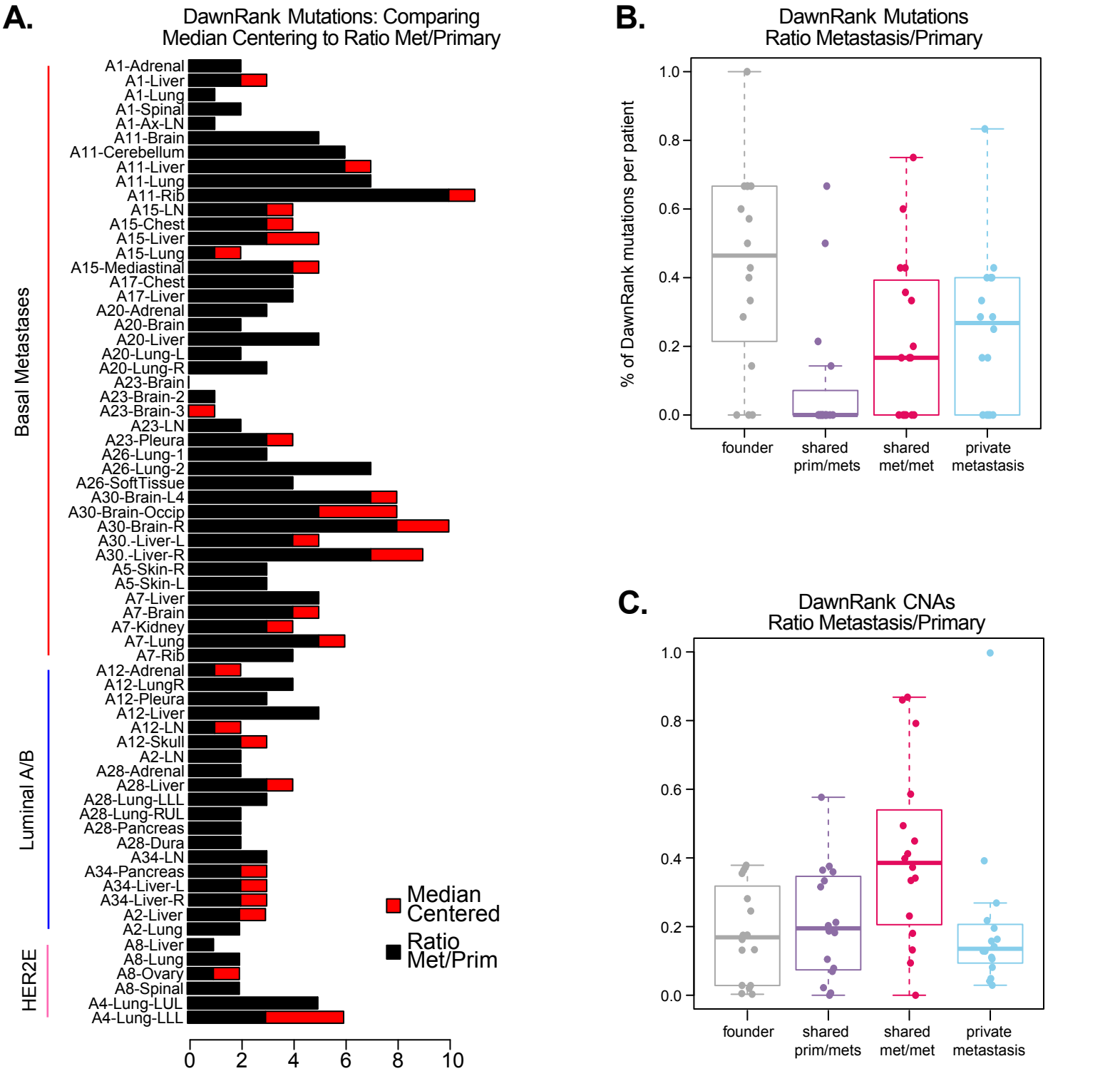


Supplemental Figure 6. Timing of somatic mutations, DawnRank mutations, and DawnRank copy number alterations. Representation of when during the development of metastasis each category of alteration could have occurred: founder alterations established in the original breast cancer and maintained in all metastases (gray); alterations observed in the primary and ≥1 metastasis (purple); metastasis-shared observed in ≥1 metastasis but not the primary (hot pink); or metastasis-private mutations, acquired at the final site of metastasis (blue). Proportion of **A.** all mutations and predicted DawnRank drivers from somatic mutation, and **B.** predicted DawnRank drivers from somatic copy number alterations.



