SUPPLEMENTAL INFORMATION

Subclones Dominate at MDS Progression Following Allogeneic Hematopoietic Cell Transplant

Meagan A. Jacoby,[1#] Eric J. Duncavage,[2#] Gue Su Chang,[3#] Christopher A. Miller,[1,3] Jin Shao,[1] Kevin Elliott,[1] Joshua Robinson,[1] Robert S. Fulton,[3] Catrina C.Fronick,[3] Michelle O'Laughlin,[3] Sharon E. Heath,[1] Iskra Pusic, [1] John S. Welch[1], Daniel C. Link,[1] John F. DiPersio,[1] Peter Westervelt,[1] Timothy J. Ley,[1] Timothy A. Graubert,[4] Matthew J. Walter[1*]

[1] Washington University School of Medicine, Department of Medicine, Division of Oncology, St. Louis, MO, USA

[2] Washington University School of Medicine, Department of Pathology & Immunology, St. Louis, MO, USA

[3] Washington University School of Medicine, McDonnell Genome Institute, St. Louis, MO, USA

[4]Massachusetts General Hospital Cancer Center, Boston, MA, USA

[#] These authors contributed equally to this work.

* Corresponding author

Table of Contents

Supplemental Experimental Methods

Supplemental Data

Supplemental Experimental Methods

A. Somatic variant analysis and validation

Exome sequencing was aligned to the human reference sequence build GRCh37-lite-build37 using bwa(1) version 0.5.9 (params: -t 4 -q 5), then merged and deduplicated using picard version 1.46 (https://broadinstitute.github.io/picard/).

All variants were called from tumor samples with their matched skin (or normal) samples. Single nucleotide variants (SNVs) were detected using the union of four callers: 1) samtools(2) version r982 (params: mpileup -BuDS) intersected with Somatic Sniper version 1.0.2(3) (params: -F vcf -q 1 -Q 15) and processed through false-positive filter v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15 --min-mapping-quality 40 --min-somatic-score 40), 2) VarScan(4) version 2.3.6 (params: --nobaq --version r982) filtered by varscan-high-confidence filter version v1 and processed through false-positive filter v1 (params: --bam-readcount-version 0.4 --bam-readcount-min-base-quality 15), 3) Strelka(5) version 1.0.11 (params: isSkipDepthFilters = 1), and 4) MuTect version 1.1.4 (--number-of-chunks 50) with COSMIC v54 (6).

Insertions and deletions (INDELs) were detected using the union of four callers: 1) GATK(7) somatic-indel version 5336 filtered with false-indel v1 (--bam-readcount-min-base-quality 15 --bam-readcount-version 0.4) (2) pindel(8) version 0.5 filtered with pindel-somatic-calls and VAF filters (params: --variant-freq-cutoff=0.08), and with pindel-read-support filter, 3) VarScan(4) version 2.3.6 filtered by varscan-high-confidence-indel and false-indel filters (params: --bam-readcount-min-base-quality 15 --bam-readcount-version 0.4) and 4) Strelka(5) version 1.0.11 (params: isSkipDepthFilters = 1).

SNVs and INDELs were further filtered by removing artifacts found in a panel of 151 normal exomes, removing sites that exceeded 1% frequency in the TCGA exome sequencing

project cohort, and then using a bayesian classifier (https://github.com/genome/genome/blob/master/lib/perl/Genome/Model/Tools/Validation/IdentifyOutliers.pm) and retaining variants classified as somatic with a binomial log-likelihood of at least 3 (params: --llr-cutoff 3 --tumor-purity 0.95) and 20x sequencing depth.

Resulting SNVs and INDELs were manually inspected by using the Integrative Genomics Viewer (IGV). As a result, germline mutations, low-quality mutations, sequencing artifacts owing to paralogous DNA sequences for example, and ambiguous variant calling were removed.

B. Ultra-deep error-corrected sequencing

Mutations for ultra-deep error-corrected sequencing were chosen from exome or RMG based sequencing data. Genomic DNA was first digested using a mixture of restriction endonucleases found in the Agilent Haloplex kit (Agilent Technologies, Santa Clara, CA). Probes (designed to cover both DNA strands), 10bp degenerate barcodes, and sample-specific indexes were synthesized by Agilent Technologies. Following 16 hour hybridization and magnetic bead enrichment, libraries were amplified for 23 PCR cycles using primers with tailed Illumina sequencing motifs. Approximately 8-12 samples were pooled and sequenced on Illumina HiSeq 2500 lanes using V3 chemistry (Illumina, Carlsbad CA). Barcoded FASTQ data was demultiplexed using a custom python script that adds degenerate barcode information to FASTQ files. Data was then aligned to build GRCh37-lite-build37 using bwa mem (version 7.9.a) with default parameters. Aligned BAMs were then analyzed using the Barcrawler pipeline (Abel and Duncavage, unpublished data), a GATK locus walker written in Java, with the following parameters *-mmq 20 -mbq 20 -minCtBC 3 -dcov 500000 -discardN 1 -minOffset 5 -maxNM 3*. The resulting list of variants were filtered to include only mutations present in at least

2 unique read families (each with a minimum of 3 members of which the variant was present in >90% of member reads); mutations with VAFs <0.1% were discarded.

We calculated whether each detected variant was above the background "noise" level on a per-assay and position-by-position basis as follows: For each variant, readcounts were gathered from all other samples (tumor and normal) at the same location. Samples with variant VAF above 20% were excluded from the background calculation. A p-value was obtained via fisher's exact test, comparing the reference and variant reads at a site to the number of reference and variant reads in all other samples. Multiple testing correction was applied with the R p.adjust function (default params), then variants with an adjusted p-value of less than 0.1 were retained. The process was then repeated iteratively, with subsequent background calculations excluding all variants retained in previous rounds. This was repeated until no new variants were identified (3 rounds).

C. Copy number analysis

Copy number analysis was performed on EES data using the Copy Change Assessment Tool 2 (CopyCAT2) R software package with default parameters (Abel and Duncavage, unpublished data, https://github.com/abelhj/cc2). Coverage was first calculated using bedTools coveragBed with default parameters.(9) To call somatic copy number alterations (CNAs), pooled skin data (minimum 5) from patients in this study was used as a normal control. To ensure that constitutional CNAs were not called as somatic variants, CNA analysis was also performed on each patient's skin sample using the same series of pooled normal skin controls; CNAs that appeared in both paired skin and paired pre-treatment marrow samples were considered constitutional variants are removed from the analysis. Somatic CNAs were called if determined to be significant by the software using a p-value cut off of $1x10^{-6}$ for allele frequency and $1x10^{-6}$ for coverage. Further, for regions identified by SNP array as copy variable, but not called as significant by CopyCAT2 (likely due to differences in covered regions between the array and

RMG panel), normalized log 2 coverage data corresponding to those regions was manually reviewed.

Genomic DNA samples from the normal and bone specimens were hybridized to Illumina HumanOmniExpress-24 Bead Chip array and iScan System (Illumina, Inc.). Analysis of copy number alterations and copy neutral loss of heterozygosity was performed using the Partek Genomics Suite (Partek, Inc), as previously described (10). Only mutations occurring in diploid regions of the genome (copy number = 2) were used to perform mutation clustering, which were used to define clones. We also excluded mutations occurring in areas of uniparental disomy, or the X and Y chromosome from clustering analysis.

D. Analysis of clonal architecture and evolution from sequencing data.

In order to determine tumor clonality and track the temporal patterns of subclonal evolution during treatment, we characterized the subclonal composition of each patient using the SciClone package for R (11). This bioinformatics tool infers clonal architecture by analyzing the variant allele fractions (VAFs) and clustering somatic mutations, and we used it with default parameters. We primarily focused on somatic variants in copy-number neutral loci (i.e., diploid, and excluding haploid sex chromosomes) without loss of heterozygosity (LOH) with a few exceptions, which allows more accurate estimation of VAF for high-confidence inference of tumor clonality. In those exceptions, we corrected the VAF of somatic variants using their copy number estimate. For short insertion and deletions, only indels of less than or equal to 4 bp in length were included in analysis.

Structural variants were identified as above, and the percentage of cells harboring the structural variants was estimated by using the formula [(50- mean deleted allele frequency) x 2]. The percentage cells harboring structural variants was assigned a clone after manual review and comparison to VAFs.

Tumor evolution models were generated using ClonEvol (12) with phylogeny plots and “monoclonal” as the “cancer.initiation.model" parameter. The evolutionary models of each patient was visualized using the fishplots package for R (13).

Supplemental Data

A. Supplemental Tables 1-5. See Excel files.

B. Supplemental Figures 1-8. See below.

C. Supplemental Figure Legends

Supplemental Figure 1. Sequencing metrics. (**A**). Sequencing coverage depths obtained for enhanced exome cases. (**B**) Box and whiskers plot of coverage depths for enhanced exomes. The line represents the median coverage depth; boxes represent upper and lower quartiles; lines represent minimum and maximum coverage values. Skin (normal DNA source) and bone marrows samples are denoted by their unique patient number (UPN) and the time-point post-banking. (**C**). Sequencing coverage depths obtained for standard validation capture-based sequencing. (**D**) Box and whiskers plot of coverage depths for validation capture sequencing. The line represents the median coverage depth; boxes represent upper and lower quartiles; lines represent minimum and maximum coverage values. Skin (normal DNA source) and bone marrows samples are denoted by their UPN and the time-point post-banking. All samples were sequenced deeply, with mean tumor coverage of 246x in the exomes, 1146x over key MDS genes, and 537x in the validation sequencing. These data provided three key pieces of information: 1) highly confident somatic mutation calls; 2) the variant allele fraction (VAF) of each mutation; and 3) copy number alteration (CNA) calls, which also incorporated data from single-nucleotide polymorphism (SNP) arrays (Supplemental Table 4).

Supplemental Figure 2. Structural variants pre- and post-transplant. (A, D-H). Dynamic changes in the B (reference allele determined by NCBI refseq) allele frequency (BAF) in MDS cells for selected chromosomes at first sampling (Day 0 pre)**,** pre-transplant (pre), and post-transplant (post). The BAF was determined using heterozygous (HET) single-nucleotide polymorphisms (SNPs) identified in normal tissue (skin) by whole exome sequencing (grey).

The BAF was then plotted for MDS samples (red). Divergence of the BAF away from 50% is directly proportional to the percentage of cells harboring loss of heterozygosity (LOH) due to a copy number alteration or uniparental isodisomy (**Supplemental Table 4**). Post- transplant samples were sorted for MDS myeloblasts to reduce donor SNP contamination. For UPN 145094 **(B)**, a small interstitial deletion on chromosome 2 identified with HumanOmniExpress-24 Bead Chip Array, and (**C**) copy number alterations of chromosome 20 as determined using exome sequencing data and analysis by CopyCAT2 are additionally shown.

Supplemental Figure 3. Mutation clustering at first sampling, pre-transplant, and relapse. (**A-H**) The variant allele fractions (VAFs) of all validated mutations in diploid genomic regions at relapse (y-axis) are compared to first sampling (Day 0 pre) (left panel, x-axis) and pre-transplant (x-axis, right panel), if available. Somatic mutations in 285 genes known to be recurrently mutated in myeloid malignancies are labeled (recurrently mutated genes, RMG), if detected. Unsupervised clustering of individual mutation VAFs was performed and mutation clusters are shown. Mutations are reported using build NCBI build 37 coordinates. The '++' sign indicates canonical hot spot mutations in an RMG gene. VAFs represent the summation of read counts from exome and validation standard capture-based sequencing. SNV, single nucleotide variant; indels, insertion or deletion. RMG, recurrently mutated gene.

Supplemental Figure 4. Clonal evolution summary models. (**A-F**) Summary of clonal evolution from first sampling (D0) to relapse for selected UPNs. Days are numbered relative to first sampling, and selected post-transplant days are shown in parentheses. Selected somatic alterations, some copy number corrected, are shown. UPN 145094 has two equally possible evolution models (A and B) that cannot be resolved. Chr2* (**A and B**) small interstitial deletion; CBL* (**D**) variant allele fraction could not be definitively quantified and assigned to a clone.

Supplemental Figure 5. Tumor phylogeny models. (**A-T**) Models of tumor phylogeny as inferred from the Clonevol package. Left panel shows the evolutionary tree and the right panel is the proportion of the tumor cells composed of each subclone at each time-point. UPN 145094 has two equally possible evolution models (A and C) that cannot be resolved.

Supplemental Figure 6. Loss of heterozygosity analysis for the HLA, PD-L1, and PD-L2 loci. Panels **(A-R)** (left and middle panel) show copy number alterations of the HLA and PD-L1/PD-L2 loci as determined using exome sequencing data and analysis by CopyCAT2 at first sampling (Day 0) and relapse. A paired skin sample was used as a source of normal DNA. (Right panel) To identify potential loss of heterozygosity caused by uniparental isodisomy, the B (reference allele determined by NCBI refseq) allele frequency (BAF) was determined at first sampling and relapse using heterozygous (HET) single-nucleotide polymorphisms (SNPs) identified in normal tissue (skin) by exome sequencing (grey). The BAF was then plotted for MDS samples (red). Divergence of the BAF away from 50% is directly proportional to the percentage of cells harboring loss of heterozygosity (LOH) due to a copy number alteration or uniparental isodisomy **(Supplemental Table 4).** Post-transplant samples were sorted for MDS myeloblasts to reduce donor SNP contamination. Genomic coordinates were determined using Ensembl gene annotation data (GRCh37 build). In sample 145094, massive structural changes (including many whole-chromosome duplications) affect most chromosomes, including chromosome 6, but no definitive evidence of abnormalities at the HLA locus are detected.

Supplemental Figure 7. Spectrum of base substitutions in emerging clones in patients treated with decitabine. Spectrum of single nucleotide variant (SNV) base substitutions present in three patients with acute myeloid leukemia (AML) treated on a clinical trial with decitabine (14). SNV mutation spectra are shown at presentation, prior to treatment (AML), and at relapse. Decitabine therapy and sequencing methods are as previously described (14).

Patients from panels **(A-C)** received a total of 6, 7, or 8 cycles of decitabine, respectively. The relapse samples show a greater proportion of C to G transversions, consistent with decitabine induced mutations.

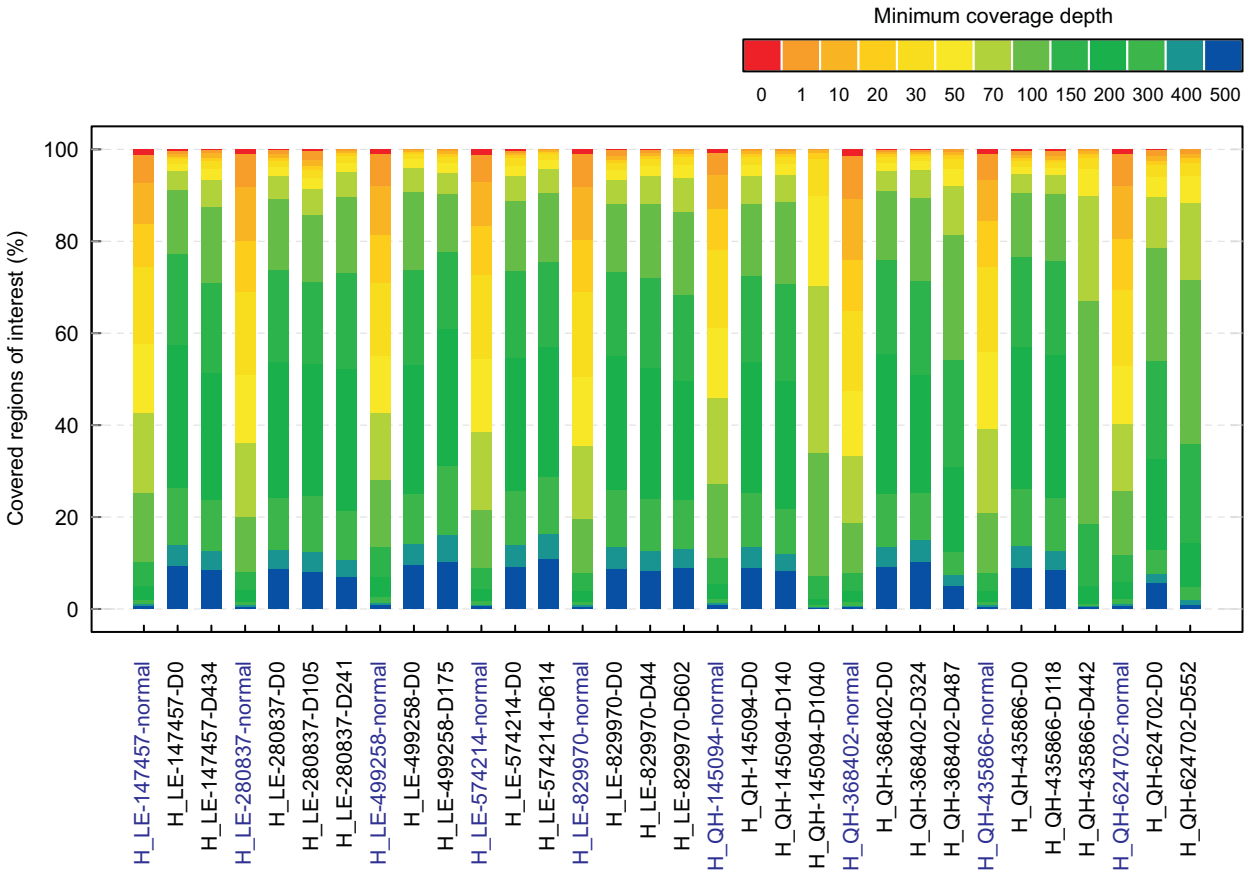
Supplemental Figure 8. Dynamic changes in mutation variant allele fractions during treatment. (A-H) Left, gene mutation variant allele fractions (VAFs) in diploid genomic regions are displayed across all time-points from first sampling (D0 pre) to relapse and labeled by mutation cluster. The heavy grey line indicates the blast percentage. Right, VAFs from 285 genes known to be recurrently mutated in myeloid malignancies (recurrently mutated genes, RMG) are displayed across all time-points and labeled by mutation cluster. Thin grey lines represent VAFs from non-RMG genes. VAFs represent the summation of read counts from exome and validation standard capture-based sequencing, if available. Mutations are reported using build NCBI build 37 coordinates. The '++' sign indicates canonical hot spot mutations in RMGs. Days are numbered relative to first sampling, and selected post-transplant days are shown in parentheses. SNV, single nucleotide variant; indels, insertion or deletion.

D. Supplemental References

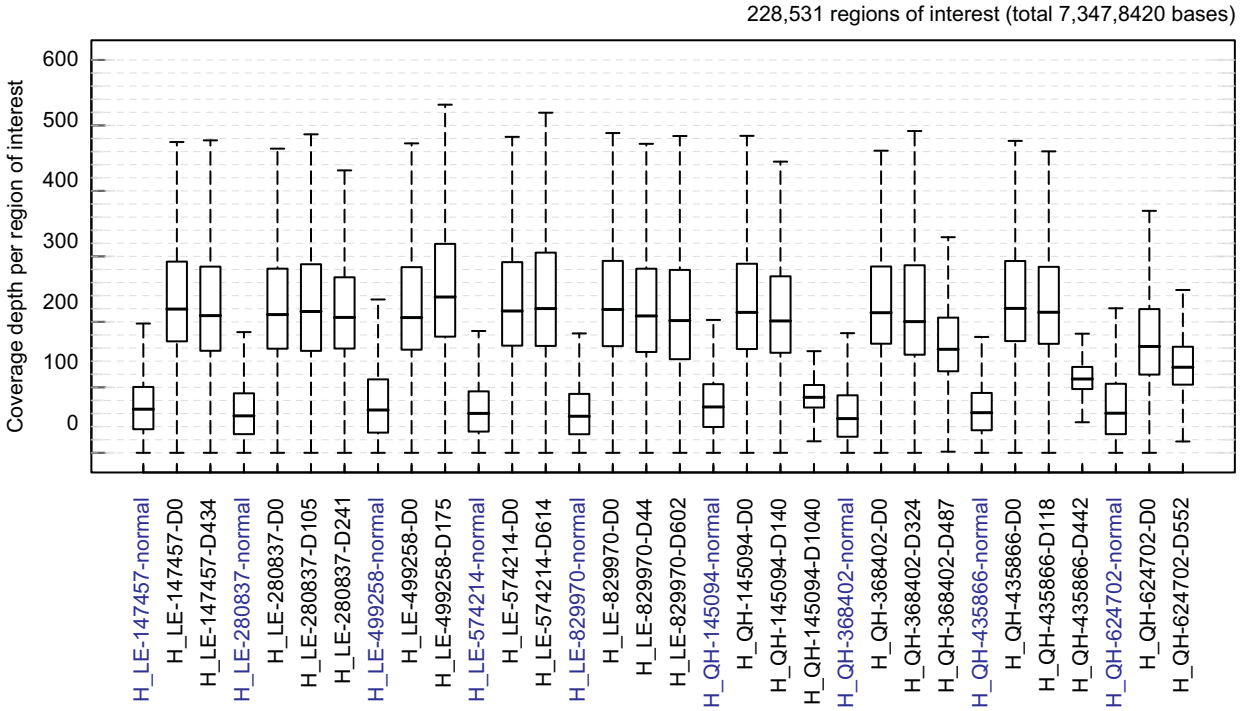
1. Li H, and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England).* 2009;25(14):1754-60.
2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
3. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics (Oxford, England).* 2012;28(3):311-7.
4. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research.* 2012;22(3):568-76.
5. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, and Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics.* 2012;28(14):1811-7.
6. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology.* 2013;31(3):213-9.
7. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research.* 2010;20(9):1297-303.
8. Ye K, Schulz MH, Long Q, Apweiler R, and Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics (Oxford, England).* 2009;25(21):2865-71.
9. Quinlan AR, and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England).* 2010;26(6):841-2.
10. Walter MJ, Shen D, Shao J, Ding L, White BS, Kandoth C, et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia.* 2013;27(6):1275-82.
11. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS computational biology.* 2014;10(8):e1003665.
12. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, et al. ClonEvol: clonal ordering and visualization in cancer sequencing. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2017.
13. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, et al. Visualizing tumor evolution with the fishplot package for R. *BMC genomics.* 2016;17(1):880.
14. Welch JS, Petti AA, Miller CA, Fronick CC, O'Laughlin M, Fulton RS, et al. TP53 and Decitabine in Acute Myeloid Leukemia and Myelodysplastic Syndromes. *The New England journal of medicine.* 2016;375(21):2023-36.

