SUPPLEMENTARY INFORMATION

Between-Region Genetic Divergence Reflects the Mode and Tempo of Tumor Evolution

Ruping Sun[1,2,3§], Zheng Hu[1,2,3§], Andrea Sottoriva[4], Trevor A. Graham[5], Arbel Harpak[6], Zhicheng Ma[1,2,3], Jared M. Fischer[7], Darryl Shibata[8], Christina Curtis[1,2,3 *]

Affiliations

[1] Department of Medicine, Stanford University School of Medicine, Stanford, California, USA

[2] Department of Genetics, Stanford University School of Medicine, Stanford, California, USA

[3] Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA

[4] Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK

[5] Barts Cancer Institute, Queen Mary University of London, London, John Vane Science Centre, Charterhouse Square, London, UK

[6] Department of Biology, Stanford University, Stanford, California, USA

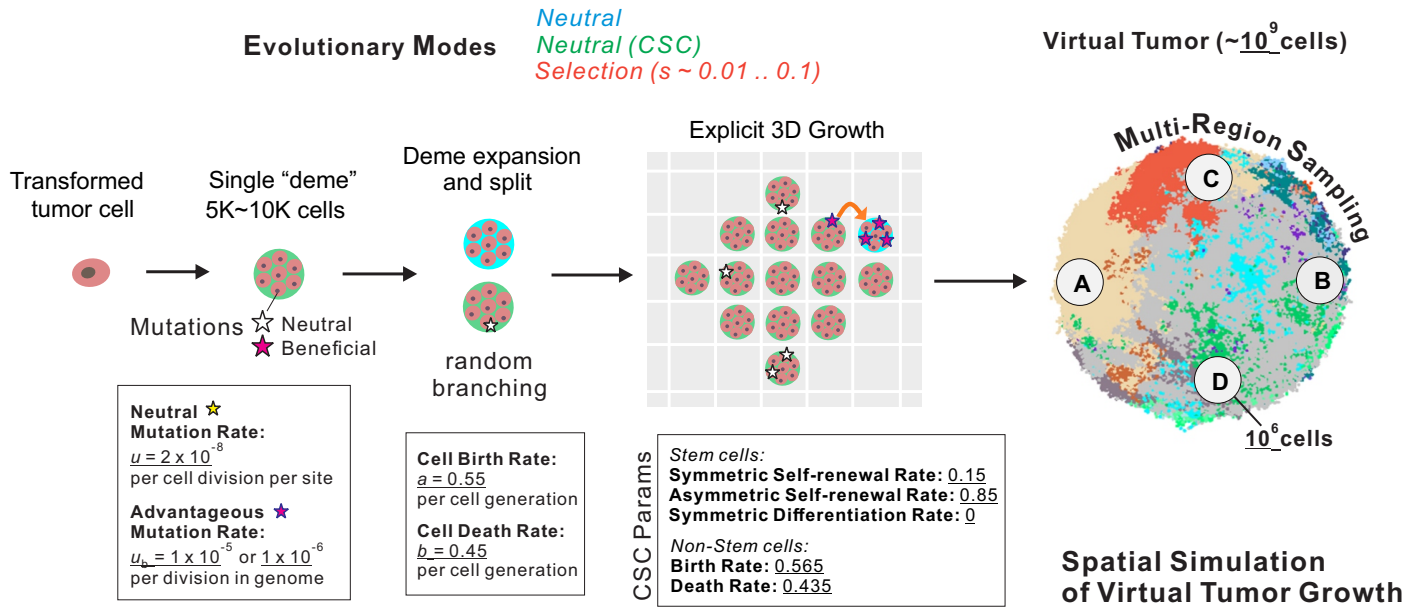
[7] Oregon Health and Science University, Department of Molecular and Medical Genetics, Portland, Oregon, USA

[8] Department of Pathology, Keck School of Medicine of the University of Southern California, Los Angeles, California, USA

[§] These authors contributed equally to this work

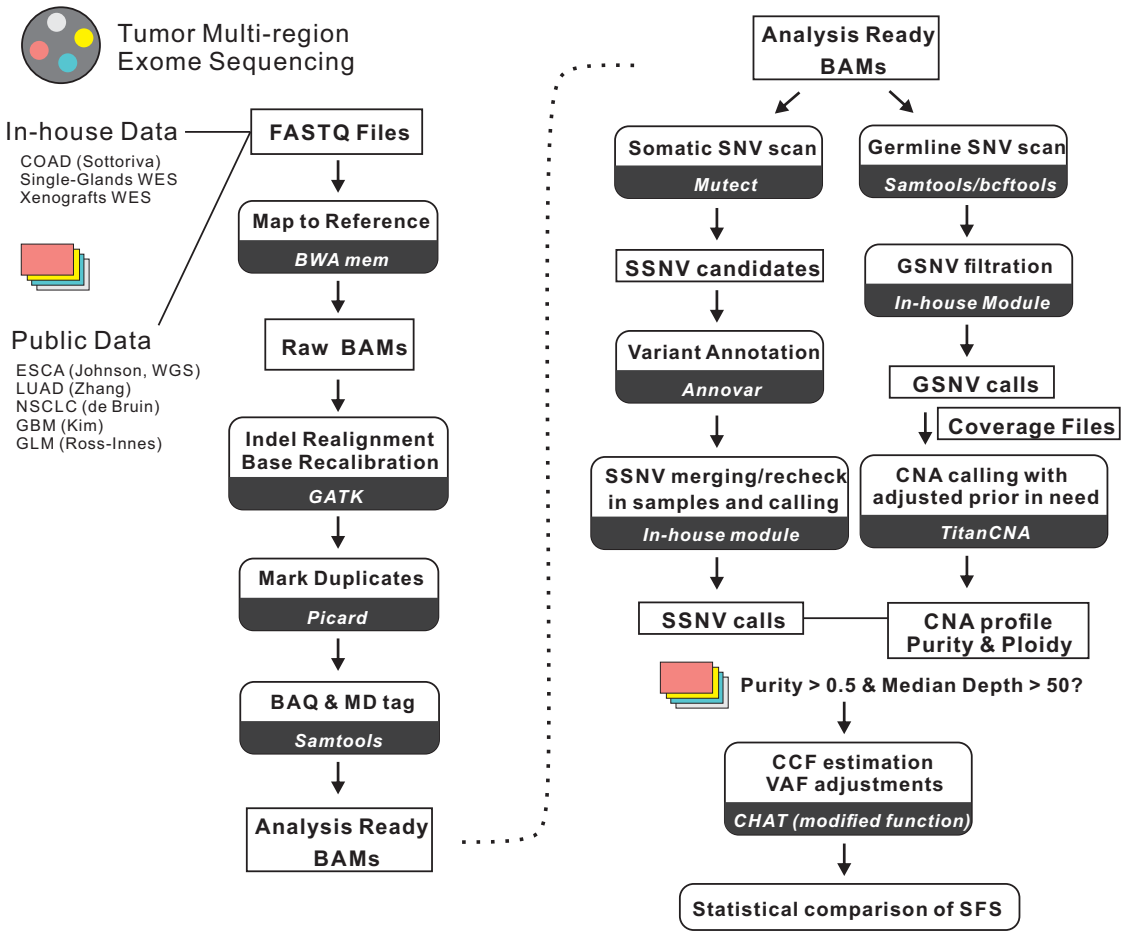
* Corresponding author: Christina Curtis, Stanford University School of Medicine, 265 Campus Drive, Lorry Lokey Building Suite G2120C, Stanford, CA 94305 Tel: 650-498-9943, Email: cncurtis@stanford.edu

Supplementary Figures



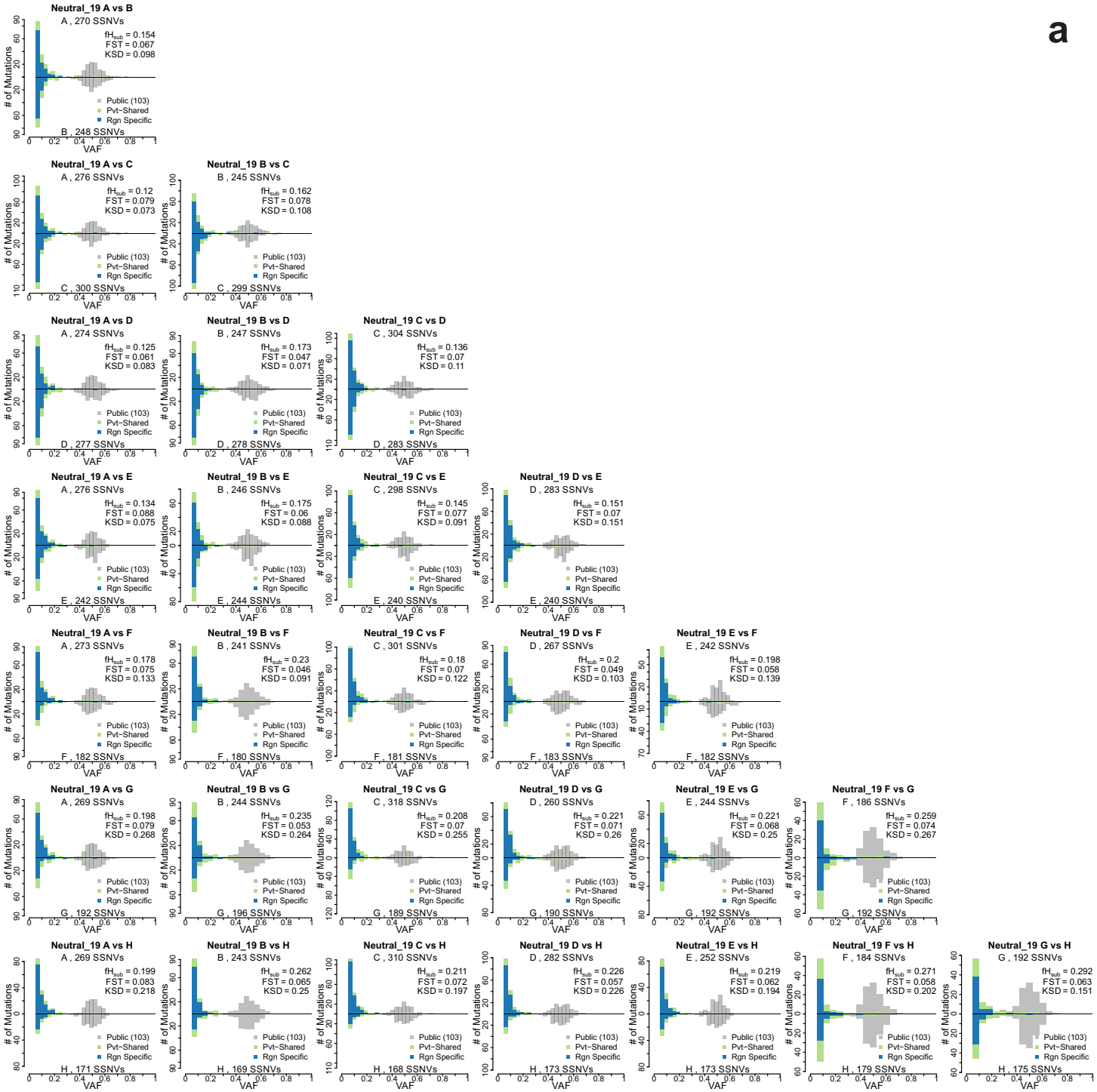
Supplementary Figure 1: An agent-based model for simulating 3D tumor growth, mutation accumulation, and multi-region sequencing data under various evolutionary modes.

Tumor growth is simulated via the expansion of deme subpopulations (composed of 5-10k cells or 0.5-1k cells) within a defined 3D cubic lattice according to explicit rules dictated by spatial constraints, where cells within each deme are well-mixed and grow via a random branching (birth-death) process (Methods). The acquisition of random mutations (neutral or beneficial) is simulated under different evolutionary modes (the null neutral model (selection coefficient, s=0), neutral-CSC (s=0), and selection models with s=0.01, 0.02, 0.03, 0.05 or 0.1), and used to trace the genealogy of each cell as the tumor expands. Subsequently, the virtual tumor is sampled and sequenced as is done experimentally after resection or biopsy. Using this approach it is possible to evaluate differences in the SFS and tumor subclonal architecture under different modes of selection and sampling strategies. Key parameters of the models are indicated in the Methods (**Supplementary Table 1**).

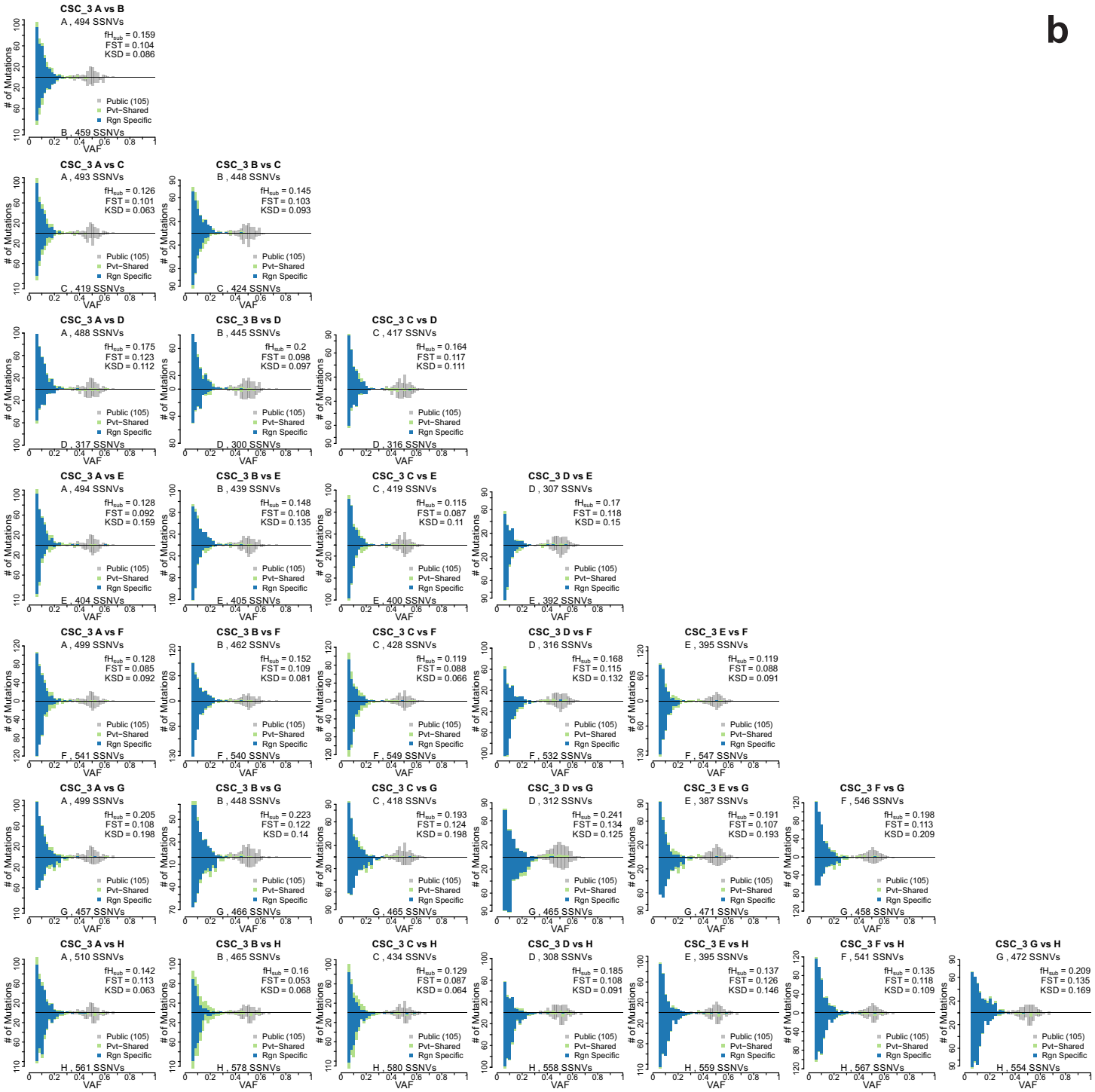


Supplementary Figure 2: Schematic overview of MRS analysis pipeline.

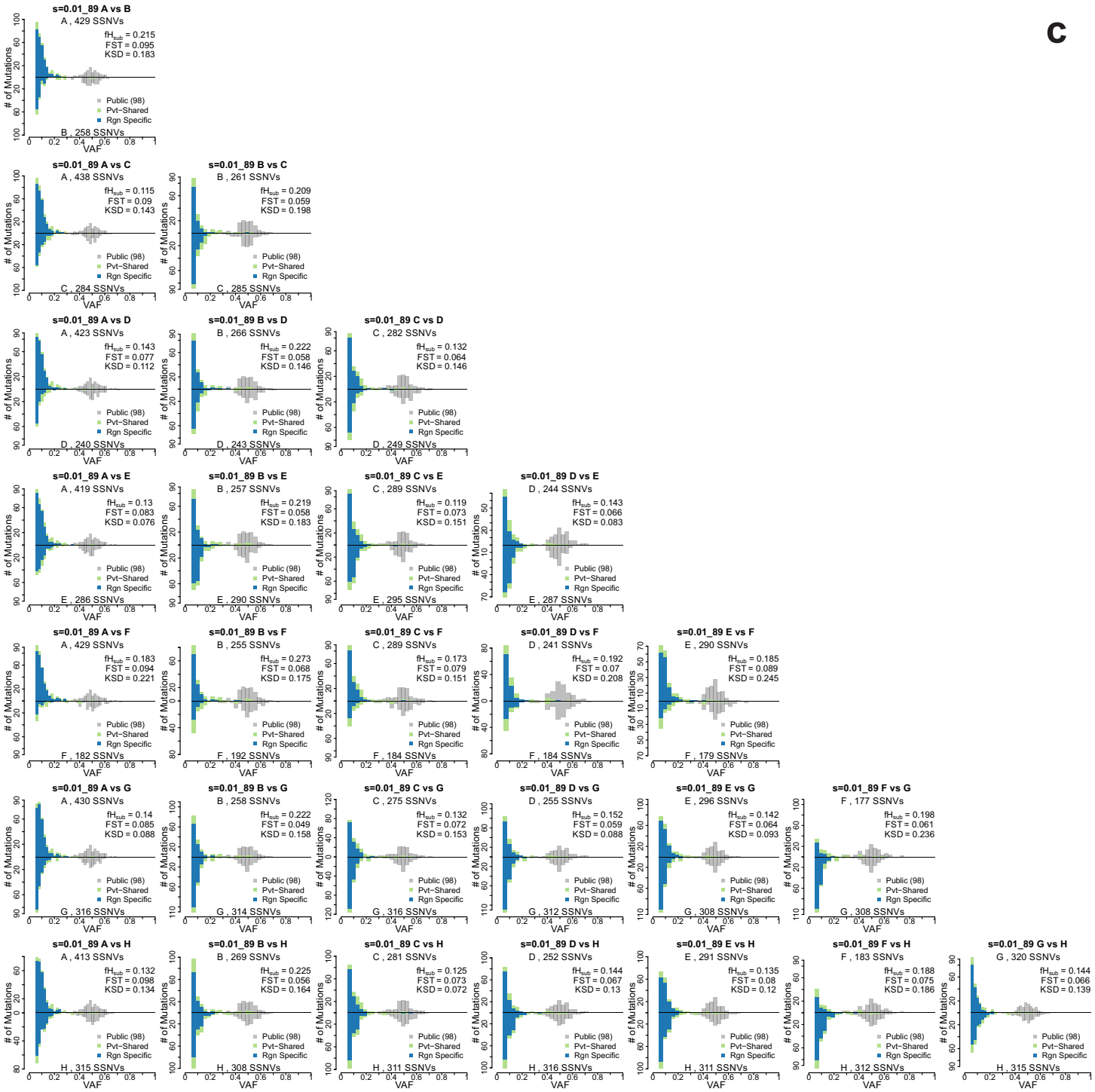
Short read data from in-house or publicly available MRS datasets were aligned to hg19 using BWA (0.7.12-r1039) [1]. The raw alignment files (BAMs) were then pre-processed following best practices, including indel realignment, base recalibration and flagging of duplicated reads. MuTect (v1.1.4) [2] and Samtools (v 1.2) [3] were employed to scan for somatic SNVs (SSNV) and germline SNVs (GSNV), respectively. The raw variant level data were then subject to analysis within a custom variant assurance pipeline (VAP) for MRS to ensure balance between sensitivity and specificity (Methods and Supplementary Note). SSNVs were determined by calculating the log odds (LOD) of the mutation at that site and the LOD of the normal being the reference allele given the sequencing reads and base qualities, as previously described [2]. To increase sensitivity, a moderately low threshold for the tumor LOD was adopted and local errors were accounted for to ensure the conservative treatment of base error rates. Additional filters were employed to exclude germline variants or sequencing artifacts (see Supplementary Note). The VAF (unadjusted) was calculated by dividing the number of reads carrying the variant by the total number of reads spanning that position. Germline SNVs (GSNVs) initially detected by Samtools were subject to the same mapping feature filter to retain high quality GSNVs. The read depth for alternative and reference alleles was calculated for heterozygous SNVs found in the normal control. CNA calls and tumor purity/ploidy estimates were generated using TitanCNA [4]. To enable more reliable adjustment of VAF and subsequent calculation of statistics related to the SFS, we set stringent QC requirements for the inclusion of tumor samples: 1) estimated purity >0.5; 2) median sequencing depth >50; 3) number of subclonal SSNVs >50. At least two primary tumor regions were required to meet these specifications. The observed VAFs were then adjusted for estimated tumor purity and local copy number as described elsewhere [5] (see Supplementary Note).



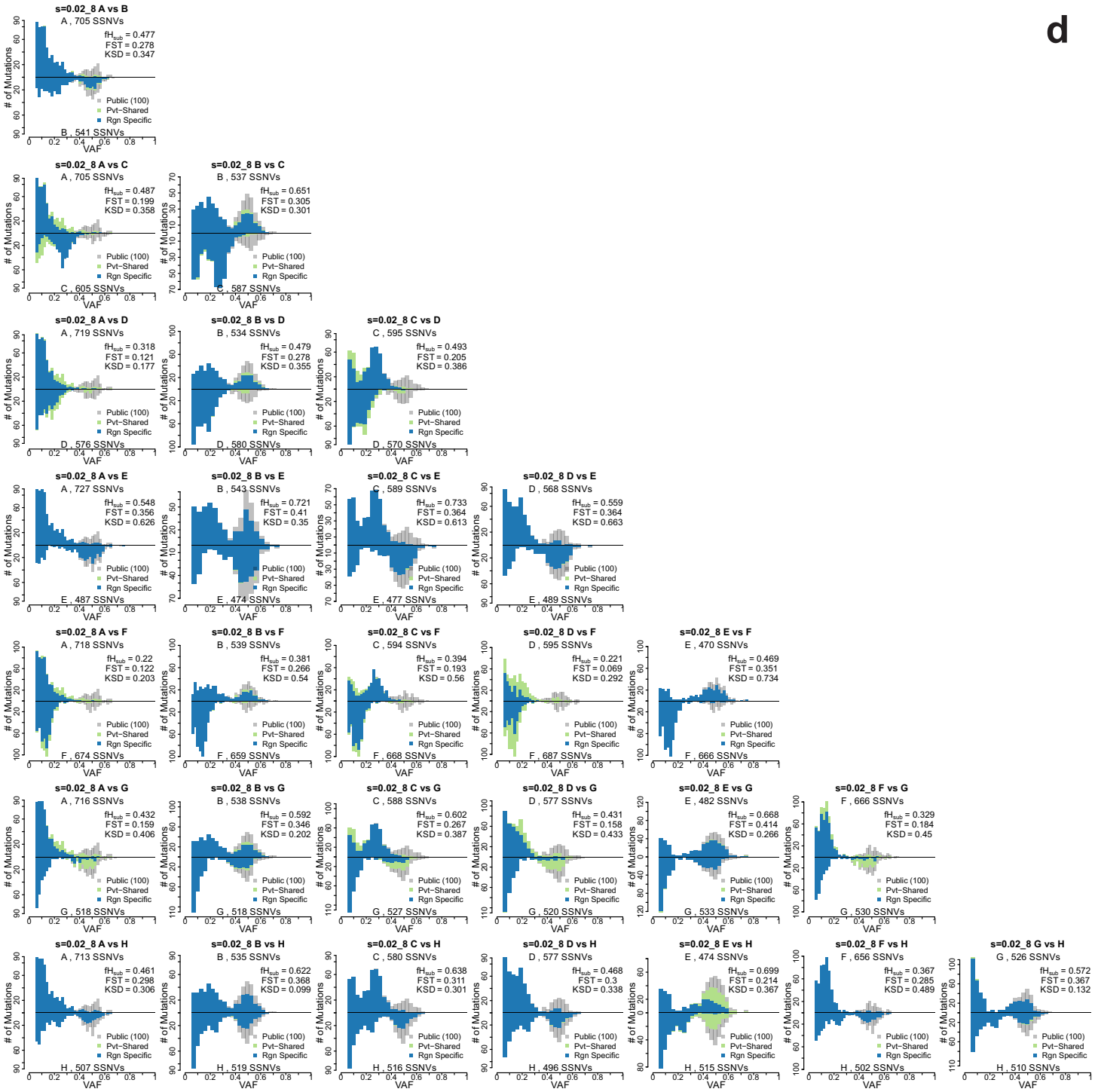
Supplementary Figure 3: Pairwise SFS histograms for representative virtual tumors simulated under different evolutionary modes.



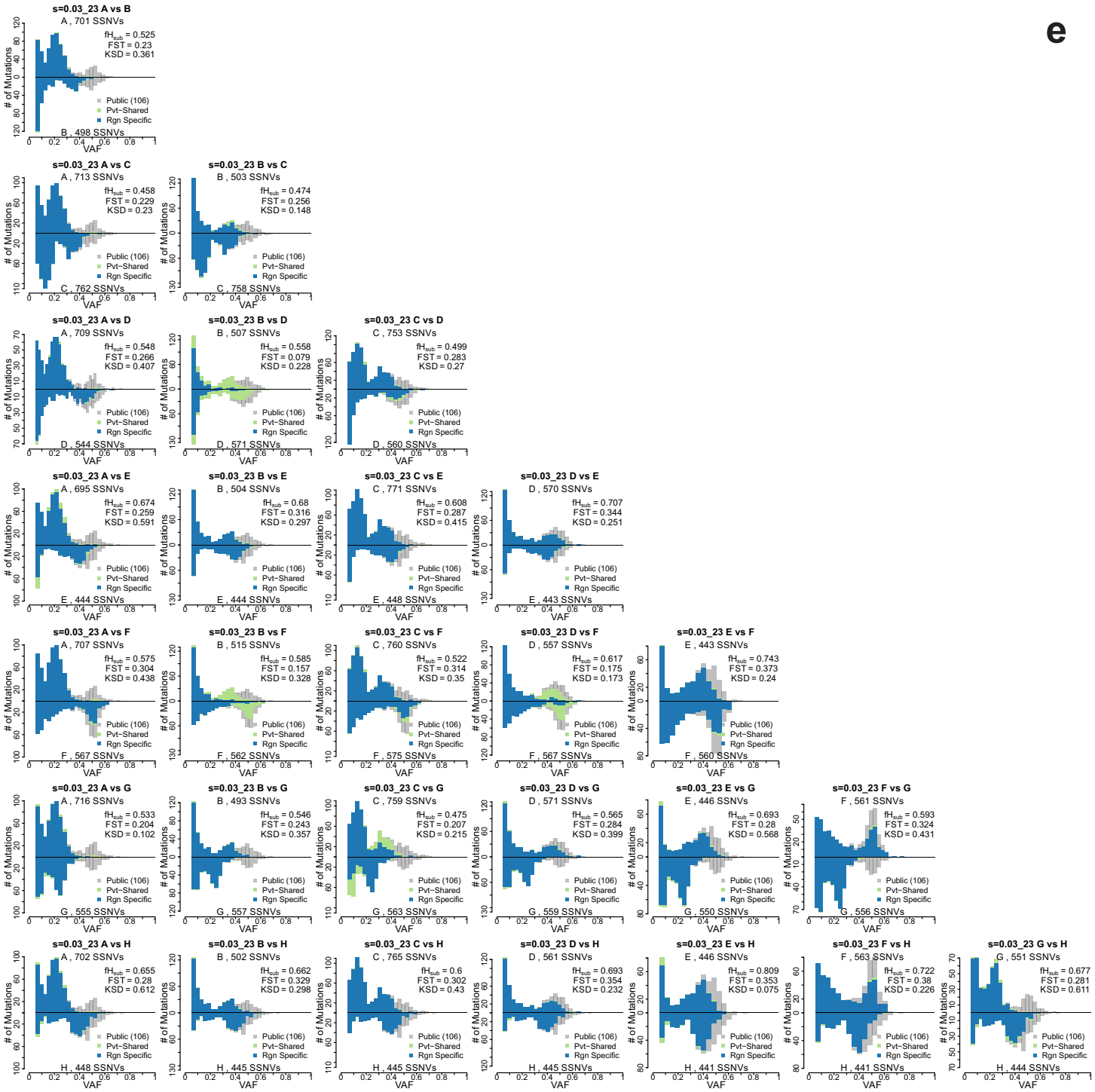
Supplementary Figure 3: Pairwise SFS histograms for representative virtual tumors simulated under different evolutionary modes.



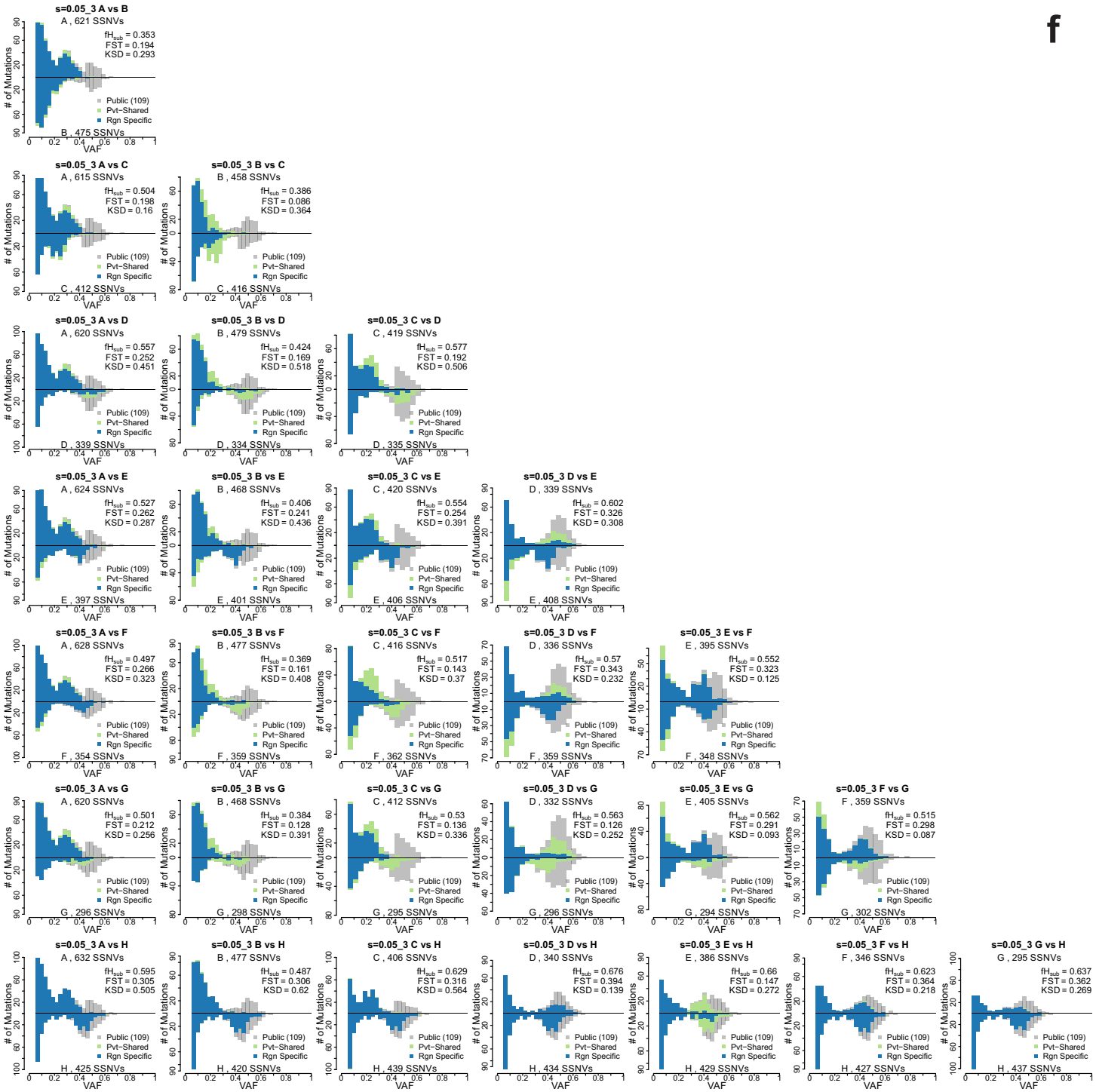
Supplementary Figure 3: Pairwise SFS histograms for representative virtual tumors simulated under different evolutionary modes.