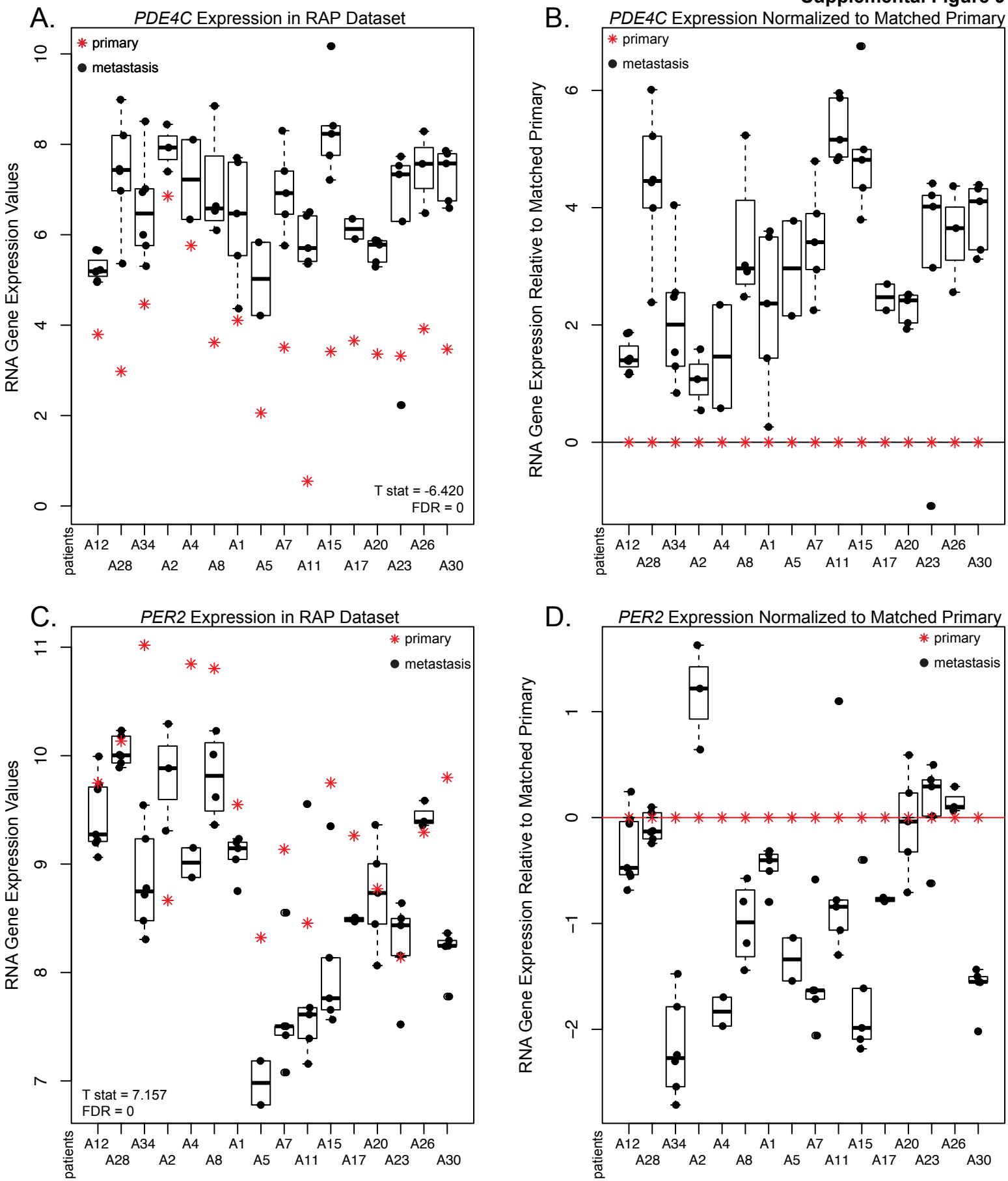
Supplemental Figure 7. Basal-specific DawnRank drivers. DawnRank drivers that were identified only in the 10 basal-like patients in the dataset plotted by frequency of driver identification adn colored by the timing with whihc the alteration was detected in the developemtn of the tumors.