SFig1

A. 23 tumor and 9 normal samples in EES (enhanced exome sequencing)

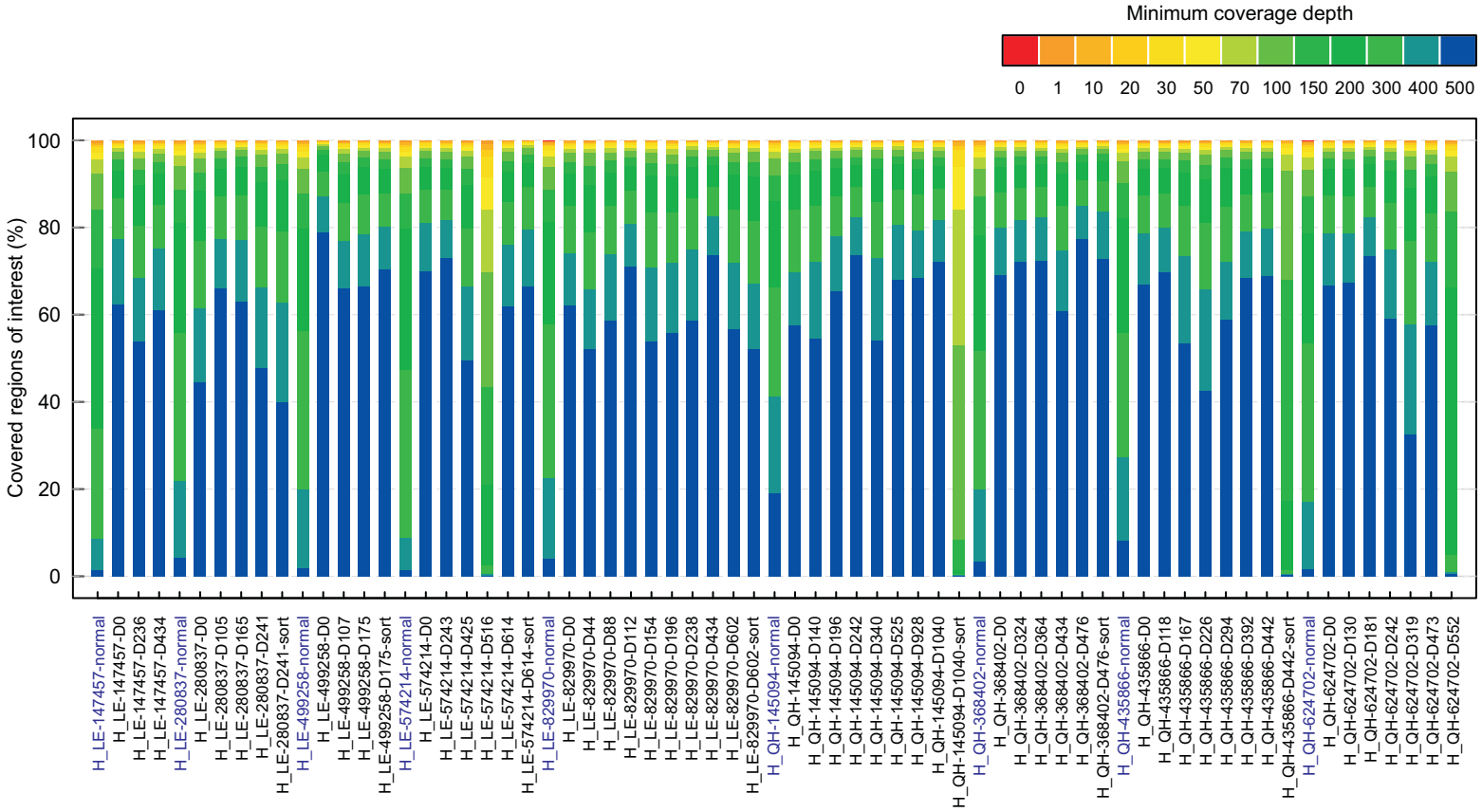


B. 23 tumor and 9 normal samples in EES (enhanced exome sequencing)

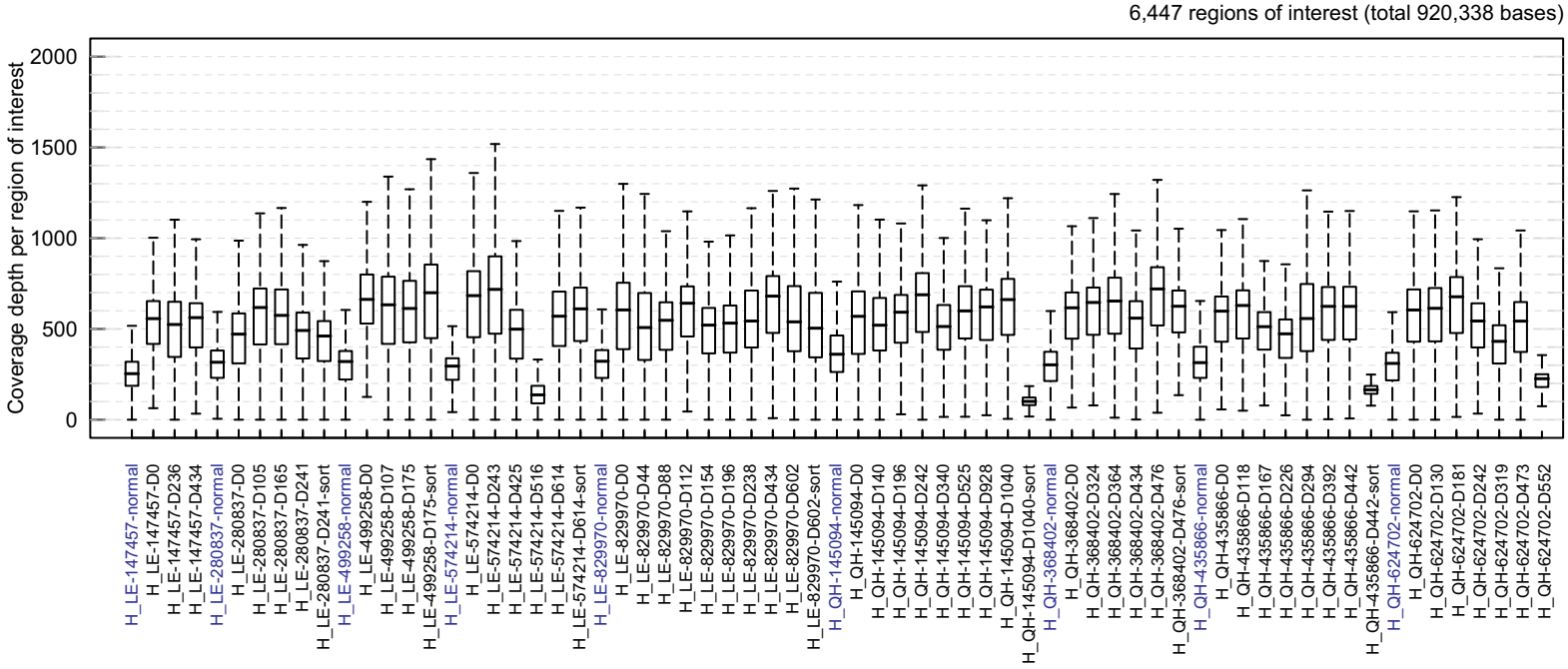


SFig1

C. 58 tumor and 9 normal samples in validation capture



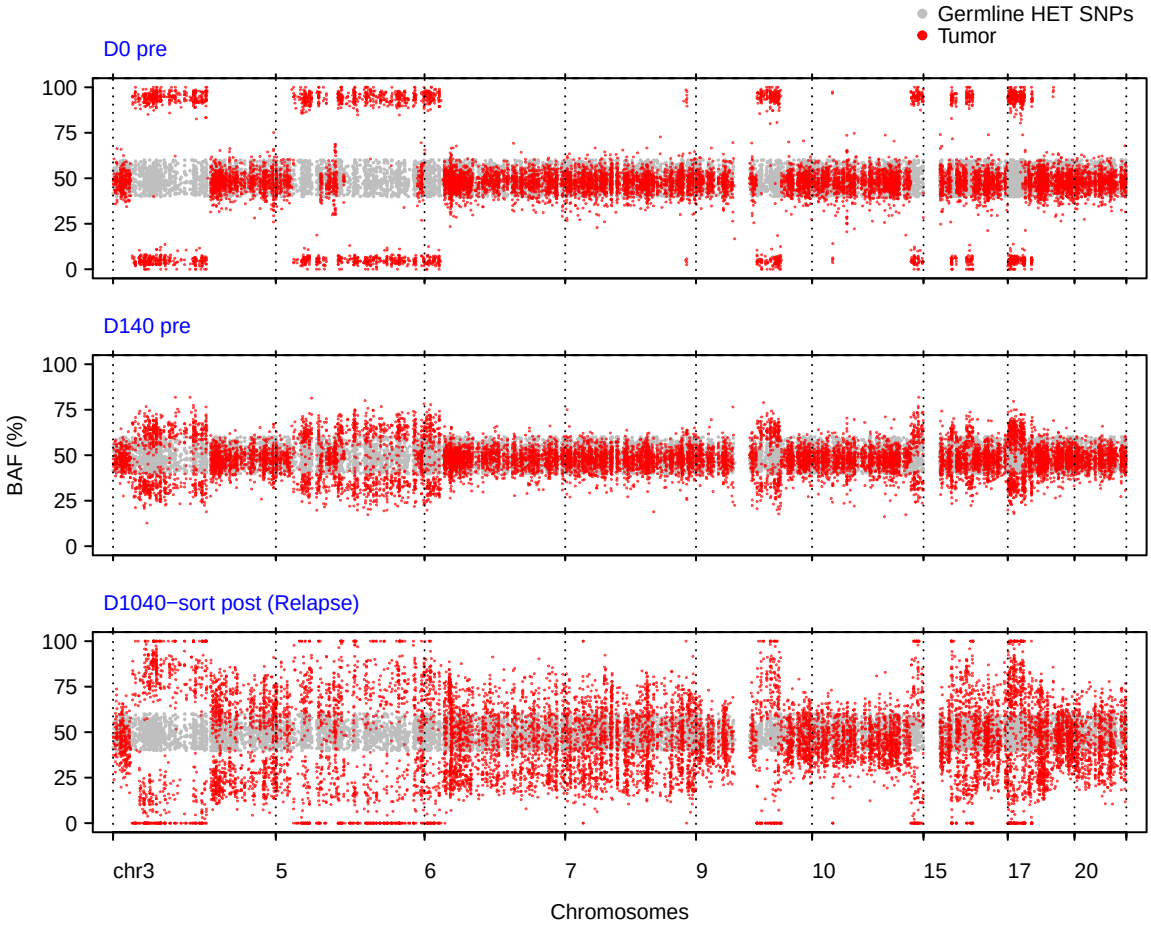
D. 58 tumor and 9 normal samples in validation capture