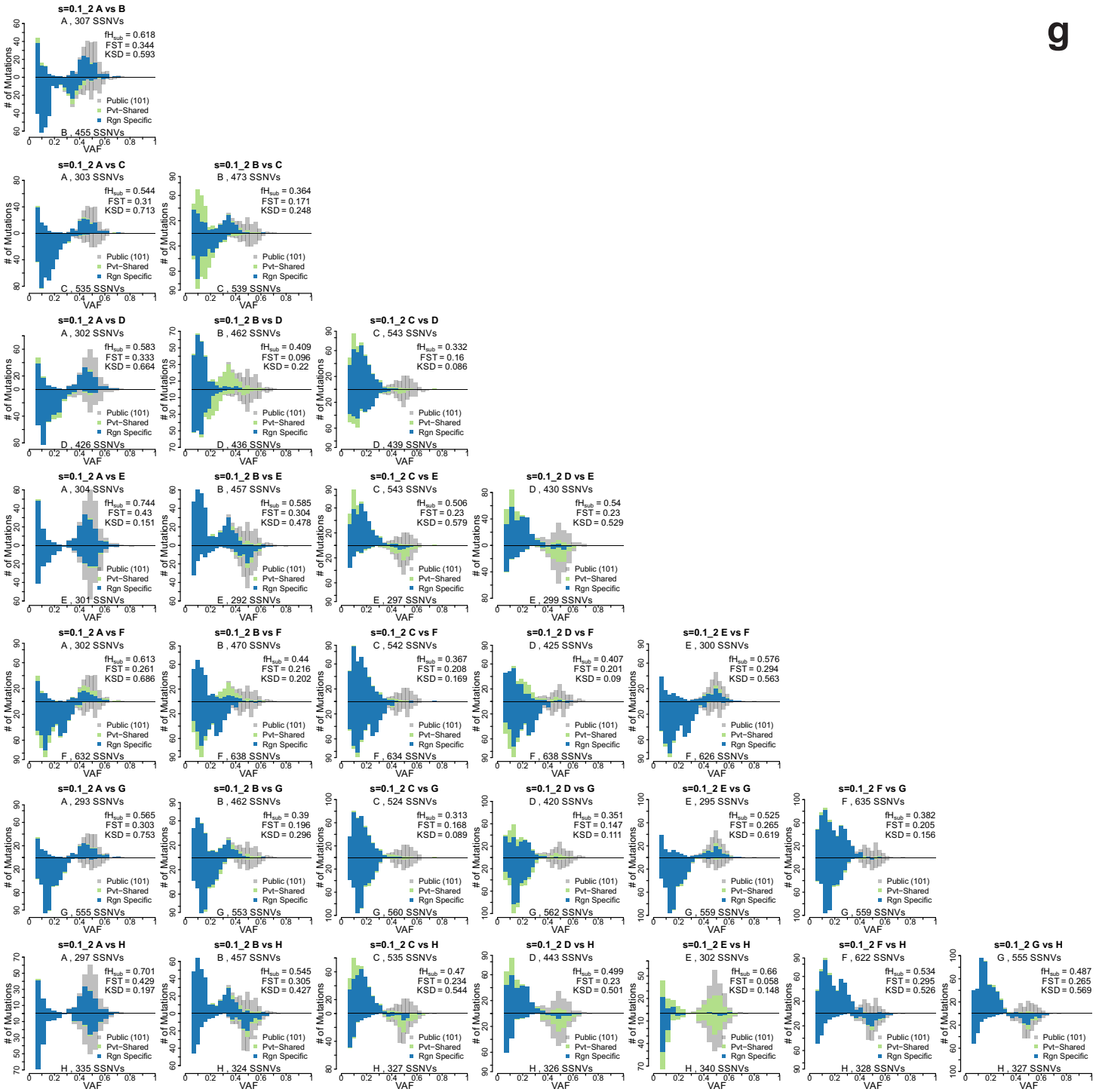
Supplementary Figure 3: Pairwise SFS histograms for representative virtual tumors simulated under different evolutionary modes.



Supplementary Figure 3: Pairwise SFS histograms for representative virtual tumors simulated under different evolutionary modes.



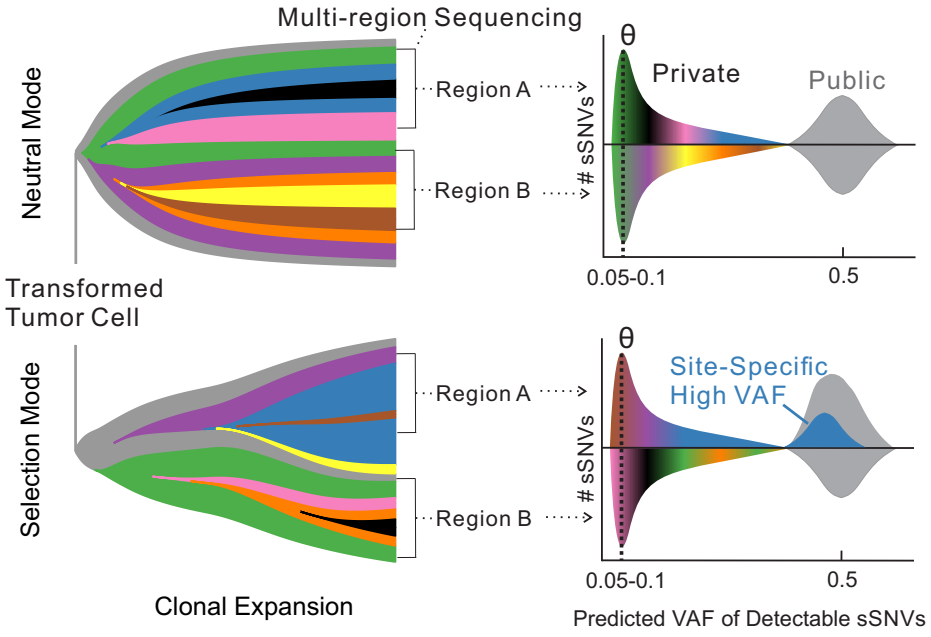
Supplementary Figure 3: Pairwise SFS histograms for representative virtual tumors simulated under different evolutionary modes.



Supplementary Figure 3: Pairwise SFS histograms for representative virtual tumors simulated under different evolutionary modes.

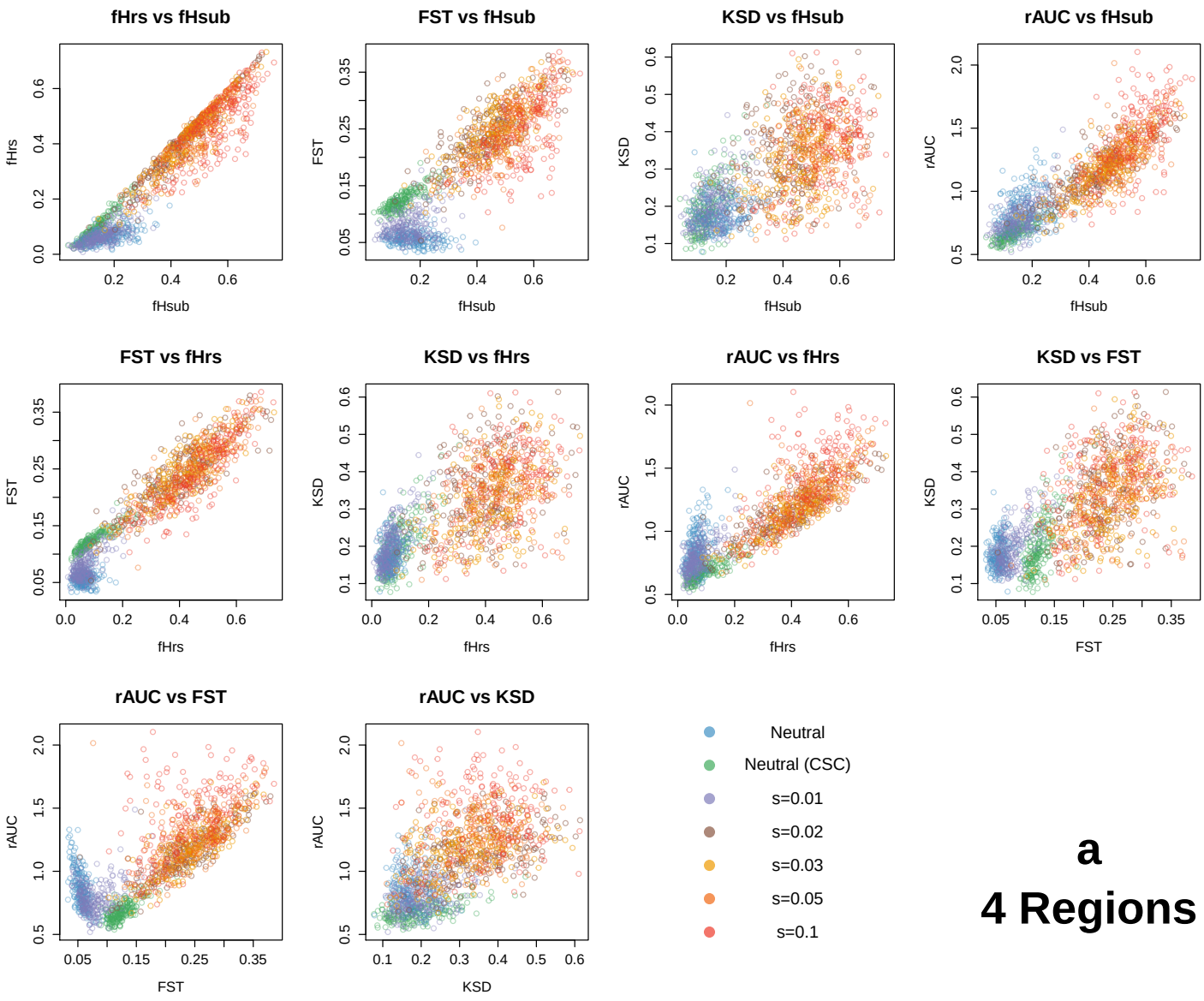
Representative pairwise SFS histograms derived from two spatially separated tumor regions (labeled as A to H) within the same tumor are shown for tumors simulated under different evolutionary modes, namely: **(a)** neutral model, **(b)** neutral (CSC), **(c)** selection with s=0.01, **(d)** s=0.02, **(e)** s=0.03, **(f)** s=0.05 and **(g)** s=0.1.

SSNVs were classified as Public (gray), Private (Pvt)-shared (green), or Pvt-region specific (blue) based on their presence in the MRS data and each pairwise SFS comparisons (Methods). The total number of SSNVs detected in each region, as well as several ITH metrics derived from the SFS are indicated, namely fHsub, FST and KSD.

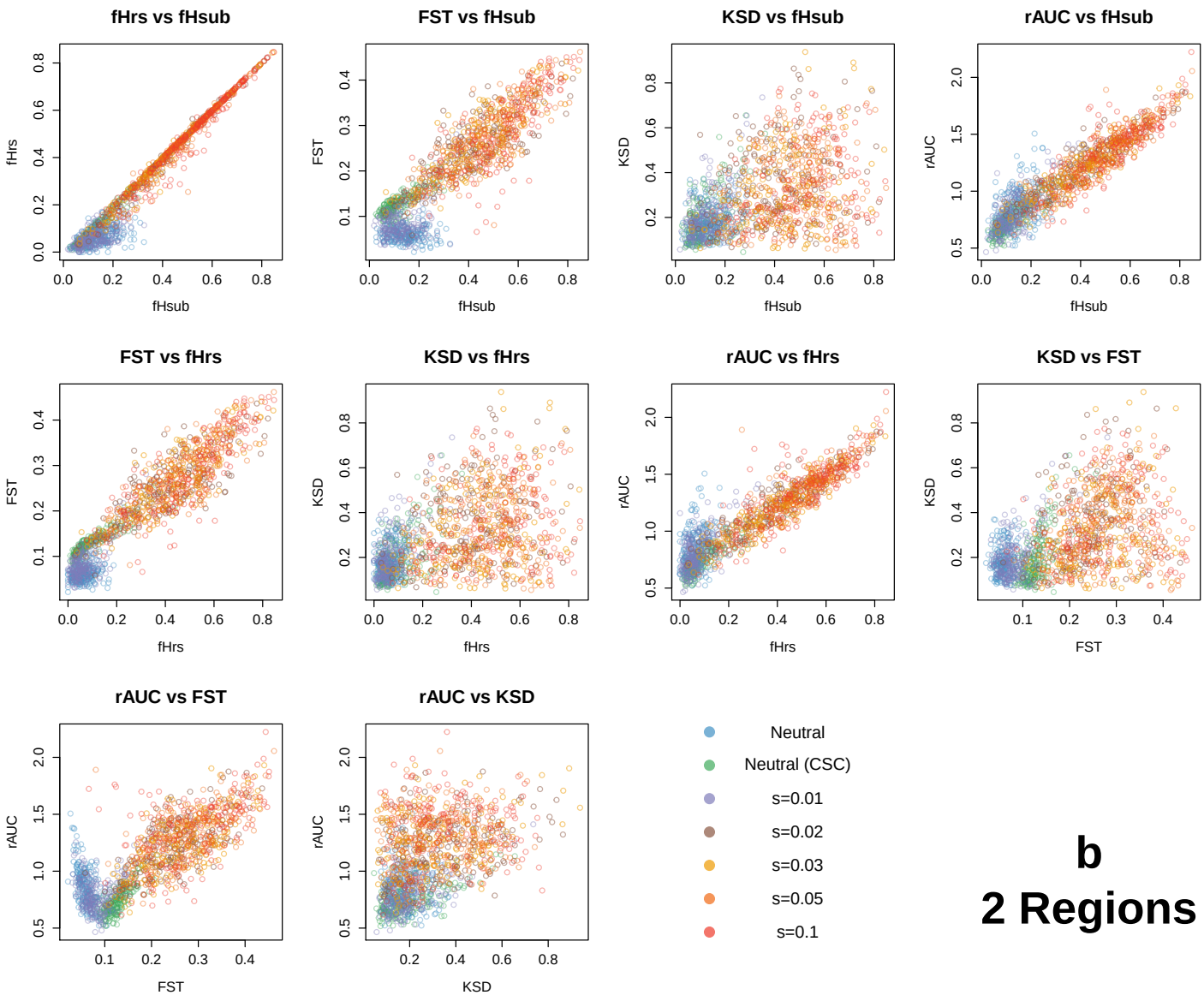


Supplementary Figure 4: Schematic representation of the expected pairwise SFS arising under effectively neutral growth and stringent selection.

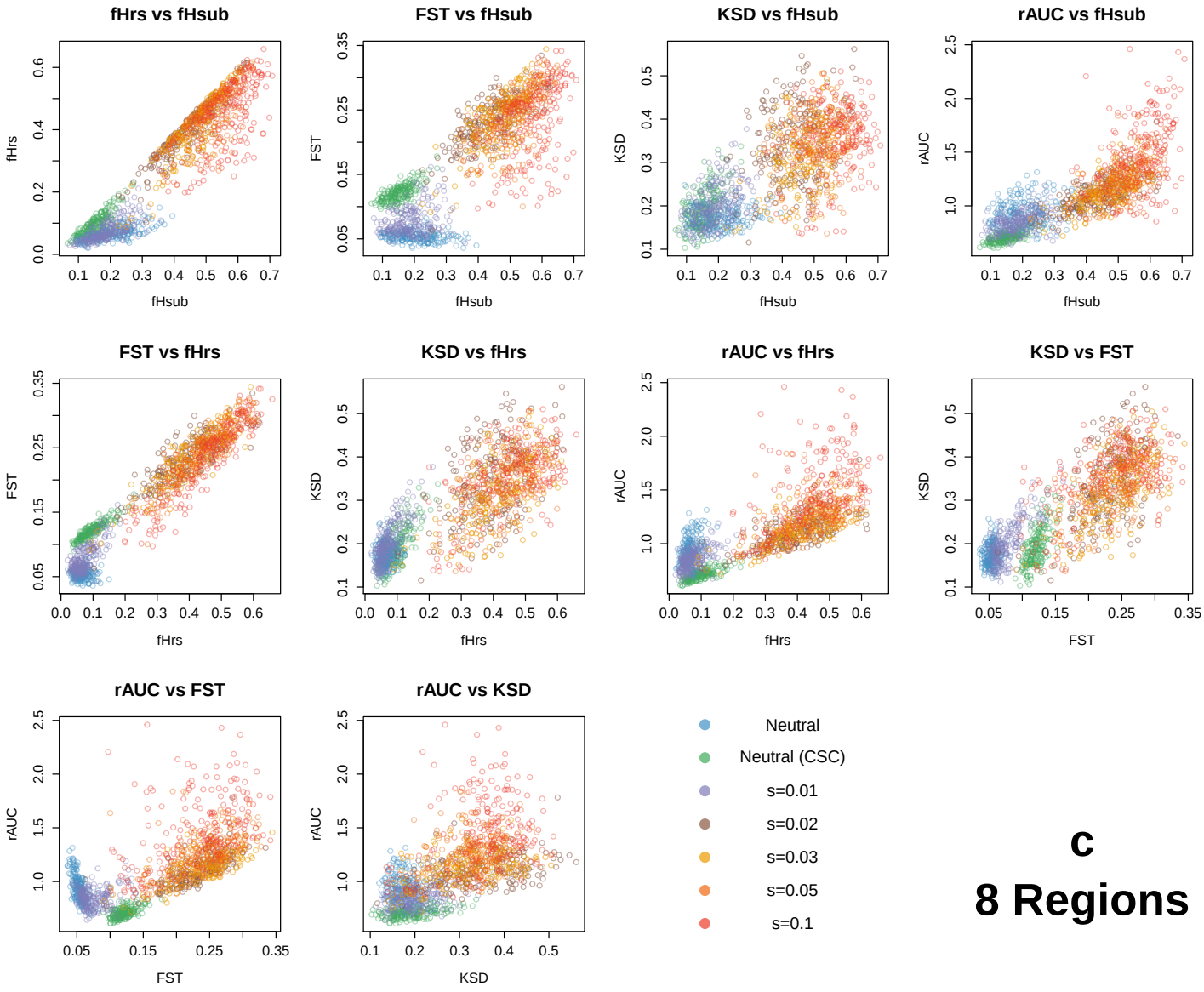
Muller plots (DOI:10.5281/zenodo.240589) were employed to illustrate tumor subclone composition under effectively neutral growth (upper panel) or strong selection (bottom panel). Descendant genotypes emerge within the parental clone, height corresponds to genotype frequency (or relative abundance) and the horizontal axis indicates time (generations). Tumors arising under effectively neutral growth are expected to exhibit similar SFS between regions with a public mutational cluster centered at 0.5 and a large right skewed distribution of low VAF subclonal mutations that accrue as the tumor expands, where θ indicates the threshold below which low frequency mutations cannot be reliably detected in bulk tumor samples. This can be contrasted with the SFS expected under strong selection in which private SSNVs can attain high frequency due to regional expansion of a clone occurring late (blue shading). These patterns highlight the need for MRS to distinguish such private high frequency SSNVs, which could appear public in single-sample sequencing.



Supplementary Figure 5: Pairwise comparisons of five ITH metrics derived from the SFS in virtual tumors.

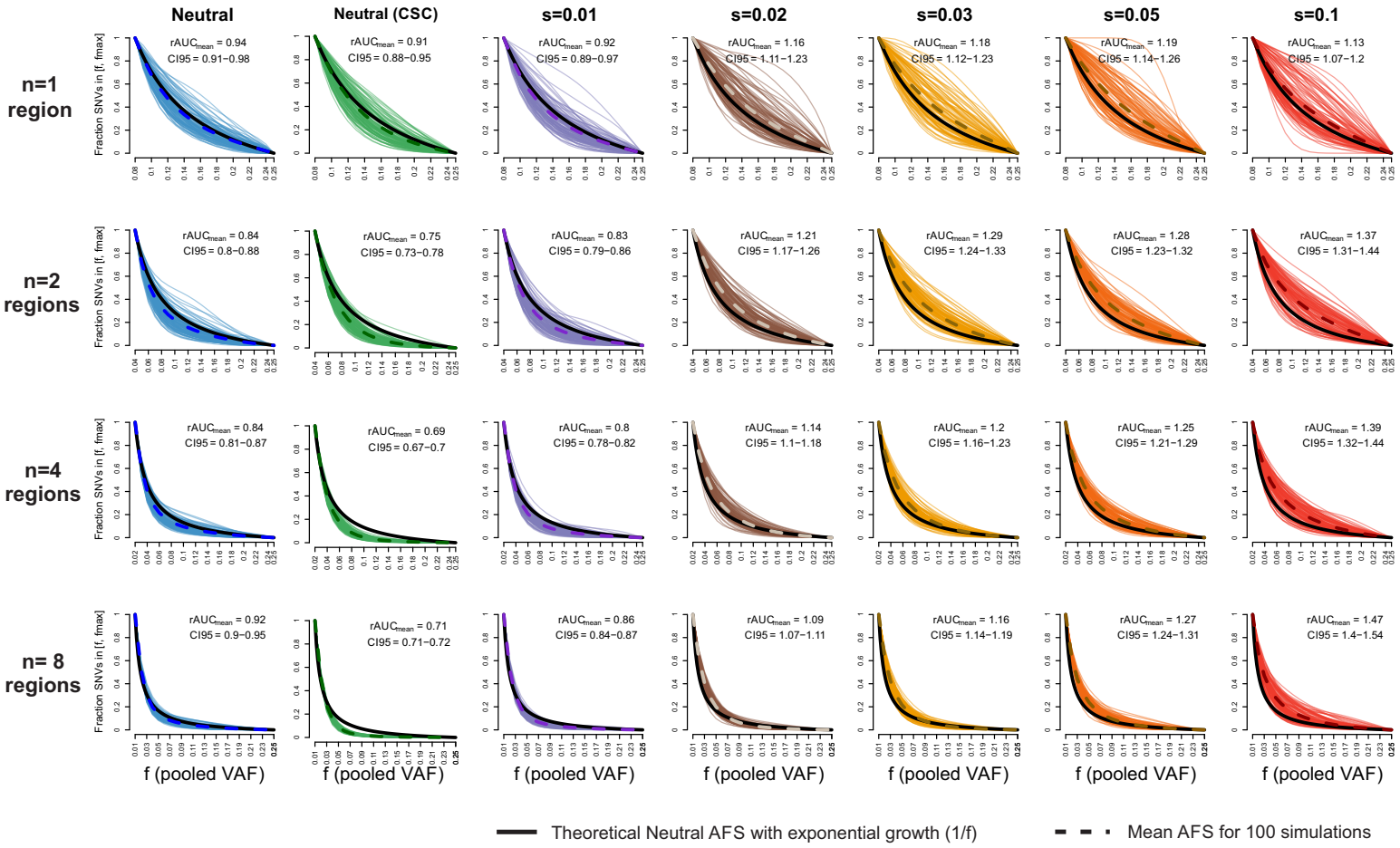


Supplementary Figure 5: Pairwise comparisons of five ITH metrics derived from the SFS in virtual tumors.



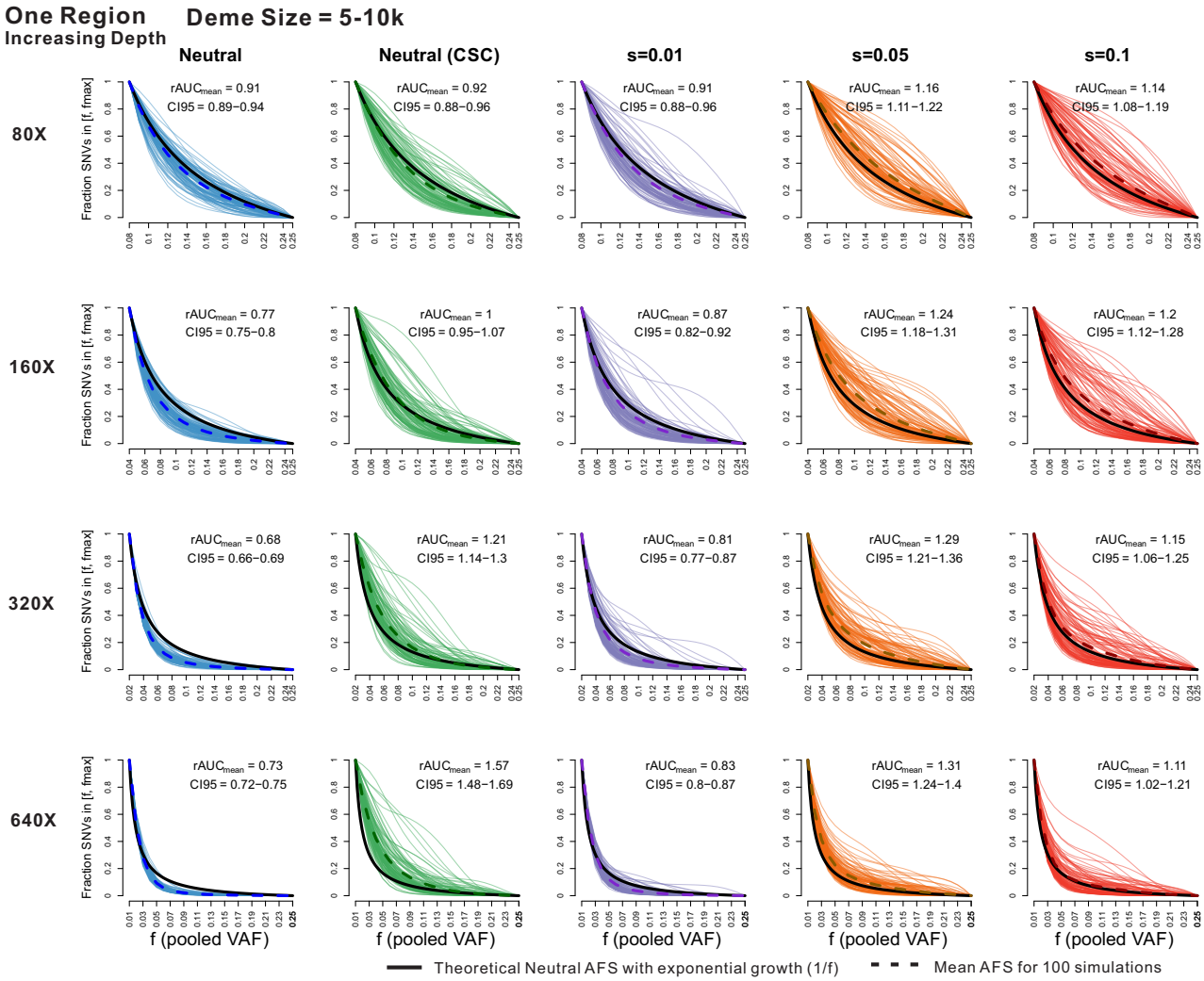
Supplementary Figure 5: Pairwise comparisons of five ITH metrics derived from the SFS in virtual tumors.

Pairwise scatterplots of the five ITH metrics (fHsub, fHrs, FST, KSD and rAUC) based on 200 virtual tumors (80X coverage) for different evolutionary modes (color coded) are shown for **(a)** 4 regions **(b)** 2 regions and **(c)** regions.



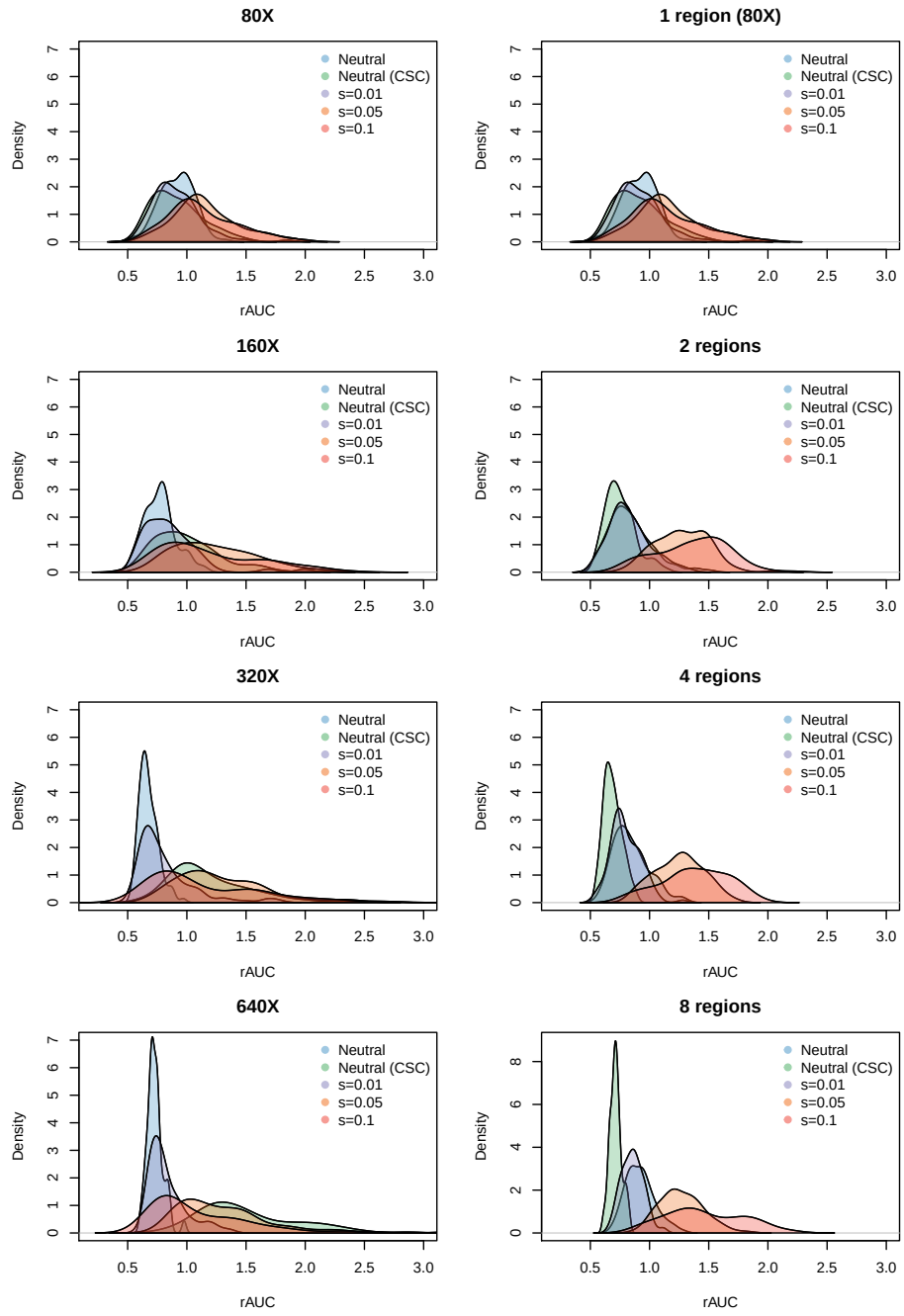
Supplementary Figure 6: Cumulative SFS for tumors subject to single region sequencing or MRS.