Supplemental Figure 8



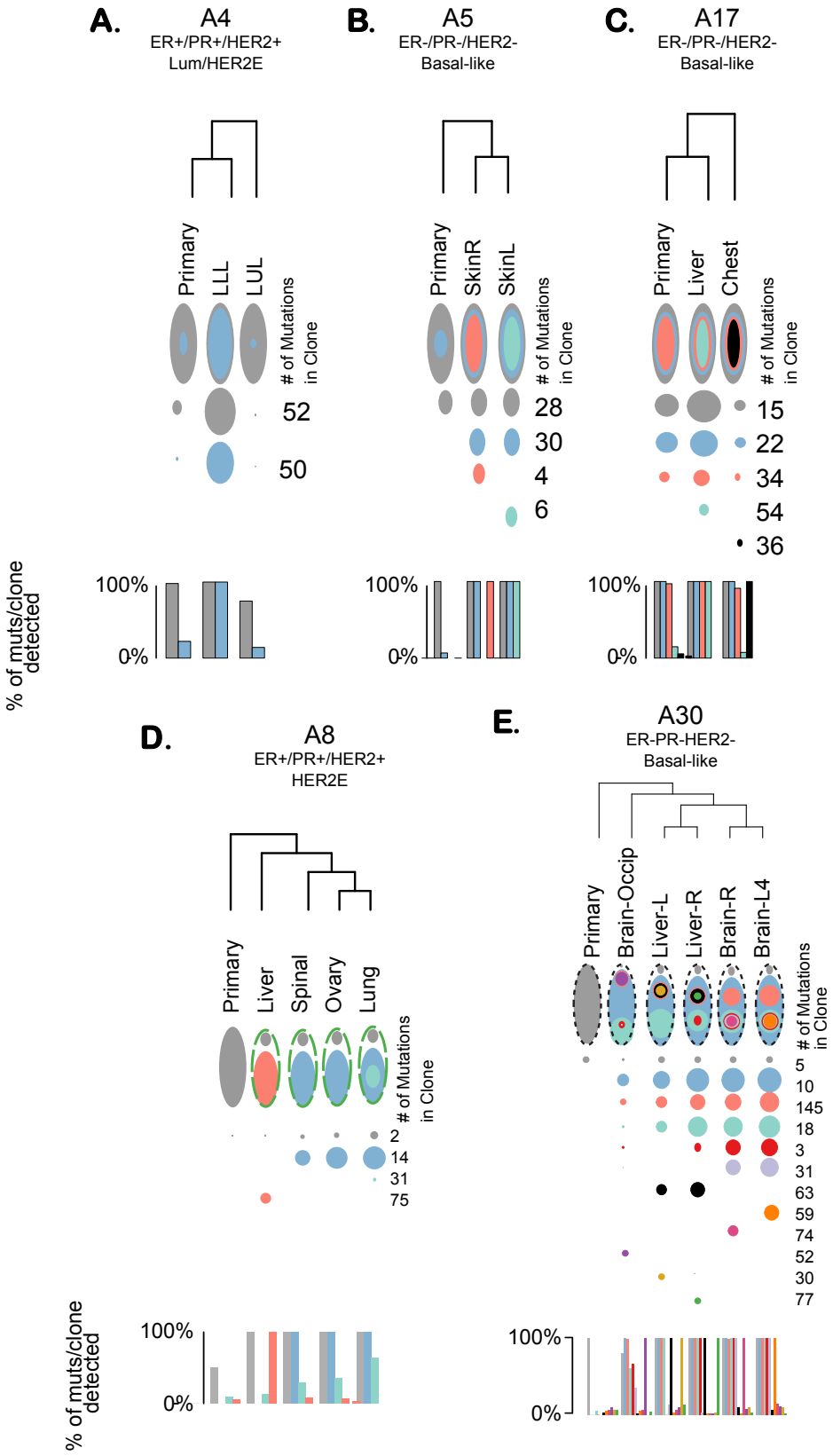
Supplemental Figure 8. DawnRank predicted drivers using input expression ratio of metastasis to matched primary. Each tumor was normalized by dividing the metastasis to the matched primary. Thus, the input to DawnRank was the gene expression ratio of metastasis to matched primary. **A.** The top 5% of DawnRank hits were then compared to the normalization method of median centering across the TCGA Lobular dataset. Timing was defined for both **B.** mutations and **C.** copy number alterations by categorizing alterations based on when they were acquired during the development of metastasis: founder (in all tumors from the patient), shared prim/mets (in the primary and at least 1 metastasis), shared met/met (in at least 2 metastases but not in the primary), or private (only identified in 1 metastasis).