SFig2

A

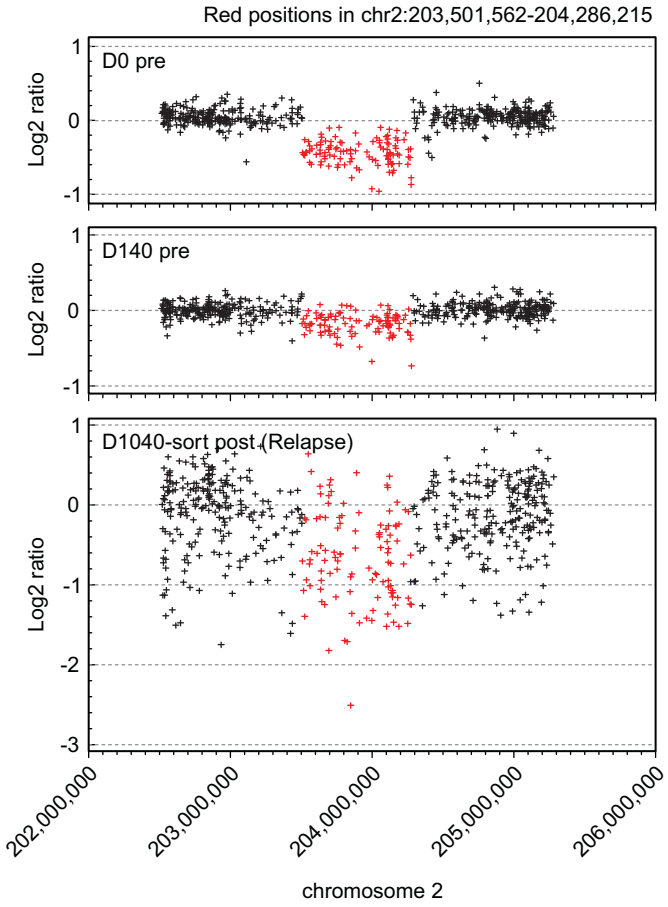
UPN 145094



SFig2

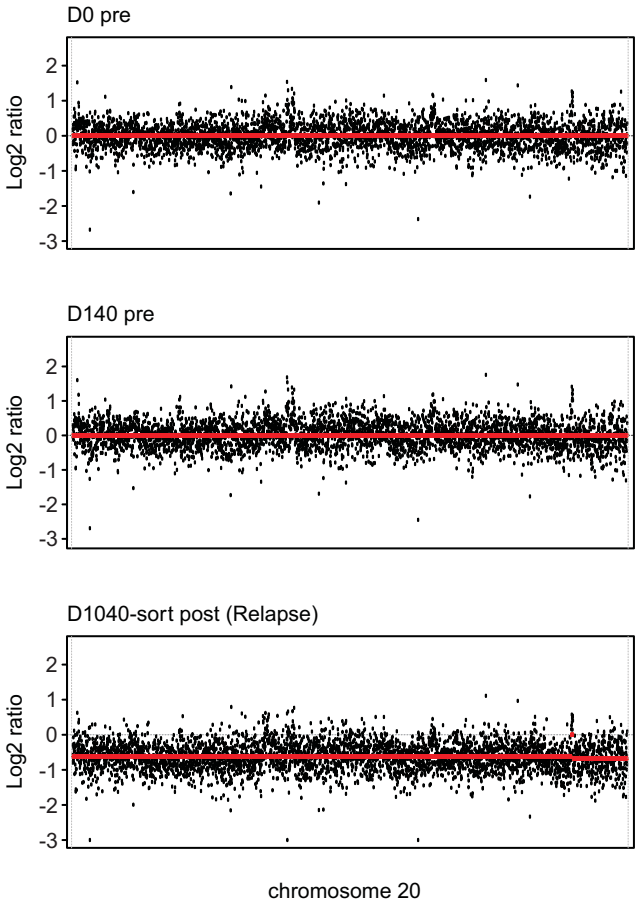
B

UPN 145094



C

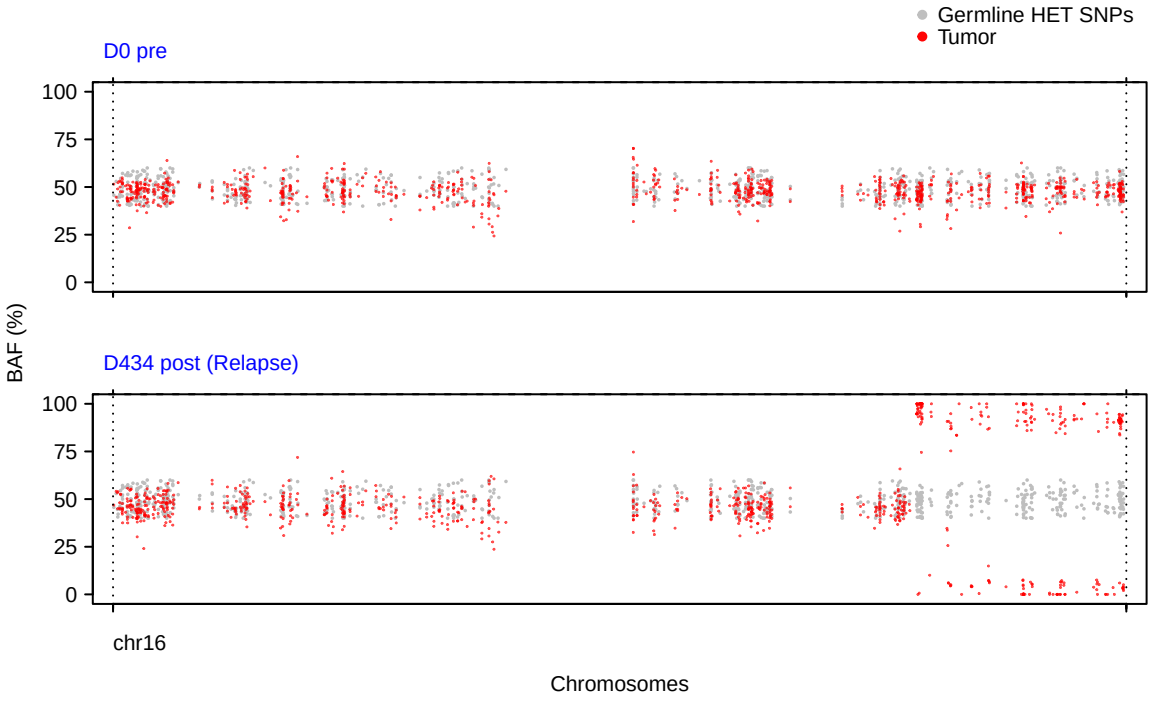
UPN 145094



SFig2

D

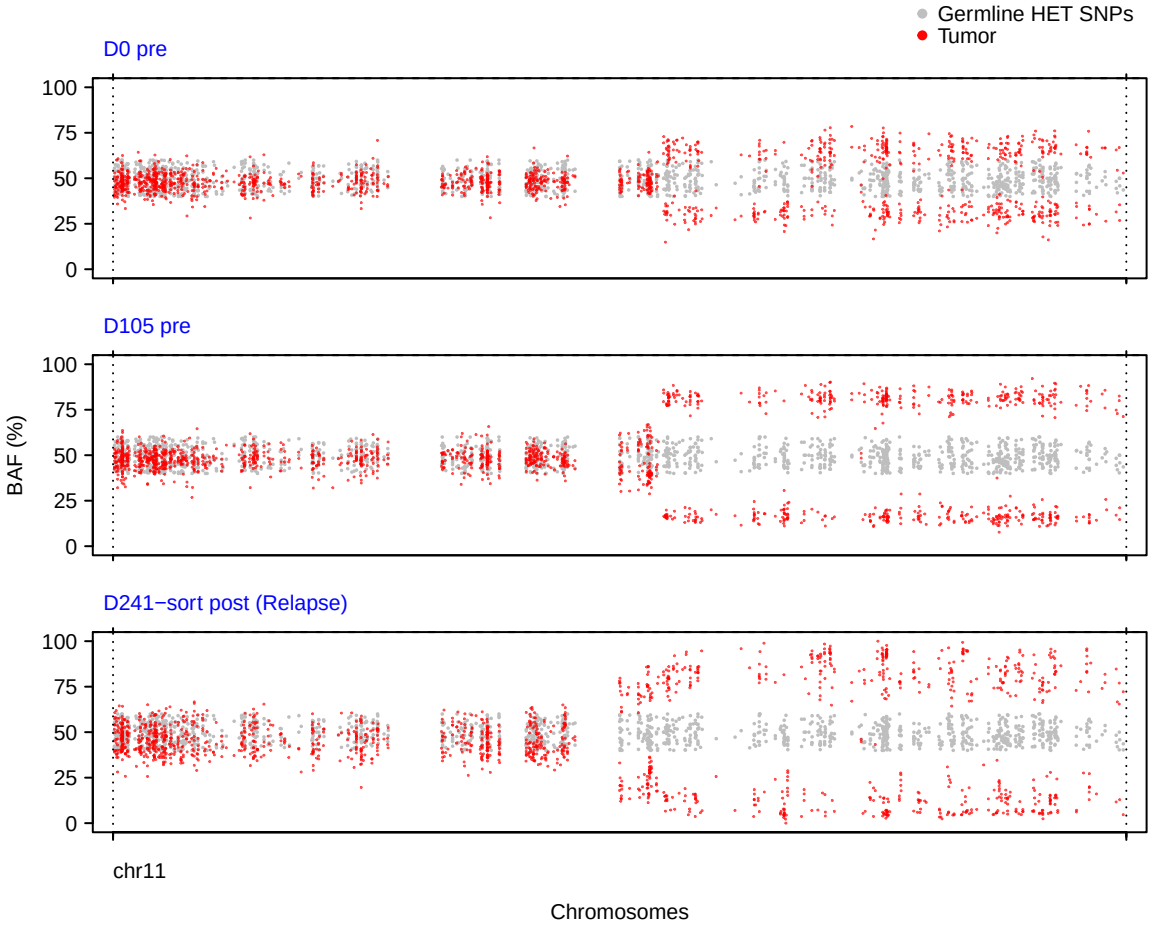
UPN 147457



SFig2

E

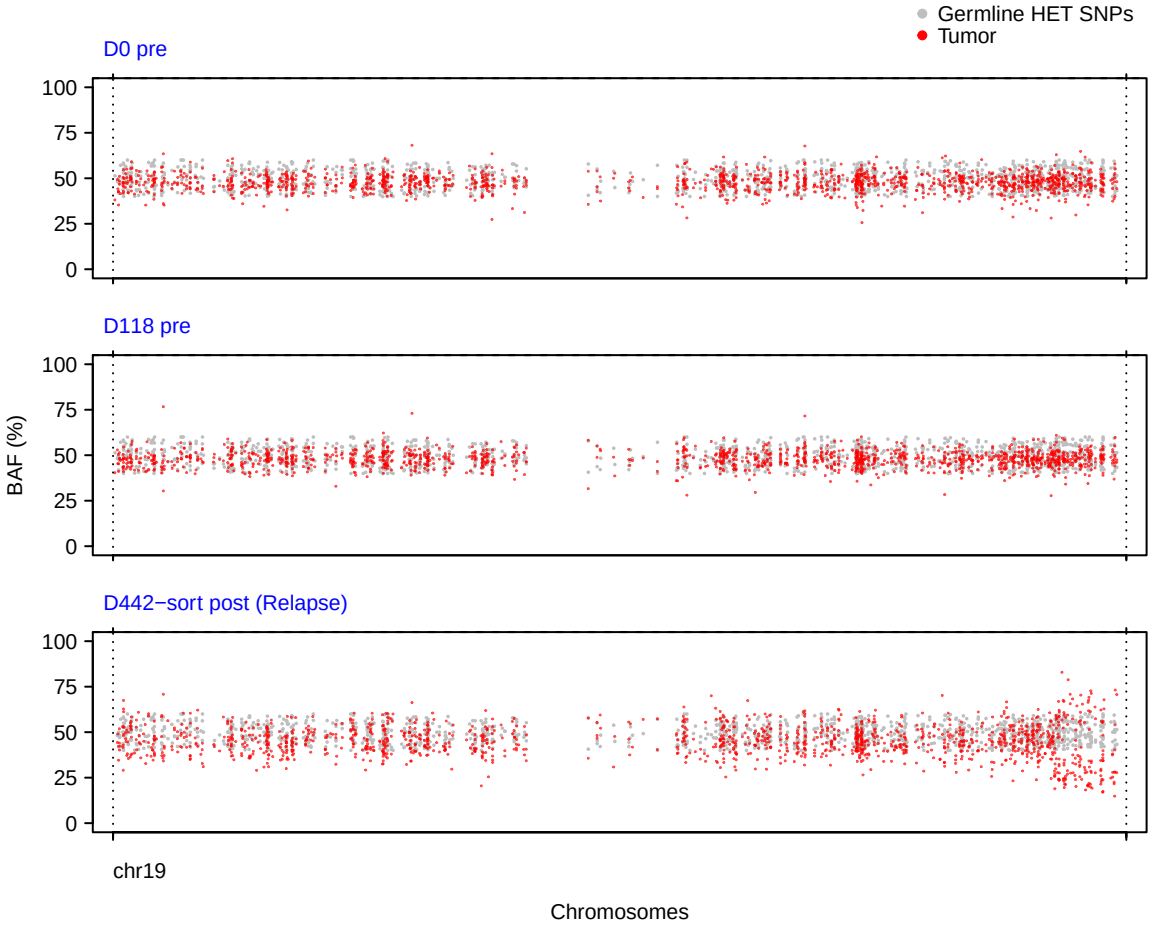
UPN 280837



SFig2

F

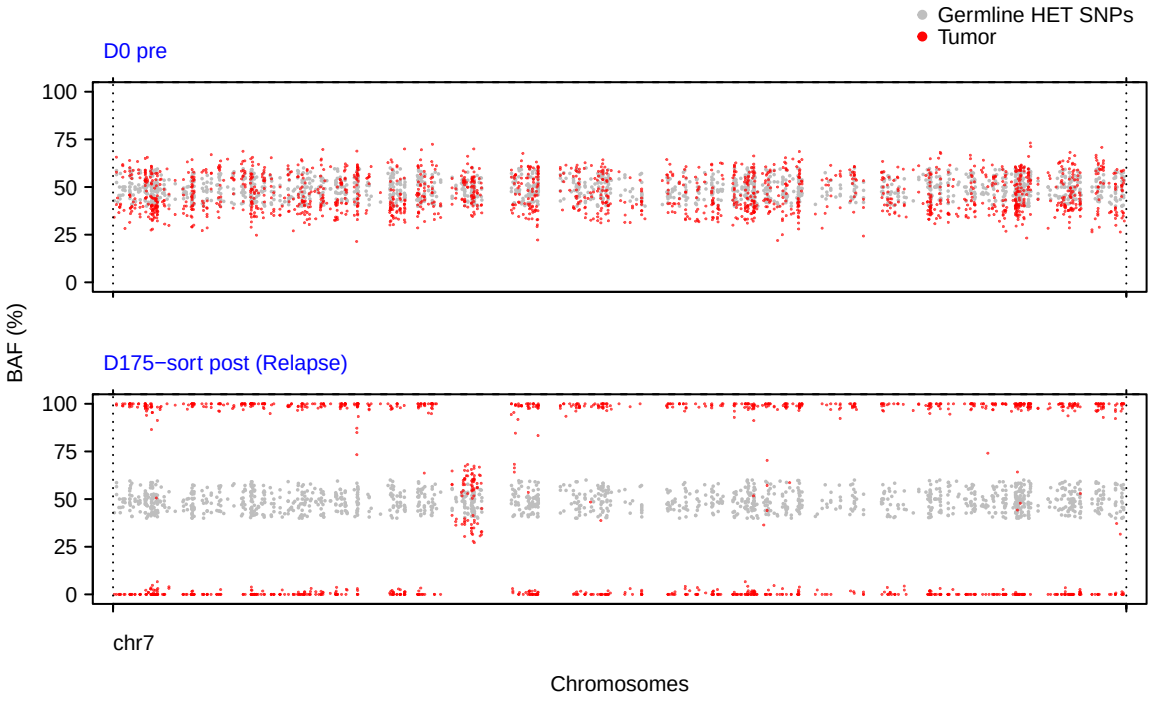
UPN 435866



SFig2

G

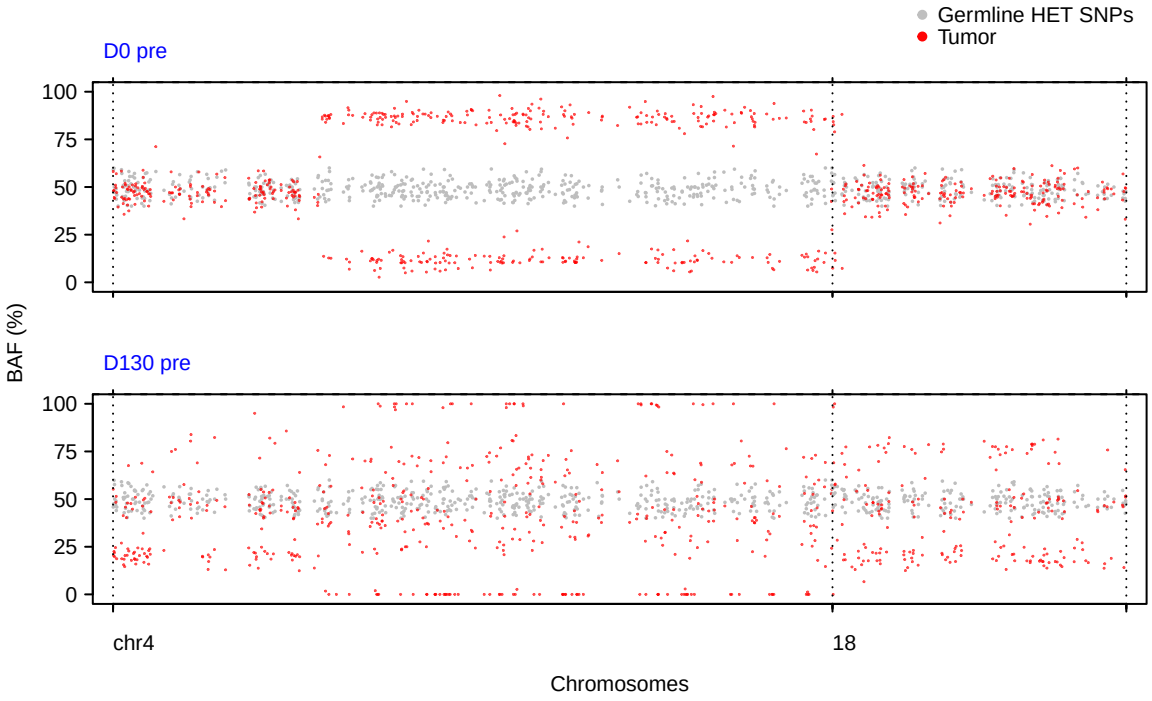
UPN 499258



SFig2

H

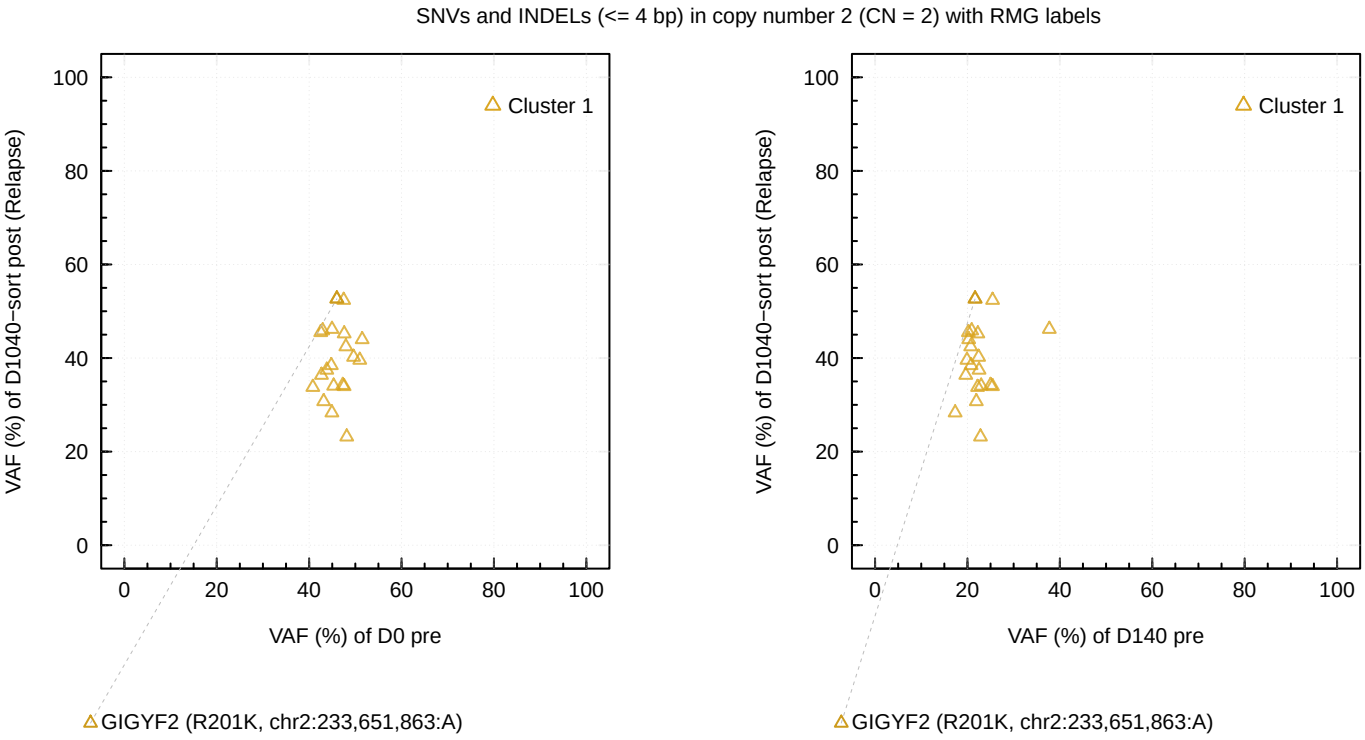
UPN 624702



SFig3

A

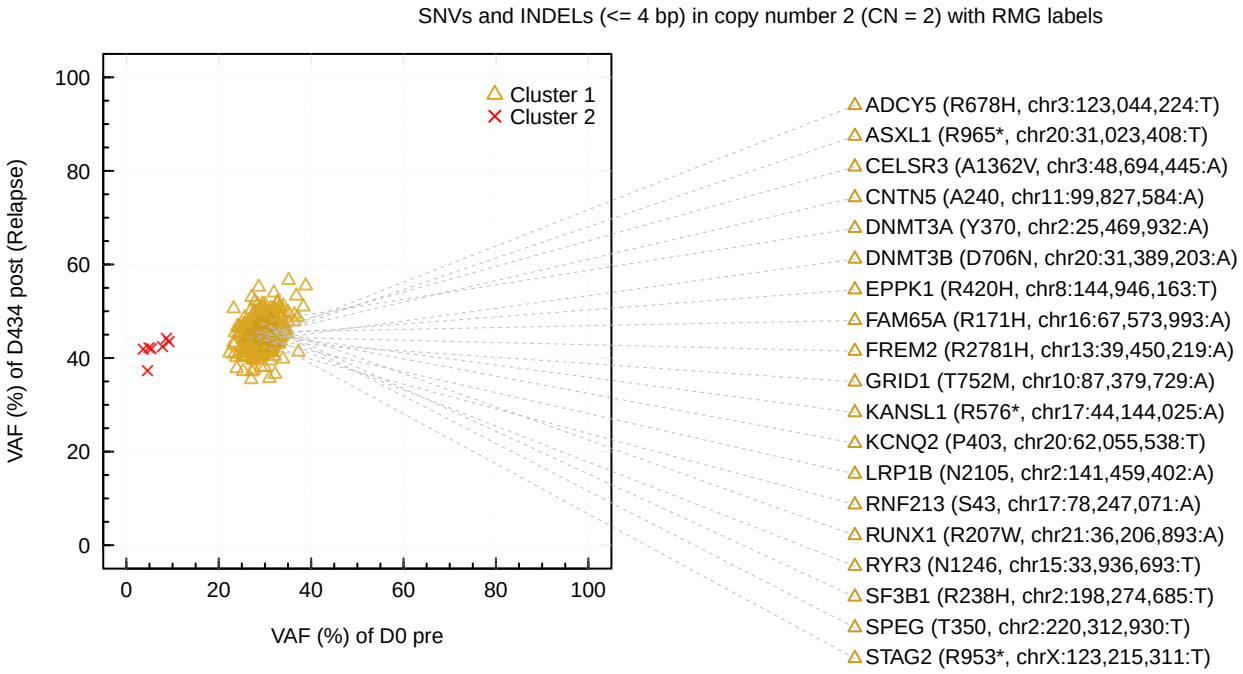
UPN 145094



SFig3

B

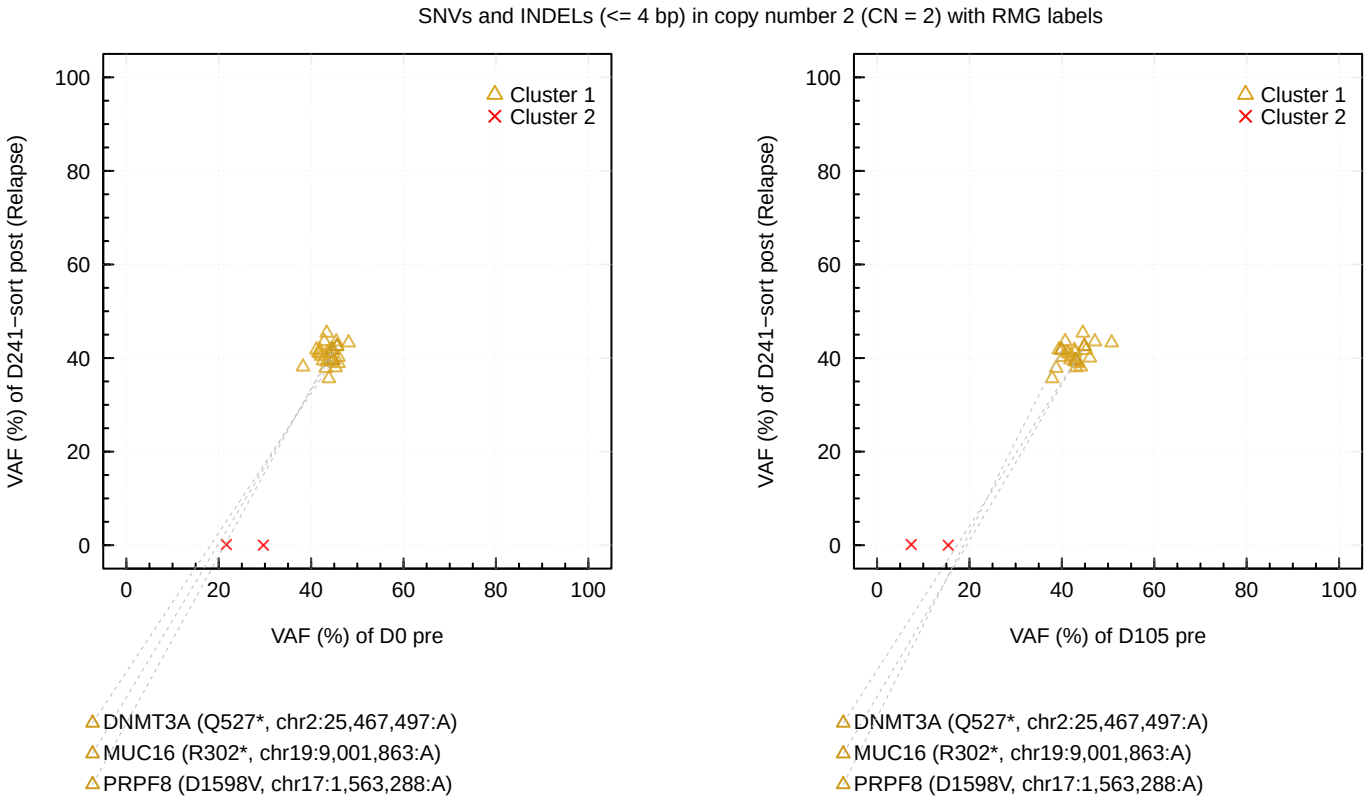
UPN 147457



SFig3

C

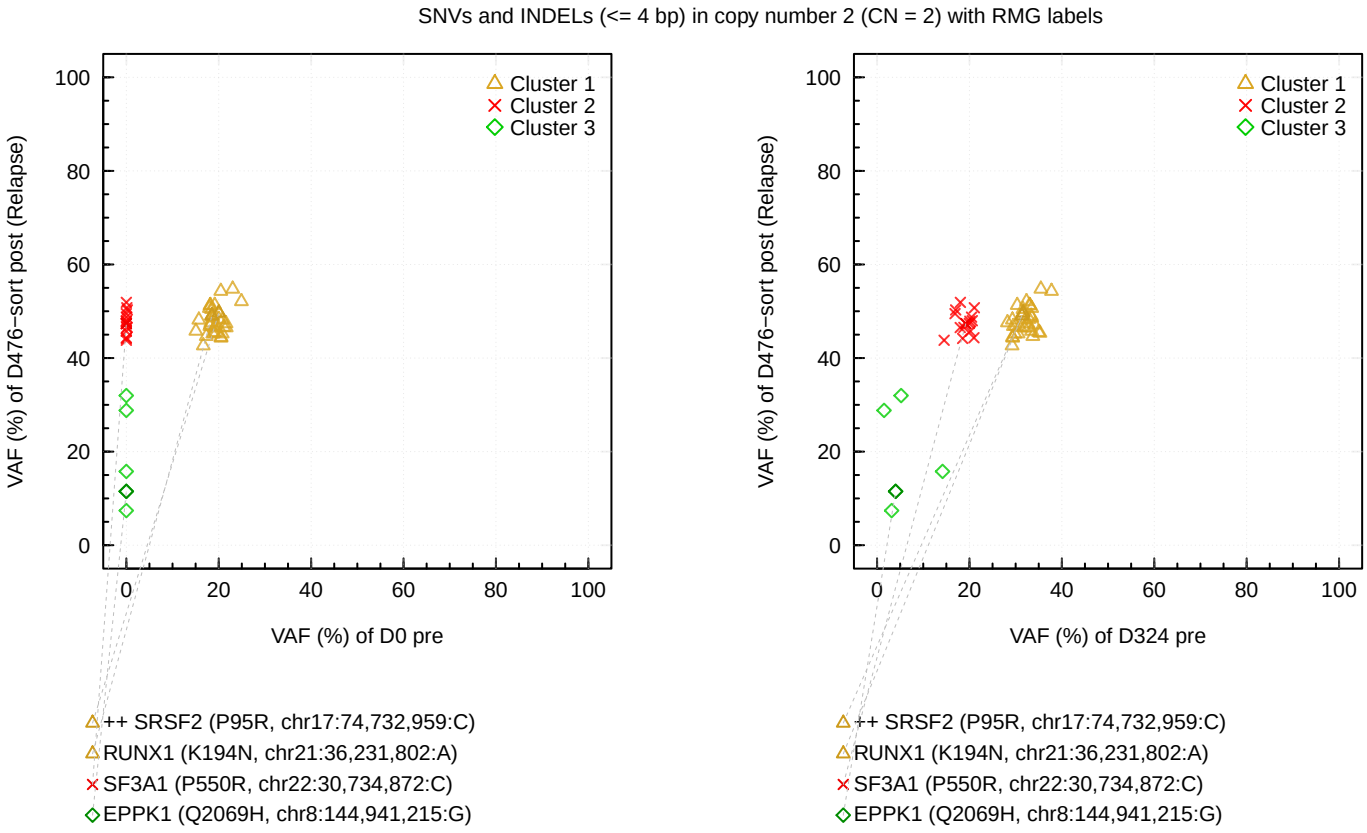
UPN 280837



SFig3

D

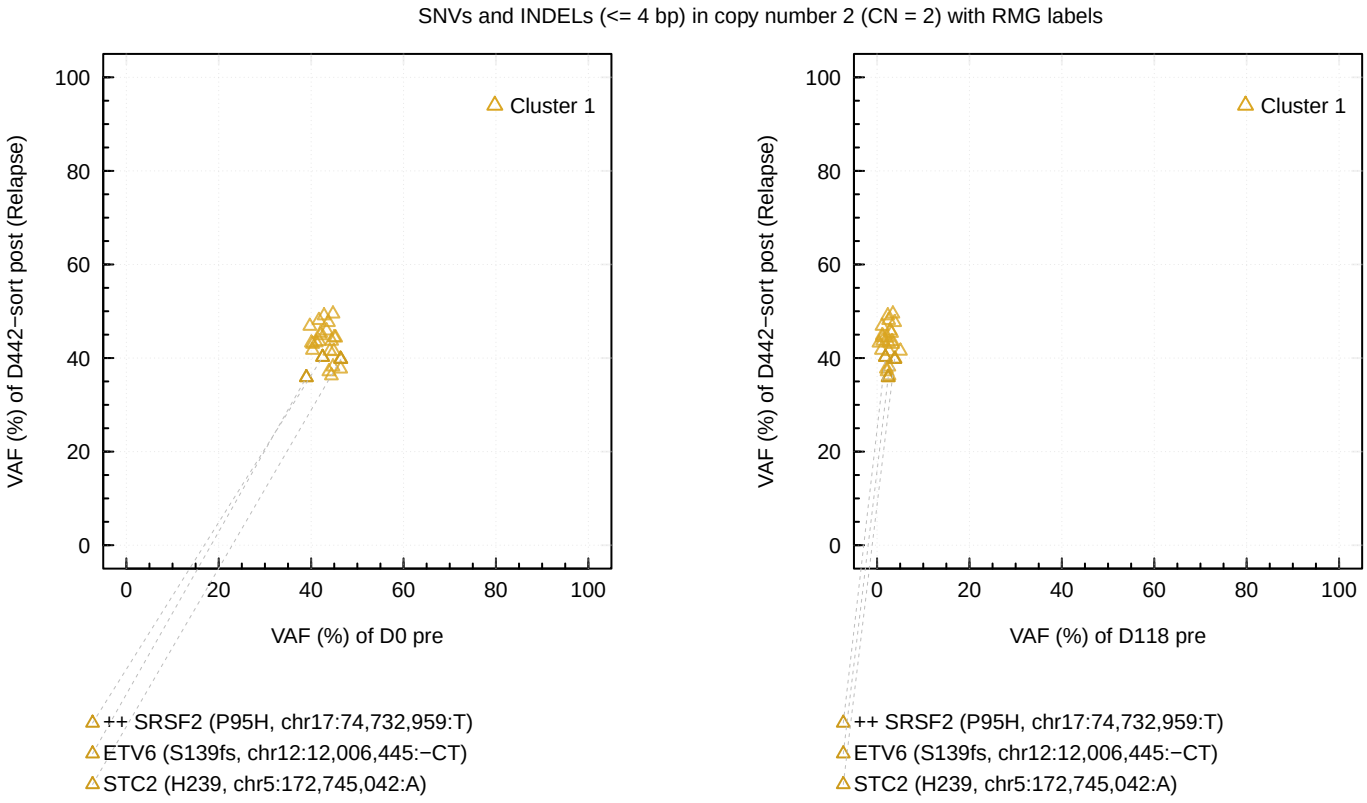
UPN 368402



SFig3

E

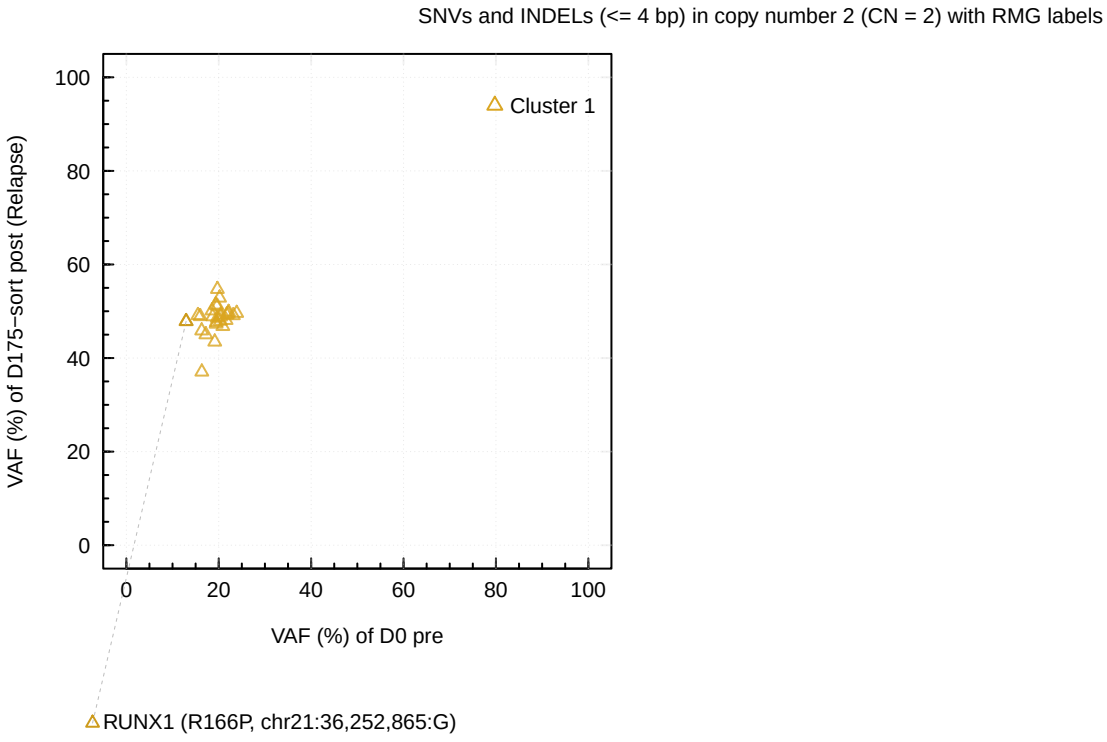
UPN 435866



SFig3

F

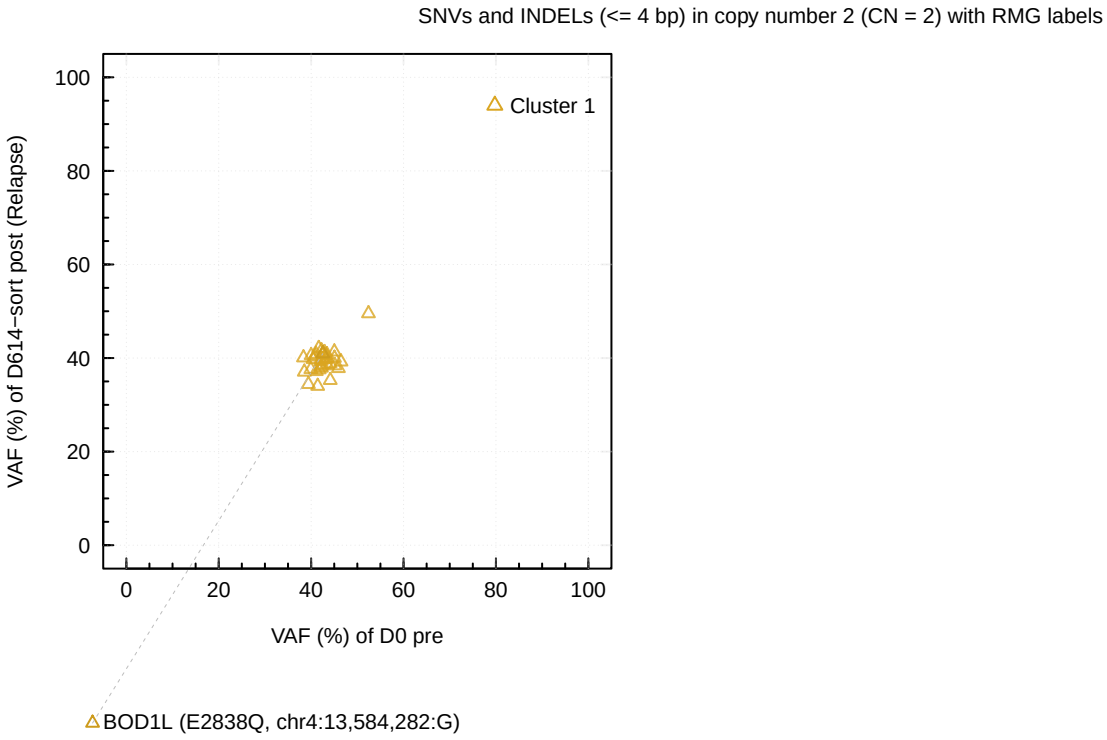
UPN 499258



SFig3

G

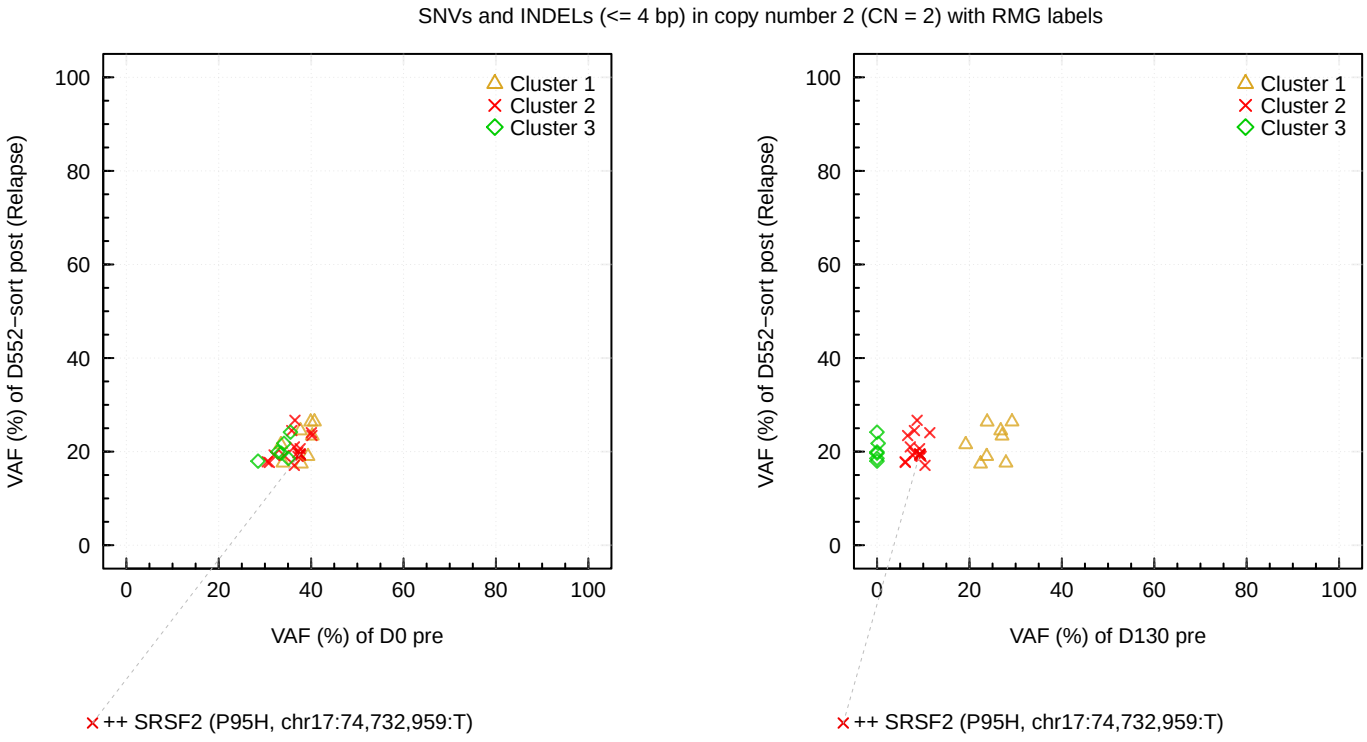
UPN 574214



SFig3

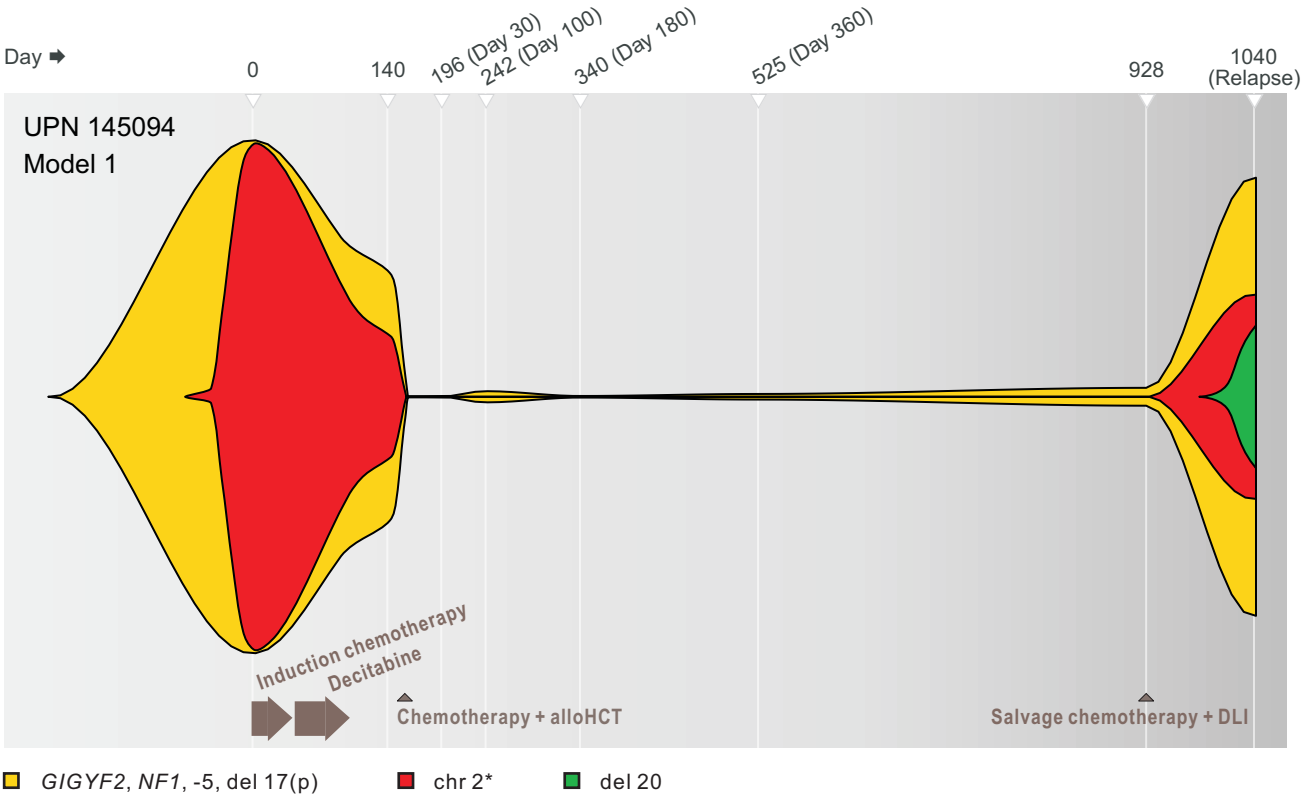
H

UPN 624702



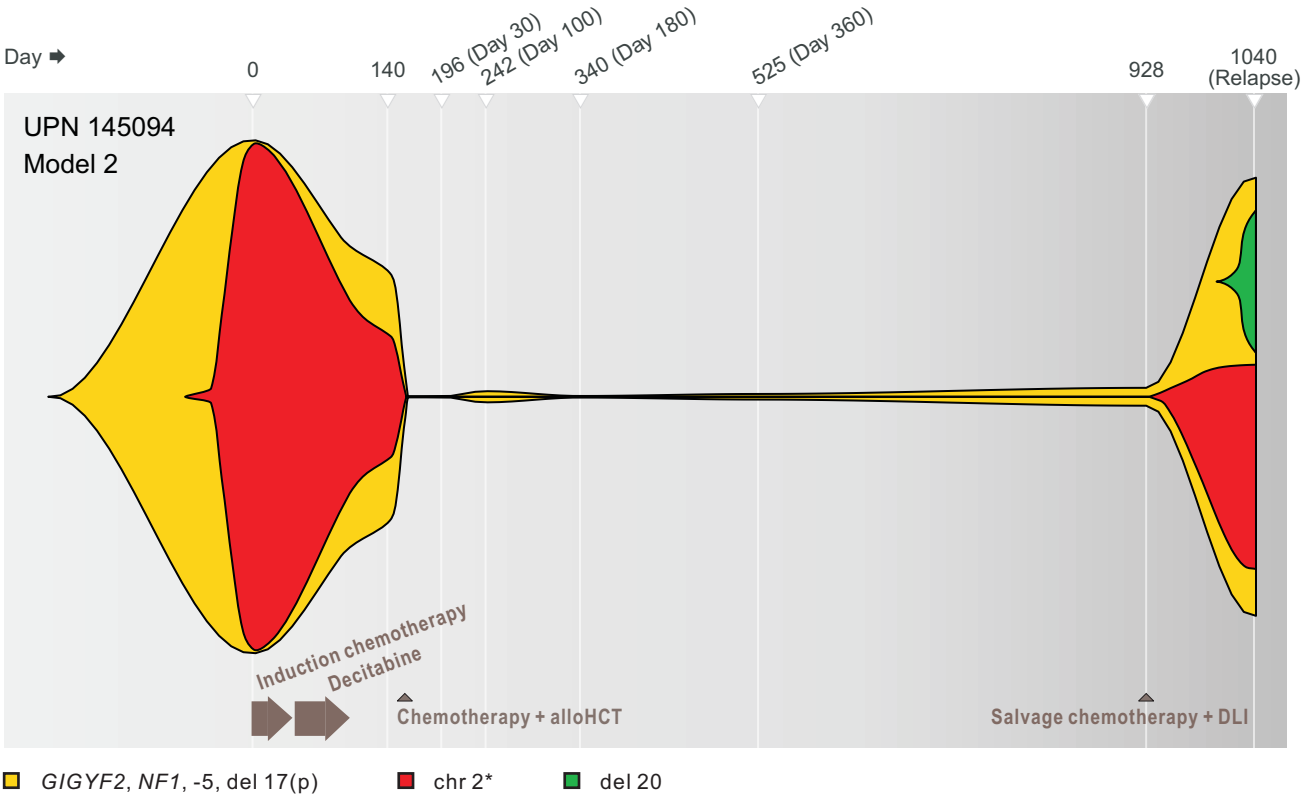
SFig4

A


Day
0
140