Cumulative SFS of simulated tumors (100 for each mode, sequenced to 80X coverage) by pooling the VAF across 1, 2, 4 or 8 regions. The cumulative fraction of mutations in the specified frequency (f) range is shown. From left to right, seven different modes are shown, i.e., neutral, neutral (CSC), s=0.01, s=0.02, s=0.03, s=0.05 and s=0.1. Colored lines correspond to simulated tumors, where the dashed line corresponds to the average. The thick black line corresponds to the theoretical cumulative SFS under neutral exponential growth in a well-mixed population. The ratio of the area under the cumulative SFS curve for virtual tumors relative to the theoretical cumulative SFS curve (denoted by rAUC) is indicated, as well as the mean of the rAUC and the mean rAUC 95% bootstrap confidence interval (based on the adjusted bootstrap percentile method).



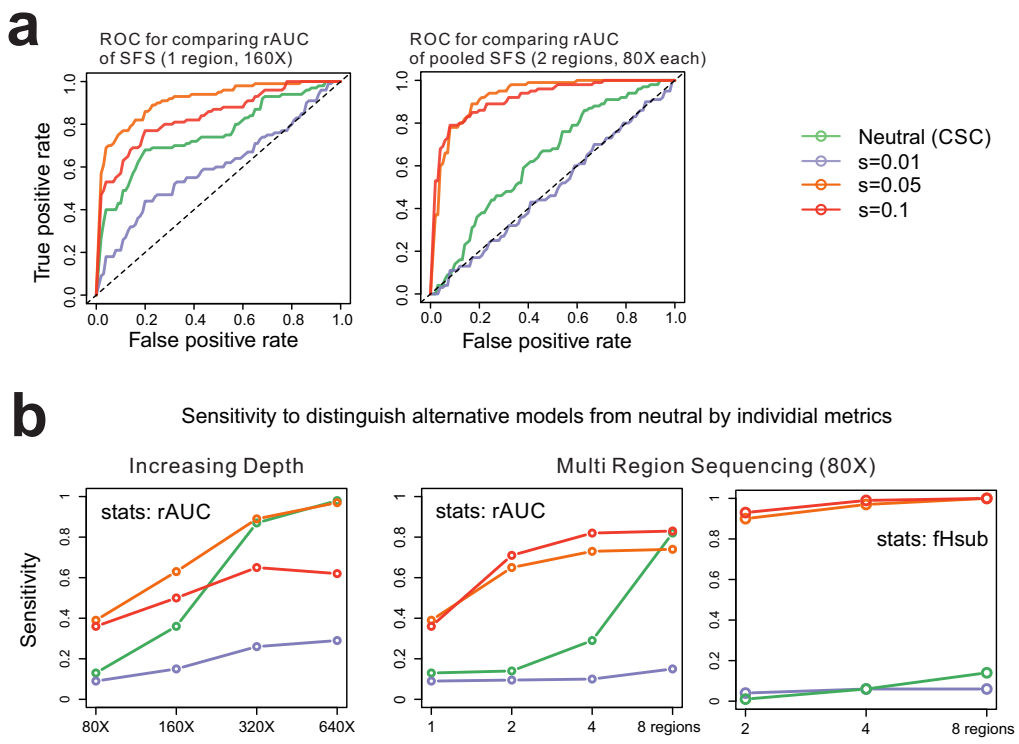
Supplementary Figure 7: Cumulative SFS for increased sequencing depth of a single bulk tumor sample.

The cumulative SFS is plotted as in **Supplementary Figure 6** for increasing sequencing depth (80X, 160X, 320X and 640X) of a single bulk sample. Five evolutionary modes are shown, i.e., neutral, neutral (CSC), s=0.01, s=0.05 and s=0.1.



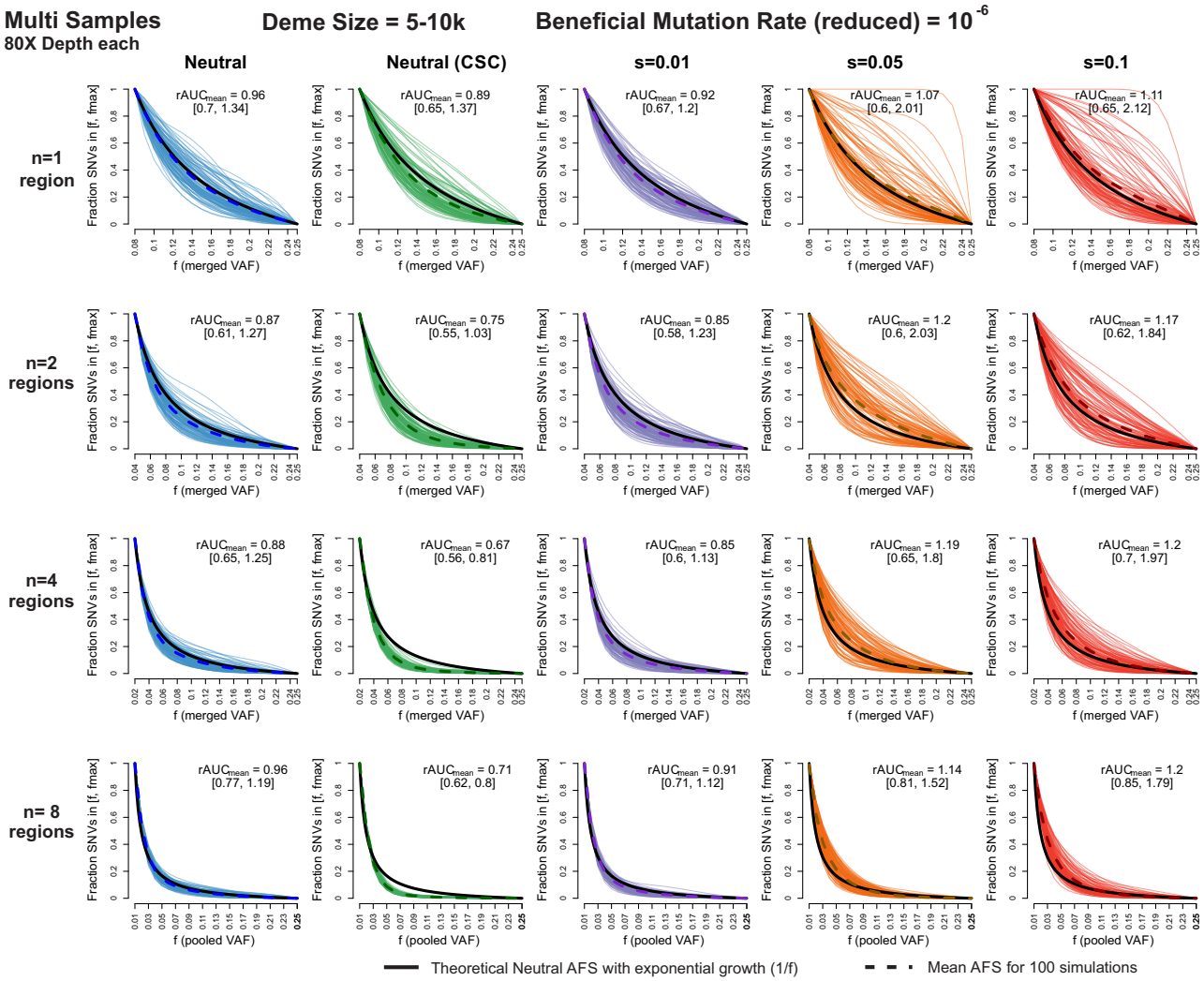
Supplementary Figure 8: MRS better distinguishes between effectively neutral growth and selection than sequencing single sample bulk tumor.

Color shaded rAUC density curves are shown for each of five evolutionary modes (neutral, neutral CSC, s=0.01, s=0.05 and s=0.1) for increasing sequencing depth for one sample (left column) or an increasing number of regions (right column). Increasing the number of regions yields better discrimination between effectively neutral growth (neutral, neutral CSC and s=0.01) and selection (s≥ 0.05) than is achieved by simply sequencing a single sample to higher depth.



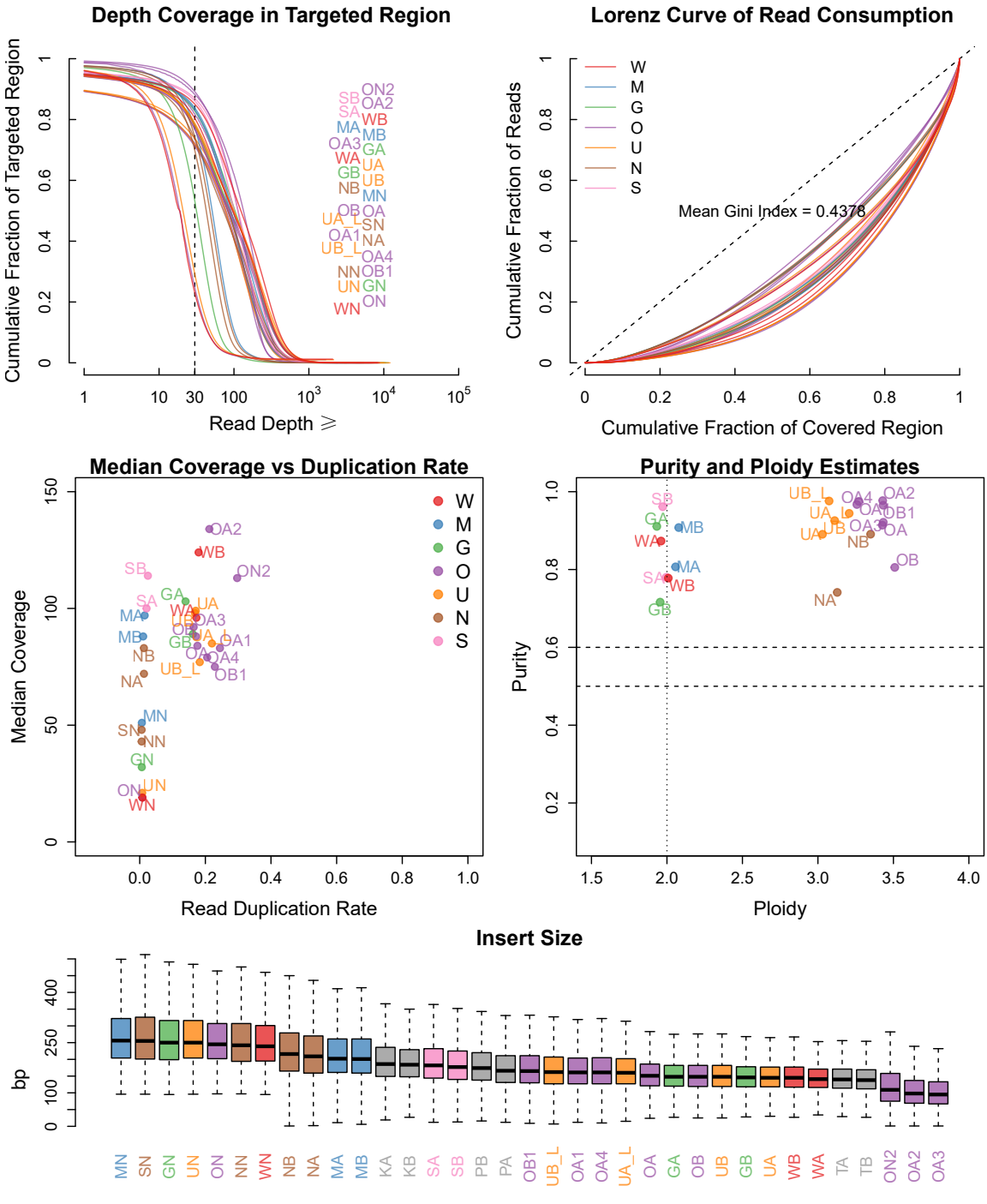
Supplementary Figure 9: Performance to distinguish alternate evolutionary modes from the simulated neutral model.

Receiver Operating Characteristic (ROC) curves illustrating the true positive versus false positive rate to distinguish four alternate evolutionary models from the simulated neutral model based on comparisons of the SFS derived from 1 tumor region (left panel) or 2 tumor regions (right panel) using the rAUC metric. The inclusion of an additional tumor region results in improved performance relative to one region for s=0.05 and especially for more stringent selection (s=0.1) owing to the ability of MRS to capture subclonal variants that are regionally fixed. In contrast, weak selection and the neutral (CSC) model cannot be robustly distinguished. (b) The power to distinguish individual alternate models (neutral-CSC, s=0.01, s=0.05 and s=0.1) from the neutral model was evaluated using rAUC (applicable to both single sample and MRS) and fHsub (specific to MRS). Power (at the 10% significance level) was computed empirically as the percentage of virtual tumors under an alternate model with the statistic (rAUC or fHsub) greater than 95% quantile or less than 5% quantile of the corresponding statistic in the neutral model (taking the larger percentage). The left panel illustrates the impact of increasing sequencing depth (80-640x) for a single sample (based on rAUC), whereas the two right panels illustrates the impact of increasing the number of tumor regions (n=1,2,4 and 8) for rAUC and fHsub (n=2,4 and 8), respectively. Of note, as deep sequencing generates accurate VAF for localized subclonal mutations in the sampled region, the tight rAUC distribution (**Supplementary Figure 8**) in the neutral model leads to high sensitivity to detect selection for $s \leq 0.05$. However, when selection increases to 0.1, deep sequencing of a single sample becomes less sensitive. This is due to the inability of single sample sequencing to distinguish high frequency subclonal mutations, which can appear public.



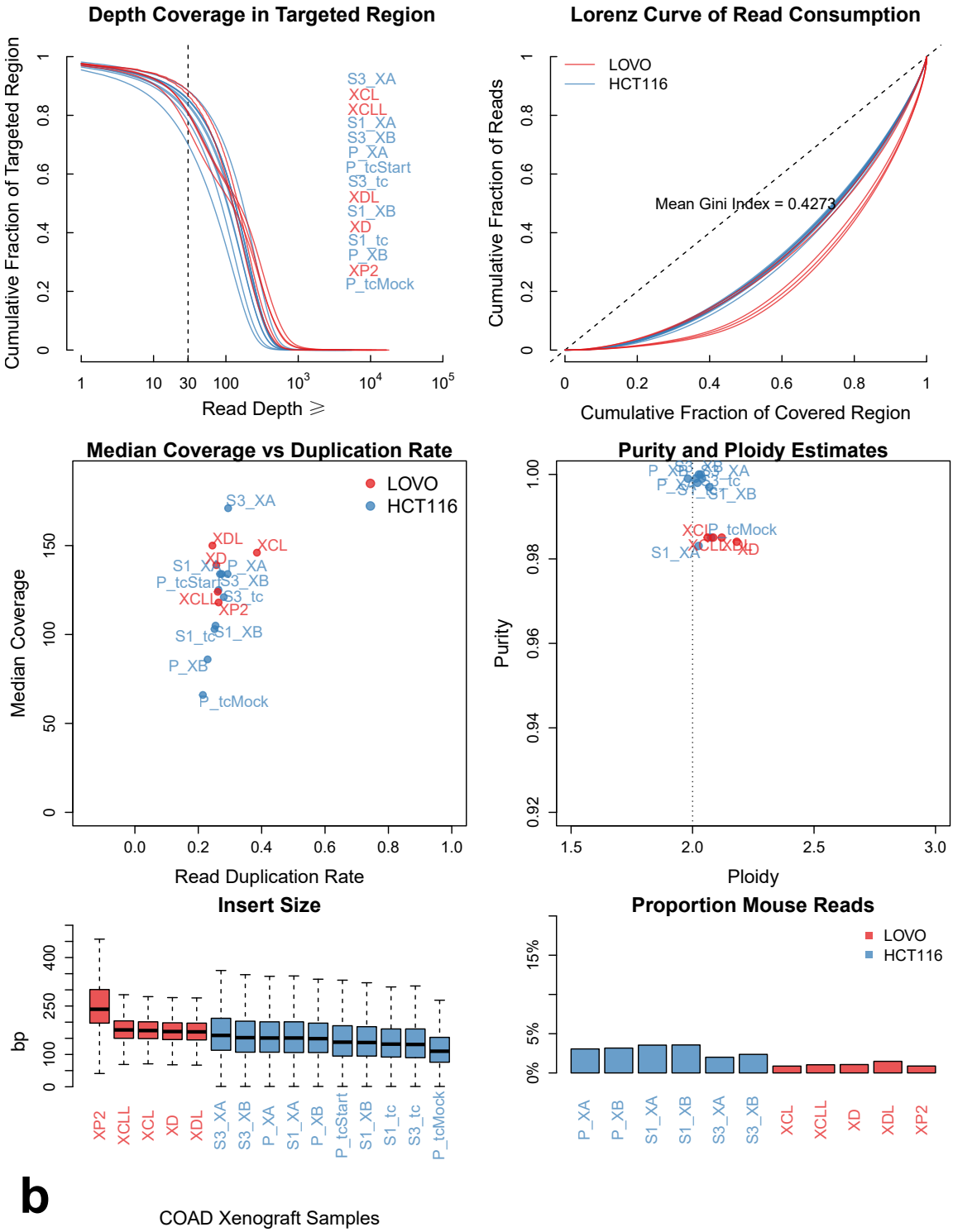
Supplementary Figure 10: Cumulative SFS under different evolutionary modes for a small advantageous mutation rate.

Tumor growth was simulated under a reduced selectively advantageous mutation rate (u_b=10^{-6} per cell division in the whole genome) in order to investigate a scenario where strong advantageous mutations are rare or occur late during tumor growth. This reduced rate is one order of magnitude smaller than that used elsewhere in the text (u_b=10^{-5}) and suggested previously [6]. The cumulative SFS for simulated tumors (100 per mode) was obtained by pooling the VAF (total alternative allele divided by total read depth for pooled regions) across 1, 2, 4 or 8 regions. The cumulative fraction of mutations within a particular frequency (f) range and an upper cutoff of pooled VAF (0.25) are shown. From left to right, five evolutionary modes are shown, i.e., Neutral, Neutral (CSC), s=0.01, s=0.05 and s=0.1. The cumulative SFS is plotted as in **Supplementary Figure 6**.

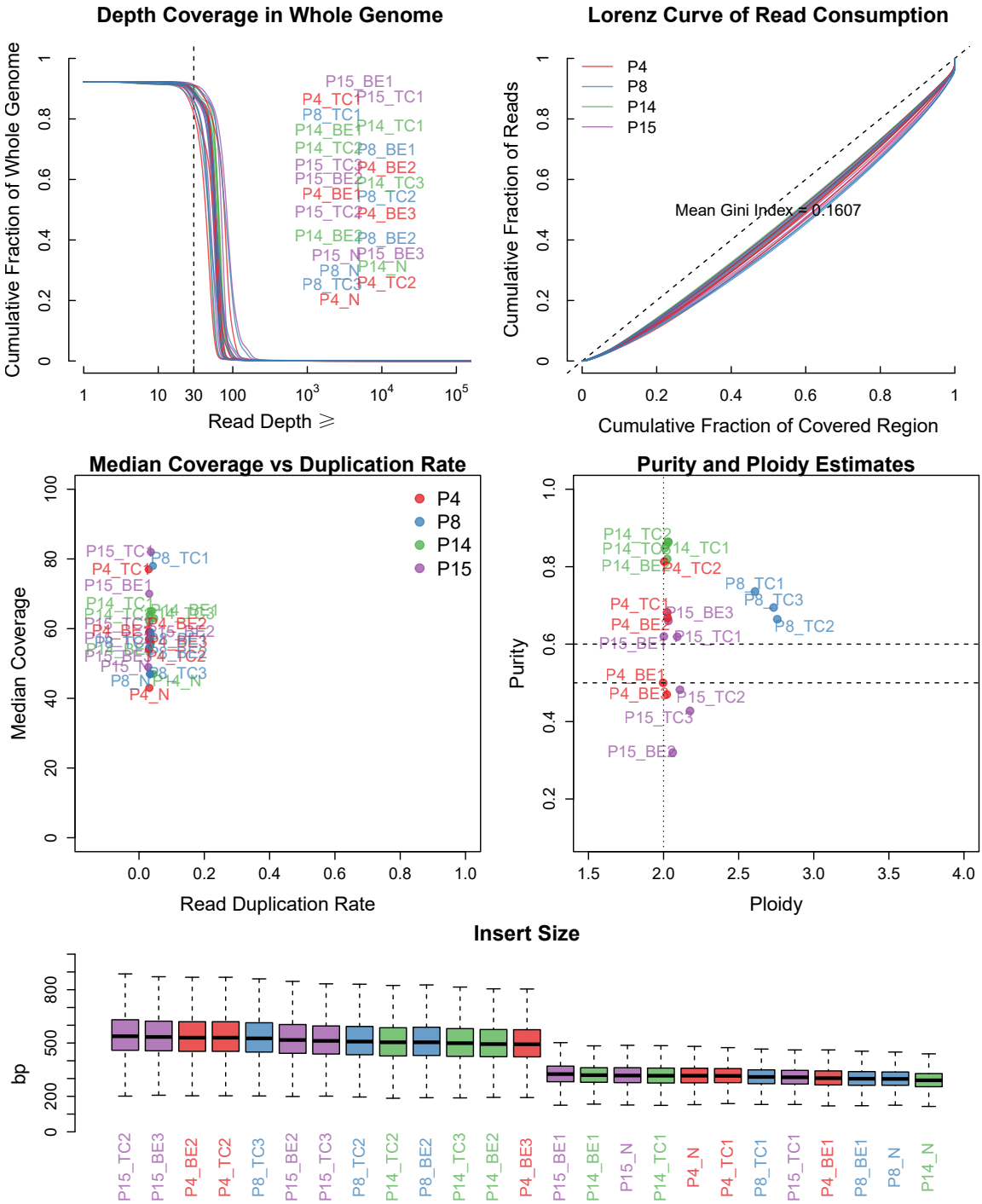


a

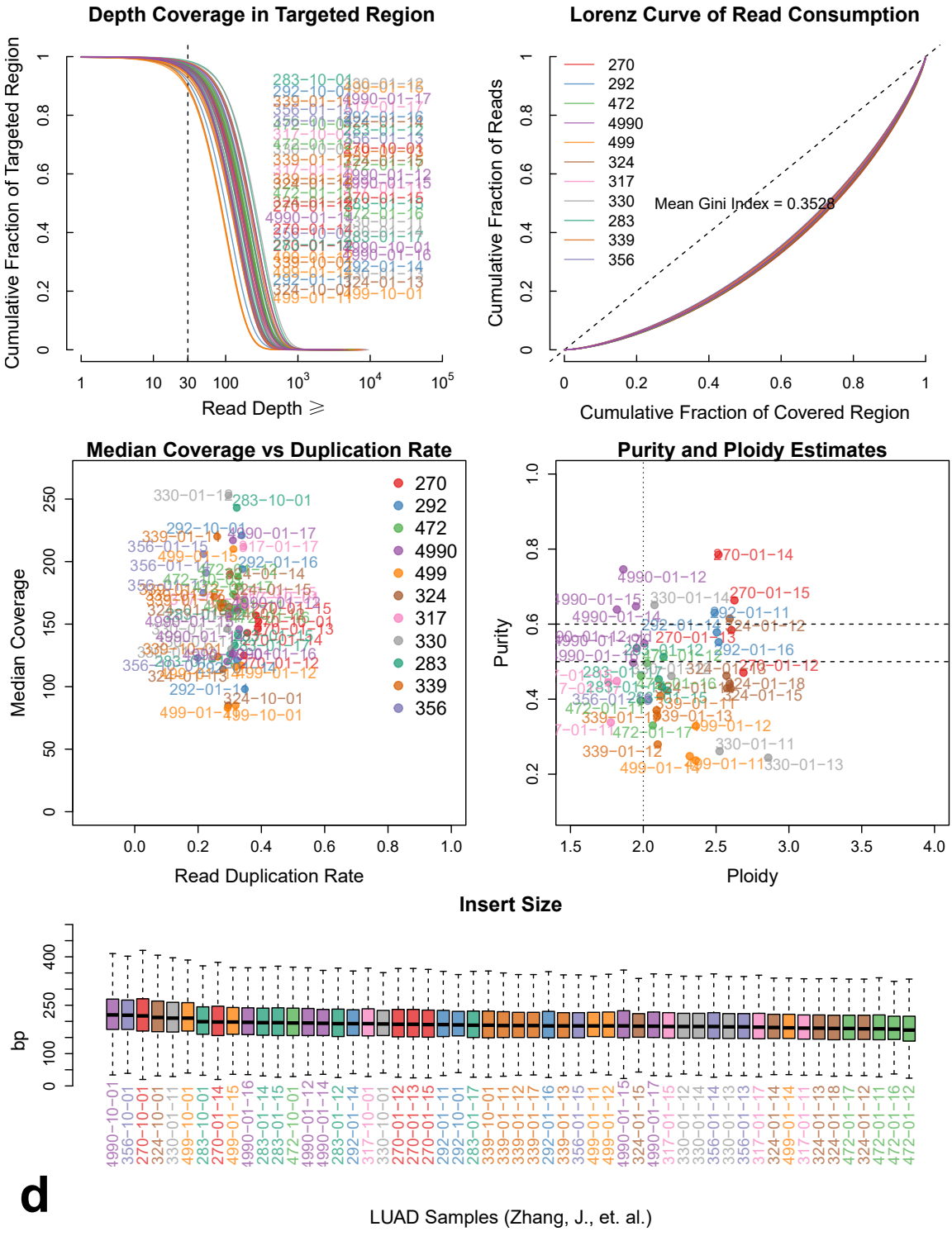
Supplementary Figure 11: Quality control assessment of MRS datasets and sample inclusion for this study.



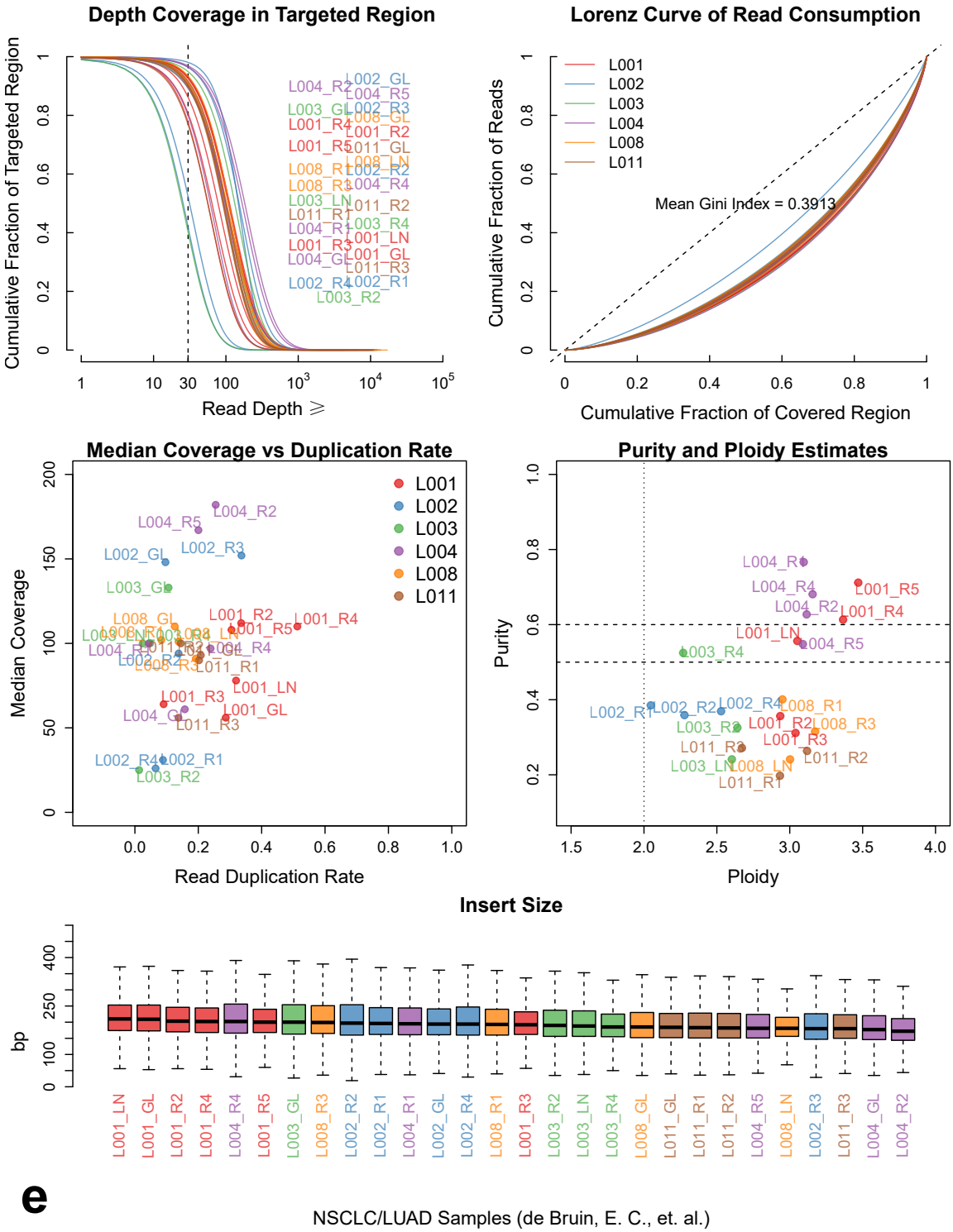
Supplementary Figure 11: Quality control assessment of MRS datasets and sample inclusion for this study.



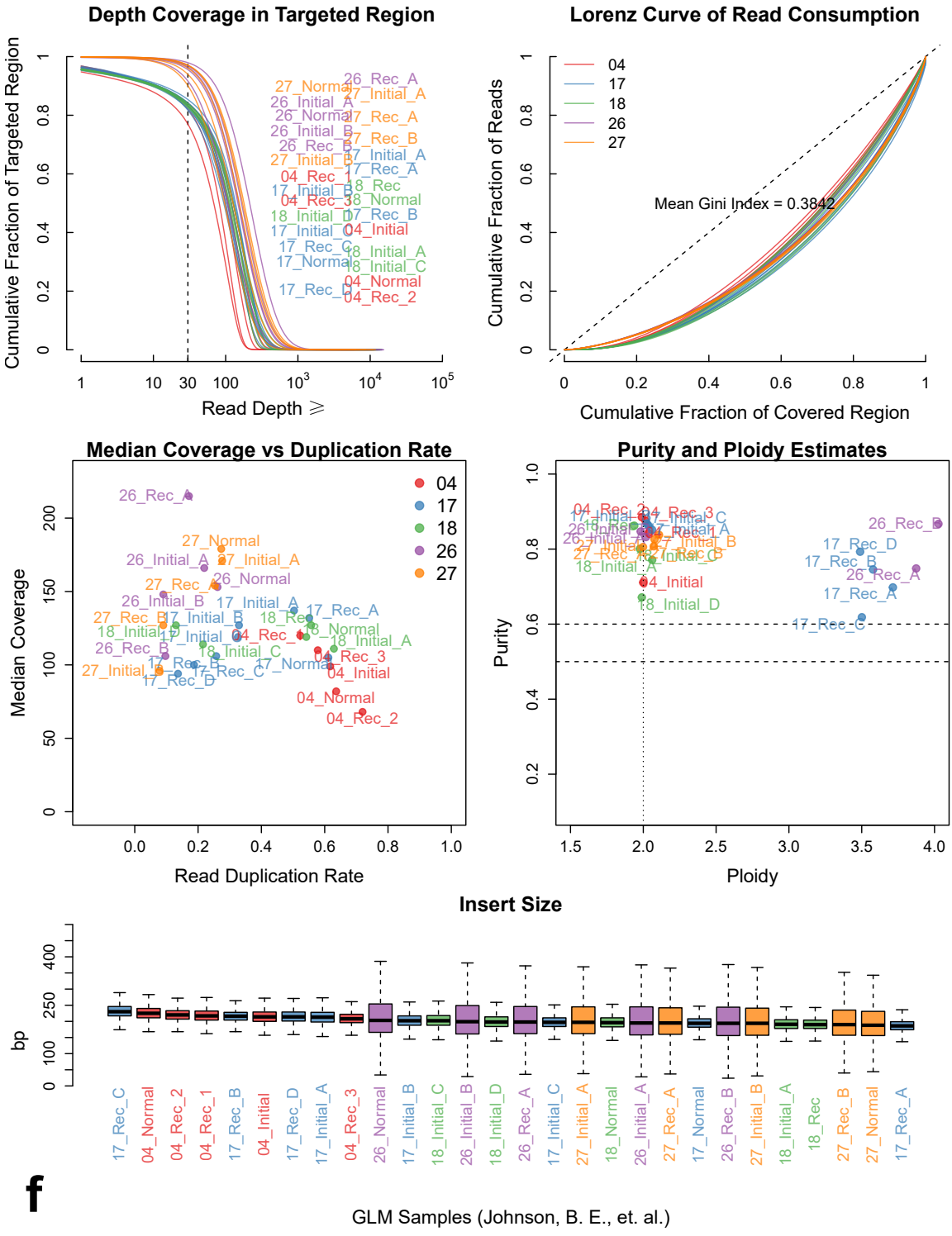
Supplementary Figure 11: Quality control assessment of MRS datasets and sample inclusion for this study.