Supplemental Figure 9



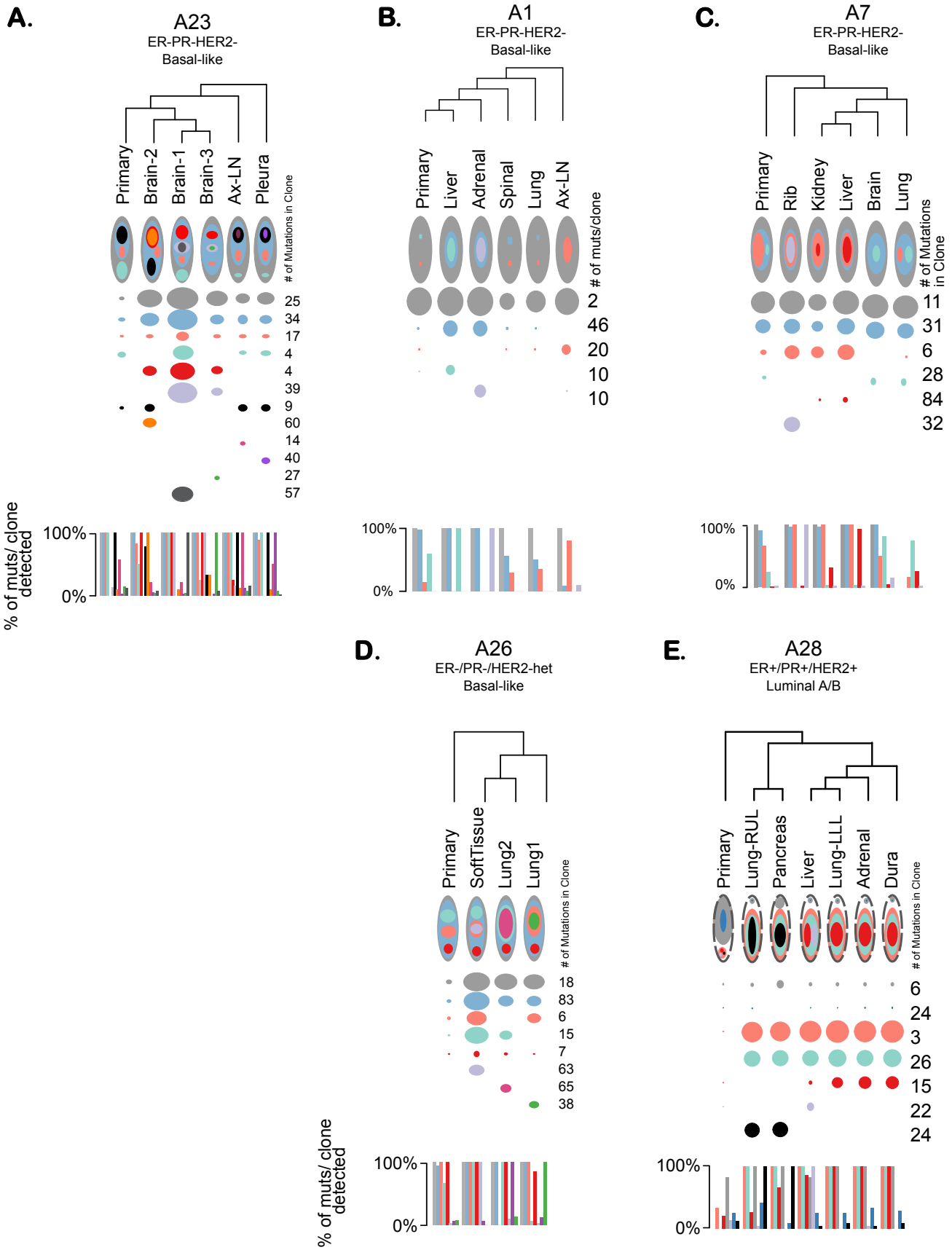
Supplemental Figure 9. Visualization of linear modeling: absolute and relative differences of RNA gene expression among matched primary and metastases of *PDE4C* and *PER2*. **A.** log2 RSEM normalized counts of RNA sequencing for *PDE4C* are plotted for each patient: boxplot indicating the distribution of metastases (black dots) with a single point for the primary tumor (red asterisk). T stat from the linear model glm and a calculated FDR from 100 permutations of the 'met' or 'primary' label. **B.** RNA values were normalized within each patient by subtracting the primary tumor from the metastases and plotted. **C.** *PER2* gene expression plotted for each patient with primaries (red asterisk) and boxplot of metastases (black dots). **D.** Relative *PER2* RNA gene expression of metastases within each patient.

Supplemental Figure 10



Supplemental Figure 10. Monoclonal primary tumors seeding both monoclonal and polyclonal metastases across subtypes. For each patient: dendrogram defined by Dollo Parsimony with bootstrapping, combined dot plots based on the keyhole principles, dot proportionate to the mean cellular prevalence of the mutations in that clone, barplot of the % of muta-tions within each clone detected for that tumor.In patients A. A4, B. A5, and C. A17 demonstrate monoclonal seeding, and D. A8 and E. A20 have monoclonal primaries and polyclonal metastases.

Supplemental Figure 11



Supplemental Figure 11. Polyclonal seeding. For each patient: dendrogram defined by Dollo Parsimony with bootstrapping, com-bined dot plots based on the keyhole principles, dot proportionate to the mean cellular prevalence of the mutations in that clone, barplot of the % of mutations within each clone detected for that tumor. A. A23, B. A1, C. A7, D. A26, and E. A28.