196 (Day 30)
242 (Day 100)
340 (Day 180)
525 (Day 360)
928
1040
(Relapse)
UPN 145094
Model 1
Induction chemotherapy
Decitabine
Chemotherapy + alloHCT
Salvage chemotherapy + DLI
GIGYF2, NF1, -5, del 17(p)
chr 2*
del 20

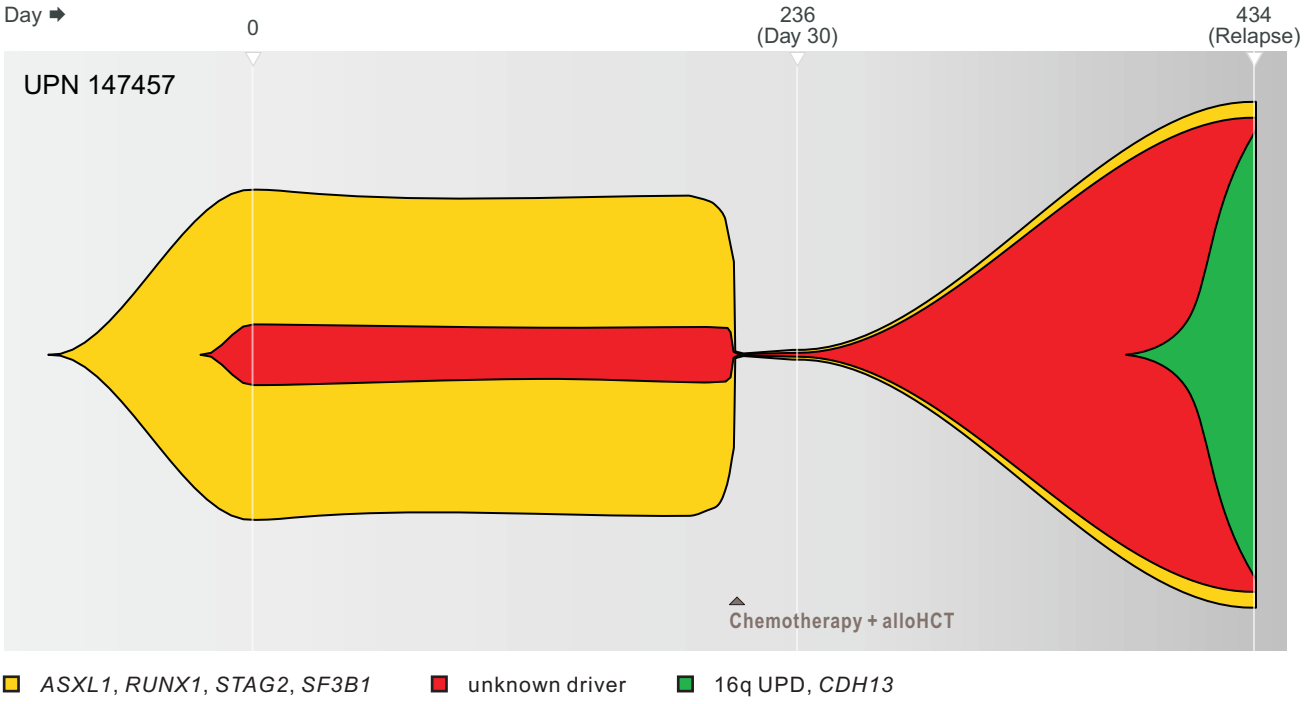
SFig4

B


Day
0
140
196 (Day 30)
242 (Day 100)
340 (Day 180)
525 (Day 360)
928
1040
(Relapse)
UPN 145094
Model 2
Induction chemotherapy
Decitabine
Chemotherapy + alloHCT
Salvage chemotherapy + DLI
GIGYF2, NF1, -5, del 17(p)
chr 2*
del 20

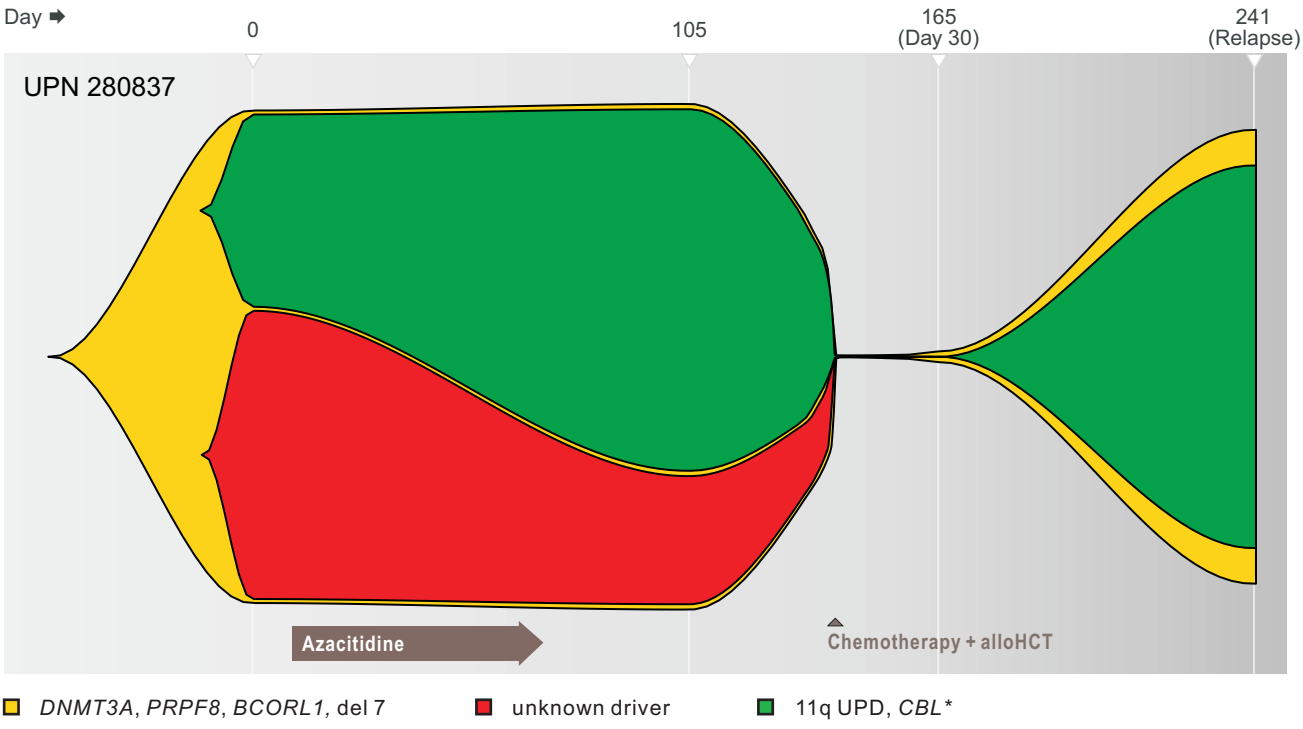
SFig4

C


Day
0
236
(Day 30)
434
(Relapse)
UPN 147457
Chemotherapy + alloHCT
ASXL1, RUNX1, STAG2, SF3B1
unknown driver
16q UPD, CDH13

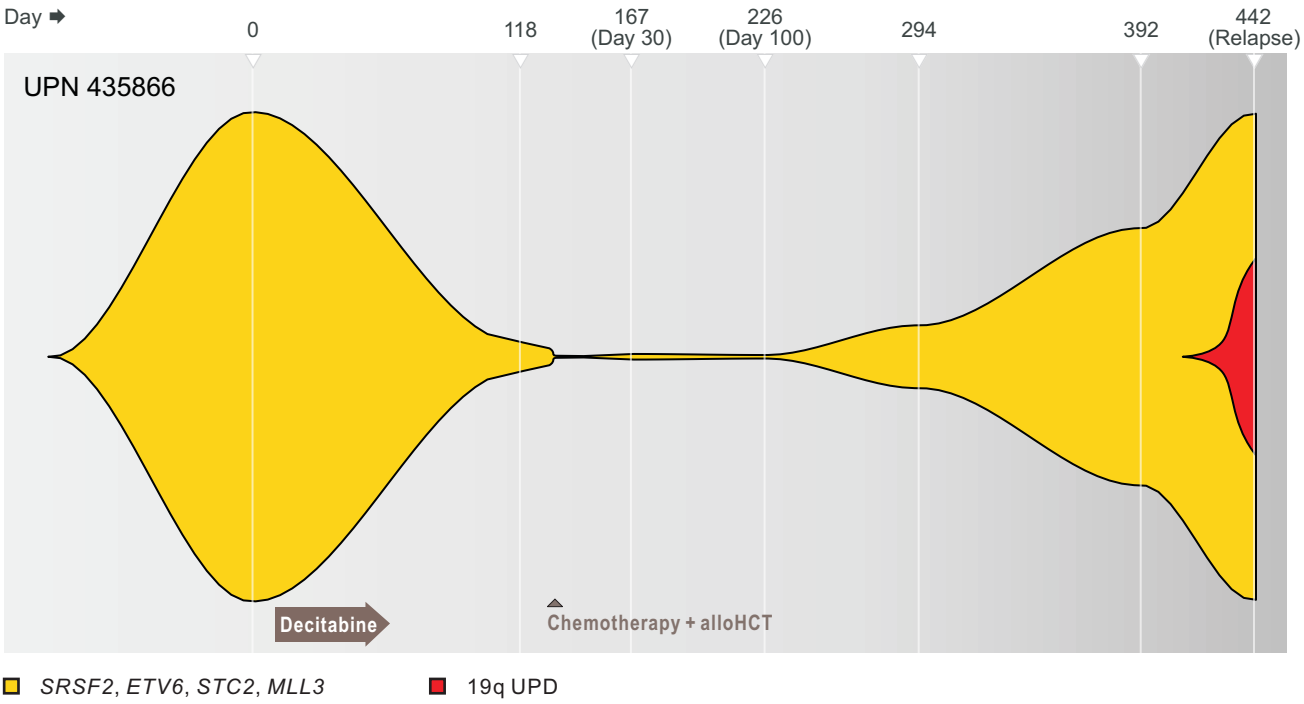
SFig4

D


Day
0
105
165
(Day 30)
241
(Relapse)
UPN 280837
Azacitidine
Chemotherapy + alloHCT
DNMT3A, PRPF8, BCORL1, del 7
unknown driver
11q UPD, CBL*

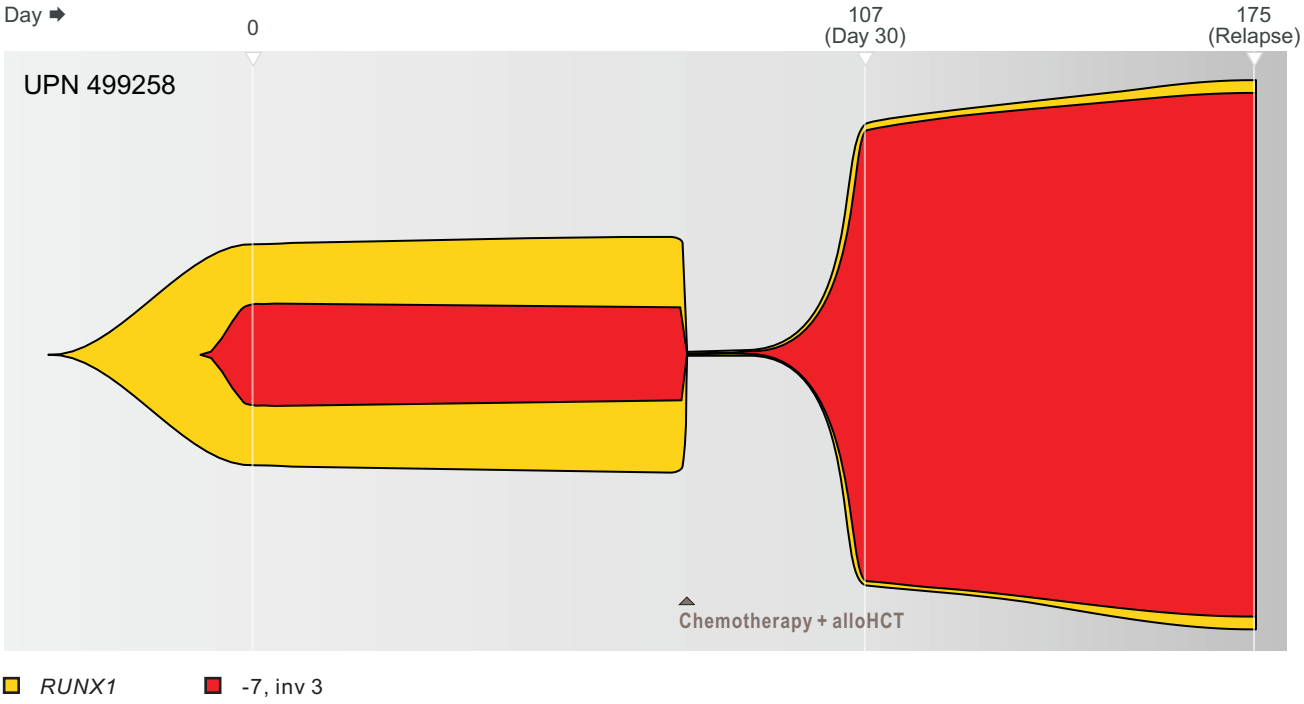
SFig4

E

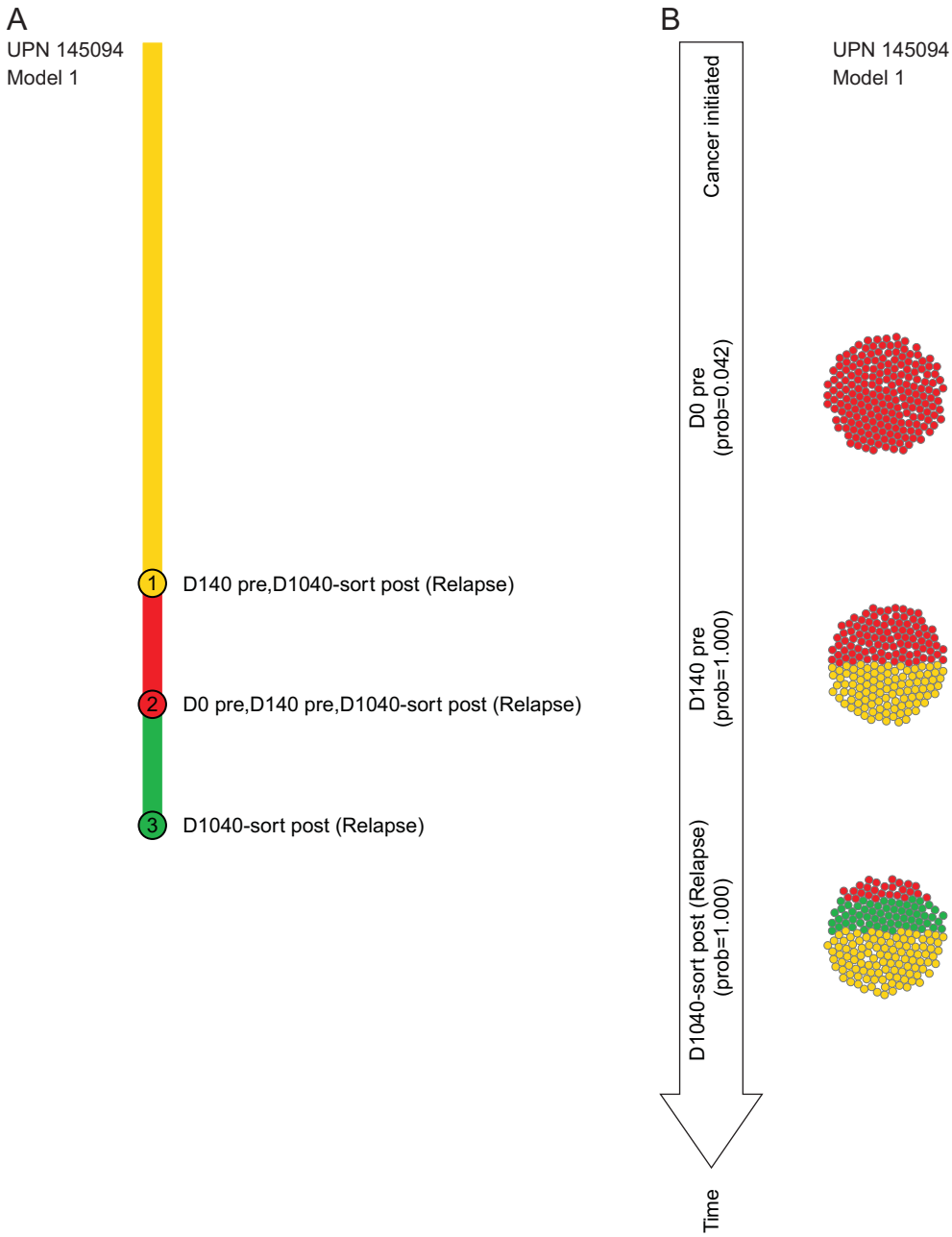

Day
0
118
167
(Day 30)
226
(Day 100)
294
392
442
(Relapse)
UPN 435866
Decitabine
Chemotherapy + alloHCT
SRSF2, ETV6, STC2, MLL3
19q UPD