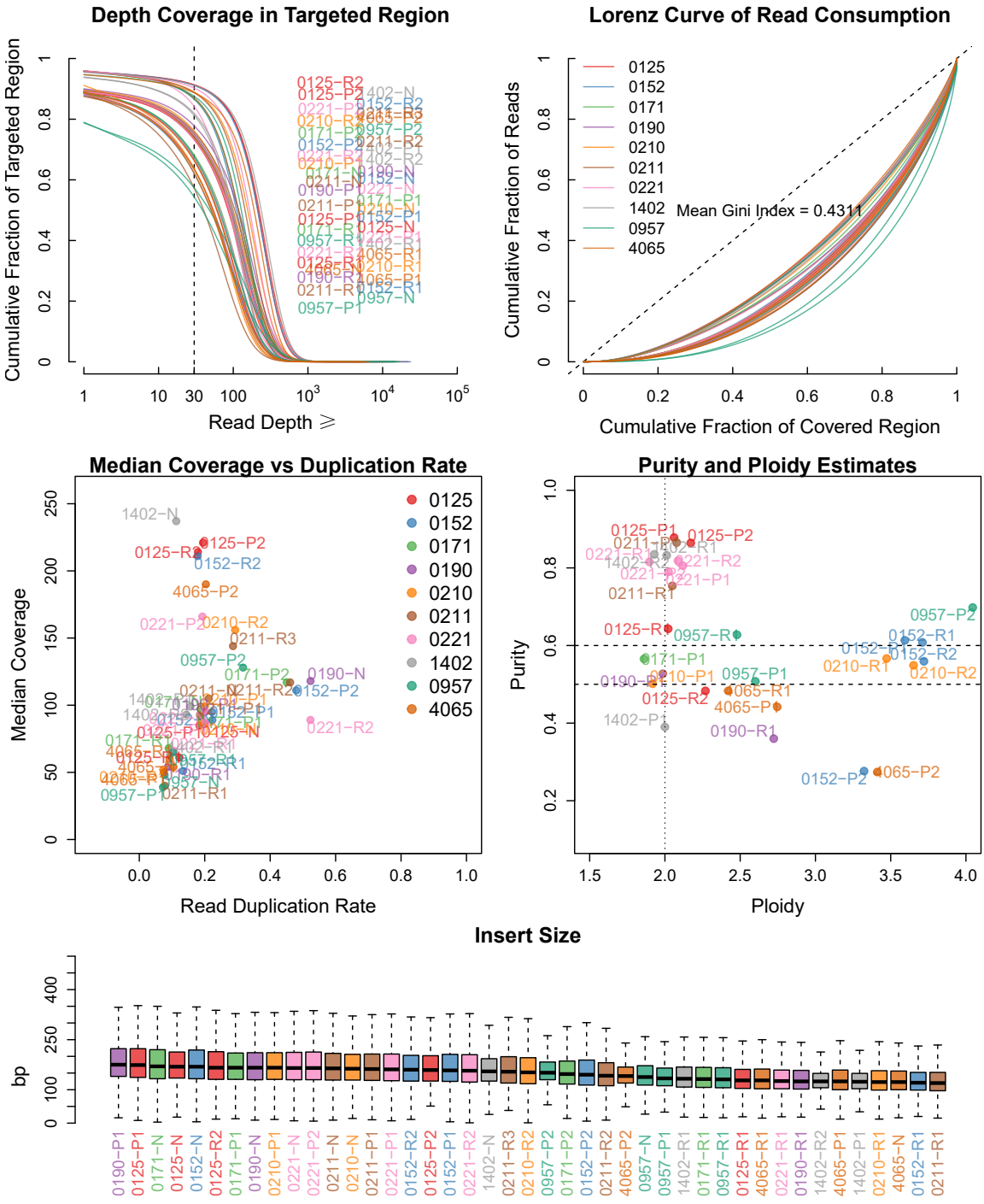
Supplementary Figure 11: Quality control assessment of MRS datasets and sample inclusion for this study.



Supplementary Figure 11: Quality control assessment of MRS datasets and sample inclusion for this study.



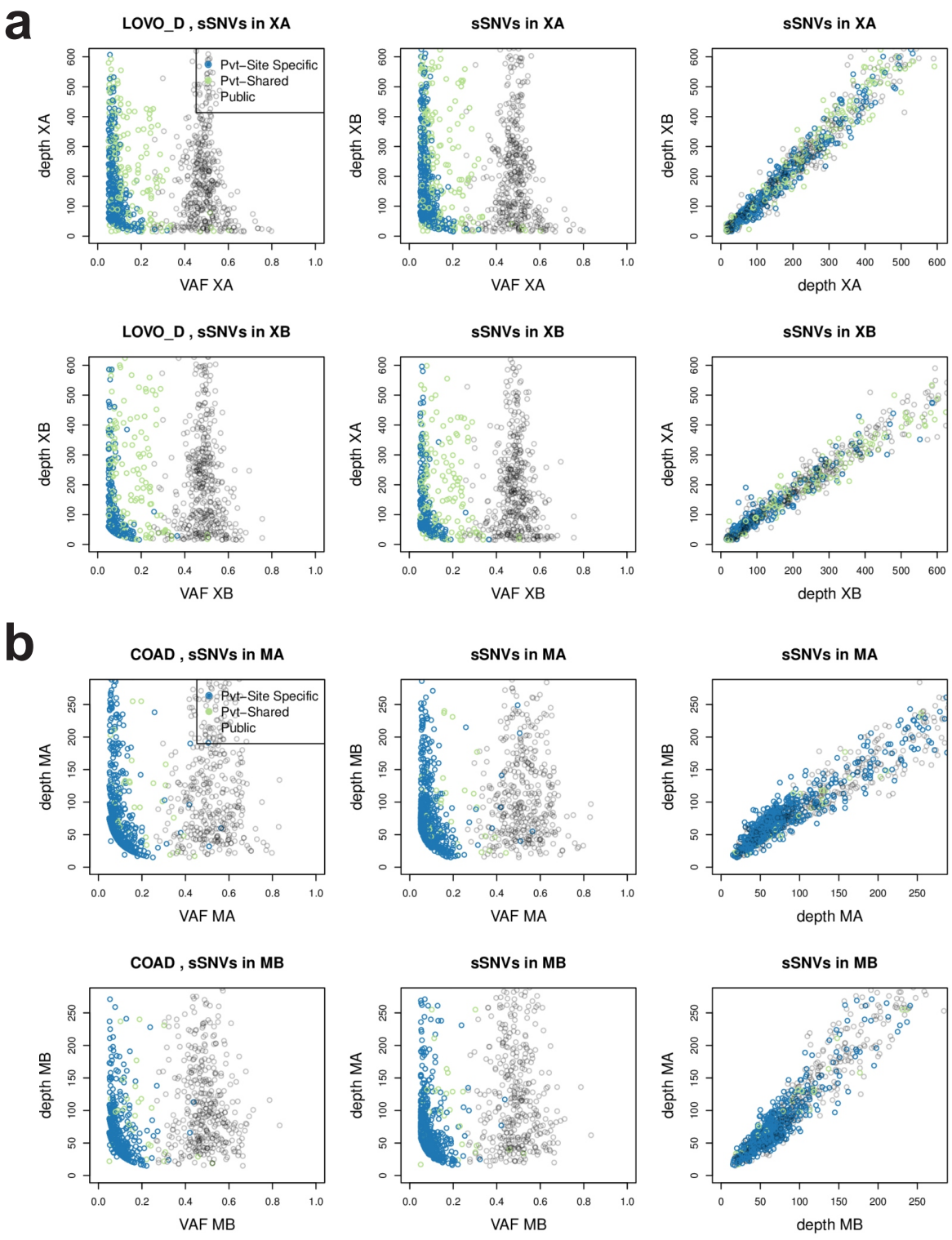
Supplementary Figure 11: Quality control assessment of MRS datasets and sample inclusion for this study.



Supplementary Figure 11: Quality control assessment of MRS datasets and sample inclusion for this study.

Seven MRS datasets were evaluated. These include **(a)** An in-house COAD whole exome MRS dataset (two spatially separated tumor regions (A and B) obtained from opposite sides of carcinomas: G, M, N, O, U, W and an adenoma S) [7] and single gland WES for carcinomas O (OA1, OA2, OA3, OA4 and OB1)

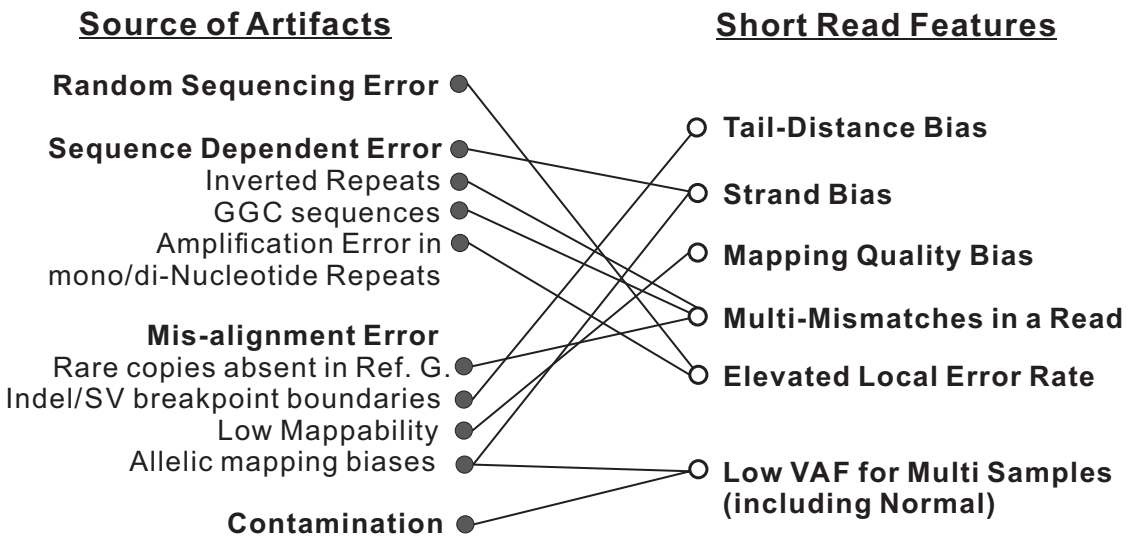
and U (UAL and UBL). **(b)** An in-house xenograft whole exome MRS dataset in which single cell cultures derived from the HCT116 and LoVo cell lines were injected into opposite flanks of a mouse and allowed to form tumors. **(c)** WGS of esophageal adenocarcinomas (ESCA) and paired Barrett's Esophagus (BE) from a subset of patients [8] (European Genome-phenome Archive, or EGA ID: EGAD00001001394). Only cases with MRS of ESCAs (Patient 4, 8, 14 and 15) were included in the analysis. **(d)** Whole exome MRS from lung adenocarcinoma [9] (LUAD, EGA ID: EGAD00001000984). **(e)** Whole exome MRS from non-small-cell-lung-cancer (NSCLC) and LUAD [10] (EGA ID: EGAD00001000900). **(f)** Whole exome MRS from low grade glioma (GLM) with recurrent tumors after treatment [11] (EGA ID: EGAD00001000714). **(g)** Whole exome MRS from glioblastoma (GBM) with post-treatment recurrences [12] (EGA ID: EGAD00001001113). Tumor regions exhibiting hypermutation were excluded. For each dataset, the cumulative faction of targeted regions (or else whole genome for the ESCA dataset) covered at or above a given read depth is shown on the top-left panel. Sample IDs are labelled in order of decreasing fraction of covered targeted regions at or above a depth of 30. The top-right panel illustrates the cumulative read consumption relative to the cumulative fraction of covered regions (Lorenz curve of read consumption) and indicates non-uniform read distributions, as expected in NGS data. The mean Gini coefficient (corresponding to the area between the diagonal line and the Lorenz curve divided by the area under the diagonal) is reported. As expected, the WGS ESCA dataset exhibited more uniform read distributions than the WES datasets. The middle-left panel shows the median sequencing depth versus the overall read duplication rate for each sample. The middle-right panel shows tumor purity relative to ploidy, where both values were estimated based on TitanCNA [4]. The insert size distribution for paired-end reads is shown in the bottom panel. For the xenograft samples in **(b)**, the proportion of reads mappable to the mouse genome is presented in the bottom-right panel. To enable the reliable adjustment of VAF and subsequent calculation of statistics from the SFS, we set stringent QC requirements for the inclusion of tumor samples. 1) estimated purity >0.5; 2) median sequencing depth >50; 3) number of subclonal SSNVs >50. We only inspected tumors for which multiple (≤ 2) tumor regions were available and met the above three requirements.



Supplementary Figure 12: Sequencing depth of SSNVs in paired regions from representative tumors.

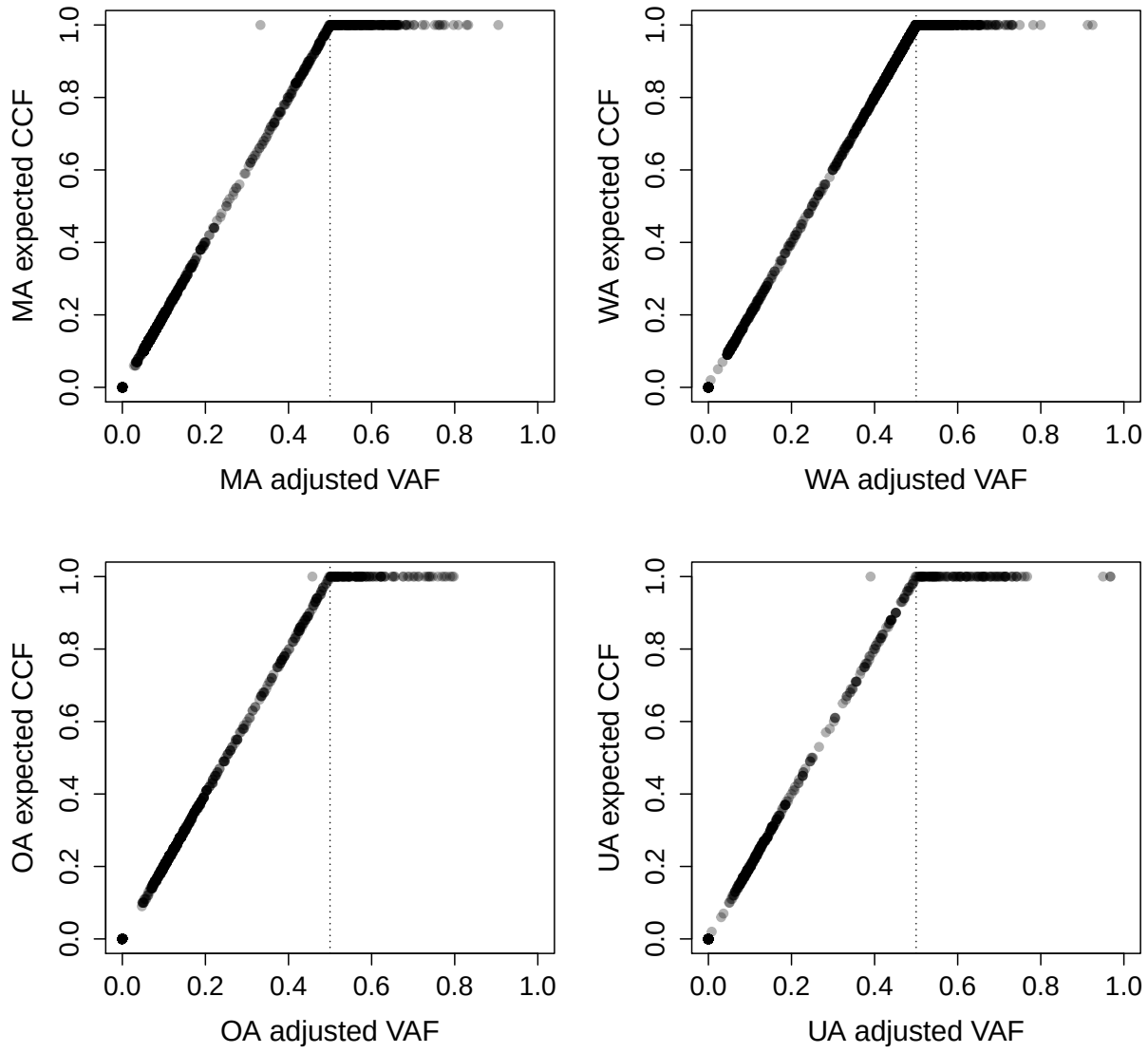
Scatterplots show the read depth for SSNVs identified as public (gray), private but shared (Pvt-Shared

in green) and region-specific (blue) for two tumor regions from representative tumors, namely **(a)** LoVo xenograft D (XA versus XB) and **(b)** COAD-M (MA versus MB). The depth (in both regions) at the genomic coordinates for each SSNV is plotted against the adjusted VAF in the region in which it was detected (left and middle panels). The depth at each SSNV for the two regions was also compared (right panel). These plots illustrate several points: 1) for SSNVs with VAF below 0.2, the depth increases as the VAF decreases, attributable to the LOD calculation in the variant assurance pipeline (VAP, Supplementary Note); 2) the depth for region-specific SSNVs in the two tumor regions is comparable due to their conservative definition (Methods); 3) the depth from two tumor regions is correlated, an intrinsic property of WES data, presumably due to sequence dependent exon capture rates.



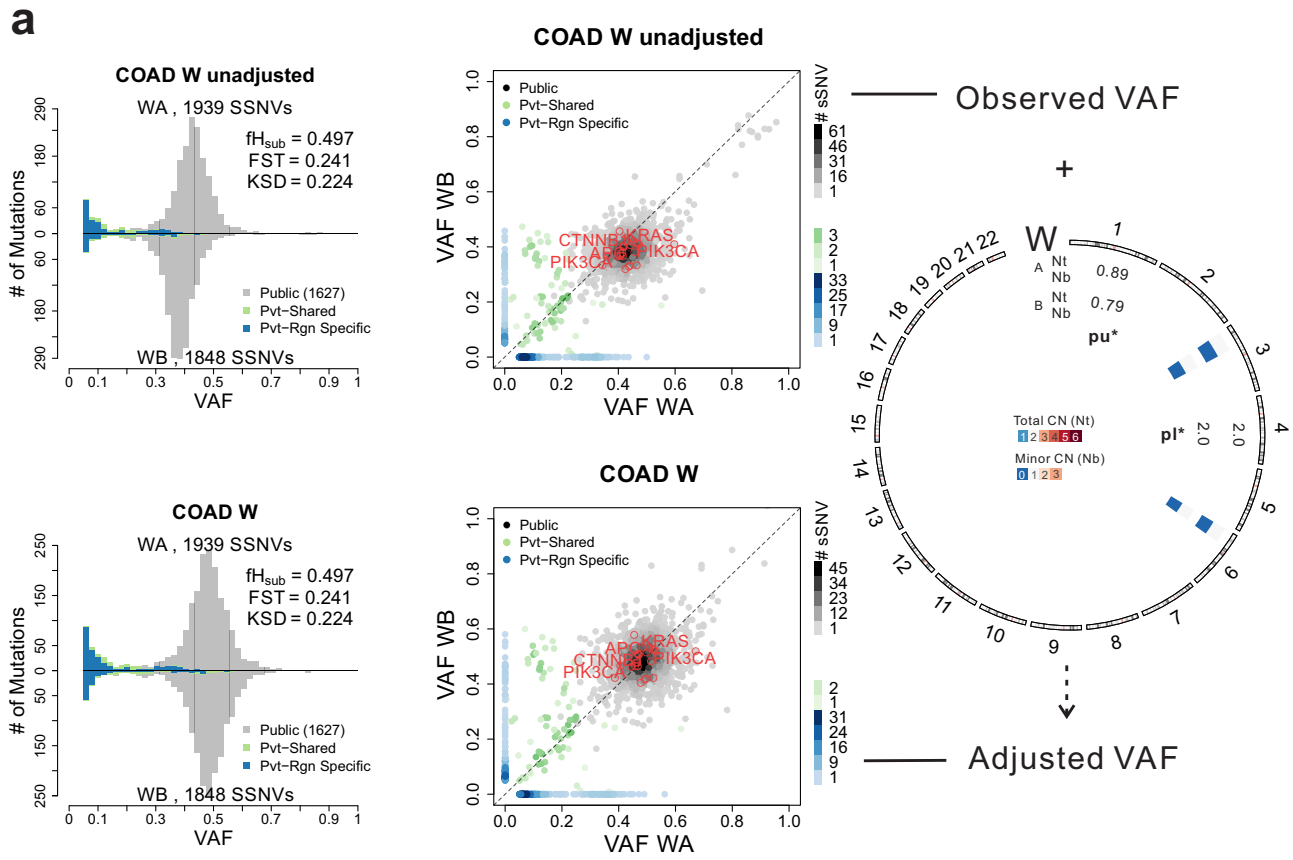
Supplementary Figure 13: Features derived from short read alignments reflect sequencing artifacts.

Sequencing artifacts are reflected in the alignment features of short reads, providing a rationale for their use in flagging false positive SSNVs.

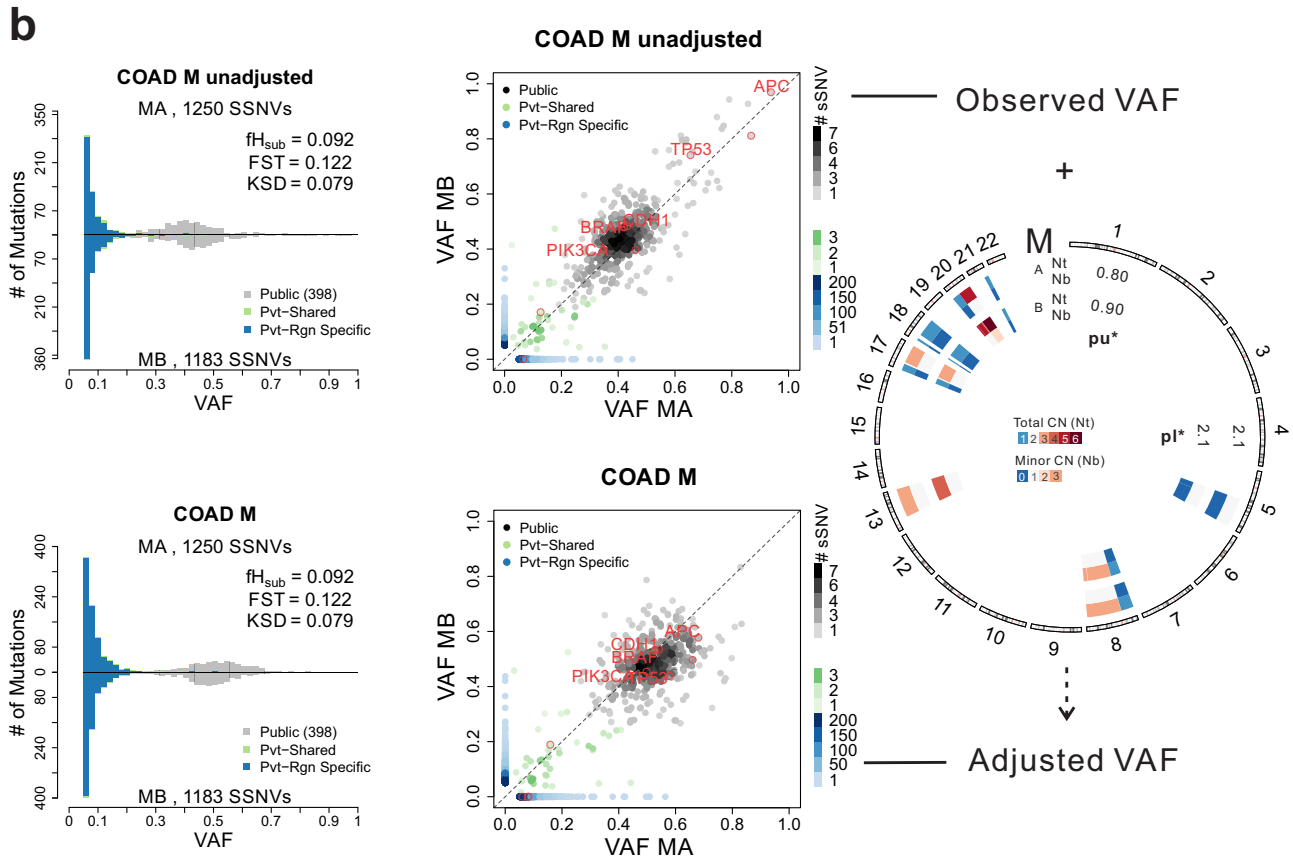


Supplementary Figure 14: Relationship between the adjusted VAF and the CCF.

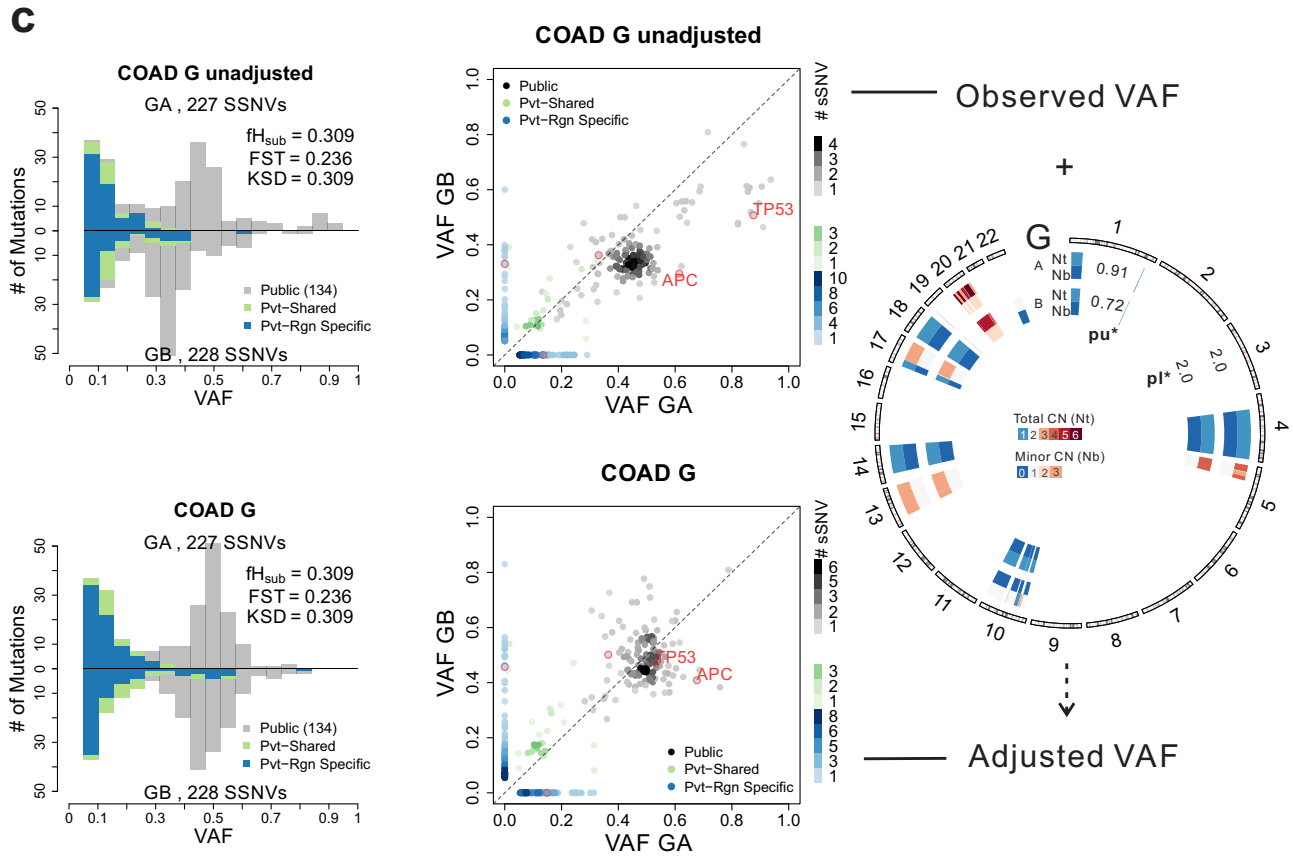
The relationship between the adjusted VAF (VAF_a) and CCF (cancer cell fraction) is illustrated with data from several representative tumors. The VAF_a can be than greater 0.5 due to the inherent noise associated with sequencing. The expected CCF, which assumes a background distribution such as a binomial, collapses to 1 when VAF_a is above 0.5, whereas VAF_a is identical to a half the CCF for values under 0.5.



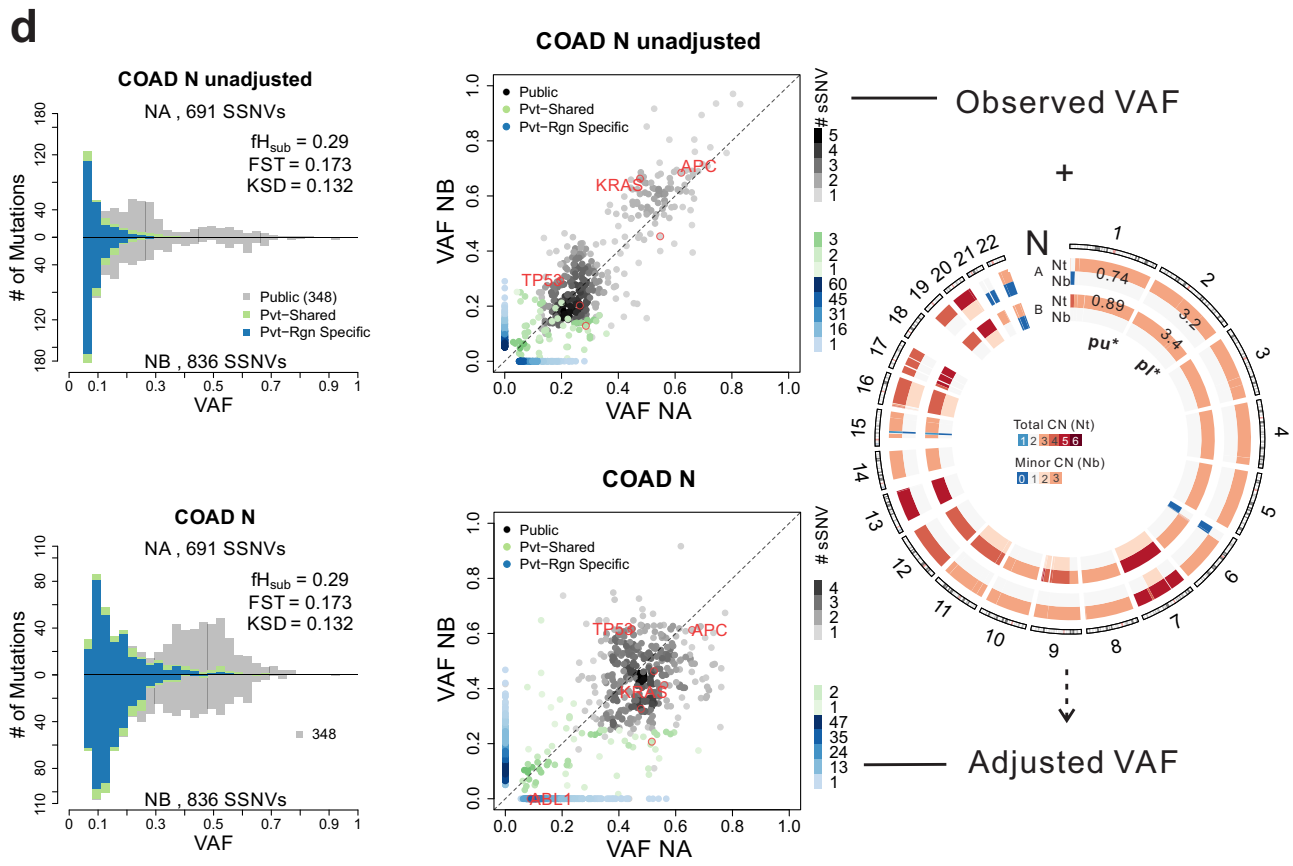
Supplementary Figure 15: Illustration of VAF adjustment on colorectal tumors.