SFig4

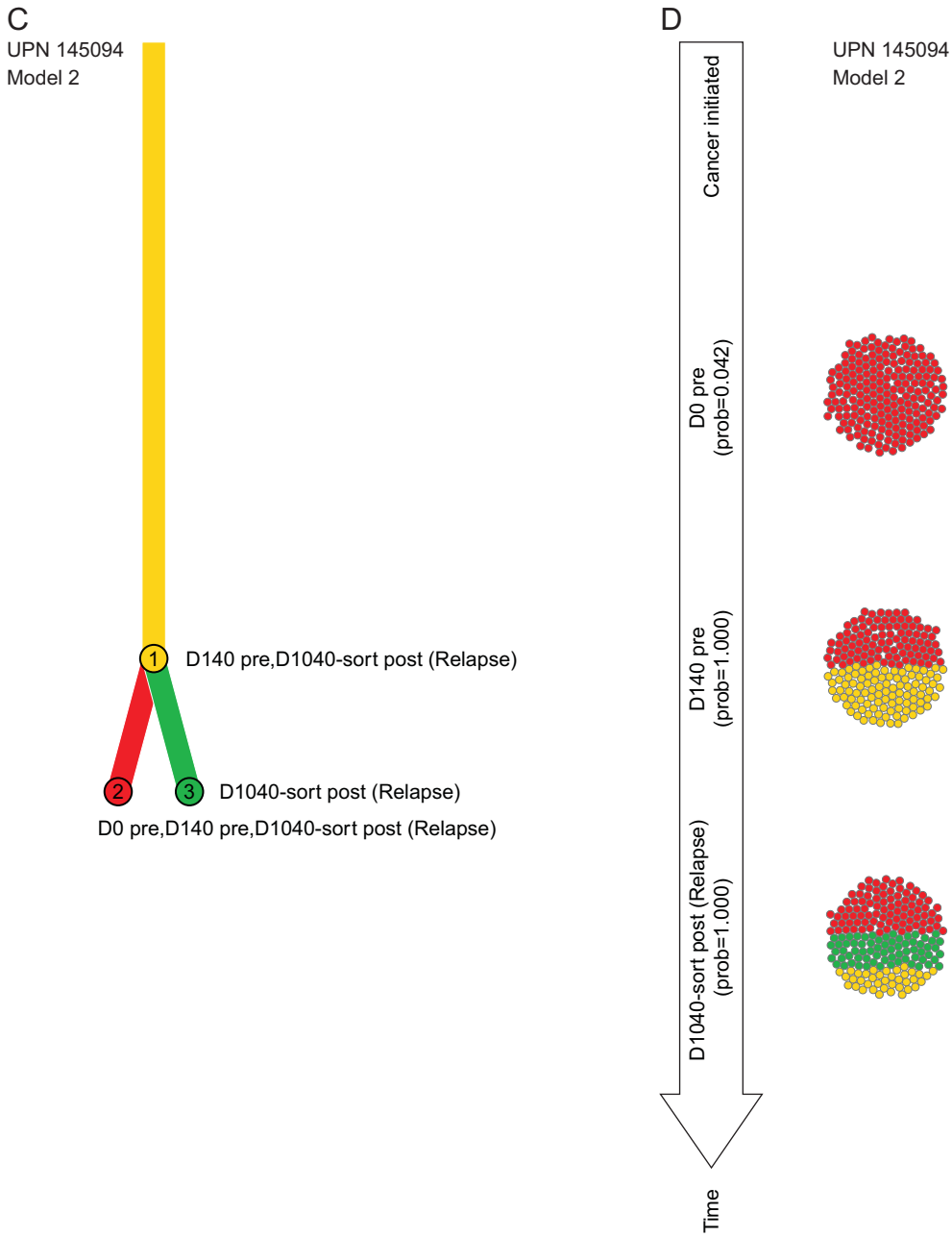
F


Day
0
107
(Day 30)
175
(Relapse)
UPN 499258
Chemotherapy + alloHCT
RUNX1
-7, inv 3

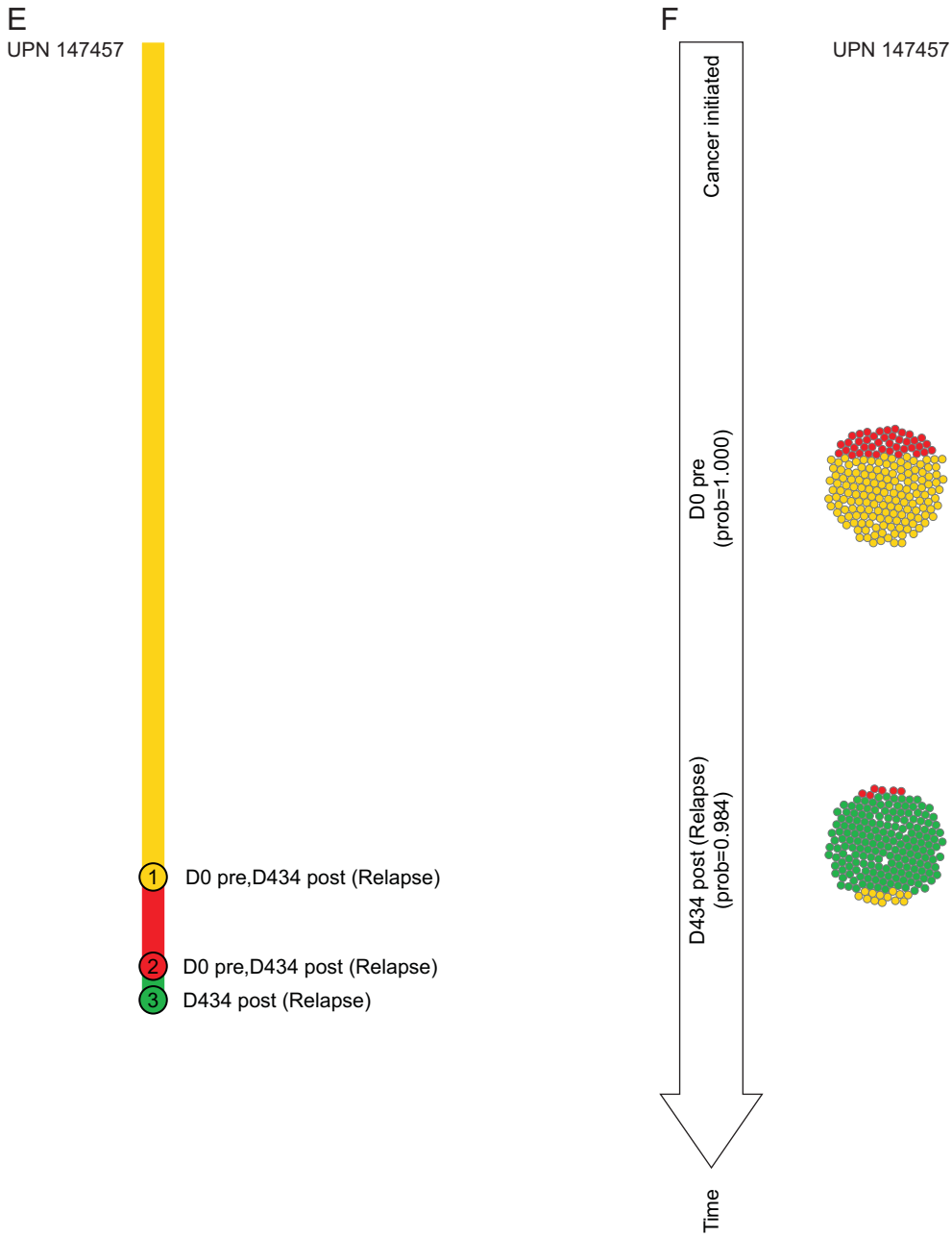
SFig5


A
UPN 145094
Model 1
1
D140 pre,D1040-sort post (Relapse)
2
D0 pre,D140 pre,D1040-sort post (Relapse)
3
D1040-sort post (Relapse)
B
UPN 145094
Model 1
Cancer initiated
D0 pre
(prob=0.042)
D140 pre
(prob=1.000)
D1040-sort post (Relapse)
(prob=1.000)
Time

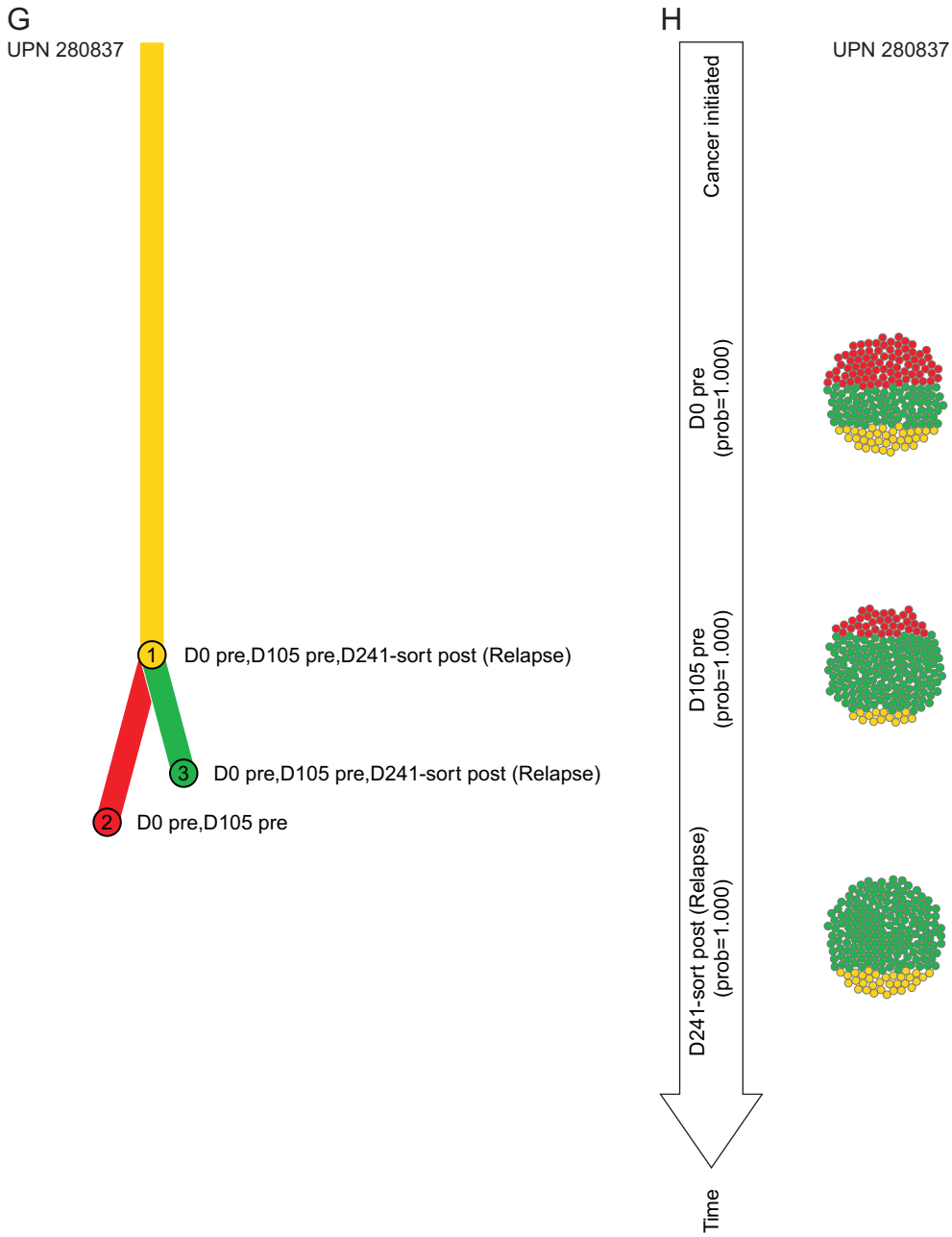
SFig5


C
UPN 145094
Model 2
1
D140 pre,D1040-sort post (Relapse)
2
3
D1040-sort post (Relapse)
D0 pre,D140 pre,D1040-sort post (Relapse)
D
UPN 145094
Model 2
Cancer initiated
D0 pre
(prob=0.042)
D140 pre
(prob=1.000)
D1040-sort post (Relapse)
(prob=1.000)
Time

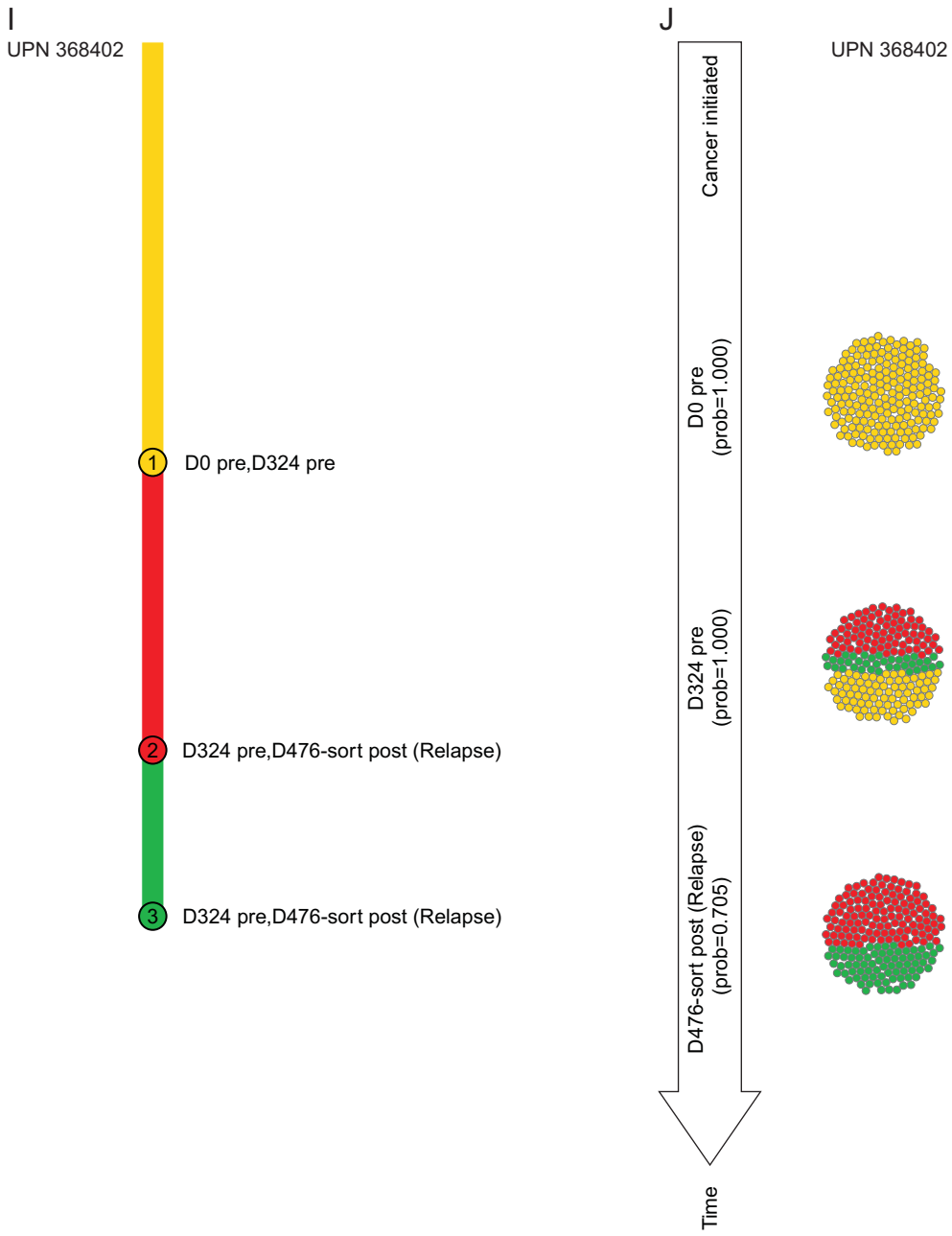
SFig5


E
UPN 147457
1 D0 pre,D434 post (Relapse)
2 D0 pre,D434 post (Relapse)
3 D434 post (Relapse)
F
UPN 147457
Cancer initiated
D0 pre
(prob=1.000)
D434 post (Relapse)
(prob=0.984)
Time

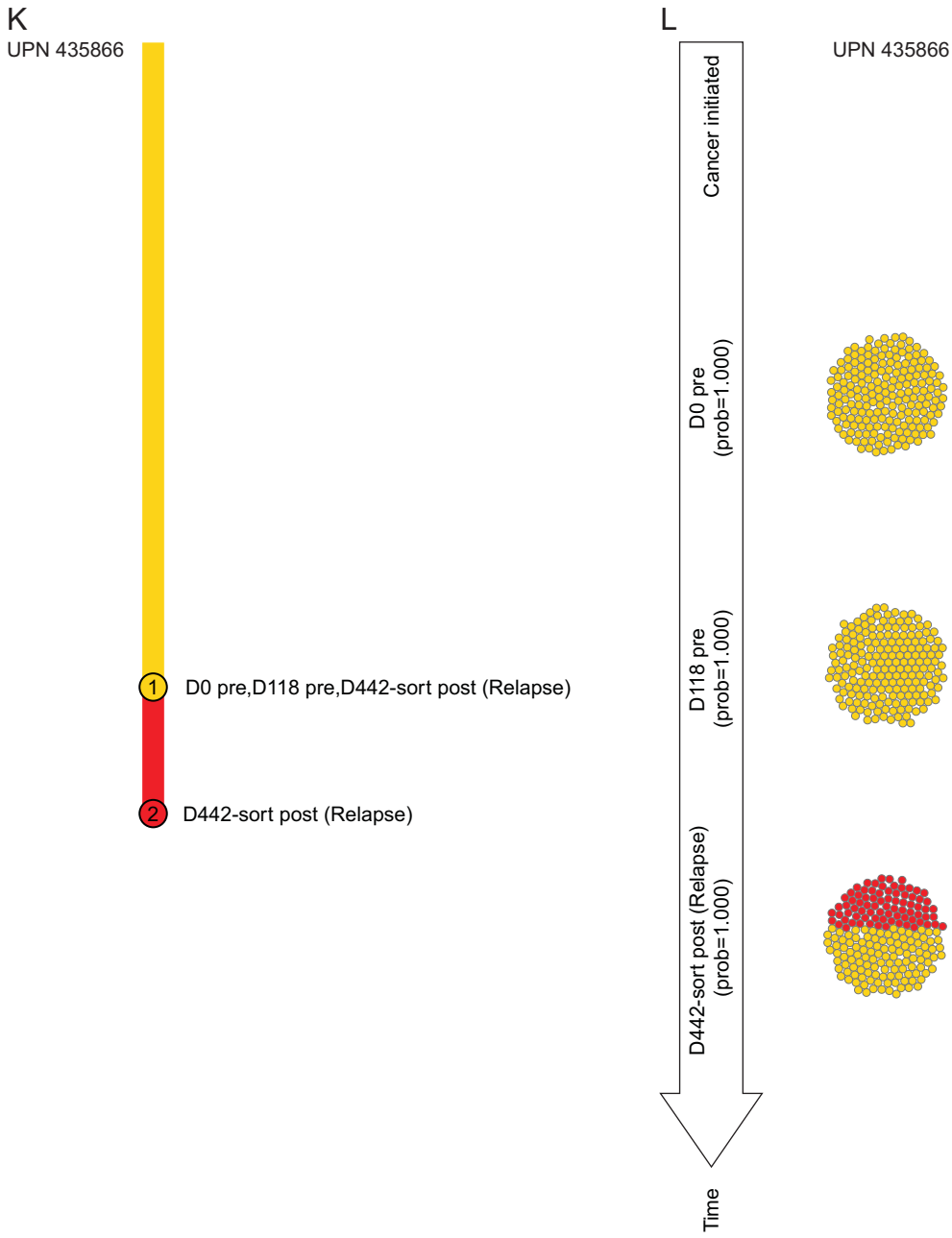
SFig5


G
UPN 280837
1
D0 pre,D105 pre,D241-sort post (Relapse)
3
D0 pre,D105 pre,D241-sort post (Relapse)
2
D0 pre,D105 pre
H
UPN 280837
Cancer initiated
D0 pre
(prob=1.000)
D105 pre
(prob=1.000)
D241-sort post (Relapse)
(prob=1.000)
Time

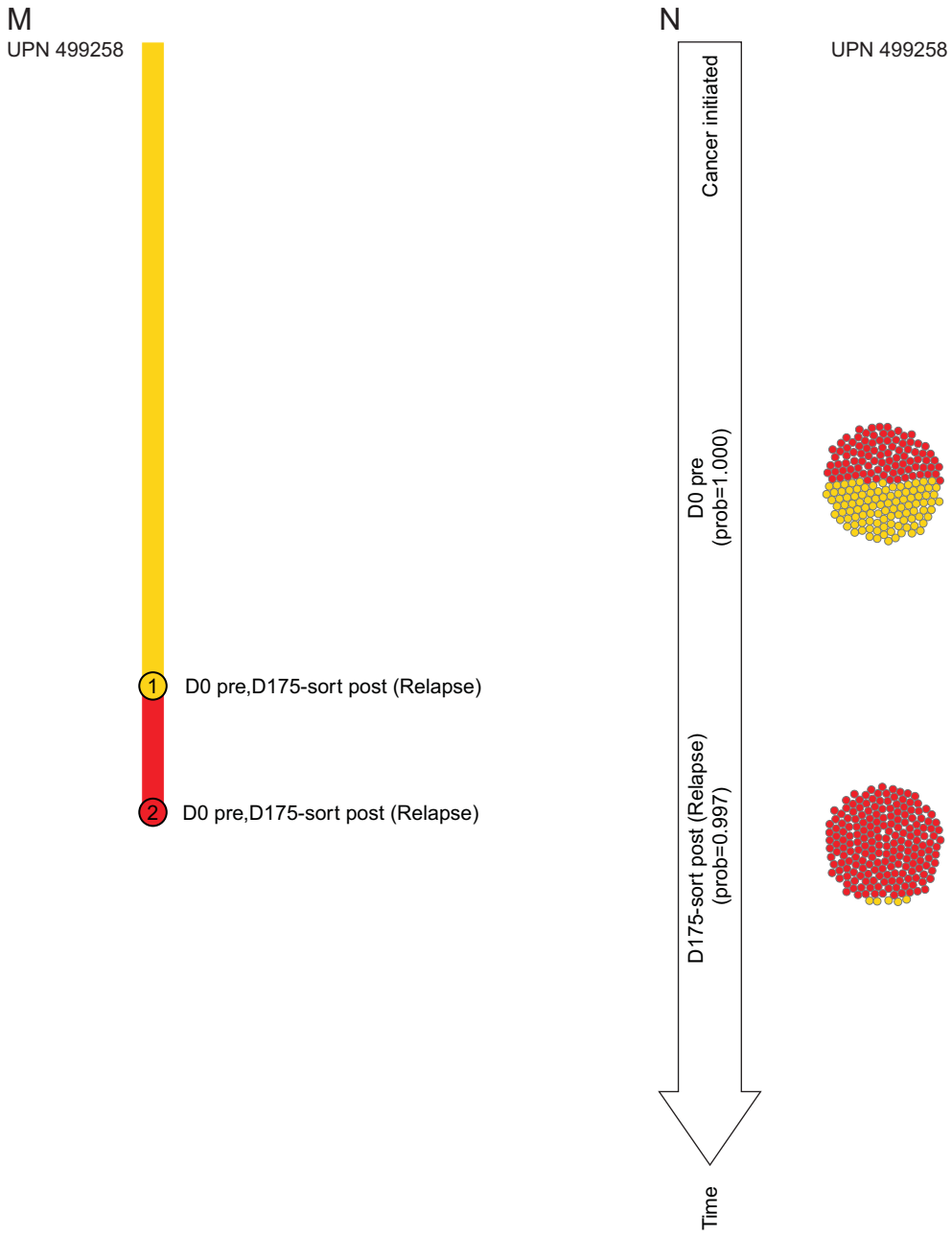
SFig5


I
UPN 368402
1
D0 pre,D324 pre
2
D324 pre,D476-sort post (Relapse)
3
D324 pre,D476-sort post (Relapse)
J
UPN 368402
Cancer initiated
D0 pre
(prob=1.000)
D324 pre
(prob=1.000)
D476-sort post (Relapse)
(prob=0.705)
Time

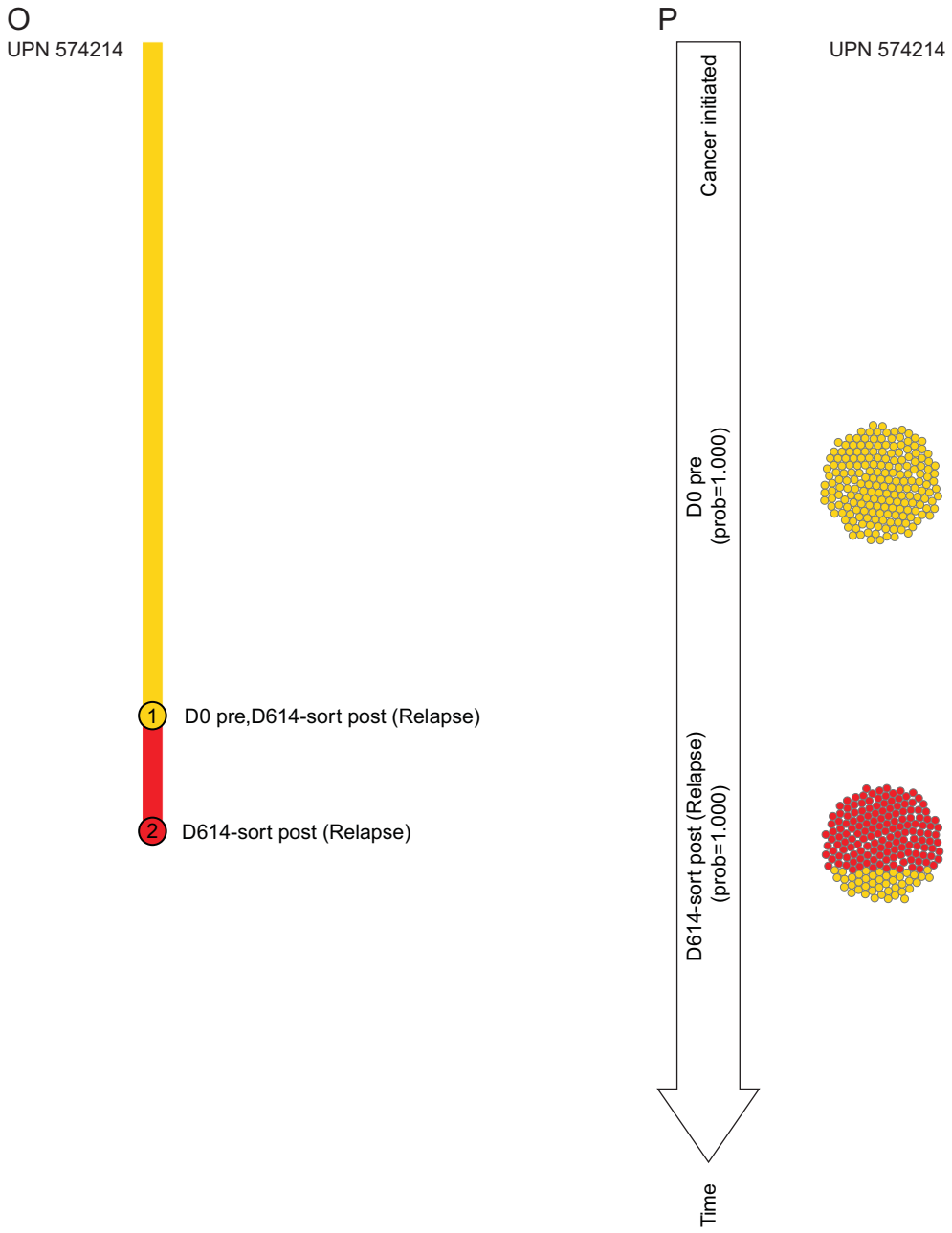
SFig5


K
UPN 435866
1
D0 pre,D118 pre,D442-sort post (Relapse)
2
D442-sort post (Relapse)
L
UPN 435866
Cancer initiated
D0 pre
(prob=1.000)
D118 pre
(prob=1.000)
D442-sort post (Relapse)
(prob=1.000)
Time

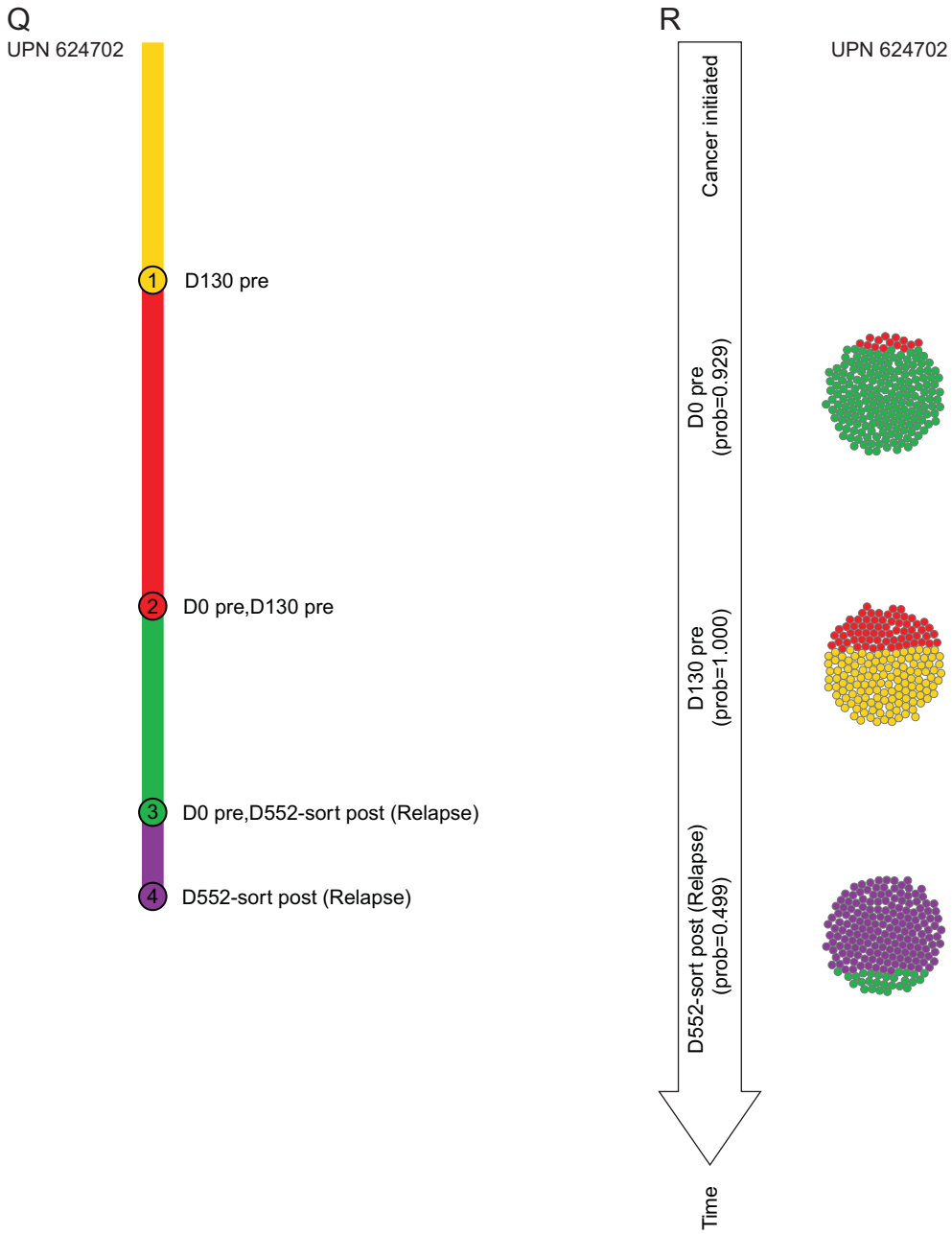
SFig5


M
UPN 499258
1
D0 pre,D175-sort post (Relapse)
2
D0 pre,D175-sort post (Relapse)
N
UPN 499258
Cancer initiated
D0 pre
(prob=1.000)
D175-sort post (Relapse)
(prob=0.997)
Time

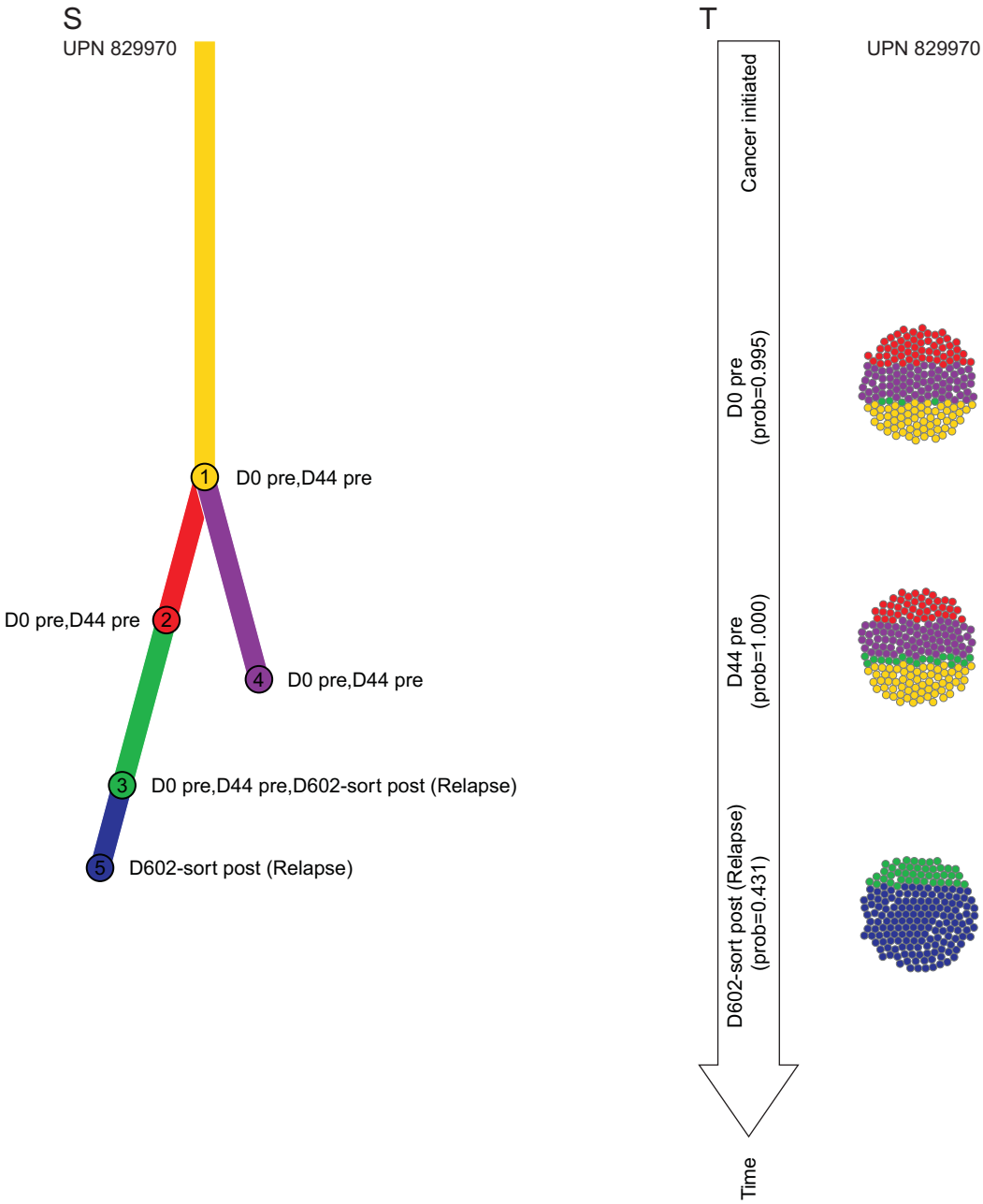
SFig5


O
UPN 574214
1
D0 pre,D614-sort post (Relapse)
2
D614-sort post (Relapse)
P
UPN 574214
Cancer initiated
D0 pre
(prob=1.000)
D614-sort post (Relapse)
(prob=1.000)
Time

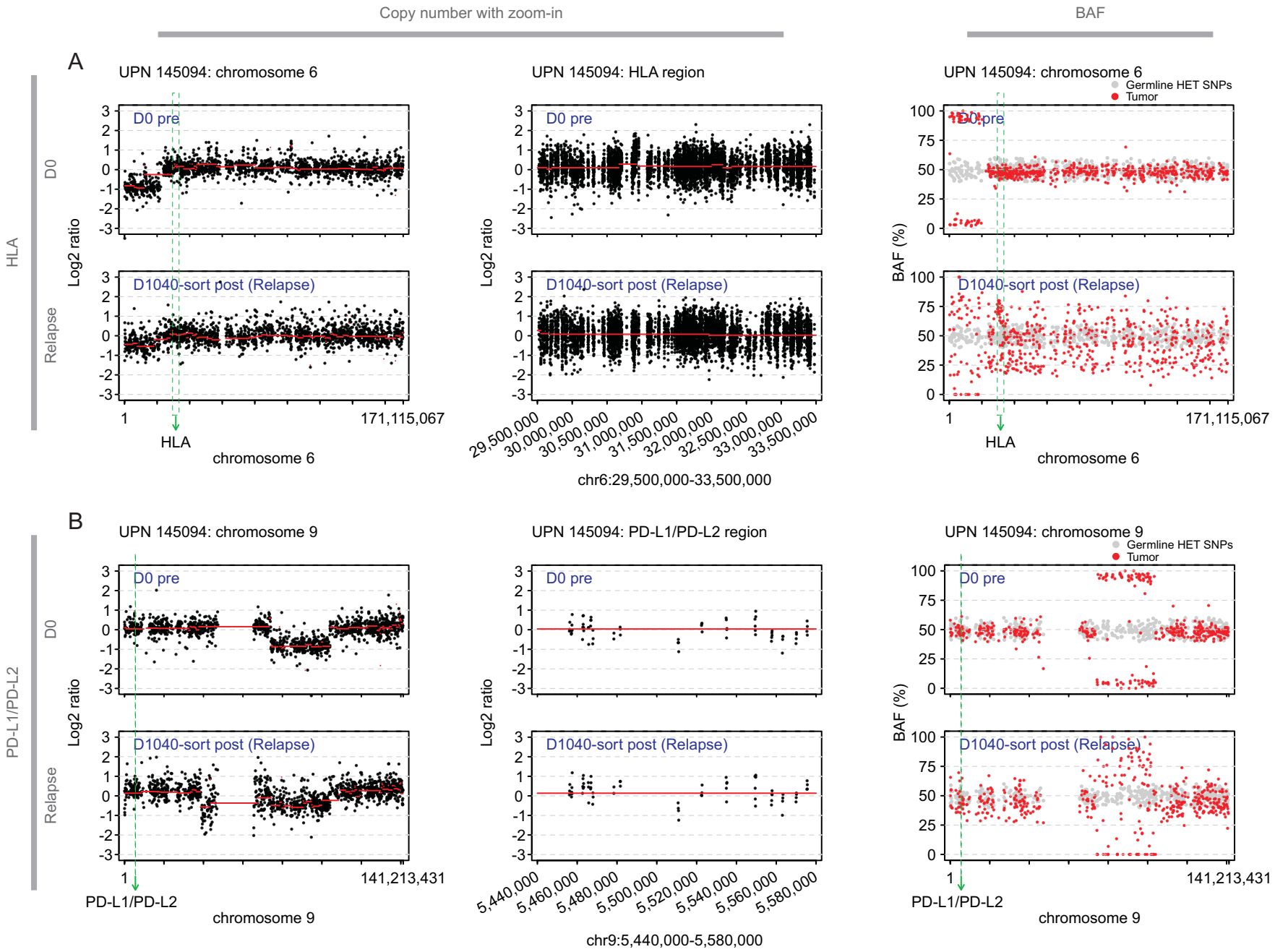
SFig5


Q
UPN 624702
1 D130 pre
2 D0 pre,D130 pre
3 D0 pre,D552-sort post (Relapse)
4 D552-sort post (Relapse)
R
UPN 624702
Cancer initiated
D0 pre (prob=0.929)
D130 pre (prob=1.000)
D552-sort post (Relapse) (prob=0.499)
Time

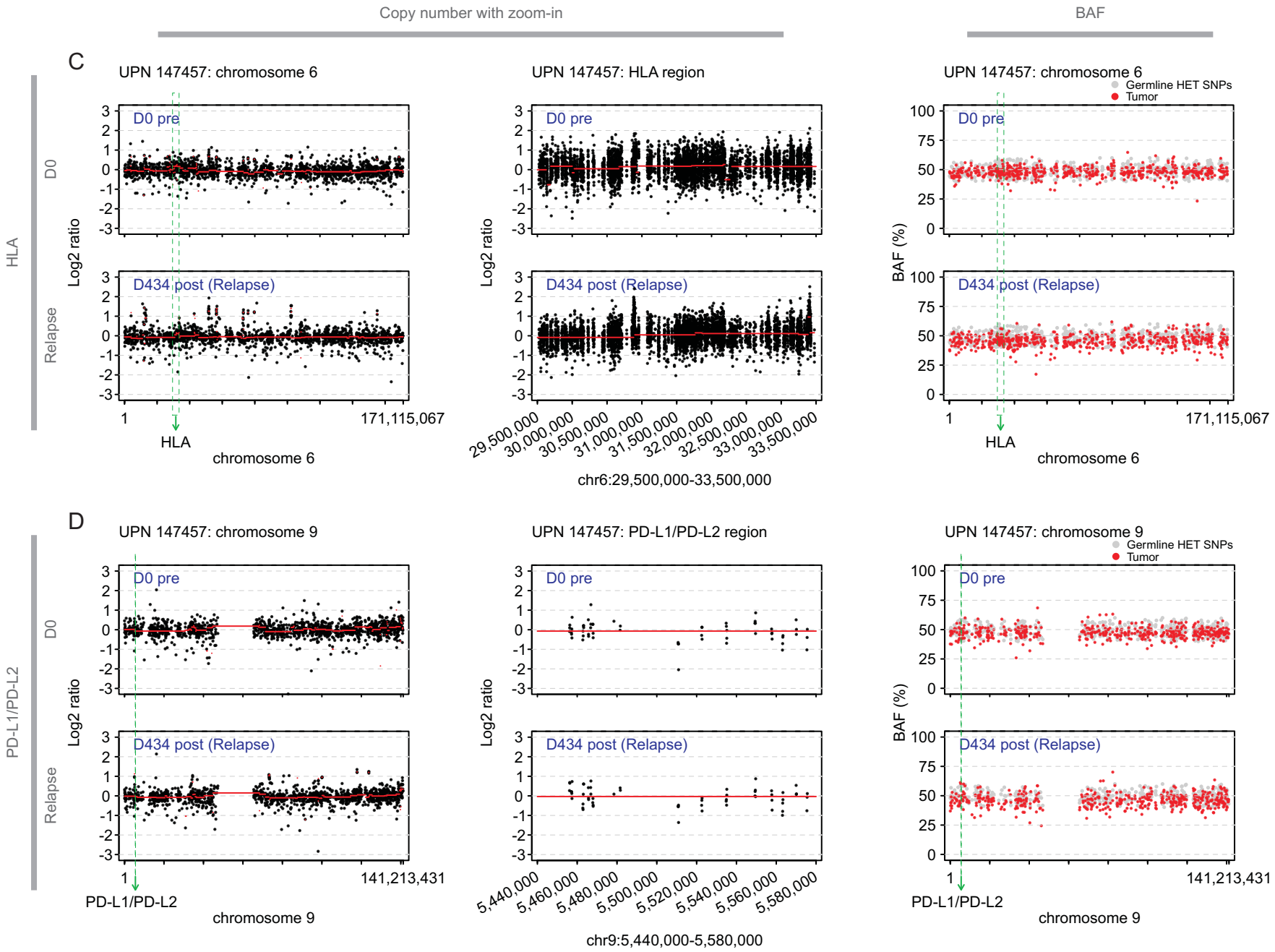
SFig5


S
UPN 829970
1
D0 pre,D44 pre
D0 pre,D44 pre
2
4
D0 pre,D44 pre
3
D0 pre,D44 pre,D602-sort post (Relapse)
5
D602-sort post (Relapse)
T
UPN 829970
Cancer initiated
D0 pre
(prob=0.995)
D44 pre
(prob=1.000)
D602-sort post (Relapse)
(prob=0.431)
Time

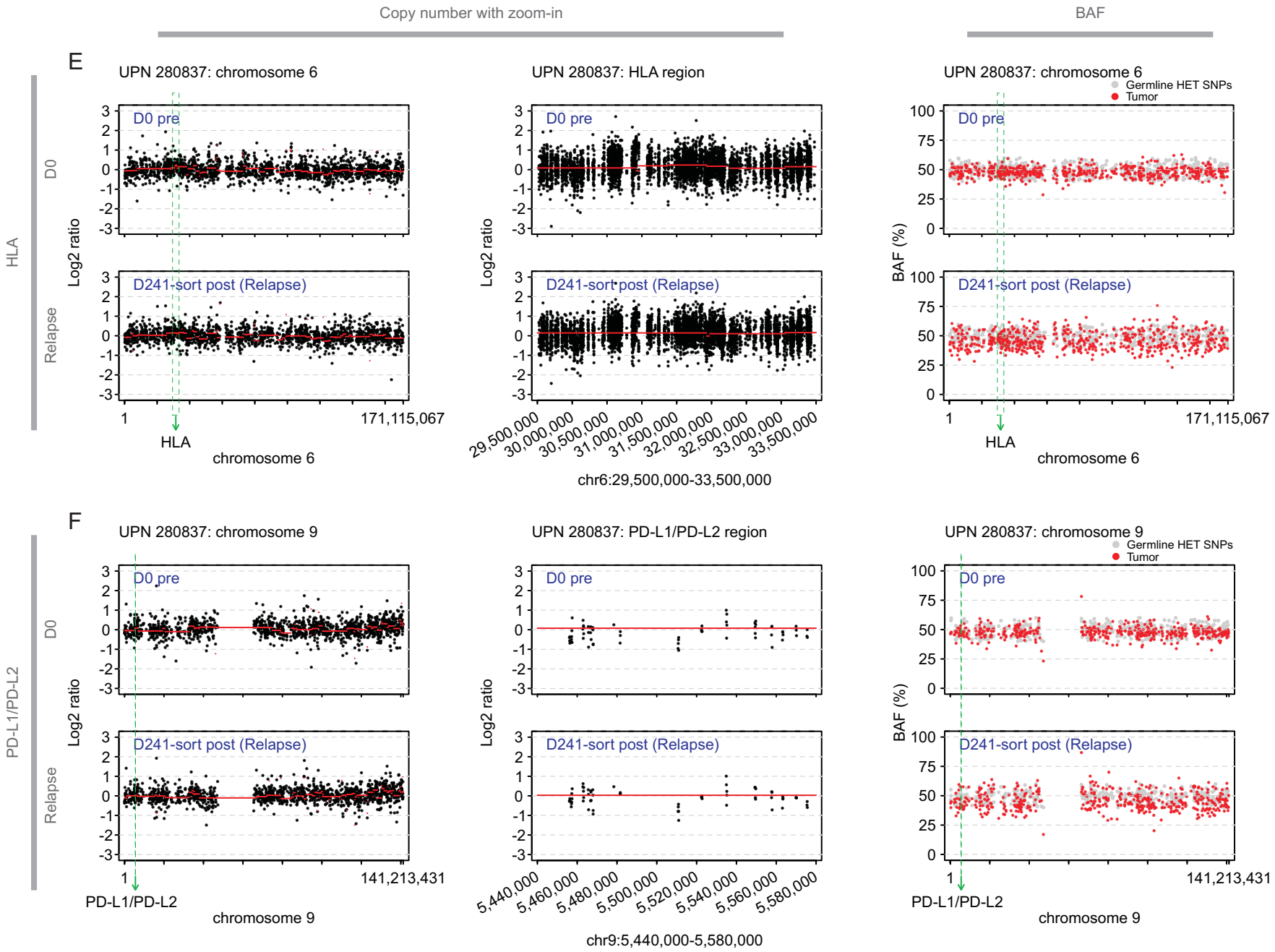
SFig6


Copy number with zoom-in
BAF
A
UPN 145094: chromosome 6
UPN 145094: HLA region
UPN 145094: chromosome 6
Germline HET SNPs
Tumor
HLA
D0
Relapse
Log2 ratio
BAF (%)
D0 pre
D1040-sort post (Relapse)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
B
UPN 145094: chromosome 9
UPN 145094: PD-L1/PD-L2 region
UPN 145094: chromosome 9
PD-L1/PD-L2
141,213,431
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

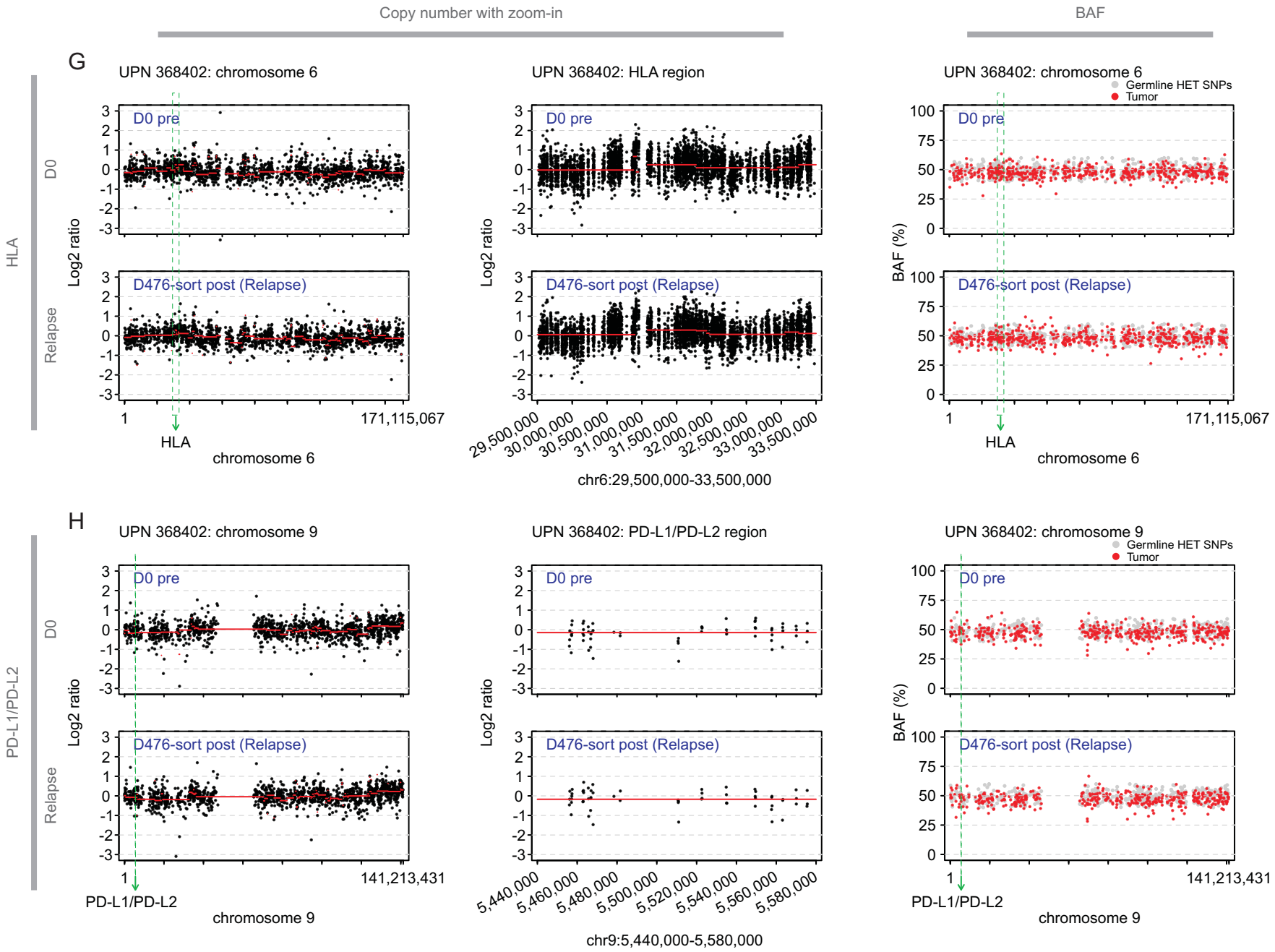
SFig6


Copy number with zoom-in
BAF
C
UPN 147457: chromosome 6
UPN 147457: HLA region
UPN 147457: chromosome 6
Germline HET SNPs
Tumor
HLA
D0
Relapse
Log2 ratio
BAF (%)
D0 pre
D434 post (Relapse)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
D
UPN 147457: chromosome 9
UPN 147457: PD-L1/PD-L2 region
UPN 147457: chromosome 9
PD-L1/PD-L2
141,213,431
PD-L1/PD-L2
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