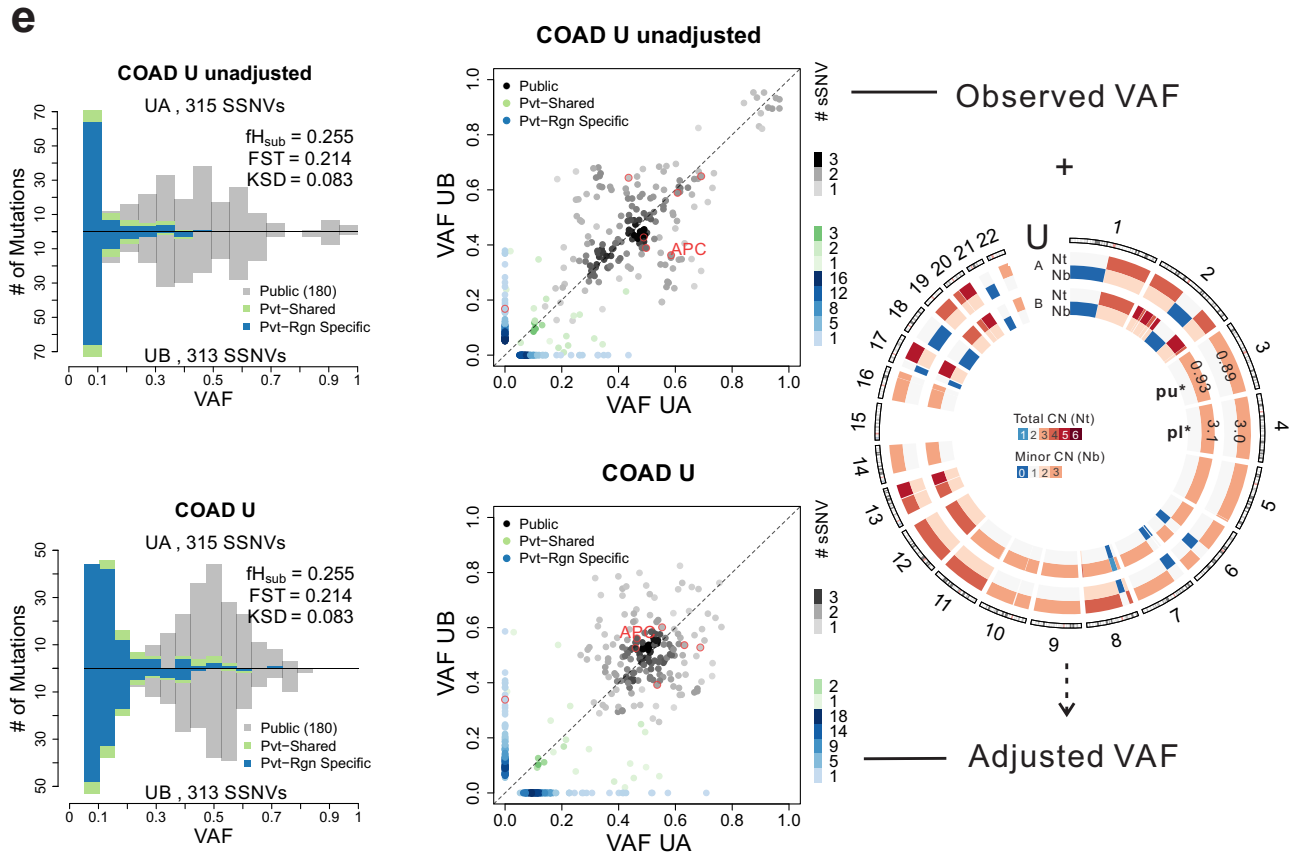
Supplementary Figure 15: Illustration of VAF adjustment on colorectal tumors.



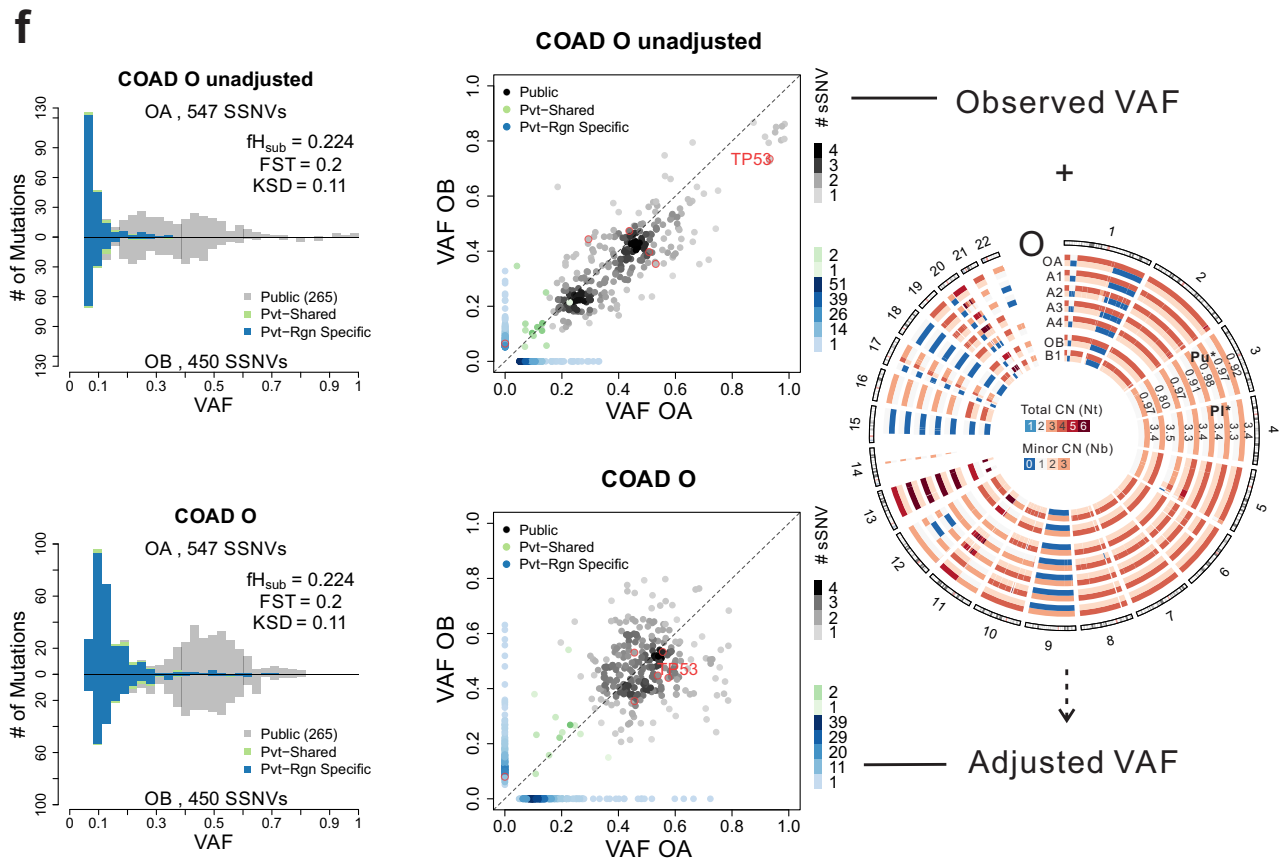
Supplementary Figure 15: Illustration of VAF adjustment on colorectal tumors.



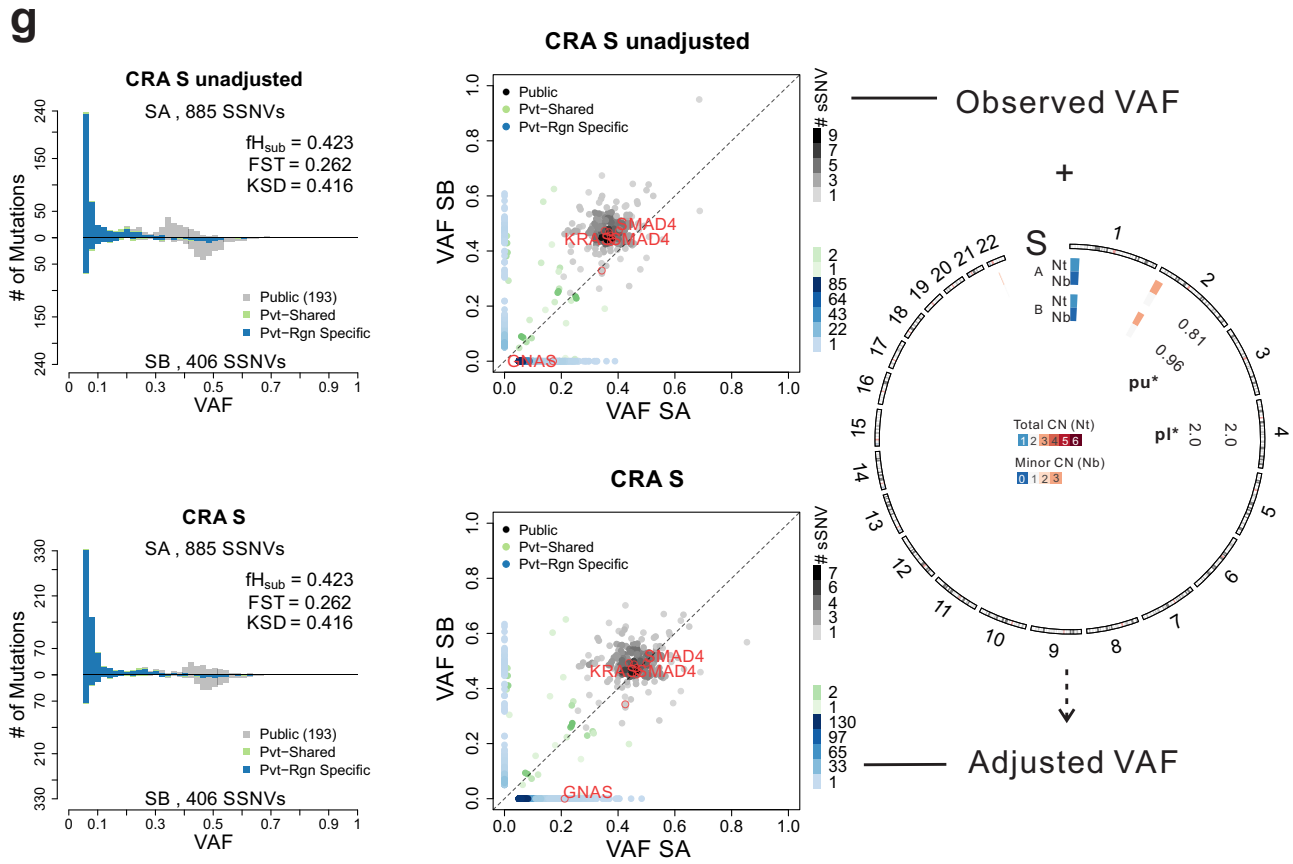
Supplementary Figure 15: Illustration of VAF adjustment on colorectal tumors.



Supplementary Figure 15: Illustration of VAF adjustment on colorectal tumors.

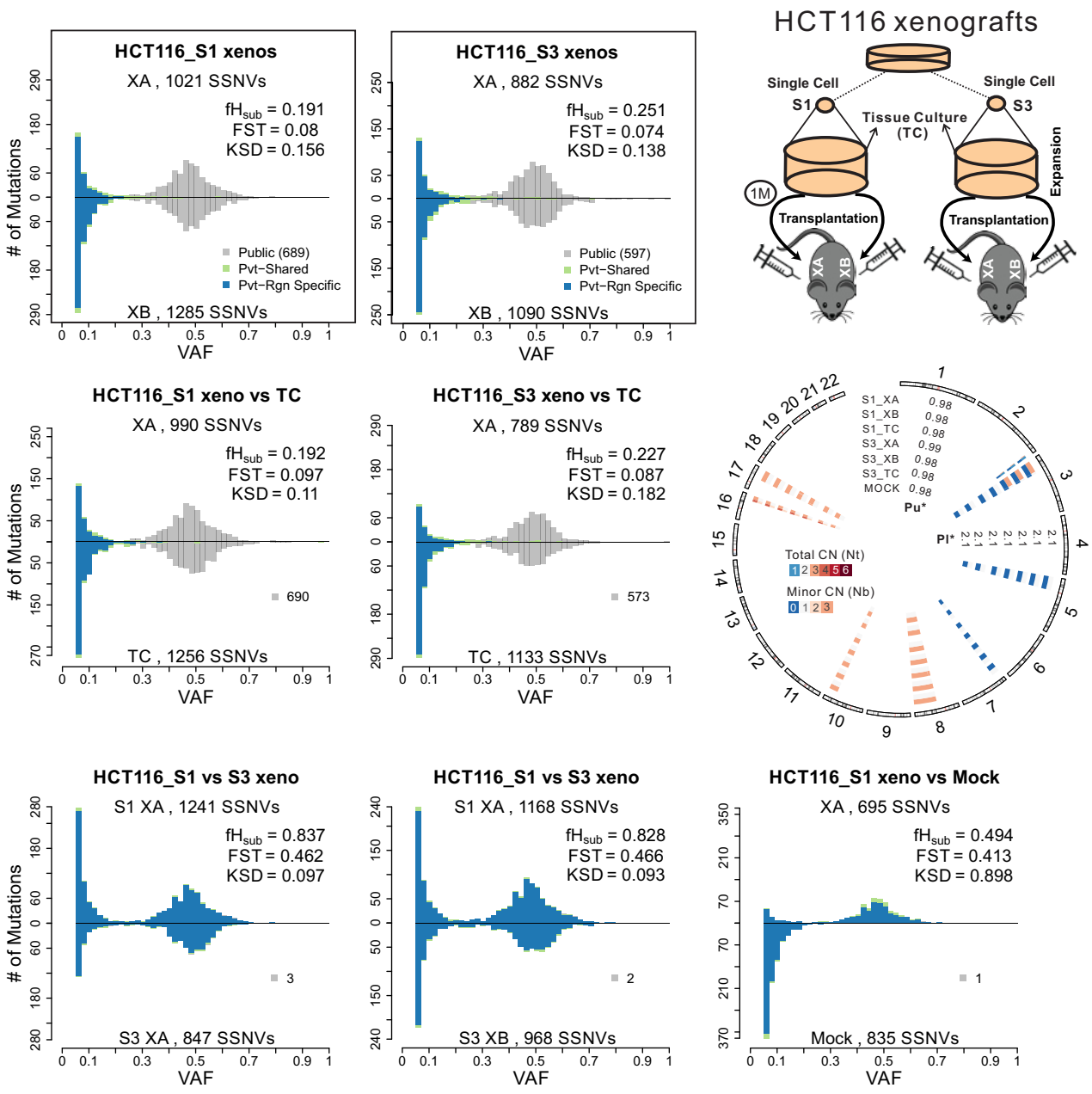


Supplementary Figure 15: Illustration of VAF adjustment on colorectal tumors.



Supplementary Figure 15: Illustration of VAF adjustment on colorectal tumors.

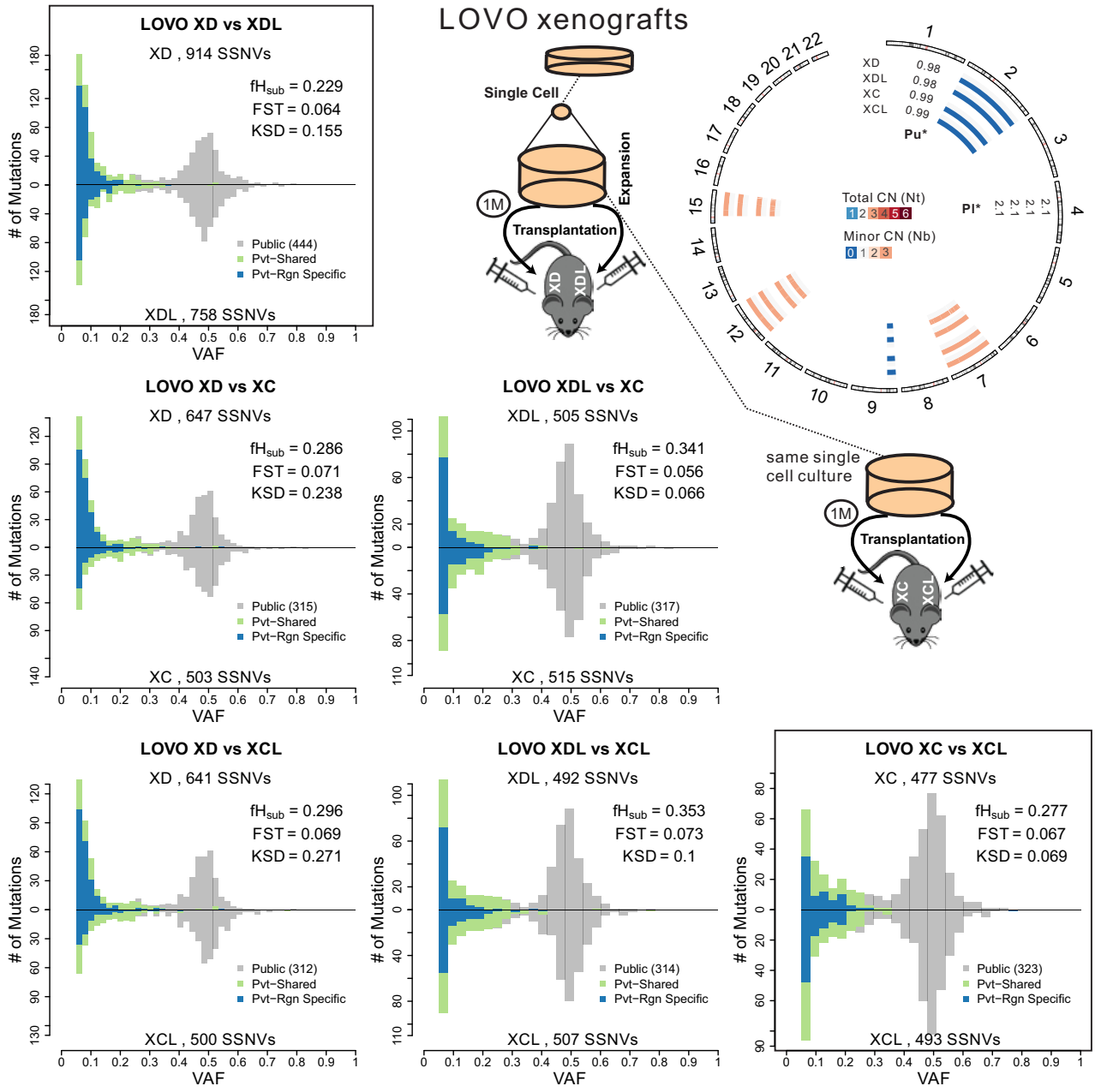
Histograms and two-way scatter plots are shown for each COAD before (upper panel) and after (lower panel) adjusting the VAF estimates for tumor purity and local copy numbers (Supplementary Note). The circos plot (right panel) summarizes the CNA profiles across tumor regions. In the pairwise histograms, the number of SSNVs detected at a given VAF for the two tumor regions is shown above and below the x-axis to enable visual comparisons. Mutations are grouped into Public (gray bars), Private (Pvt)-shared (green) and Private-region specific (blue) events (Methods). A scatterplot of the pairwise VAF is shown in the middle panel to enable comparisons of the SSNVs present in each lesion at a given VAF. The color scale bar indicates the number of SSNVs in a square (0.02 on a side) region centered on each SSNV for each group of alteration. Nonsilent SSNVs including both known and predicted COAD drivers (IntOgen version 2016.5) are indicated by red circles with known drivers labeled. The circos plot illustrates the predicted absolute total CN (Nt) and minor allele CN (Nb) for each tumor region. Wild-type diploid segments are indicated by white for Nt (two copy) and Nb (one copy), whereas segments with copy number gain and loss are shown in red and blue, respectively. Purity and ploidy estimates are labeled in the corresponding concentric rings. The shift in the public mutation cluster below 0.5 due to impurity is particularly evident in near diploid tumors **(a)** COAD-W, **(b)** COAD-M and **(c)** COAD-G and **(g)** CRA-S (colorectal adenoma), whereas in triploid tumors: **(d)** COAD-N; **(e)** COAD-U; **(f)** COAD-O, public mutations exhibit a greater range of values. However, in both cases robust adjustment is achieved by accounting for purity and local copy number.



Supplementary Figure 16: The SFS histogram in HCT116 xenografts.

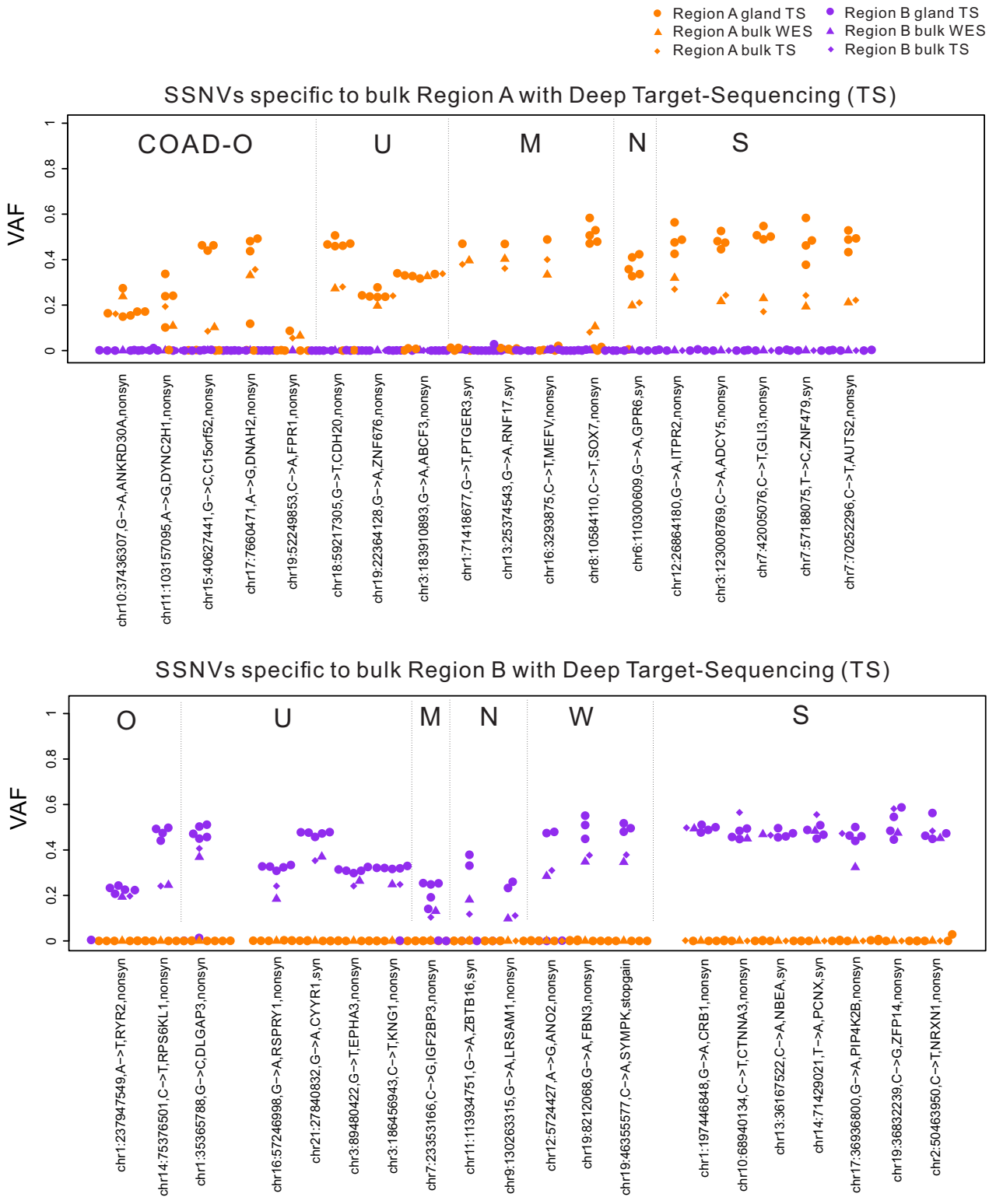
HCT116 cells were expanded *in vitro* and two single cells (S1 and S3) from this population were cloned and expanded prior to transplantation into the right and left flanks of two NSG mice, as illustrated in the schema. Xenografted tumors were allowed to develop to a size of $\sim$1 billion cells (1 cm^3) prior to sampling them and subjecting them to WES. Polyclonal tissue culture of the starting HCT116 cell line was employed as a control for copy number analysis and somatic mutation calling. A summary of the copy number profiles across samples is shown in a circos plot. WES of the two tumors (XA and XB) from each of two single cell (S1 and S3) cultures consistently resulted in a bimodal SFS histogram that lacked enrichment for high frequency private SSNVs (primary comparisons of interest shown in boxes). In contrast, two single cells cultures had distinct lineages with completely different clonal SSNVs (S1 versus S3, bottom panel). In addition, tissue culture (TC) of cells sampled at the time of injection had similar patterns to their corresponding xenografts

(xeno versus TC, middle panel), suggesting a lack of stringent selection in the expansion of the xenografted tumors.



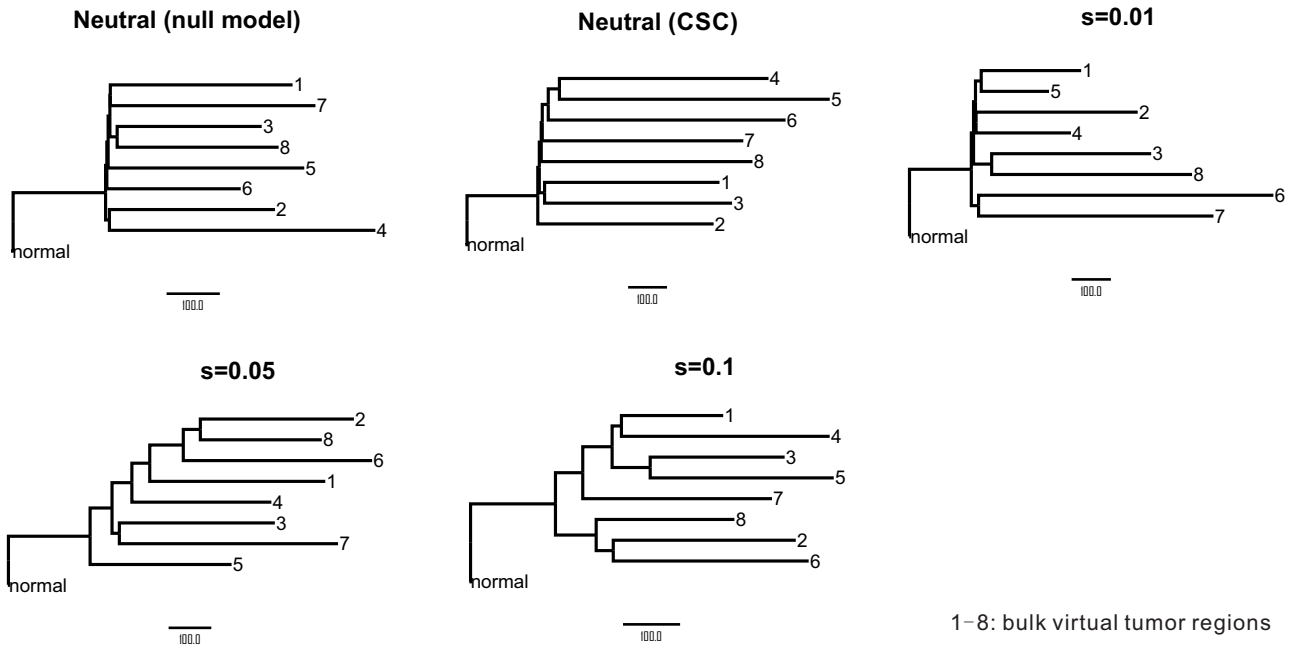
Supplementary Figure 17: The SFS histogram in LoVo xenografts.

LoVo cells were expanded *in vitro* and a single cell from this population was cloned and expanded prior to transplantation into the right and left flanks of a SCID mouse, as illustrated in the schema. WES of the two tumors resulted in a bimodal SFS histogram that lacked enrichment for high frequency private SSNVs (primary comparisons of interest shown in boxes). Replicate xenografts from the same single cell LoVo culture showed similar patterns in the left and right tumors (XD versus XDL, XC versus XCL) from a given mouse as well as between mice (XD/L versus XC/L). Circos plot summarizing the CNA profiles from different samples is shown. Of note, SSNVS in the low VAF range were mostly unique to each injection in the HCT116 xenografts, whereas for the LoVo xenografts, the fraction of shared private SSNVs was larger. This may be due to a greater reduction in the initial population size during HCT116 xenograft initiation leading to fewer shared private mutations.



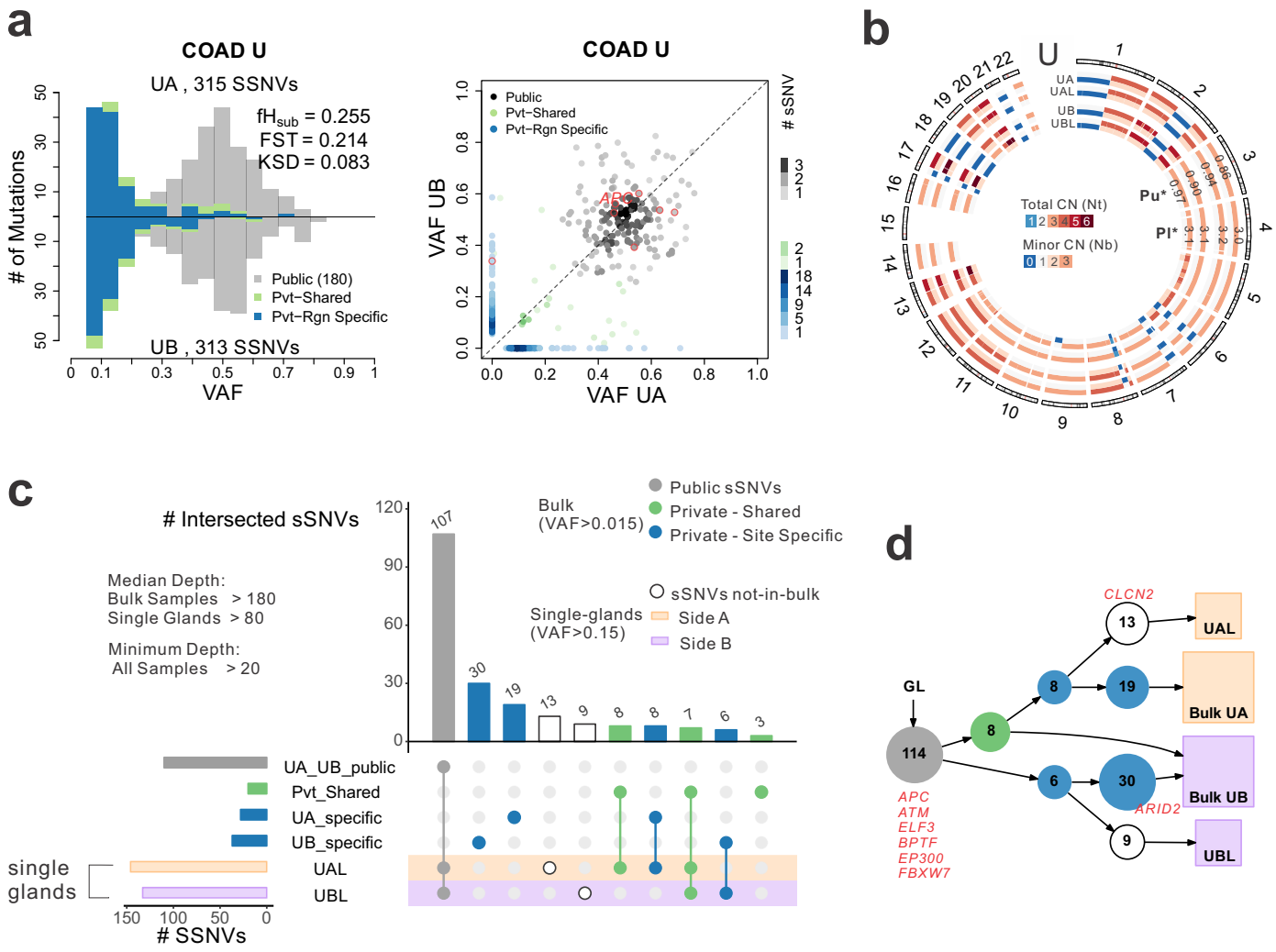
Supplementary Figure 18: Targeted deep sequencing of single glands reveals spatial constraints amongst region-specific SSNVs in bulk tumor samples.

WES was performed on each of two bulk regions (sampled from opposite tumor sides >3cm apart) per tumor, followed by targeted deep sequencing of a custom mutational panel in single glands. The VAF for region-specific SSNVs that were included in the targeted panel is shown and color-coded according to the tumor region (A - orange, B - purple). All such region specific mutations found in the bulk region sample were only detected in the single glands from that region, consistent with the spatial constraints observed in the single gland WES of Tumor O (**Figure 4**) and U (**Supplementary Figure 20**). The depth for single glands from regions A and B were not significantly different (p = 0.67, Wilcoxon rank sum test, mean depth >900 for both tumor sides).



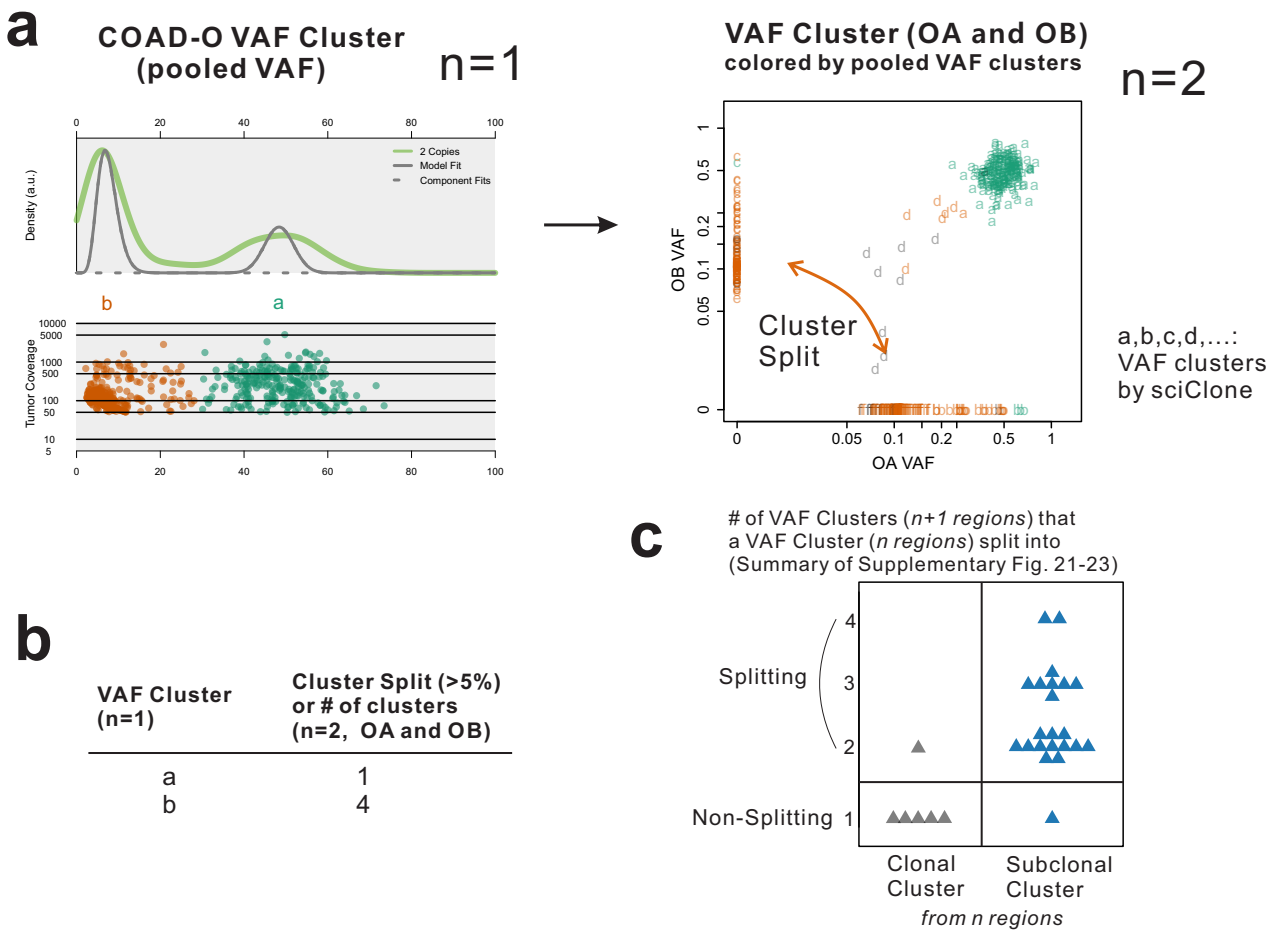
Supplementary Figure 19: Representative phylogeny for virtual tumors simulated under different evolutionary modes.

8 bulk virtual tumor regions (each composed of $\sim 10^6$ cells) sampled respectively from the 8 quadrants of a virtual tumor ($\sim 10^9$ cells) were evaluated for each of the five evolutionary modes (neutral, neutral CSC, s=0.01, s=0.05 and s=0.1). The parameters used in the simulation are detailed in **Supplementary Figure 1**, **Supplementary Table 1** and the Methods. Mutations were called if they had ≥ 15X sequencing depth and ≥ 3 variant reads. The neighbor-joining method [13] was used to reconstruct the phylogenetic relationship amongst the 8 regions of each tumor based on mutational presence/absence. The resulting representative phylogenetic trees highlight the expected star-like phylogenies under effectively neutral growth. They also highlight the challenges in constructing an accurate tree from a single bulk sample.



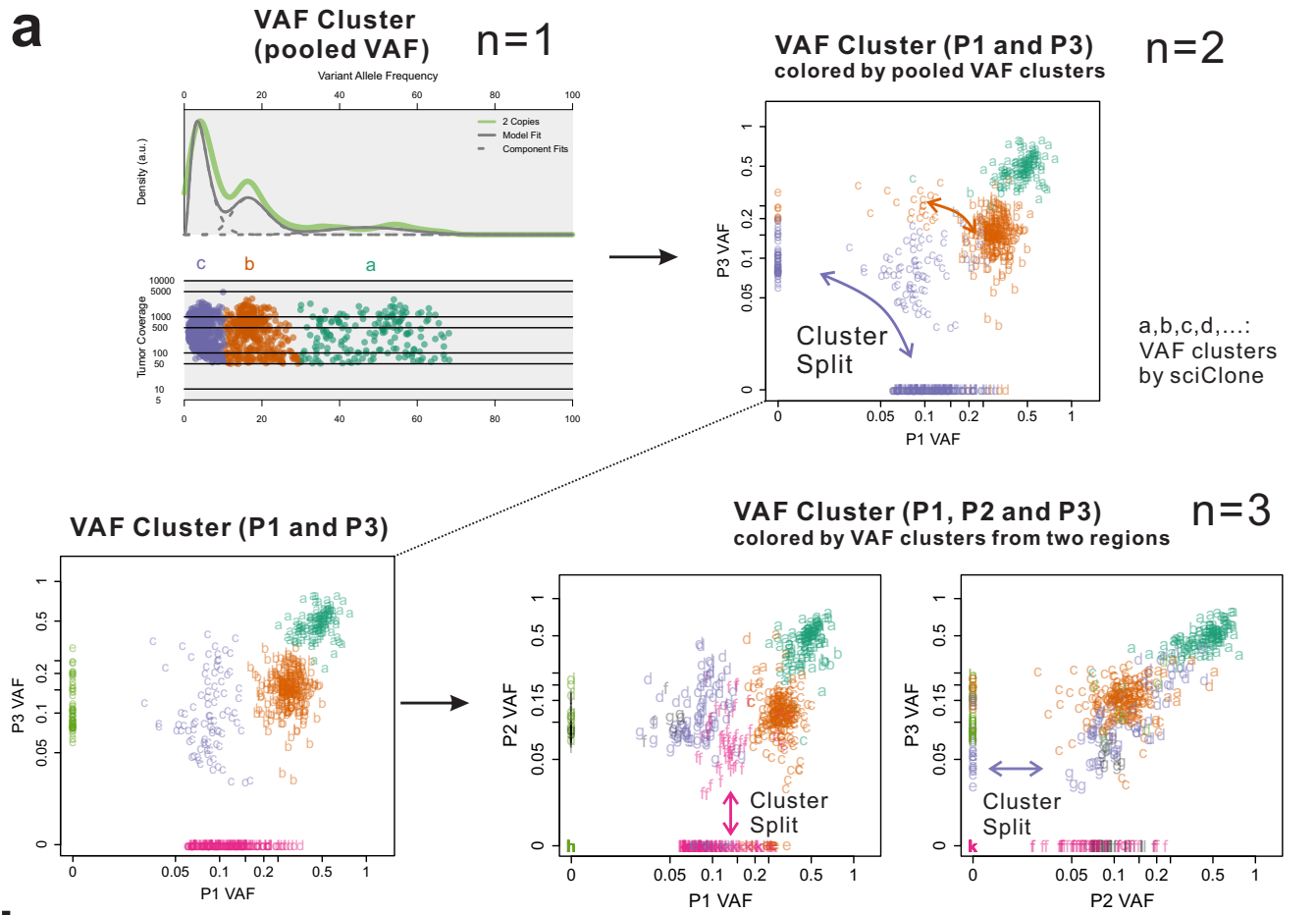
Supplementary Figure 20: Single-gland sequencing in COAD-U reveals spatial constraints amongst subclonal SSNVs.

(a) Pairwise histograms and scatterplots of the SFS for SSNVs found in two bulk regions from tumor COAD-U (UA and UB). **(b)** CNA profiles for the bulk samples and two single glands (UAL and UBL) are shown in circos plot. Purity (Pu) and Ploidy (Pl) estimates for each sample are labeled on the corresponding concentric rings. **(c)** Intersection plot based on the UpSetR package[14] for SSNVs found in bulk and single glands includes mutations that are i) covered by at least 20 reads in each of the samples; ii) with a VAF above 1.5% in the bulk sample or above 15% in the single-glands and iii) not from regions with varying LOH patterns among samples. Gray, green and blue bars/points are for those mutations classified as public, private-shared and region-specific in the comparison between two bulk samples (Methods). Orange and purple bars/shades in the matrix indicate single glands from tumor sides A and B, respectively. **(d)** SSNV presence/absence tree constructed using LICHeE. The bulk sample and single glands from the same tumor side are in the same lineage, suggesting the presence of spatial constraints during tumor expansion. SSNVs in known driver genes are labeled in red.



Supplementary Figure 21: VAF clusters do not correspond to unique subclones (COAD-O).

SciClone [15] (default settings) was employed to cluster VAF values for COAD-O, taking as input the pooled VAF from two regions (OA, OB), which served as a composite for clustering (n=1). **(a)** In the composite one sample setting, two VAF clusters were identified (a and b; left panel). However, the low VAF cluster b separated into 4 distinct clusters when two regions were included in the clustering (scatterplots, VAF shown on log2 scale). SSNVs in the green clonal VAF cluster centered at 0.5 remain clustered (persisted) when two regions were included. Only SSNVs with depth >30 across all regions and VAF >0.05 in the detected regions were considered. **(b)** The number of distinct VAF clusters (containing >5% of SSNVs in cluster X inferred from n regions) when an additional region was included for clustering (n+1) was used to determine the number of "cluster splits". A cluster split of 4 was observed for the orange cluster b, indicating that the VAF cluster observed for n=1 sample does not correspond to a unique subclone. **(c)** Summary of the SciClone [15] analysis. Subclonal clusters derived from 'n' regions separate into additional clusters (with more than 5% SSNVs in the original cluster) when sequencing data from 'n+1' tumor regions are employed in analysis. Whereas clonal clusters were persistent (SSNVs remained grouped) upon the inclusion of data from another region, subclonal clusters generally did not and this difference was significant across tumors (p<0.001, Fisher's exact test). Of note, one tumor (ESCA-8, **Figure 5b**, **Supplementary Figure 23**) has a group of subclonal mutations approaching fixation in two out of the three regions, and appeared 'clonal' using the pooled VAF, but split upon the introduction of a third tumor region in the analysis.



b

VAF Cluster (n=1)	Cluster Split (>5%) or # of clusters (n=2, P1 and P3)
a	1
b	2
c	3

VAF Cluster in two regions (n=2, P1 and P3)	Cluster Split (>5%) in three regions (n=3, P1, P2 and P3)
a	1
b	2
c	3
d	2
e	3

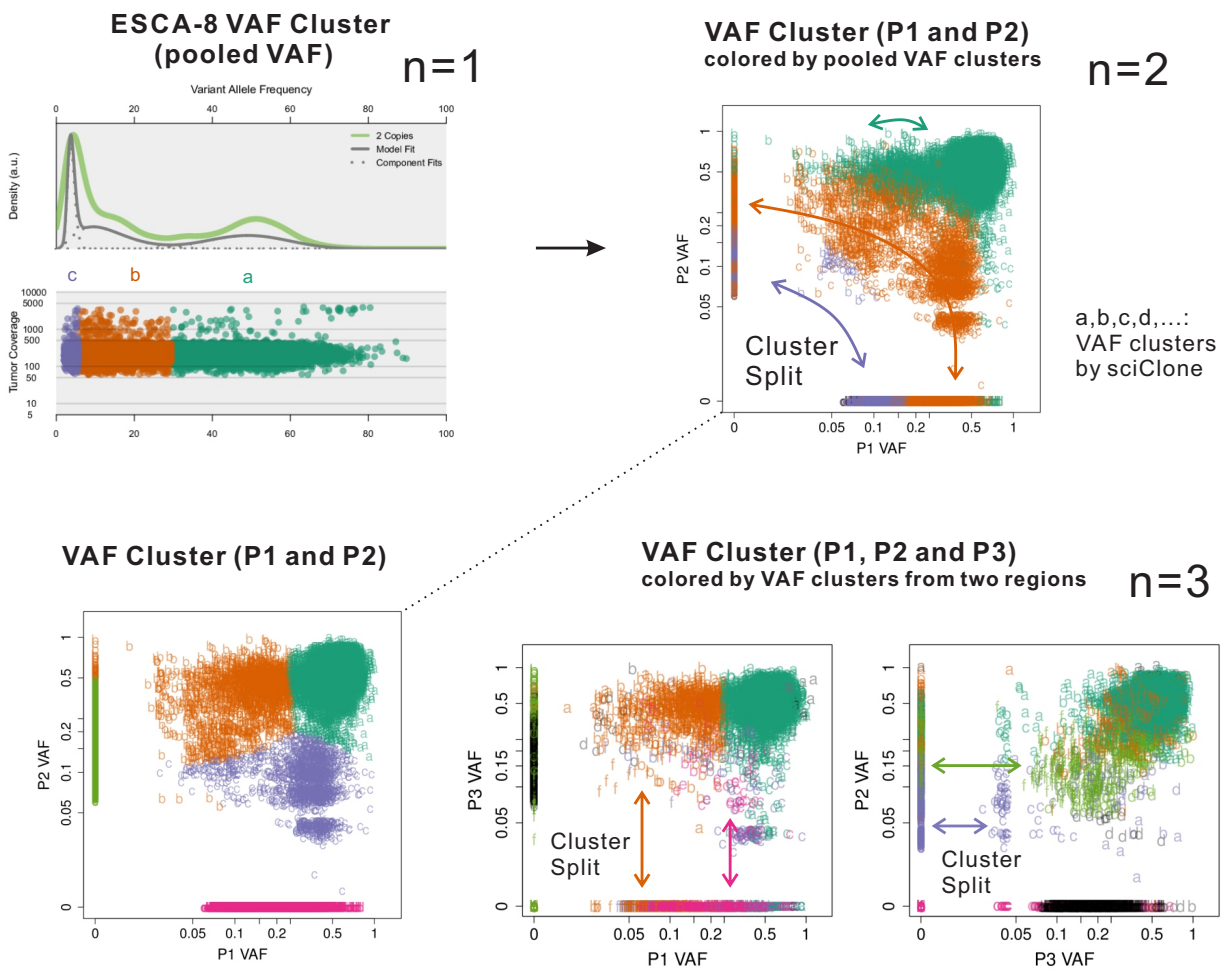
VAF Cluster in three regions (n=3, P1, P2 and P3)	Cluster Split (>5%) in four regions (n=4, P1, P2, P3 and P4)
a	1
b	3
c	1
d	3
e	2
f	2
g	2
h	2
k	3
l	2

Supplementary Figure 22: VAF clusters do not correspond to unique subclones (LUAD-4990).

As in **Supplementary Figure 21**, SciClone [15] was employed to cluster the VAF values for LUAD-4990, taking as input the pooled VAF from four regions (P1, P2, P3 and P4), which served as a composite for clustering (n=1). **(a)** In the composite one sample setting, three VAF clusters were identified (a, b, c; left panel). The impact of including additional regions was evaluated for two regions (P1 and P3, n=2), three regions (P1, P2 and P3, n=3) and four regions (P1, P2, P3 and P4, n=4). SSNVs in the green clonal VAF cluster centered at 0.5 remain clustered (persisted) when additional regions were included. In contrast, the low VAF cluster identified using n regions almost always separated into distinct clusters when n+1 regions were used for clustering (scatterplots, VAF shown on log2 scale). Only SSNVs with depth >30 across all

regions and VAF >0.05 in the detected regions were included. **(b)** The number of distinct VAF clusters (containing >5% of SSNVs in cluster X inferred from n regions) when an additional region was included for clustering (n+1) was used to determine the number of "cluster splits". The observation that cluster splits were common for low VAF clusters suggest that they do not correspond to unique subclones. Of note, it is theoretically possible to identify persistent VAF clusters across multiple regions (such as the majority of the orange cluster), which could represent subclones that arose under selection.

a



b

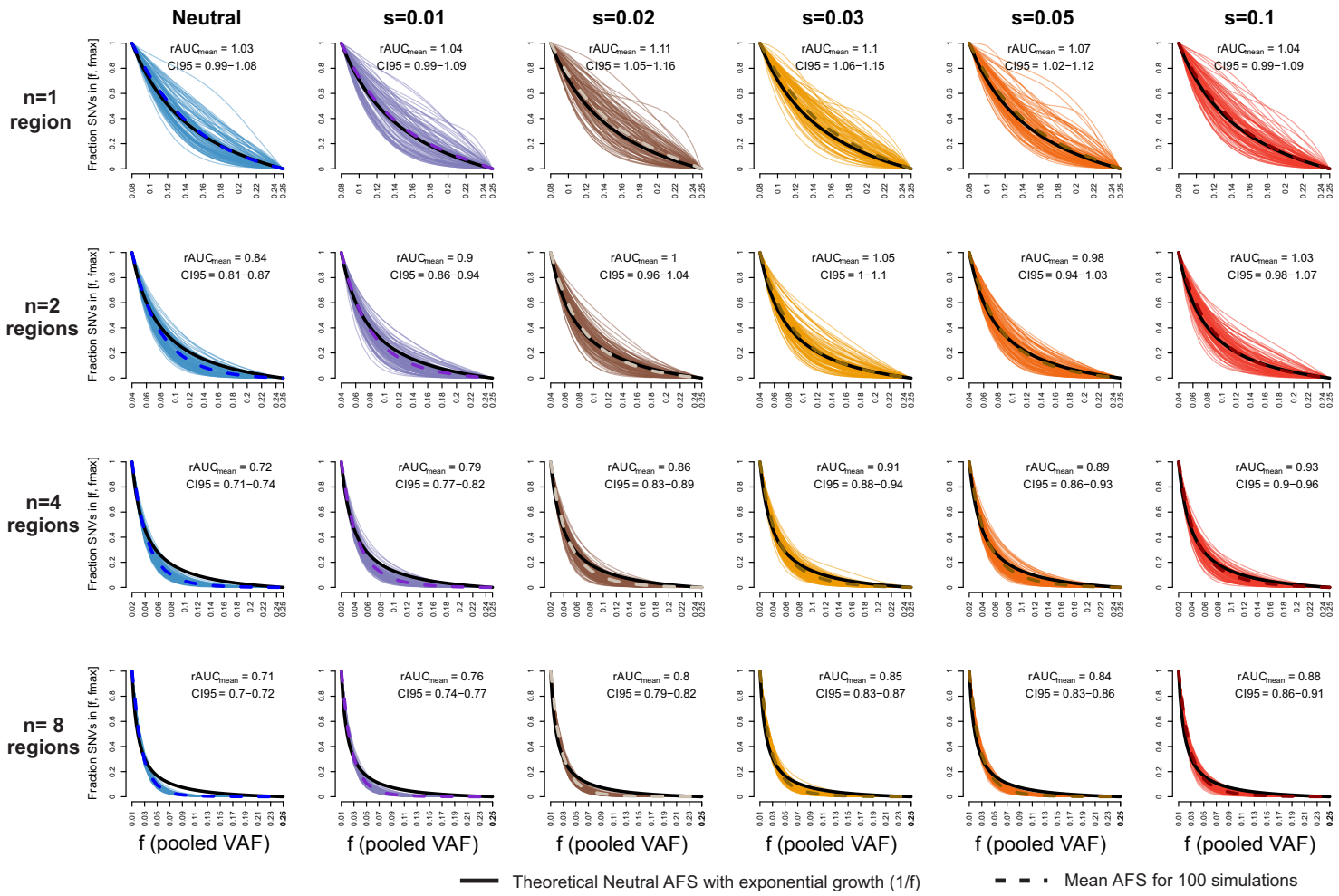
VAF Cluster (n=1)	Cluster Split (>5%) or # of clusters (n=2, P1 and P2)
a	2
b	4
c	2

VAF Cluster in two regions (n=2, P1 and P2)	Cluster Split (>5%) in three regions (n=3, P1, P2 and P3)
a	1
b	3
c	2
d	2
e	2

Supplementary Figure 23: VAF clusters do not correspond to unique subclones (ESCA-8).

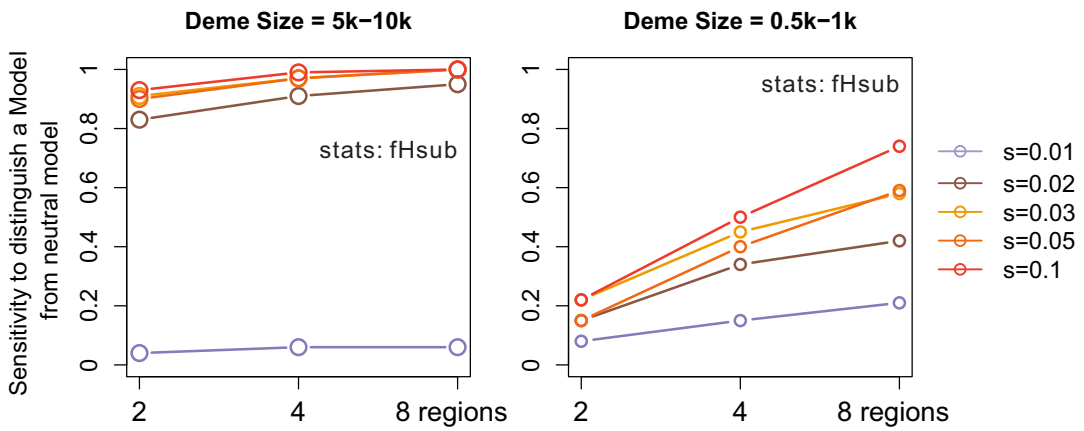
As in **Supplementary Figure 21**, SciClone [15] was employed to cluster the VAF values for ESCA-8 (for which WGS was available) taking as input the pooled VAF from three regions (P1, P2 and P3), which served as a composite sample for clustering (n=1). **(a)** In the composite one sample setting, three VAF clusters were identified (a, b, c; left panel). The impact of including additional regions was evaluated for two regions (P1 and P2, n=2) and three regions (P1, P2 and P3, n=3). Similar to COAD-O and LUAD-4990, SSNVs in the green clonal VAF cluster centered at 0.5 remain clustered (persisted) when additional regions were included. In contrast, low VAF clusters found in n regions always separate into distinct clusters when n+1 regions were used for clustering. Only SSNVs with depth >30 across all regions and VAF >0.05 in the de-

tected regions were included. **(b)** The number of distinct VAF clusters (containing >5% of SSNVs in cluster X inferred from n regions) when an additional region was included for clustering (n+1) was used to determine the number of “cluster splits”. The observation that cluster splits were common for the low VAF cluster suggest that they do not correspond to unique subclones. Of note, an apparent clonal cluster (VAF centered at 0.5) was separated into the true clonal cluster and additional clusters containing subclonal SSNVs that reached fixation in regions P2 and P3 but not in P1 (presumably due to strong selection; green cluster in the scatterplot for n=2, see also **Figure 5b**) as these SSNVs appeared clonal in a single composite sample.



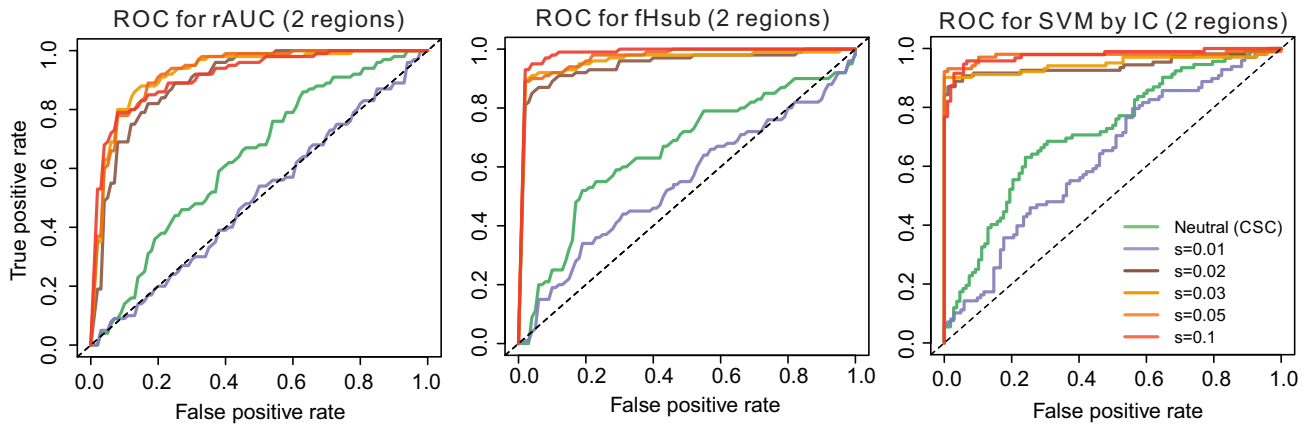
Supplementary Figure 24: Cumulative SFS of virtual tumors simulated under smaller deme size 0.5-1k.

The cumulative SFS (or pooled SFS for multiple regions) is shown for virtual tumors (100 per scenario) from which 1, 2, 4 or 8 regions were sampled, corresponding to the cumulative fraction of SSNVs with VAFs in the range f_min – 0.25. Seven models (neutral, neutral CSC, s=0.01, 0.02, 0.03, 0.05, 0.1) were simulated for a much smaller deme size (0.5-1k cells per deme). The cumulative SFS is plotted as in **Supplementary Figure 6**.



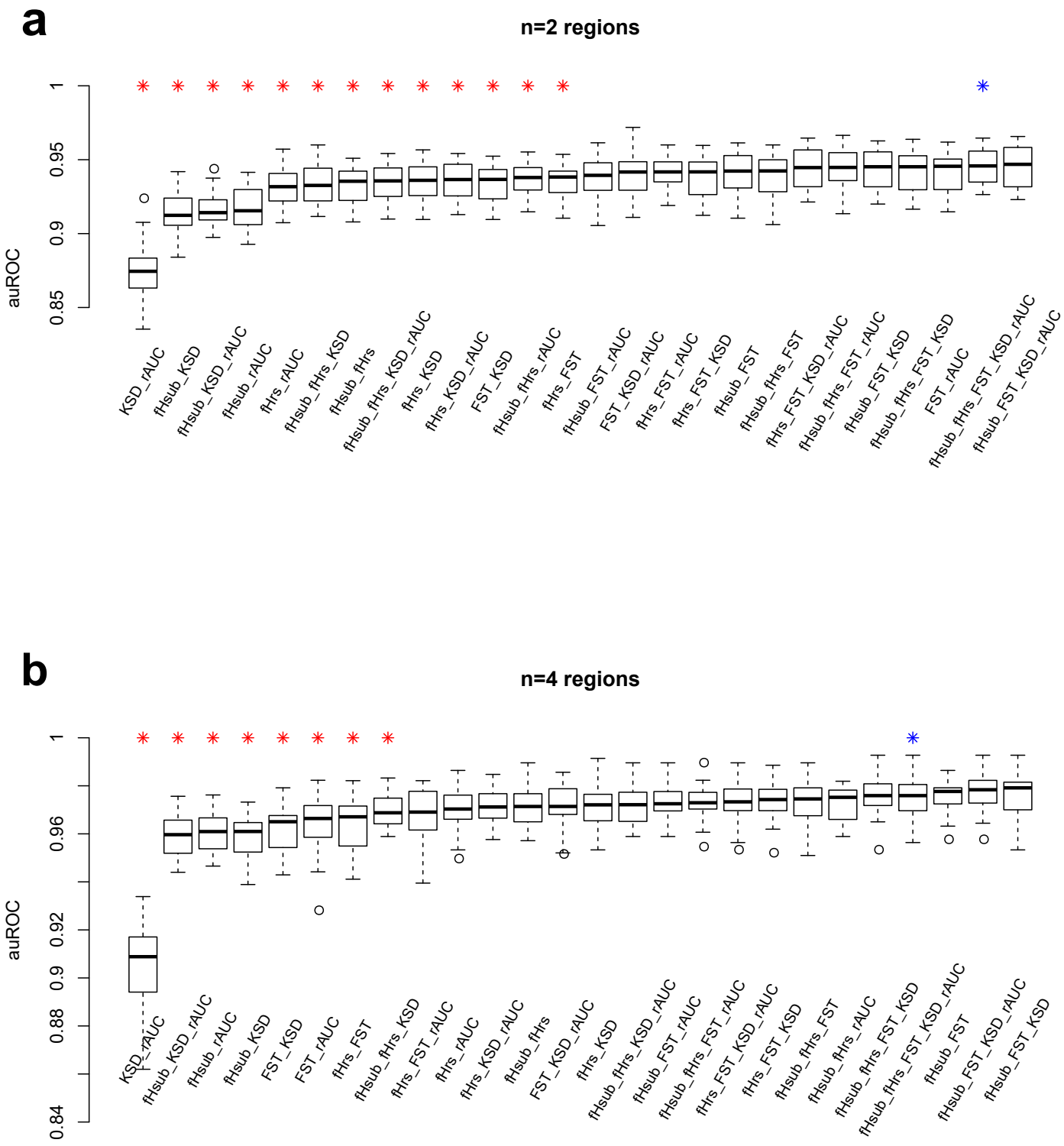
Supplementary Figure 25: Power of fHsub to distinguish selection from the neutral model under different deme sizes (5-10k versus 0.5-1k).

The power to distinguish an alternate model (with s=0.01, 0.02, 0.03, 0.05 and 0.1) from the neutral model using one of the ITH metrics, namely fHsub, are shown for models under two different deme size ranges (5-10k on the left and 0.5-1k on the right). Power (at the 10% significance level) was computed empirically as the percentage of virtual tumors under an alternate model with the statistic (fHsub) greater than 95% quantile or less than 5% quantile of the corresponding statistic in the neutral model (taking the larger percentage). Reduced deme size results in decreased sensitivity to distinguish selective models from neutrality, due to increased spatial constraints.