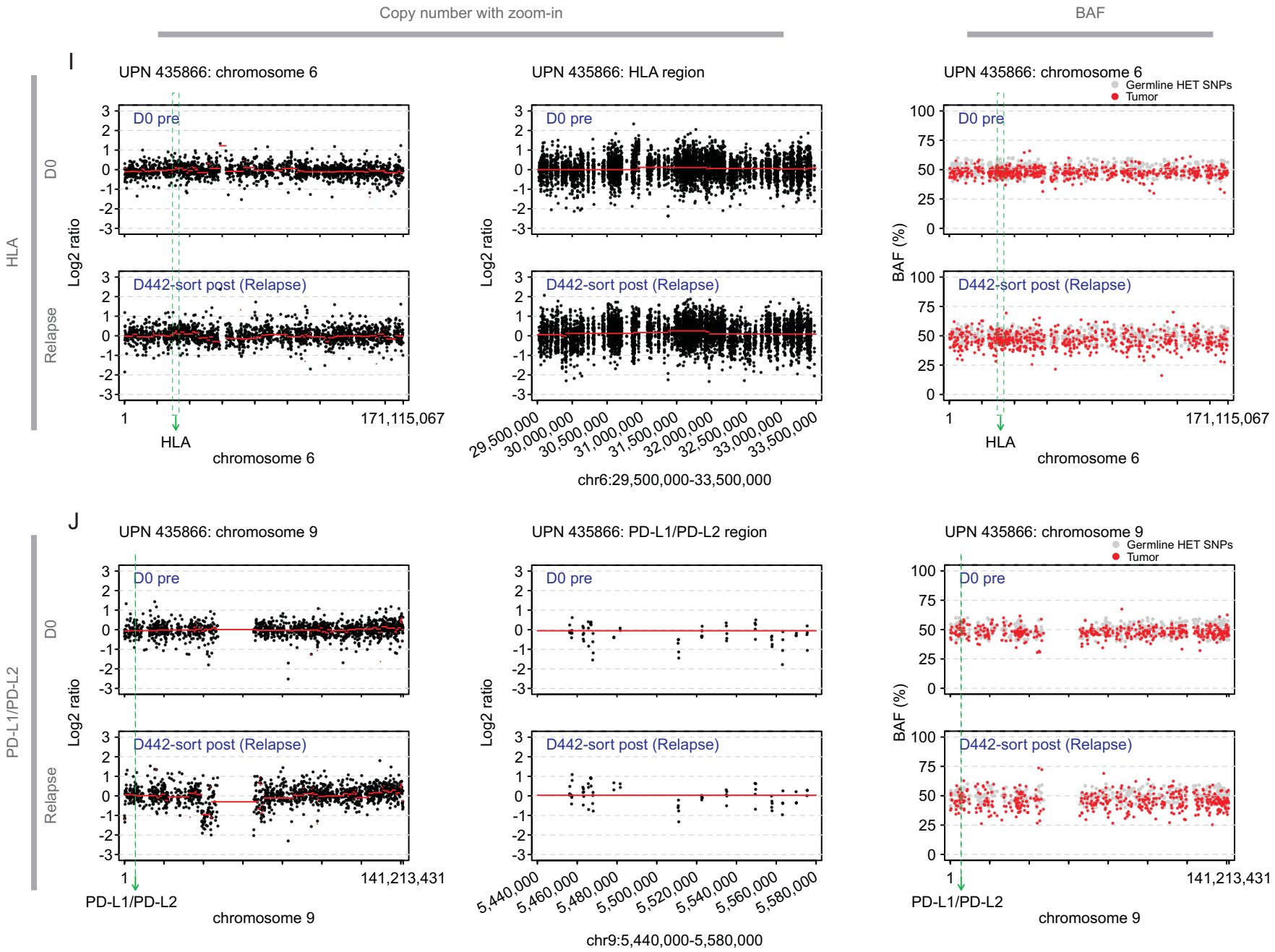
SFig6


Copy number with zoom-in
BAF
E
UPN 280837: chromosome 6
UPN 280837: HLA region
UPN 280837: chromosome 6
Germline HET SNPs
Tumor
HLA
D0
Relapse
D0 pre
D241-sort post (Relapse)
Log2 ratio
BAF (%)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
F
UPN 280837: chromosome 9
UPN 280837: PD-L1/PD-L2 region
UPN 280837: chromosome 9
PD-L1/PD-L2
141,213,431
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

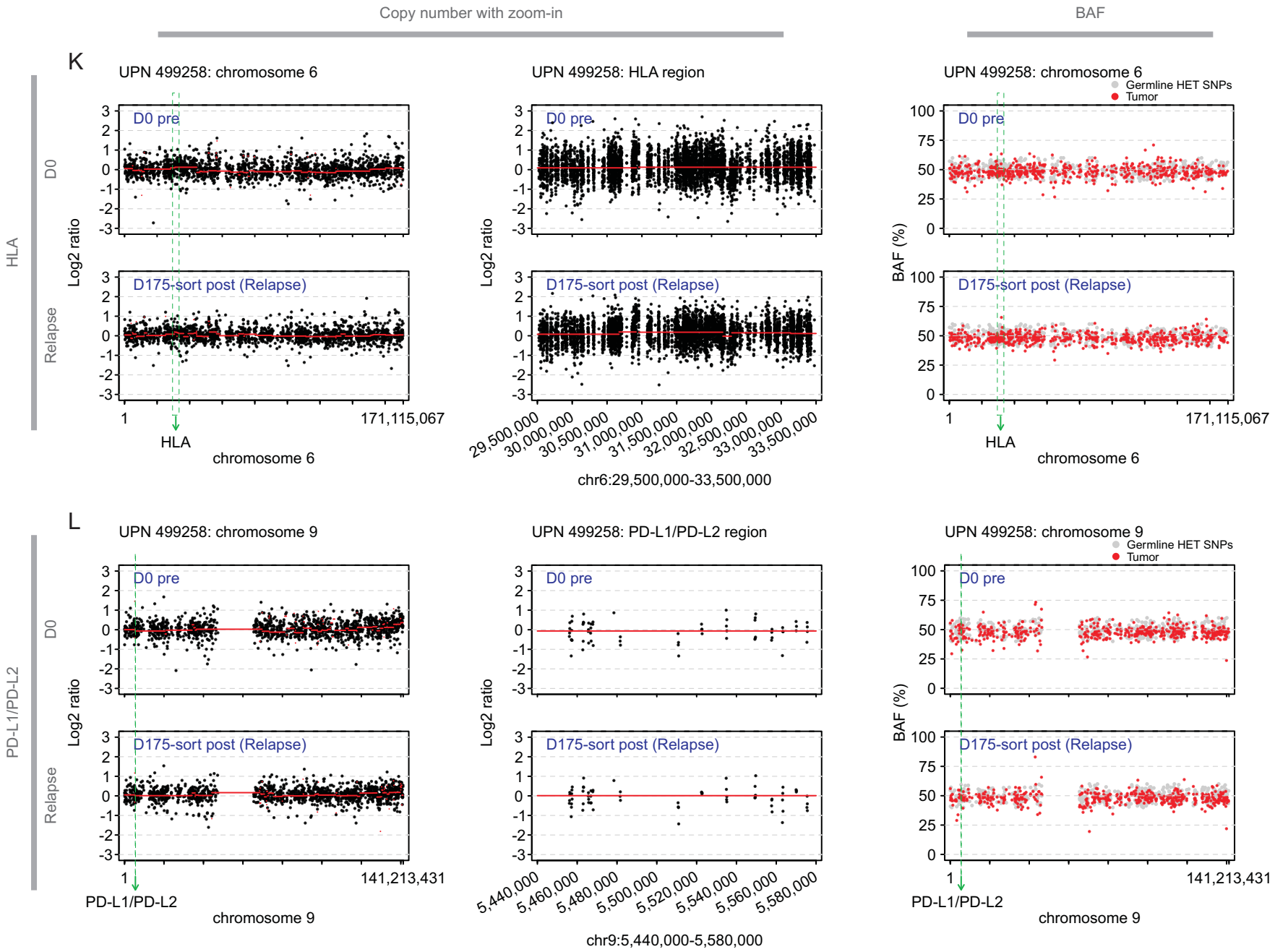
SFig6


Copy number with zoom-in
BAF
G
UPN 368402: chromosome 6
UPN 368402: HLA region
UPN 368402: chromosome 6
Germline HET SNPs
Tumor
HLA
D0
Relapse
Log2 ratio
BAF (%)
D0 pre
D476-sort post (Relapse)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
H
UPN 368402: chromosome 9
UPN 368402: PD-L1/PD-L2 region
UPN 368402: chromosome 9
PD-L1/PD-L2
141,213,431
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

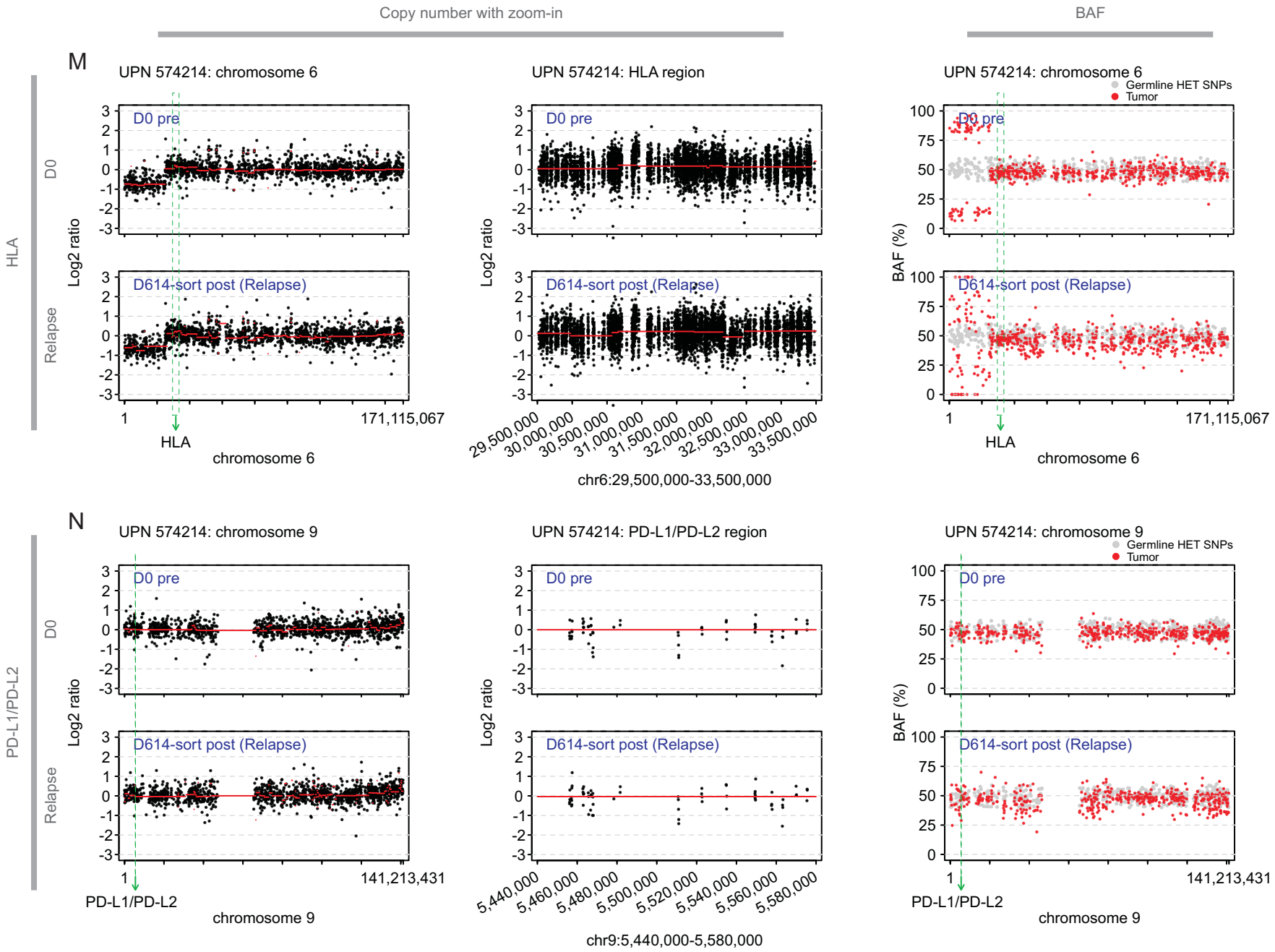
SFig6


Copy number with zoom-in
BAF
I
UPN 435866: chromosome 6
UPN 435866: HLA region
UPN 435866: chromosome 6
Germline HET SNPs
Tumor
HLA
D0
Relapse
D0 pre
D442-sort post (Relapse)
Log2 ratio
BAF (%)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
J
UPN 435866: chromosome 9
UPN 435866: PD-L1/PD-L2 region
UPN 435866: chromosome 9
PD-L1/PD-L2
141,213,431
PD-L1/PD-L2
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

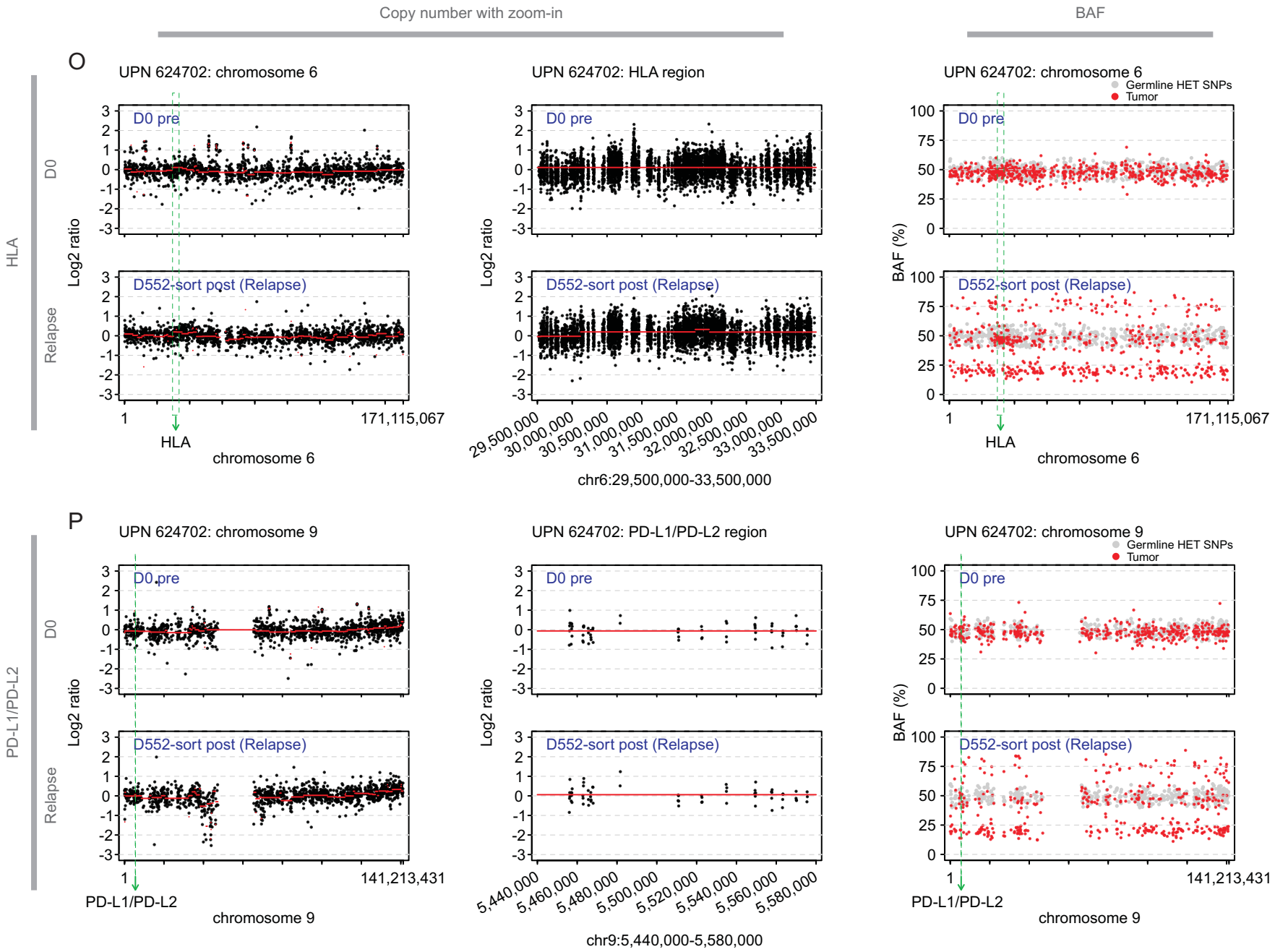
SFig6


Copy number with zoom-in
BAF
K
UPN 499258: chromosome 6
UPN 499258: HLA region
UPN 499258: chromosome 6
Germline HET SNPs
Tumor
HLA
D0
Relapse
Log2 ratio
BAF (%)
D0 pre
D175-sort post (Relapse)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
L
UPN 499258: chromosome 9
UPN 499258: PD-L1/PD-L2 region
UPN 499258: chromosome 9
PD-L1/PD-L2
141,213,431
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

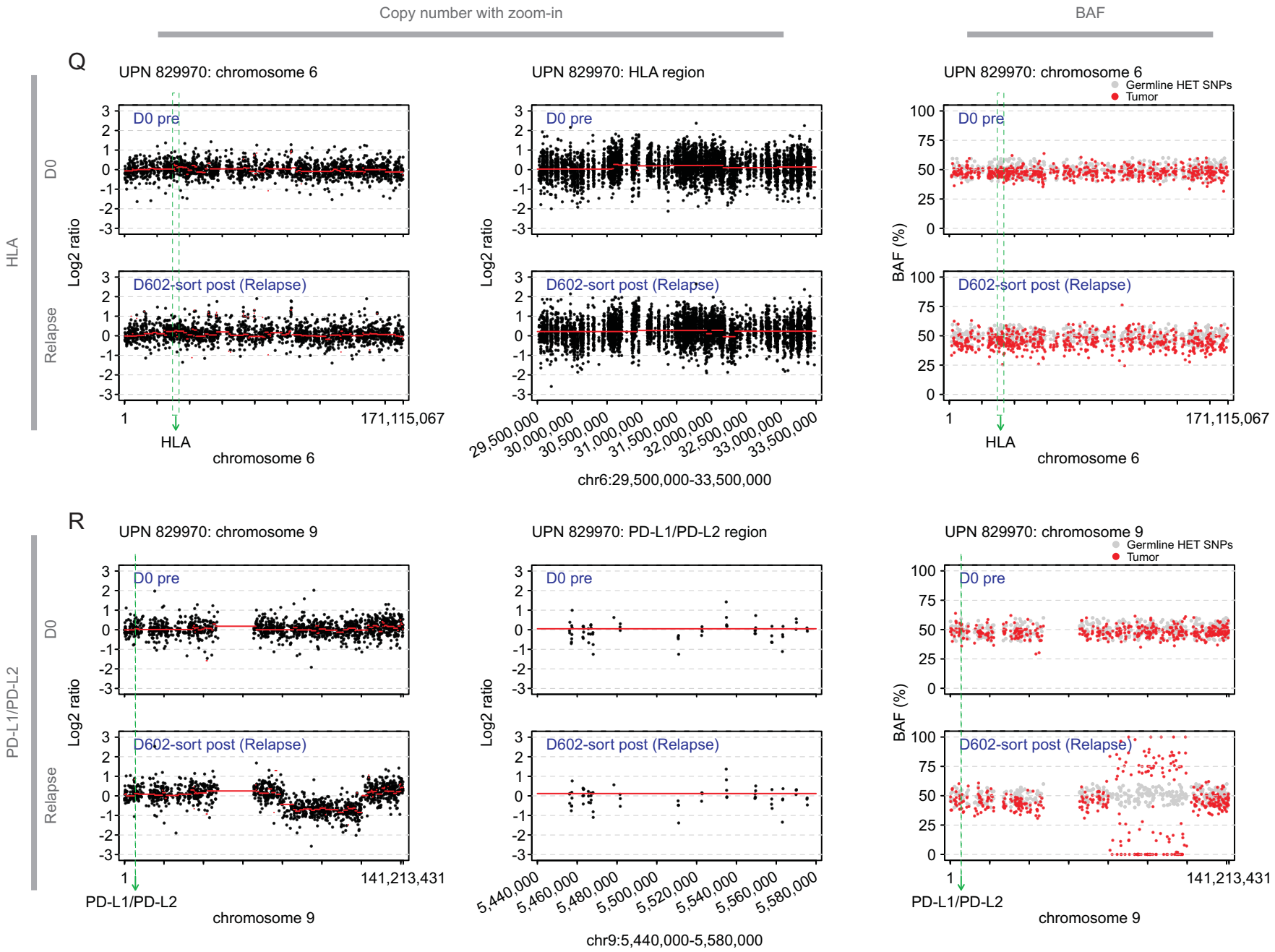
SFig6


Copy number with zoom-in
BAF
M
UPN 574214: chromosome 6
UPN 574214: HLA region
UPN 574214: chromosome 6
Germline HET SNPs
Tumor
HLA
D0
Relapse
Log2 ratio
BAF (%)
D0 pre
D614-sort post (Relapse)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
N
UPN 574214: chromosome 9
UPN 574214: PD-L1/PD-L2 region
UPN 574214: chromosome 9
PD-L1/PD-L2
141,213,431
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

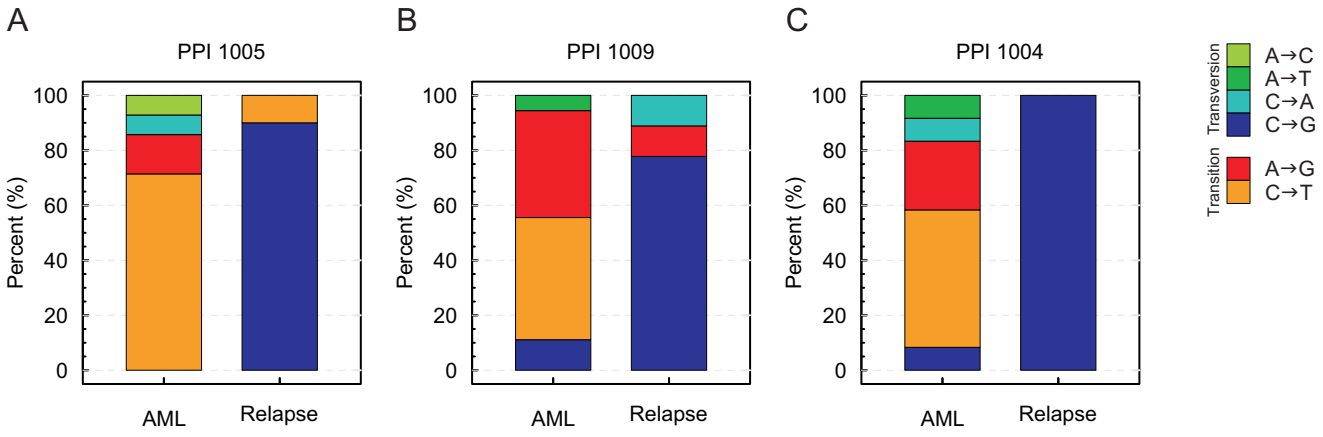
SFig6


Copy number with zoom-in
BAF
O
HLA
D0
Relapse
UPN 624702: chromosome 6
UPN 624702: HLA region
UPN 624702: chromosome 6
Germline HET SNPs
Tumor
D0 pre
D552-sort post (Relapse)
Log2 ratio
BAF (%)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
P
PD-L1/PD-L2
UPN 624702: chromosome 9
UPN 624702: PD-L1/PD-L2 region
UPN 624702: chromosome 9
141,213,431
PD-L1/PD-L2
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

SFig6


Copy number with zoom-in
BAF
Q
UPN 829970: chromosome 6
UPN 829970: HLA region
UPN 829970: chromosome 6
Germline HET SNPs
Tumor
HLA
D0
Relapse
Log2 ratio
BAF (%)
D0 pre
D602-sort post (Relapse)
1
171,115,067
HLA
chromosome 6
29,500,000
30,000,000
30,500,000
31,000,000
31,500,000
32,000,000
32,500,000
33,000,000
33,500,000
chr6:29,500,000-33,500,000
R
UPN 829970: chromosome 9
UPN 829970: PD-L1/PD-L2 region
UPN 829970: chromosome 9
PD-L1/PD-L2
141,213,431
PD-L1/PD-L2
chromosome 9
5,440,000
5,460,000
5,480,000
5,500,000
5,520,000
5,540,000
5,560,000
5,580,000
chr9:5,440,000-5,580,000

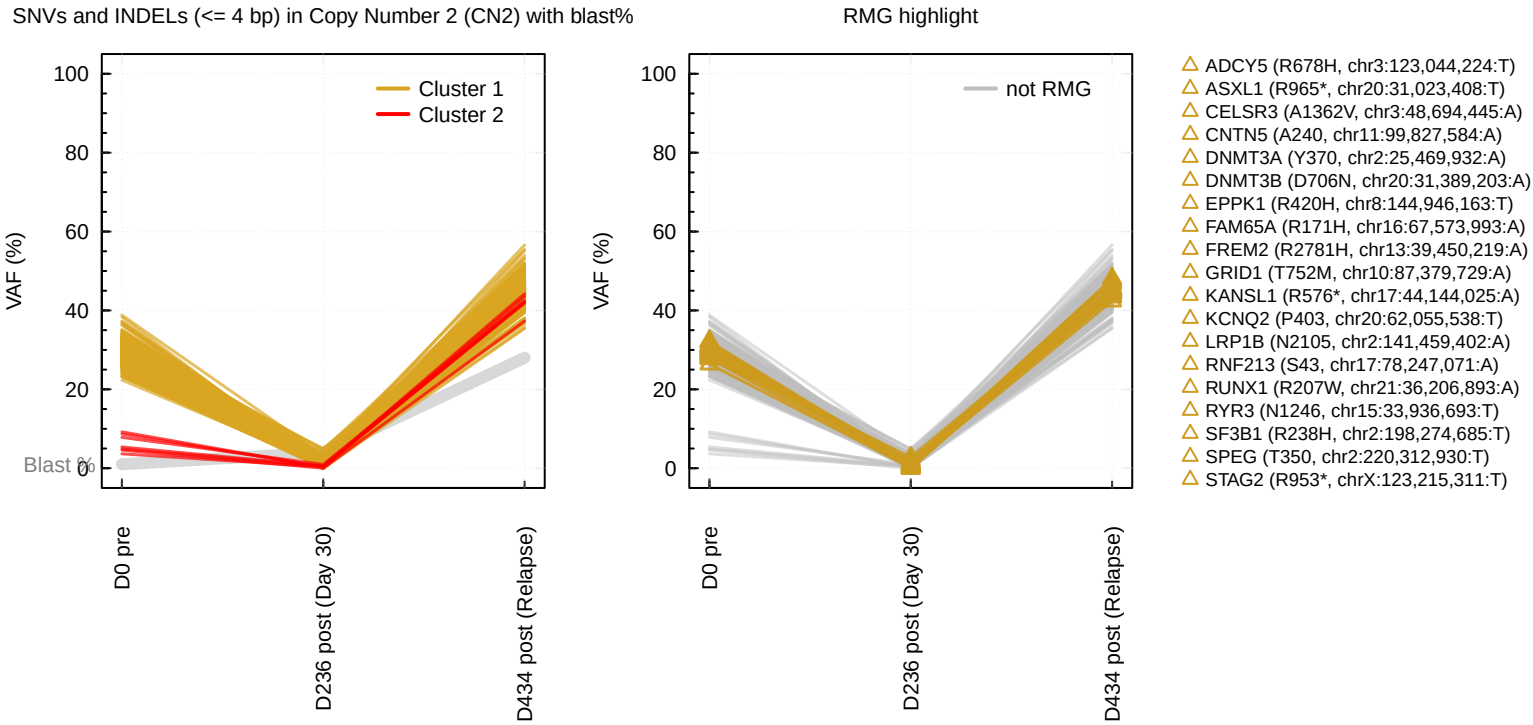
SFig7


A
PPI 1005
B
PPI 1009
C
PPI 1004
Percent (%)
100
80
60
40
20
0
AML
Relapse
Transversion
A→C
A→T
C→A
C→G
Transition
A→G
C→T

SFig8

A

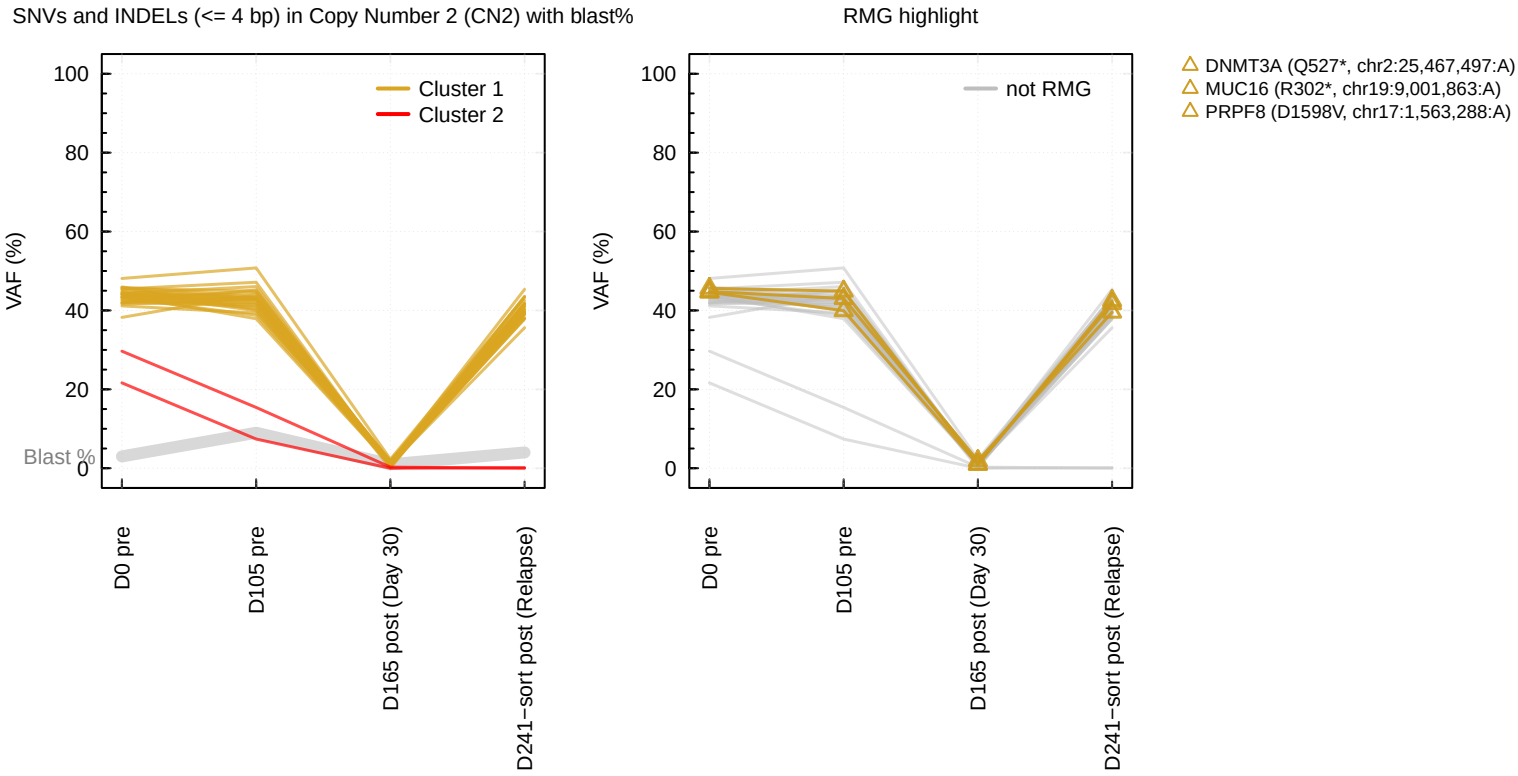
UPN 147457



SFig8

B

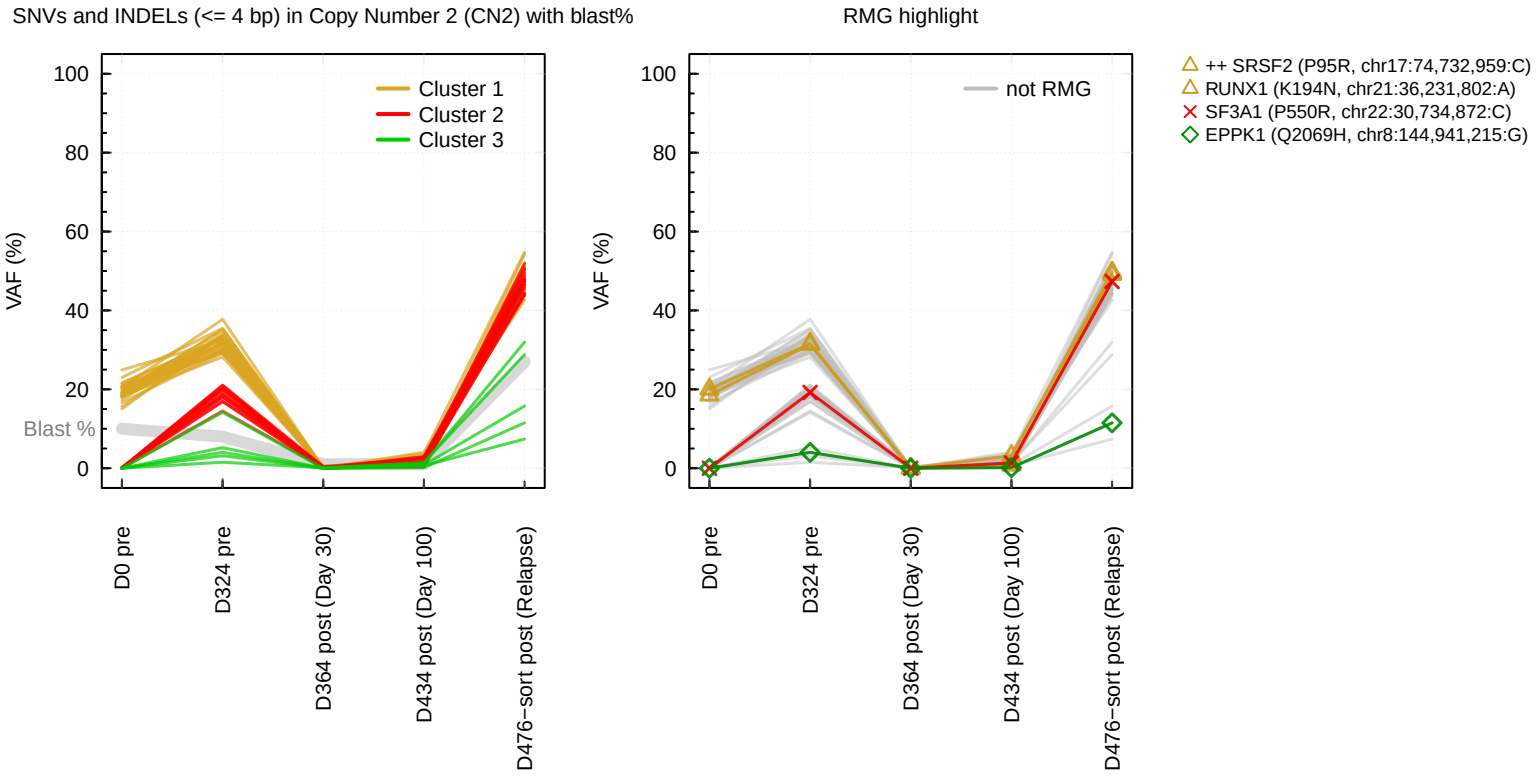
UPN 280837



SFig8

C

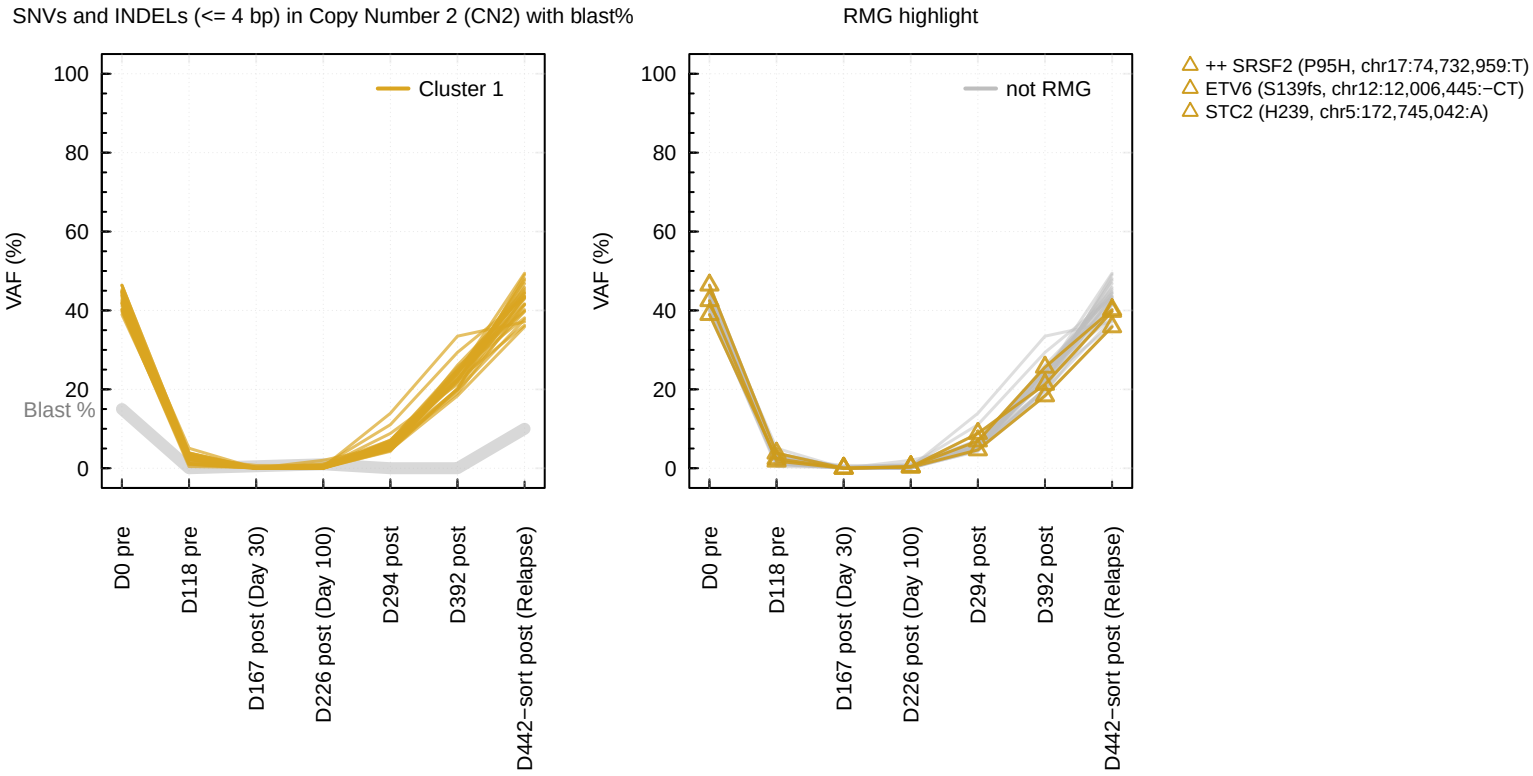
UPN 368402



SFig8

D

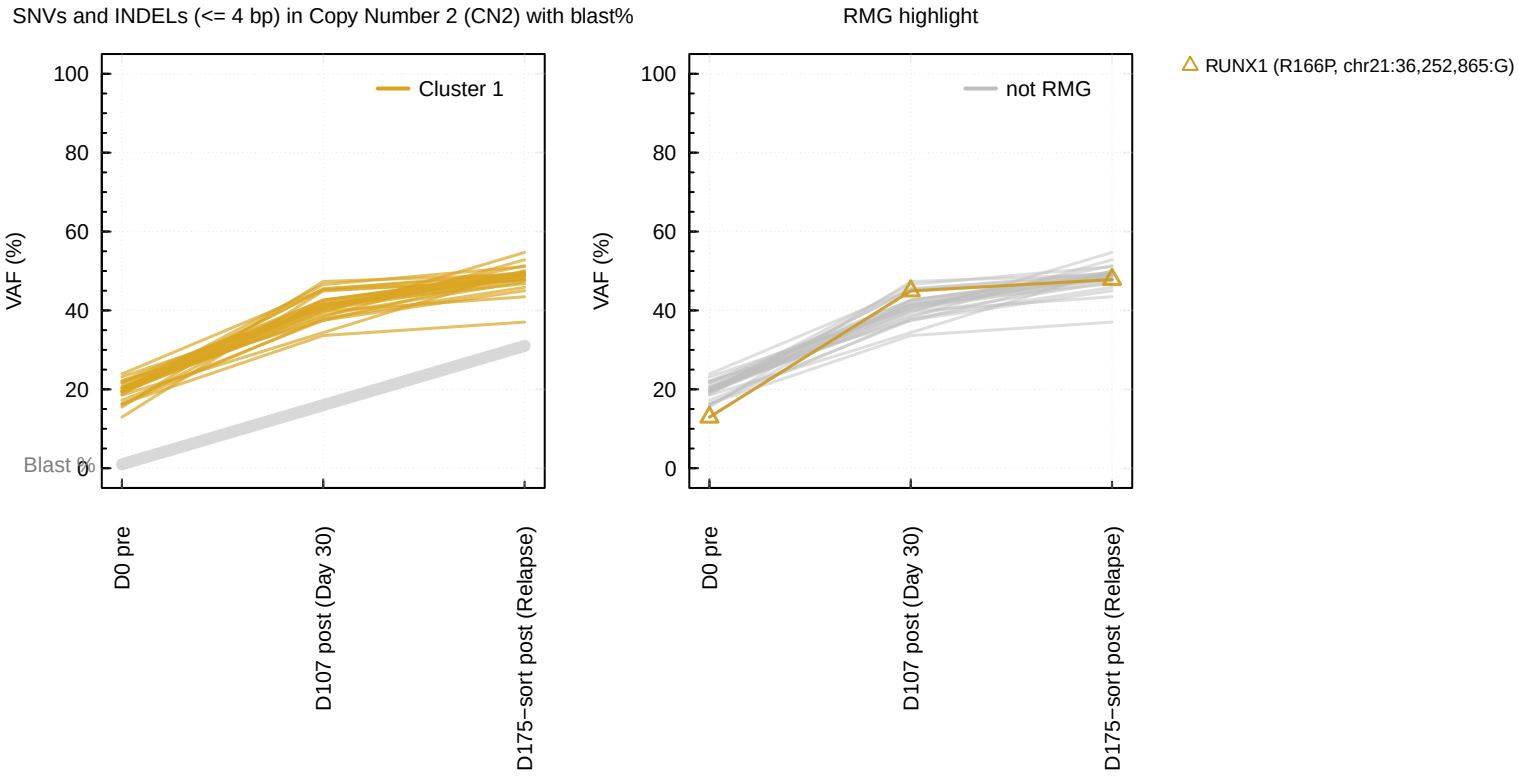
UPN 435866



SFig8

E

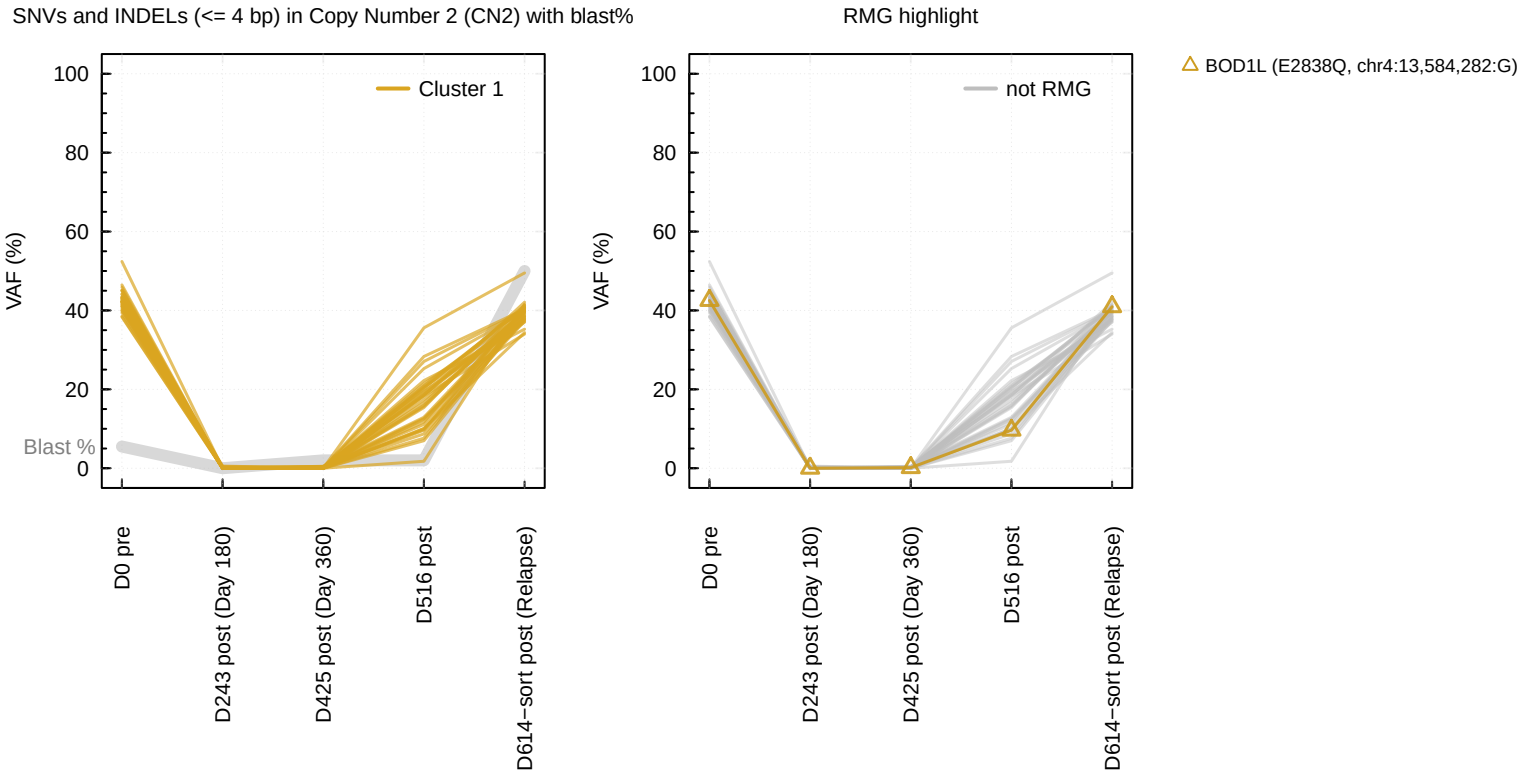
UPN 499258



SFig8

F

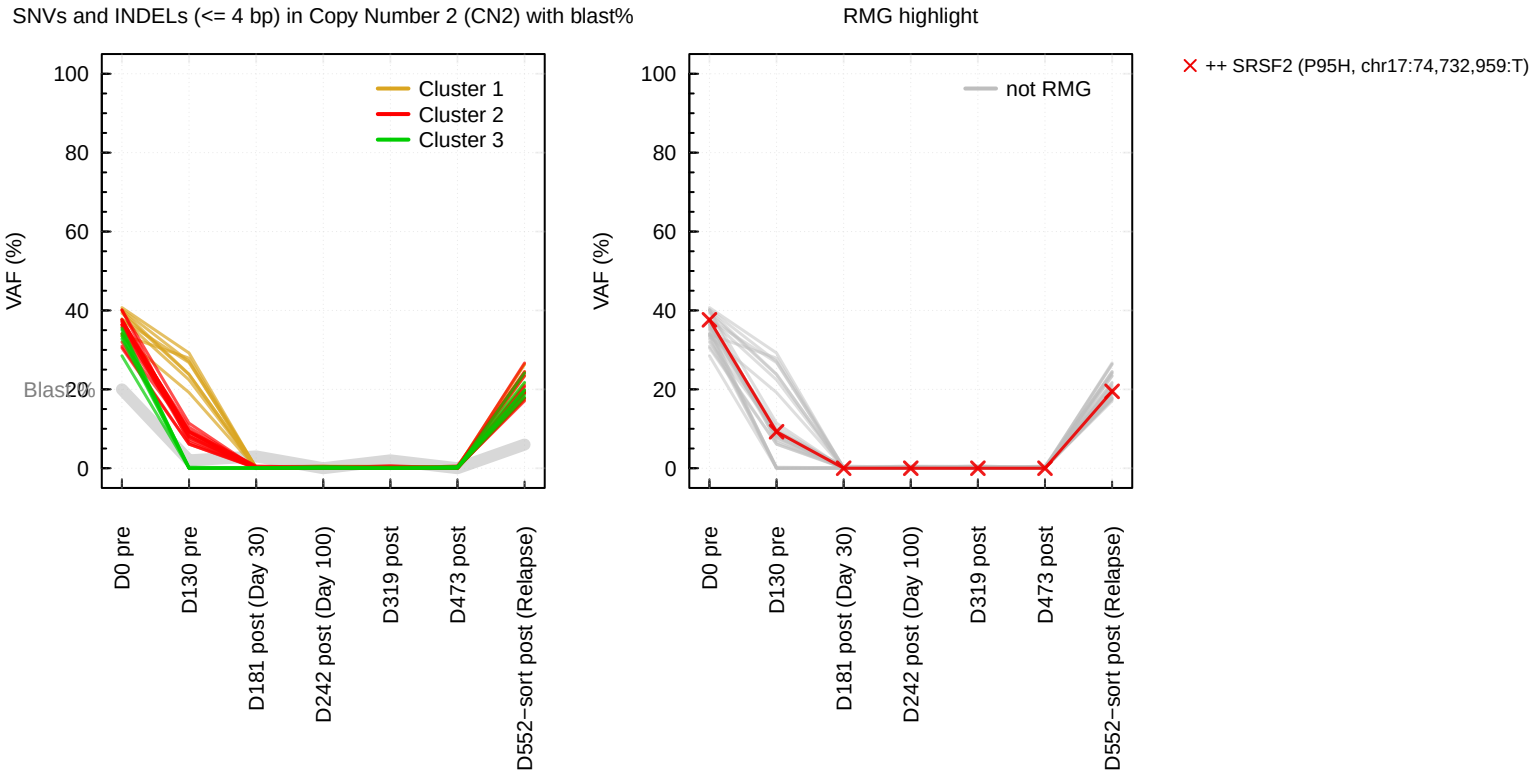
UPN 574214



SFig8

G

UPN 624702



SFig8

H

UPN 829970