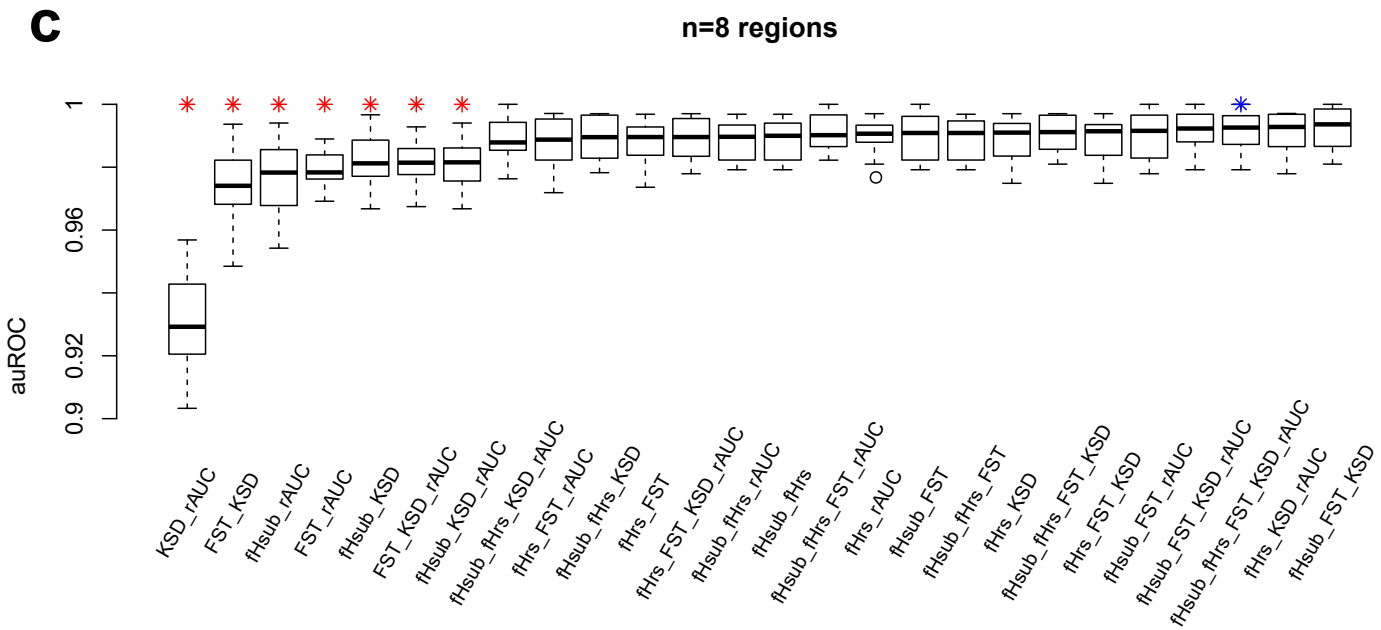


Supplementary Figure 26: Performance to distinguish alternate evolutionary models from a null neutral model based on the SVM classifier and single ITH metrics

The Receiver Operating Characteristic (ROC) curves for distinguishing alternate models from neutral by using single ITH metrics, rAUC (left), fHsub (middle) and SVM by two IC components (right). The ROC curves comparing the true positive versus false positive rate for using two tumor regions are also shown. For single metrics, ROC is generated using frequency based calculations. For the SVM, the ROC curve corresponds to the median auROC amongst 20 SVMs trained on 100 randomly chosen virtual tumors (for each of the alternate models and null neutral model) and tested on the remaining 100 virtual tumors.

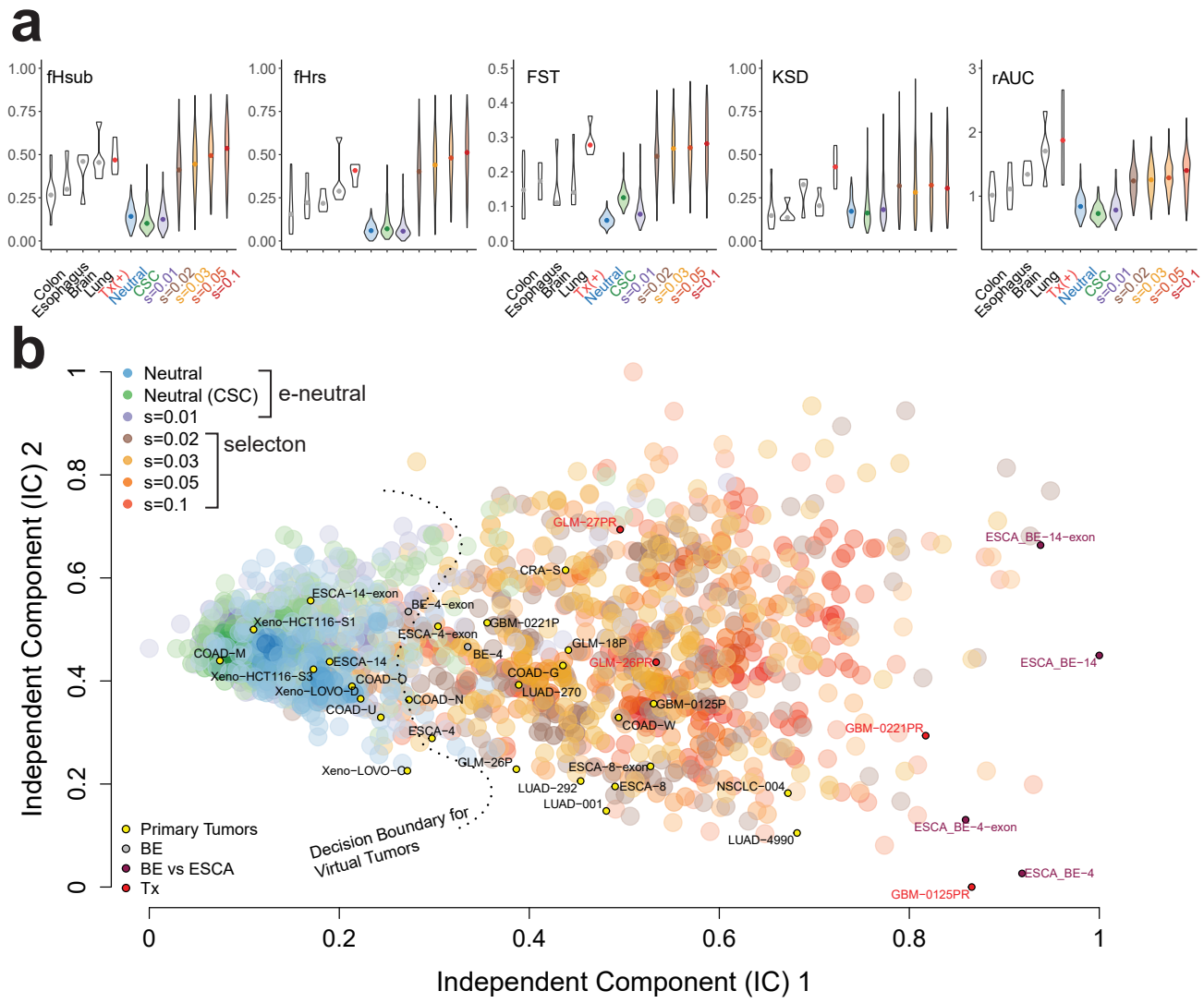

a
n=2 regions
auROC
0.85
0.9
0.95
1
KSD_rAUC
fHsub_KSD
fHsub_KSD_rAUC
fHsub_rAUC
fHrs_rAUC
fHsub_fHrs_KSD
fHsub_fHrs
fHsub_fHrs_KSD_rAUC
fHrs_KSD
fHrs_KSD_rAUC
FST_KSD
fHsub_fHrs_rAUC
fHrs_FST
fHsub_FST_rAUC
FST_KSD_rAUC
fHrs_FST_rAUC
fHrs_FST_KSD
fHsub_FST
fHsub_fHrs_FST
fHrs_FST_KSD_rAUC
fHsub_fHrs_FST_rAUC
fHsub_FST_KSD
fHsub_fHrs_FST_KSD
FST_rAUC
fHsub_fHrs_FST_KSD_rAUC
fHsub_FST_KSD_rAUC

b
n=4 regions
auROC
0.84
0.88
0.9
0.92
0.96
1
KSD_rAUC
fHsub_KSD_rAUC
fHsub_rAUC
fHsub_KSD
FST_KSD
FST_rAUC
fHrs_FST
fHsub_fHrs_KSD
fHrs_FST_rAUC
fHrs_rAUC
fHrs_KSD_rAUC
fHsub_fHrs
FST_KSD_rAUC
fHrs_KSD
fHsub_fHrs_KSD_rAUC
fHsub_FST_rAUC
fHsub_fHrs_FST_rAUC
fHrs_FST_KSD_rAUC
fHrs_FST_KSD
fHsub_fHrs_FST
fHsub_fHrs_rAUC
fHsub_fHrs_FST_KSD
fHsub_fHrs_FST_KSD_rAUC
fHsub_FST
fHsub_FST_KSD_rAUC
fHsub_FST_KSD



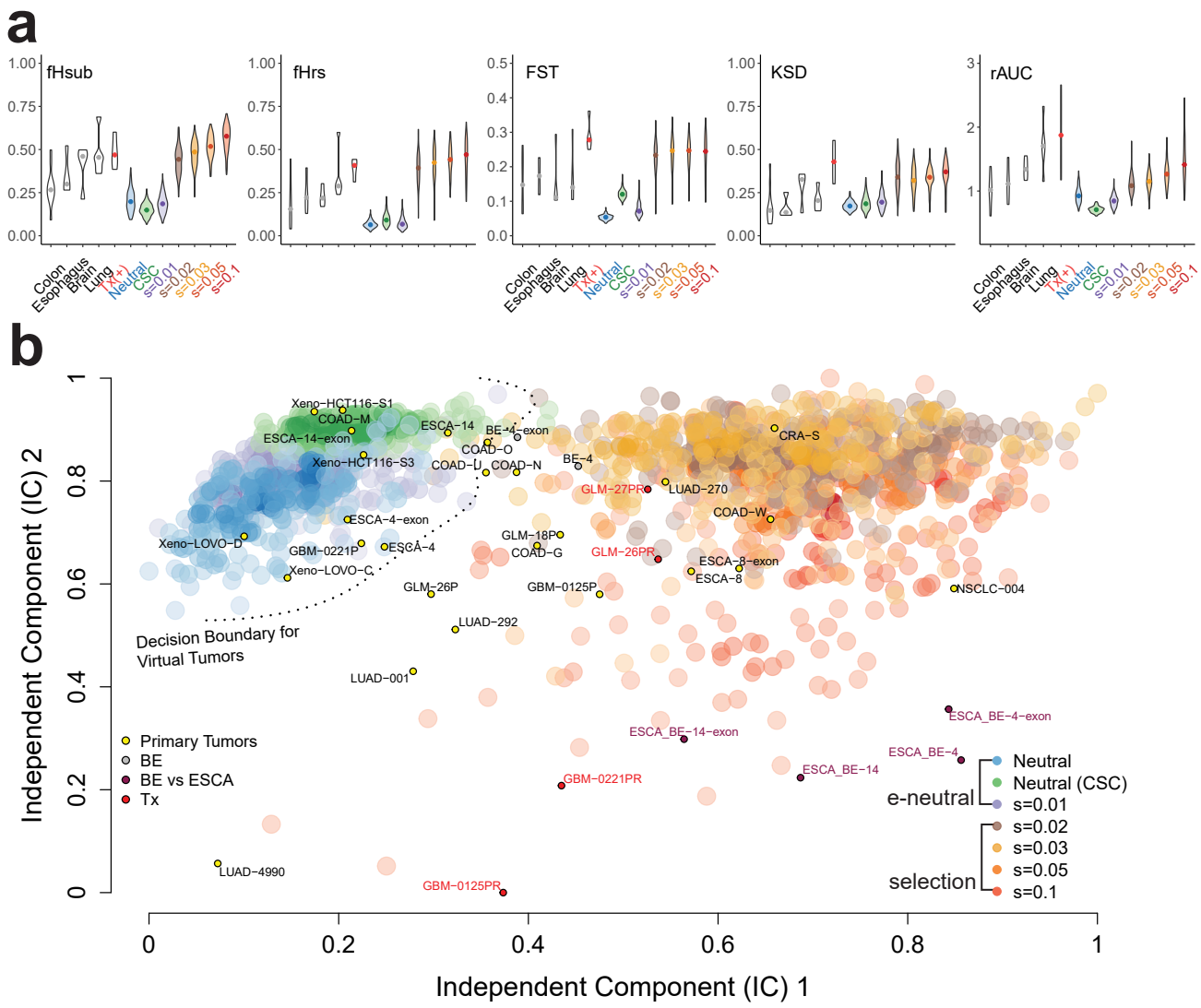
Supplementary Figure 27: ROC based on SVM classification of various combinations of five ITH metrics.

To evaluate the relative importance of different combinations of the five ITH metrics for classification, SVMs were run 20 times for each of 26 possible combinations of the five metrics with the same seed for random splitting. A total of 1120 simulated tumors were used for training and 280 for testing. The resulting auROCs were compared for **(a)** 2 regions, **(b)** 4 regions and **(c)** 8 regions. Blue stars indicate SVMs trained using all five metrics. Red stars indicate SVMs trained with various combinations of the five metrics that have significantly smaller auROCs than the SVM trained with all five metrics (Wilcoxon rank sum test).



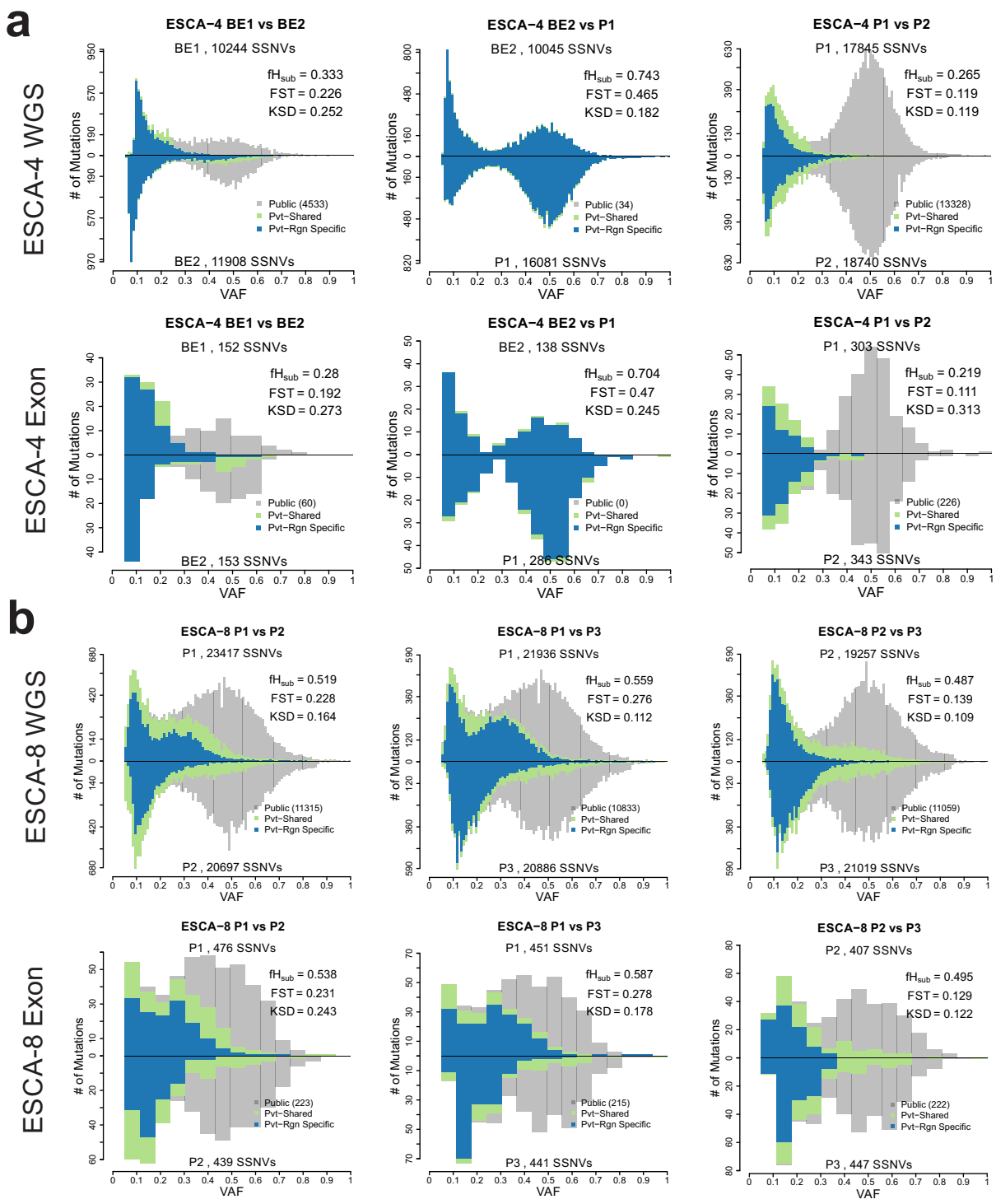
Supplementary Figure 28: Projection of patient tumors onto distinct evolutionary modes based on ITH metrics derived from the SFS (2 virtual tumor regions).

(a) Violin plots for each of the five ITH metrics, namely, fHsub, fHrs, Fst, KSD and rAUC. Colored violin plots correspond to virtual tumors simulated under different evolutionary modes, whereas the white violins correspond to clinical samples. Paired pre-treatment primary and post-treatment recurrent brain tumors are denoted by "Tx" and serve as a positive control for selection. **(b)** Independent component analysis (ICA) of virtual and real tumors based on the five aforementioned metrics. The two independent components (IC) clearly separate the simulated tumors under *effectively (e) neutral growth* (neutral, neutral-CSC and s=0.01) versus *positive selection* (s$\geq$ 0.02). The decision boundaries for an SVM trained on the two IC components based on the simulated tumors is indicated by the dashed line. Large transparent colored circles represent values from virtual tumors simulated under different models (200 tumors from each of the seven models are shown). Small filled circles corresponding to clinical samples are labeled by their sample ID and color-coded according to the nature of the sample. COAD: colorectal adenocarcinoma; CRA: colorectal adenoma; ESCA: esophageal adenocarcinoma; BE: Barrett's esophagus; LUAD: lung adenocarcinoma; NSCLC: non-small-cell lung cancer; GLM: glioma; GBM: glioblastoma; Xeno: COAD cell line xenografts.

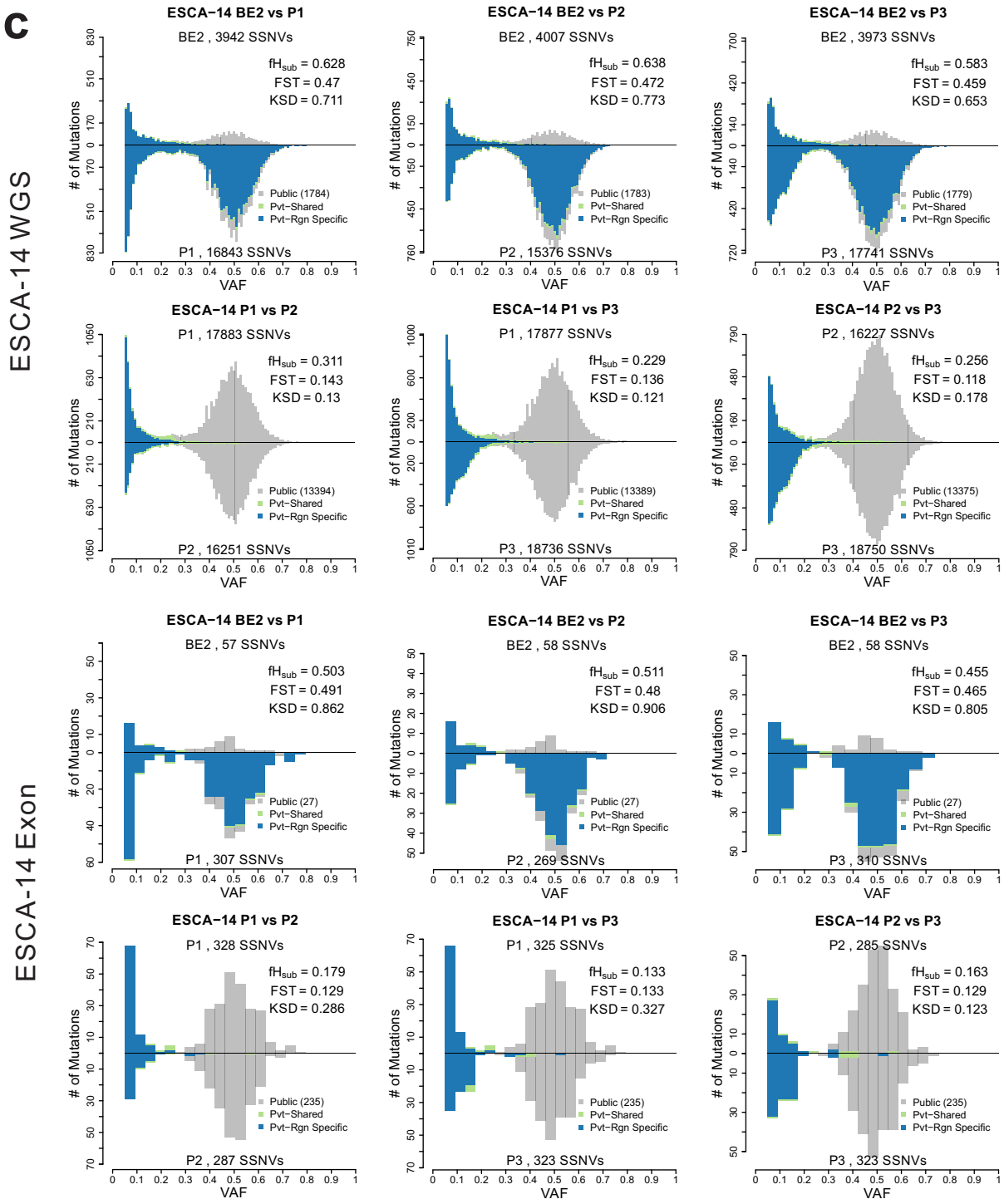


Supplementary Figure 29: Projection of patient tumors onto distinct evolutionary modes based on ITH metrics derived from the SFS (8 virtual tumor regions).

The legend is the same as for **Supplementary Figure 28**.



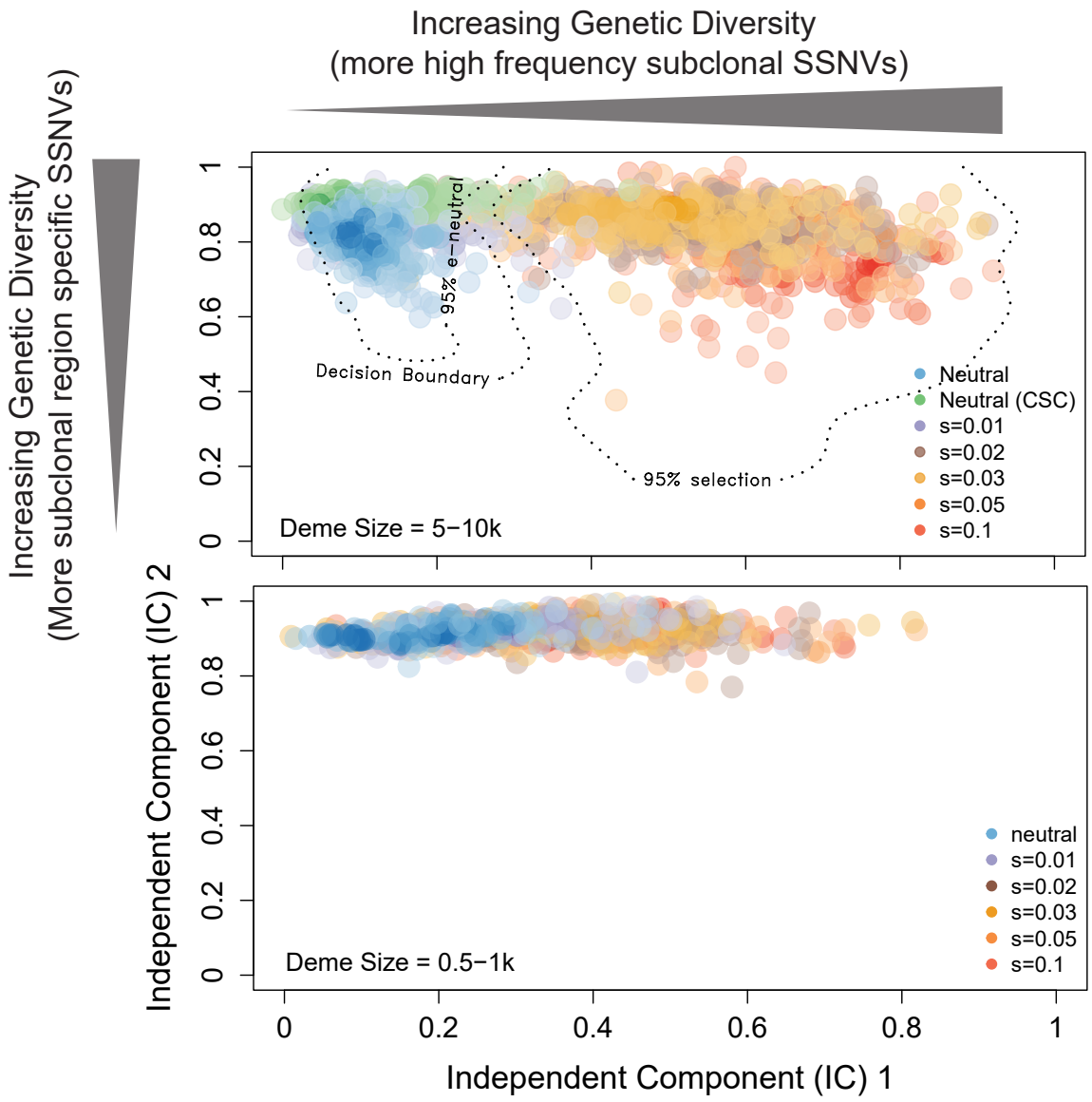
Supplementary Figure 30: The shape of the SFS histograms derived from whole-genome sequencing versus down-sampled exonic regions is comparable.



Supplementary Figure 30: The shape of the SFS histograms derived from whole-genome sequencing versus down-sampled exonic regions is comparable.

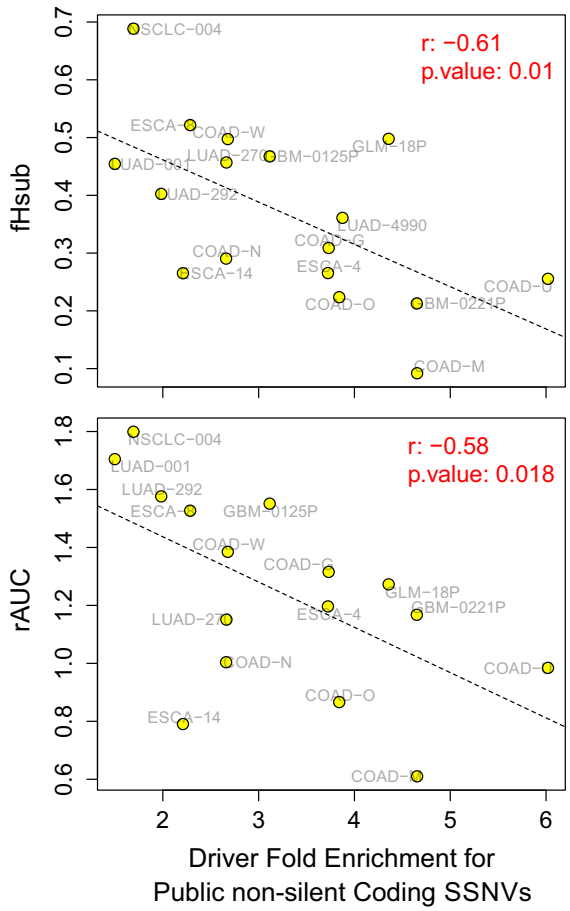
Despite the much larger number of SSNVs observed in the WGS, the overall shape of SFS is comparable for down sampled SSNVs residing in exonic region (UTRs included) for three ESCA patients: **(a)** ESCA-4; **(b)** ESCA-8 and **(c)** ESCA-14. For ESCA-4 and ESCA-14, the Barrett's esophageal (BE) lesions from the

same patient were also compared with their matched carcinomas (ESCA), where strong genetic divergence (indicating selection) was observed. A marginal signal of selection was observed in the comparison of BE1 and BE2, as can be seen in WGS data from ESCA-4. Whereas the reduced number of SSNVs in the exonic regions dilutes the signal slightly, genetic diversity between BE1 and BE2 measured by down-sampled exonic data is still mapped onto selection space (**Figure 6b**). This highlights the utility of WES to capture tumor dynamics. The non-proportional decrease in exonic SSNVs as compared to whole genome SSNVs is potentially due to negative selection for more functional sites.



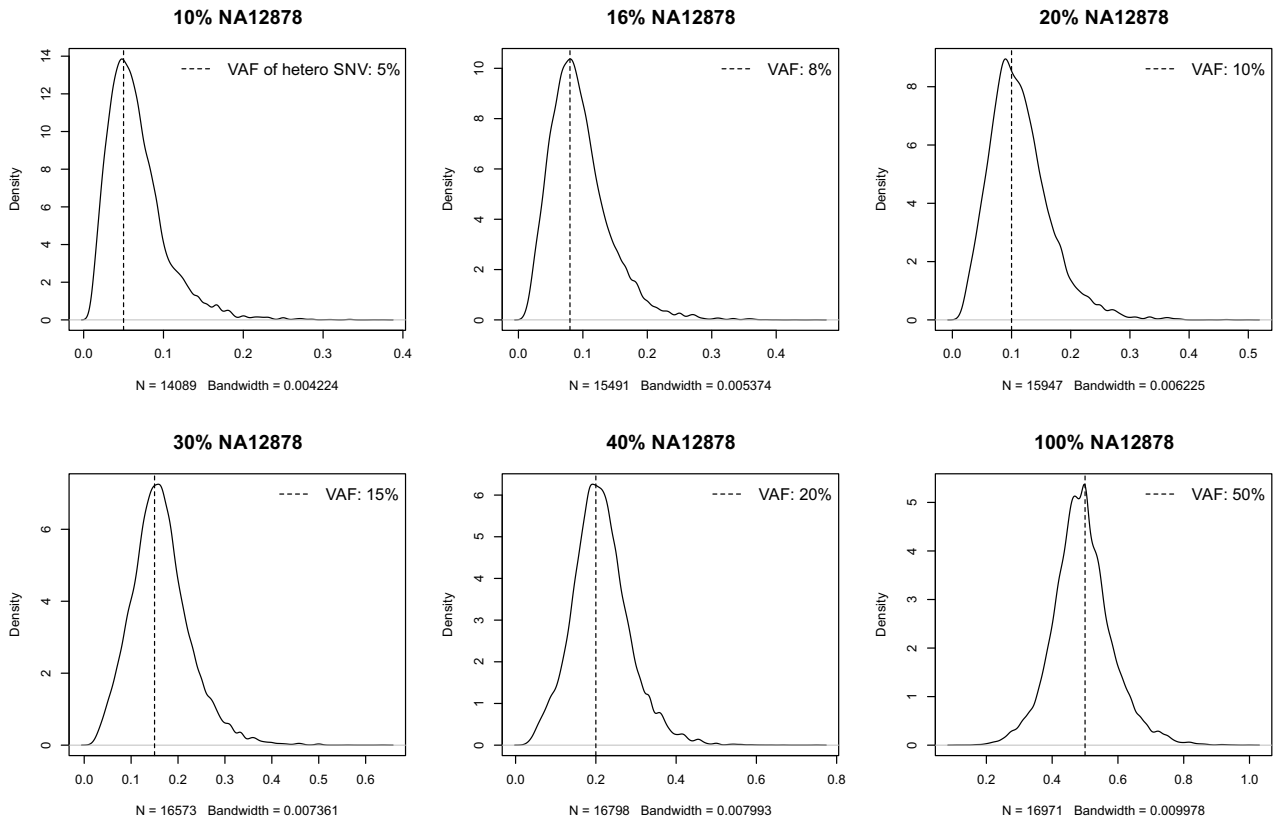
Supplementary Figure 31: Independent component (IC) analysis for models with deme sizes of 5-10k and 0.5-1k, respectively.

The five ITH metrics calculated under models with deme sizes of 5-10k (upper panel) and 0.5-1k (lower panel) were used in an independent component (IC) analysis (4 tumor regions). For the 5-10k deme size, the ICs clearly separate tumors simulated under *effectively (e) neutral growth* (neutral, neutral-CSC and s=0.01) versus *positive selection* (s≥ 0.02). The decision boundary for a SVM trained on the virtual tumors based on the two ICs is indicated by dashed lines as are the 95% CIs for classification of virtual tumors into *effectively neutral (e-neutral)* versus *positive selection* modes. Transparent colored circles represent values from virtual tumors simulated under different models (200 virtual tumors from each of the seven models are shown). In contrast, under a deme size an order of magnitude smaller, stronger spatial constraints blur the distinction between *effective neutrality* and *positive selection* (100 virtual tumors from each model). Although general trends between selection and the levels of detectable between-region genetic divergence were observed, we note that specific patterns could be model dependent.



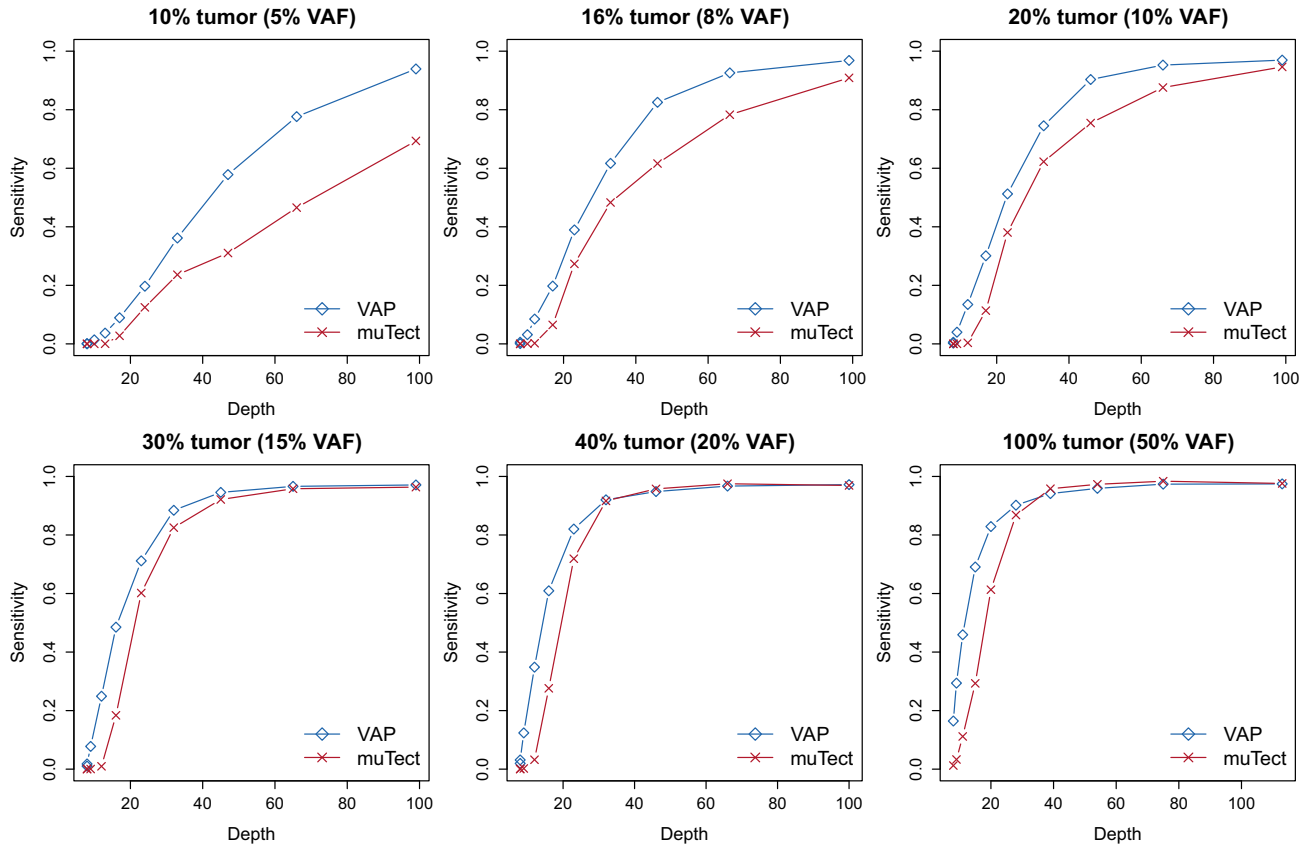
Supplementary Figure 32: Tumors enriched for SSNVs in public driver genes are associated with effectively neutral growth.

A significant negative correlation was observed between two ITH metrics (fHsub and rAUC) and the fold enrichment of non-silent public SSNVs in tumor type specific driver genes based on IntOGen. Driver fold enrichment was determined based on the number of driver genes in the corresponding tumor type harboring non-silent coding SSNVs out of the total number of genes with non-silent coding SSNVs. The resulting metric was normalized by the fraction of tumor type specific driver genes out of all genes. Only invasive primary tumors with >30 non-silent coding public SSNVs were included.



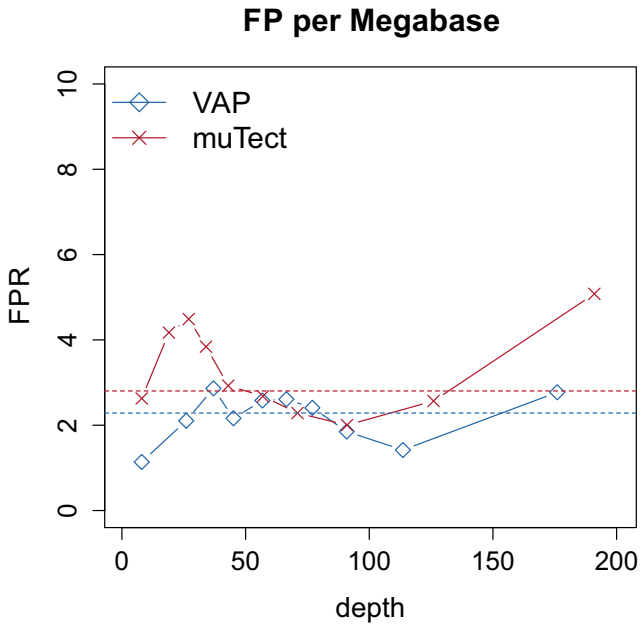
Supplementary Figure 33: Mixing of NA12878 and NA18489 sequencing data to generate virtual tumor samples.

BAM files for two individual whole exome sequencing from 1,000 genome project were downloaded and mixed to generate the following mixture concentrations of NA12878 (as tumor): 10%, 16%, 20%, 30%, 40% and 100%, resulting in an expected VAF of heterozygous SNVs of NA12878 at 5%, 8%, 10%, 15%, 20% and 50% in the mixed sequencing data. The VAF density curves for variants detected in the virtual tumor samples (using NA18489 as normal control), which are also heterozygous SNVs in the gold standard variant list provided by GIAB-NIST for NA12878, indicate that the mixing was as intended.



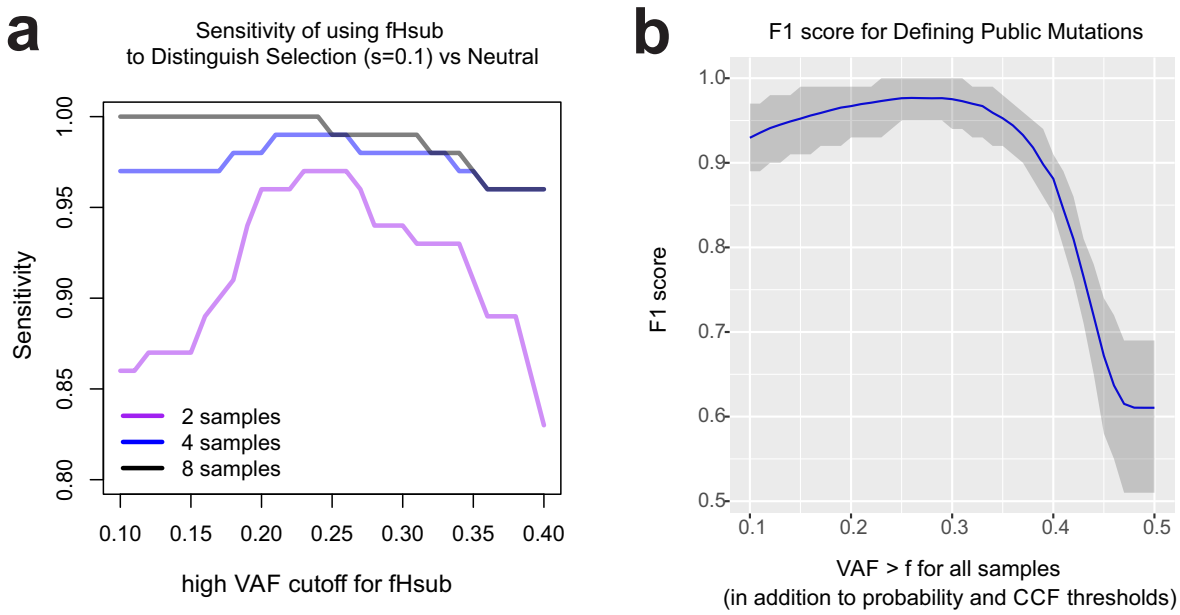
Supplementary Figure 34: Sensitivity of VAP and MuTect based on a virtual tumor samples dilution series.

Heterozygous SNVs in the gold standard list for NA12878 were divided into 10 groups with equal size according to the sequencing depth in each virtual tumor samples. The detection sensitivity for VAP and MuTect for each group of gold standard variants are shown against the group's median depth. Each panel shows a virtual tumor with an intended VAF obtained via *in silico* mixing.



Supplementary Figure 35: False Positive Rate per million base pairs for VAP and MuTect for virtual tumor samples.

False positive calls aggregated across the virtual tumors were divided into 10 groups with equal size according to sequencing depth. The false positive rate (FPR) per million base pairs for VAP and MuTect in each depth group were generated by dividing the FP within each group to the total bases designed in SureSelect protocol that are covered with the same range of depth as for the FPs. FPR were then plotted against the group's median depth.



Supplementary Figure 36: Impact of VAF cutoffs on the sensitivity of fHsub to detect selection and the identification of public mutations.

(a) The sensitivity to distinguishing selection (s=0.1) from neutrality using fHsub calculated using various "high" VAF cutoffs (0.1-0.4). A plateau is apparent where the highest sensitivity is achieved for VAFs between 0.2 and 0.25. We selected a cutoff of 0.2 to ensure a sufficient number of mutations (counts) for the nominator when considering the variability in real sequencing data. **(b)** The performance to identify public mutations was evaluated when considering different cutoffs for the third criterion (i.e. at least one region has adjusted VAF <the cutoff, Methods). Here 100 virtual tumors subject to multi-region sampling (n=4 regions) were used for evaluation. The F1 score (harmonic mean of precision and recall) is shown for cutoffs ranging from 0.1 to 0.5. The mean F1 is the blue line and the gray shade indicates the 95% quantile of the F1 scores for 100 tumors under each cutoff. A plateau is apparent with the highest performance obtained between 0.25 and 0.3. 0.25 was chosen as the cutoff considering the noise in real sequencing data where public mutations can have an adjusted VAF as low as 0.25.

References Cited in Supplementary Figure Legends

These references are cited in Supplementary Figure legends[1-15].

Supplementary Tables

Supplementary Table 1. Parameters in the deme model of spatial tumor growth.

Parameters	Values	Justifications/Remarks
Final tumor size, N	N~10^9 cells	There are ~10^9 or more cells in a typical solid tumor.
Deme size, K	K= 5–10k cells or 0.5–1k cells	The demes recapitulate the glandular structure often found in epithelial tumors such as colorectal cancer in which the gland size is approximated at 2,000-10,000 cells [16]. The deme size models the degree of spatial constraints and clone mixing during tumor growth. For instance, small deme size imposes stringent spatial constraints and reduces subclone mixing, thereby hindering the efficacy of selection. In contrast, large deme sizes result in relaxed spatial constraints and elevated subclone mixing.
The birth and death probabilities for each cell at each time step during deme expansion, p and q, respectively	p=0.55 and q=0.45	The death-birth rate ratio h=q/p=0.82 falls between that reported for rapidly growing colorectal cancer metastases (h=0.72) [17] and early tumors (h=0.99) [17].
Neutral point mutation rate per cell division in the whole exonic region, u	u=1.2	The point mutation rate *per cell division in the exonic region* (u) can be variable among tumors [18]. Given the genomic instability due to inactivation of DNA repair mechanisms in many tumors, the per-cell division mutation rate for cancer is assumed to be significantly higher than somatic normal cells which is on the order of 10^{-9} per base pair per cell division [19]. We assume a mutation rate u=1.2 per cell division for the 60M exonic region of diploid human genome (equivalent to 2×10^{-8} /base pair/division) for the simulations, giving rise to 100-1000 subclonal SNVs (VAF >0.08) in each bulk sample, which is comparable to that observed in clinical samples in the current study.
Driver mutation rate per cell division in the genome, u_b	u_a=10^{-5} or 10^{-6}	The driver mutation rate is also unknown, although it is assumed that there are ~3-7 drivers per tumor [20]. Bozic et al. [21] estimated u_b to be 10^{-5} per cell division in the genome. Smaller u_b can result in rarer or later occurrence of driver mutations and weaker sweeps even when the selection coefficient is large.
Selection coefficient, s	s=0, 0.01, 0.02, 0.03, 0.05 or 0.1	We explored variable selection coefficients, where s=0 is equivalent to a neutral (null) model. Experimental estimates of selection coefficients during human tumorigenesis and post transformation are lacking.

Supplementary Table 2. Terminology and Descriptions

Term	Description
MRS	Multi-Region Sequencing
WES	Whole-Exome Sequencing
WGS	Whole-Genome Sequencing
Sample/Region	Refers to a region or sample of a tumor
ITH (Intra Tumor Heterogeneity)	Between region subclonal genetic divergence based on MRS
Public/Clonal Mutations	Mutations predicted to be present in all tumor cells
Private/Subclonal Mutations	Mutations predicted to be present only in a part of tumor cells
Region-Specific Mutations	Mutations predicted to be only present in one of two regions
VAF (Variant Allele Frequency)	Variant Allele Frequency adjusted for local copy number and tumor purity
SFS	Site Frequency Spectrum or VAF distribution
Cumulative SFS	Cumulative distribution of Site Frequency Spectrum
Pooled VAF/SFS	For MRS, reads are pooled across regions for computing the VAF/SFS

Supplementary Table 3. Summary of the multi-region sequencing (MRS) datasets included in this study, including the study accession ID, tumor type, sample ID, and sample types.

Tumor Type	Database	Accession Code	Tumors passed QC	Sample IDs	Sample Type
COAD[7]	ArrayExpress	E-MTAB-2247	COAD-W	WA, WB	Primary
			COAD-M	MA, MB	Primary
			COAD-G	GA, GB	Primary
			COAD-N	NA, NB	Primary
			COAD-U	UA, UB	Primary
			COAD-O	OA, OB	Primary
			CRA-S	SA, SB	Primary
COAD Xenografts	ArrayExpress	E-MTAB-5547	Xeno-HCT116-S1	S1_XA, S1_XB	Xenograft
			Xeno-HCT116-S3	S3_XA, S3_XB	Xenograft
			Xeno-LoVo-D	XD, XDL	Xenograft
			Xeno-LoVo-C	XC, XCL	Xenograft
ESCA[8]	EGA	EGAD00001001394	BE-4	BE1, BE2	Preneoplasia
			ESCA-4	P1, P2	Primary
			ESCA-8	P1, P2, P3	Primary
			ESCA-14	BE2, P1, P2, P3	Primary
Glioma[11]	EGA	EGAD00001000714	GLM-18P	P1, P2, P3	Primary
			GLM-26P	P1, P2	Primary
			GLM-26R	R1, R2	Tx
			GLM-27R	R1, R2	Tx
NSCLC[10]	EGA	EGAD00001000900	LUAD-001	P1, P2	Primary
			NSCLC-004	P1, P2, P3, P4	Primary
LUAD[9]	EGA	EGAD00001000984	LUAD-270	P1, P2, P3	Primary
			LUAD-292	P1, P2, P3	Primary
			LUAD-4990	P1, P2, P3, P4	Primary
GBM[12]	EGA	EGAD00001001113	GBM-0125P	P1, P2	Primary
			GBM-0125R	R1, R2	Tx
			GBM-0221P	P1, P2	Primary
			GBM-0221R	R1, R2	Tx

Supplementary Table 4. SFS derived ITH metrics for virtual tumors.
Five ITH metrics (fHsub, fHrs, FST, KSD and rAUC) based on pairwise comparisons of the SFS for subclonal variants in different tumor regions are shown for each virtual tumor. Statistics derived from MRS of 2, 4 and 8 tumor regions are indicated in separate tabs.

Supplementary Table 5. SFS derived ITH metrics and SVM based classification of patient samples.
Five ITH metrics (fHsub, fHrs, FST, KSD and rAUC) based on pairwise comparisons of the SFS for subclonal SSNVs in different tumor regions were calculated for each patient's tumor. The classification probabilities for the SVM trained on virtual tumors with MRS from four regions is shown for i) the SVM based on five statistics or ii) the two primary independent components (IC) derived from these five statistics.

Supplementary Note

Somatic SNV calling

Short-reads produced by WES (or WGS) on the Illumina platform were aligned to hg19 using *BWA* (version 0.7.12-r1039)[1]. The raw alignment files (BAMs) were then pre-processed through indel realignment, base recalibration and flagging of duplicated reads. Raw SSNV calls were made using *MuTect* (v1.1.4) [2] and VCF files were annotated using *Annovar* [22]. Nonsilent SSNVs in predicted and known 'driver' genes were annotated with *IntOGen* (v.2016.5) [23]. To facilitate quantitative comparisons of the SFS, we devised a unified variant assurance (filtering and rescuing) pipeline (VAP, manuscript in preparation) to achieve a balance in sensitivity and specificity when MRS is available such that information can be borrowed across regions. False positive calls due to misalignments or other technical artefacts tend to exhibit suggestive local mapping features where manual inspection is often used to ensure the quality of calls. On the other hand, variants can also be missed due to limited sensitivity [24] in the face of non-uniform read/quality distribution in different samples. Thus, for each raw variant call (unfiltered), the sequencing data from all samples was re-inspected to assess the confidence (in detected samples) and evidence (in un-detected samples) for the alternative allele. We removed variant calls supported by reads (in the detected sample) having (*a*) strong strand bias (one strand only unless the same bias is seen for the reference allele), (*b*) overall low mapping quality (60% or above with less than MAQ 30), (*c*) tail-distance bias (90% or above are within 15 bp from the read-ends), (*d*) equal or better mapping quality to alternative locations by another mapper [25] and (*e*) multiple mismatches and indels (n ≥ 3 on average). These summarized biases are associated with different types of artefacts (**Supplementary Figure 13**). Of note, segmental duplications or misalignments often lead to consistent mismatched clusters in the same read captured by the multi-mismatch filter. Variants were rescued for high-quality reads (n ≥ 3) supporting the alternative allele in initially undetected samples with none of the aforementioned biases.

SSNVs were called by requiring (*a*) the tumor LOD score (log odds probability of the tumor having the SSNV) exceeds 4.3 and the normal LOD score (log odds for the absence of germline variants in the normal) exceeds 2.3, where the scores were calculated as previously described [2]. The base error rates were set to the maximum of the corresponding Phred base-error or a local error rate computed from the number of non-consensus mismatches[26]; (*b*) read depth above eight for normal control; (*c*) a Fisher's exact p value < 0.05 in comparing alternative allele presence between normal and tumor samples and VAF for normal is less than 0.05 (*d*) not a known SNP found in multiple individuals, some of which can be regarded as germline variants. The VAF (unadjusted) was calculated by dividing the number of reads carrying the variant by the total number of reads spanning that position. Benchmarking (section "Benchmarking of Variant Assurance Pipeline" in this Supplementary Note) shows that our pipeline achieves better sensitivity at moderate read depth (<100x) for SSNVs with VAF between 5-20% than the *MuTect* default setting with a comparable false positive rate (**Supplementary Figure 33-35**).

For the xenograft samples, a combined human (hg19) and mouse (mm10) reference genome was adopted for read alignment. Reads that mapped to the mouse genome with equal or better quality than to the human genome were removed (less than 5% of the total reads, **Supplementary Figure 11b**). As matched normal cell lines

do not exist, for both the LoVo and HCT116 xenograft studies, an ancestral polyclonal tumor cell mixture (taken prior to the single cell expansion) was used as a reference. This ancestral population harbors many shared public somatic SSNVs with the descendent cells, which are therefore removed in a tumor-normal comparison, such that the remaining public SSNVs arose after the imposed population bottleneck.

The enrichment of subclonal region-specific mutations in the low VAF range is unlikely to be fully explained by coverage bias since: 1) the lower the VAF, the higher the depth required to achieve the same LOD (log-likelihood odds ratio for calling); 2) for region-specific SSNVs, the variant absent region is required to have sufficient depth to achieve a binomial probability of missing variant reads less than 0.05 given the observed VAF; 3) depth from different tumor regions is typically correlated (**Supplementary Figure 12**).

CNA detection and VAF adjustment

Copy number and tumor purity (p) were estimated with *TitanCNA* [27] (version 1.8.0) in exome-seq mode (except for the ESCA dataset where WGS was available). The one clone solution was generally the best fit to the data, with the exception of case LUAD-4990 for which the two clones model was a better fit. Germline heterozygous SNVs used as input to *TitanCNA* were identified using *Samtools* [28] (version 1.2) and subject to the same filtering strategy as was applied to SSNVs. Homozygous SNVs were specifically retained for germline LOH regions in the starting HCT116 and LoVo cell lines for the xenograft experiments to enable the recovery of such events. Read depth controls were obtained from the paired normal except for the ESCA cohort as the paired normal samples exhibited highly variable depth that was not seen in the tumors. The Barrett's esophageal lesion from patient 8 (BE-8) was adopted instead as a depth control as it appears to be polyclonal and diploid. For xenograft experiments, germline LOH or copy number gains present in the starting cell lines were identified by segmenting homozygous SNPs and corresponding normalized read depth using *HMMcopy* [27]. The read depth files for the starting cell lines were set to diploid for regions with copy number gains (chromosomes 7,12 and 15 for LoVo and 8q, 10q, 16q and 17q for HCT116) to enable the recall of these events in xenograft samples.

The observed VAF, VAF_o, was adjusted as previously described [29] in order to map the VAF_o to a pure/diploid state to enable comparisons of SFS between samples/tumors. In brief, for an SSNV residing in a genomic segment with a total copy number of N_t, minor allele copy number of N_b and cellular prevalence P_{CNA} of the CNA in the tumor content, the adjusted variant allele frequency (VAF_a) is half of the point estimate of cancer cell fraction (CCF_{est}) given VAF_o is true:

$$VAF_a = \frac{CCF_{est}}{2}, \qquad CCF_{est} = \begin{cases} N_c \times \frac{VAF_o}{p'} - P_{CNA} \times (N_t - N_b - 1) & Early\ Major \\ N_c \times \frac{VAF_o}{p'} - P_{CNA} \times (N_b - 1) & Early\ Minor \\ N_c \times \frac{VAF_o}{p'} & Late/Independent \end{cases}$$

where $N_c = N_t \times P_{CNA} + 2 \times (1 - P_{CNA})$ and the effective purity $p' = \frac{N_t \times p}{N_t \times p + 2 \times (1-p)}$ (p is the estimated tumor purity). Point estimates of VAF_a can be greater than 0.5 due to the inherent variability associated with sampling the observed VAF_o. The VAF_a is equivalent to half the expected CCF for values below 0.5, which collapses to 1 when VAF_a is above

0.5 (**Supplementary Figure 14**). We used VAF_a in keeping with the use of allele frequencies in population genetics and to enable visualization of the public mutation cluster centered at 0.5 (public mutations are not informative for the quantification of ITH). The temporal ordering and background composition of SSNVs and SCNAs was inferred by comparing the conditional probabilities of the observed number of mutant reads out of total reads, under each scenario and CNA configuration (N_t, N_b, P_{CNA}) as follows: *Early Major or Minor:* SSNV in the major or minor allele occurred before the CNA; *Late:* SSNV occurred after the CNA; *Independent*: the SSNV and CNA occurred in independent lineages [29]. Region-specific variants were assumed to have occurred after clonal CNAs ($P_{CNA} > 0.9$ for all regions).

Benchmarking of Variant Assurance Pipeline

Dataset

The Variant Assurance Pipeline (VAP) is described within aims to: (1) achieve higher sensitivity than existing tools such as *MuTect*[2] (with default setting), for moderate to low (5%-20%) VAF at moderate sequencing depth (30-100x) which represents the bulk of current multi-region sequencing (MRS) data (**Supplementary Figure 11**); (2) exploit variants that are confidently called in one MRS sample to salvage high quality alternative reads for samples where the corresponding variant was initially not called; (3) control the false positive rate (FPR) by considering patterns of artefactual calls in the sequence alignment. To evaluate the performance of the VAP, we generated *in silico* tumor-normal WES data by mixing WES data for two individuals (NA12878 and NA18489) from the 1,000 Genome Project[30] generated using the Agilent SureSelect V2 capture kit, as was employed for many samples in this study. NA12878 was treated as the "*in silico* tumor" for which GIAB-NIST has generated a gold standard variant list (version 3.3.2 downloadable through GIAB-NIST ftp server addressed as ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/release/NA12878_HG001/latest/GRCh37/), NA18489 was used as the matched *in silico* normal sample. Fully processed read alignment files (using BWA against GRCh37) in BAM format were downloaded from 1,000 Genome FTP site (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data/). The BAM files for NA12878 and NA18489 contain 209 and 202 million aligned reads, respectively and were post-processed with the same steps as adopted in our pipeline, namely, indel realignment, base quality recalibration and flagging of duplicated reads. Subsequently, the two BAM files were randomly down-sampled using *Picard* tool (http://broadinstitute.github.io/picard/) to achieve the following mixture concentrations of NA12878 (*in silico* tumor): 10%, 16%, 20%, 30%, 40% and 100%, resulting in an expected VAF of heterozygous SNVs of NA12878 at 5%, 8%, 10%, 15%, 20% and 50%, respectively in the mixed sequencing data (**Supplementary Figure 33**). This dilution series also can serve as a source of multiple non-independent samples with varying "purity" to evaluate the VAP read salvaging strategy.

Sensitivity

VAP was applied to the *in silico* "tumor-normal" WES pairs, with read salvaging employed across multiple samples from the dilution series, with the same settings that were employed for the actual tumor sequencing data. *MuTect* (version 1.1.4) was run with the default setting. The resultant somatic SSNVs calls marked as "KEEP" were retained except for those with less than 3 alternate reads. Gold standard variants covered by at least 8 reads in each tumor sample, but NOT found in the normal sample

(VAF > 0.15 in NA18489) were used as true positives, where sensitivity was calculated as follows: $sensitivity = \frac{TP}{TP+FN}$. VAP exhibited greater sensitivity than *MuTect* default for all six *in silico* tumor samples with increasing median VAF. The improvement is particularly evident for variants in the range of 5%, 8% and 10% (**Supplementary Figure 34**), where sensitivity depends on read depth. While *MuTect* is sufficiently sensitive for highly covered variants (e.g., >100x), it failed to call 40% of true positives at 8% VAF with depth 50x. In contrast, VAP failed to call 18% of true positives under the same condition. The higher sensitivity is due to both the lowered threshold for, as well as the read salvaging process implemented in VAP.

False Positive Rate

To evaluate the FPR per million base pairs, we counted all targeted bases (based on Agilent SureSelect V2) covered at different depths in the *in silico* tumor samples. By dividing the number of false positives by the total number of bases at a range of depths, $FPR = \frac{FP}{FP+TN} \times 10^6$, *MuTect* and VAP achieved a median FPR per megabase of 2.82 and 2.29, respectively (**Supplementary Figure 35**) which is considered low[31]. Although the FPR is similar between VAP and *MuTect* for depths ranging from 50-100x, MuTect yielded a greater number of FPs than VAP for depths < 50x and >100x (above which the FPR is positively correlated with depth). The FPR for the VAP is controlled across varying depths by requiring clean read alignment patterns which have been broadly adopted for selecting "gold standard calls" [32]. It should be noted that in many sequencing applications, customized filters are often adopted to filter FPs after applying callers such as *MuTect*. However, simply adopting this strategy does not achieve comparable sensitivity to VAP.

General statistical analysis

R version 3.2.4 (2016-03-16 r70336) was used for all statistical analyses. A binomial test was performed using the R function *pbinom* (lower tail) with a 5% significance level to evaluate whether a SSNVs is region-specific where the null model that the same VAF is in the variant-missing region (given the sequencing depth). A Fisher's exact test was constructed to evaluate the association between mutational presence and spatial information (region) in bulk sequencing and single gland sequencing data using the R function *fisher.test* (two.sided, **Figure 4b**). Fisher's exact test was also used to evaluate the imbalance of cluster splits between clonal clusters and subclonal clusters (**Supplementary Figures 21-23**). The significance of Pearsons correlation between ITH metrics and the dMF/dLF of subclonal mutations, or between ITH metrics and driver fold enrichment of public mutations was assessed using the R function *cor.test* (two-tailed, **Figure 6c**, **Supplementary Figure 32**). The Wilcoxon rank sum test was used to evaluate whether two groups of values have the same mean for comparison of the fraction of shared subclonal SSNVs between patient tumors and virtual tumors with small deme size (0.5-1k), as well as comparisons of the auROC from SVMs trained with different combinations of the five ITH metrics (**Supplementary Figure 27**) using R function *wilcox.test* (two sided).

References

1. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v1* (2013).
2. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-9 (2013).
3. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079 (2009).
4. Ha, G. *et al.* TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome research* **24**, 1881-93 (2014).
5. Li, B. & Li, J.Z. A general framework for analyzing tumor subclonality using SNP array and DNA sequencing data. *Genome biology* **15**, 473 (2014).
6. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 18545-50 (2010).
7. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nature Genetics* (2015).
8. Ross-Innes, C.S. *et al.* Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nature Genetics* **47**, 1038-1046 (2015).
9. Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science (New York, N.Y.)* **346**, 256-9 (2014).
10. de Bruin, E.C. *et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **346**, 251-256 (2014).
11. Johnson, B.E. *et al.* Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science (New York, N.Y.)* **343**, 189-93 (2014).
12. Kim, H. *et al.* Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome research* **25**, 316-27 (2015).
13. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**, 406-25 (1987).
14. Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. & Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE transactions on visualization and computer graphics* **20**, 1983-92 (2014).
15. Miller, C.A. *et al.* SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS computational biology* **10**, e1003665 (2014).
16. Siegmund, K.D., Marjoram, P., Woo, Y.J., Tavare, S. & Shibata, D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc Natl Acad Sci U S A* **106**, 4828-33 (2009).
17. Diaz, L.A., Jr. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-40 (2012).
18. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155-60 (2014).
19. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A* **107**, 961-8 (2010).
20. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-58 (2013).

21. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A* **107**, 18545-50 (2010).
22. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
23. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods* **10**, 1081-2 (2013).
24. Robasky, K., Lewis, N.E. & Church, G.M. The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* **15**, 56-62 (2014).
25. Wu, T. & Watanabe, C. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005).
26. Sun, R. *et al.* Breakpointer: using local mapping artifacts to support sequence breakpoint discovery from single-end reads. *Bioinformatics (Oxford, England)* **28**, 1024-1025 (2012).
27. Ha, G. *et al.* TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res* **24**, 1881-93 (2014).
28. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
29. Li, B. & Li, J.Z. A general framework for analyzing tumor subclonality using SNP array and DNA sequencing data. *Genome Biol* **15**, 473 (2014).
30. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073 (2010).
31. Xu, H., DiCarlo, J., Satya, R., Peng, Q. & Wang, Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. *BMC Genomics* **15**, 244 (2014).
32. Alioto, T.S. *et al.* A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nature Communications* **6**, 10001 (2015).