

Fig. S1

a

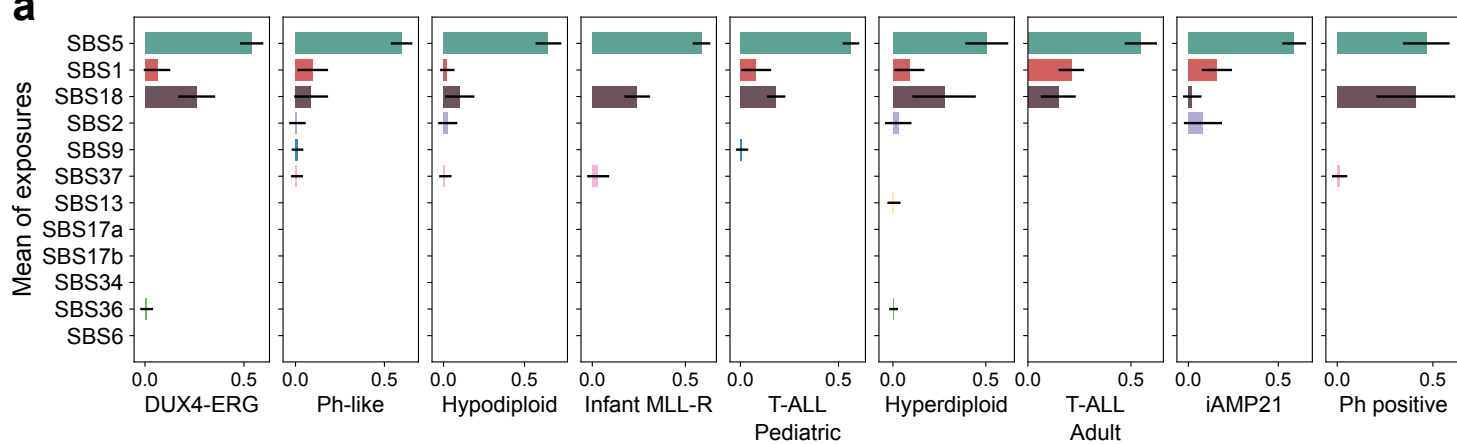


Fig. S1. Probing for active mutational processes across ALL cohorts

Mutational processes active across primary ALL cohorts, represented by their mean (and standard deviation) contribution of the mutation burden of each cohort. The list of signatures to fit was determined by their activity in any hematopoietic cancer according to COSMIC (see Supp. Methods). This linear fitting (Methods), was used to select the subset of mutational processes active in ALL tumors of the studied cohorts which are shown in the main Figures.

Fig. S2

a

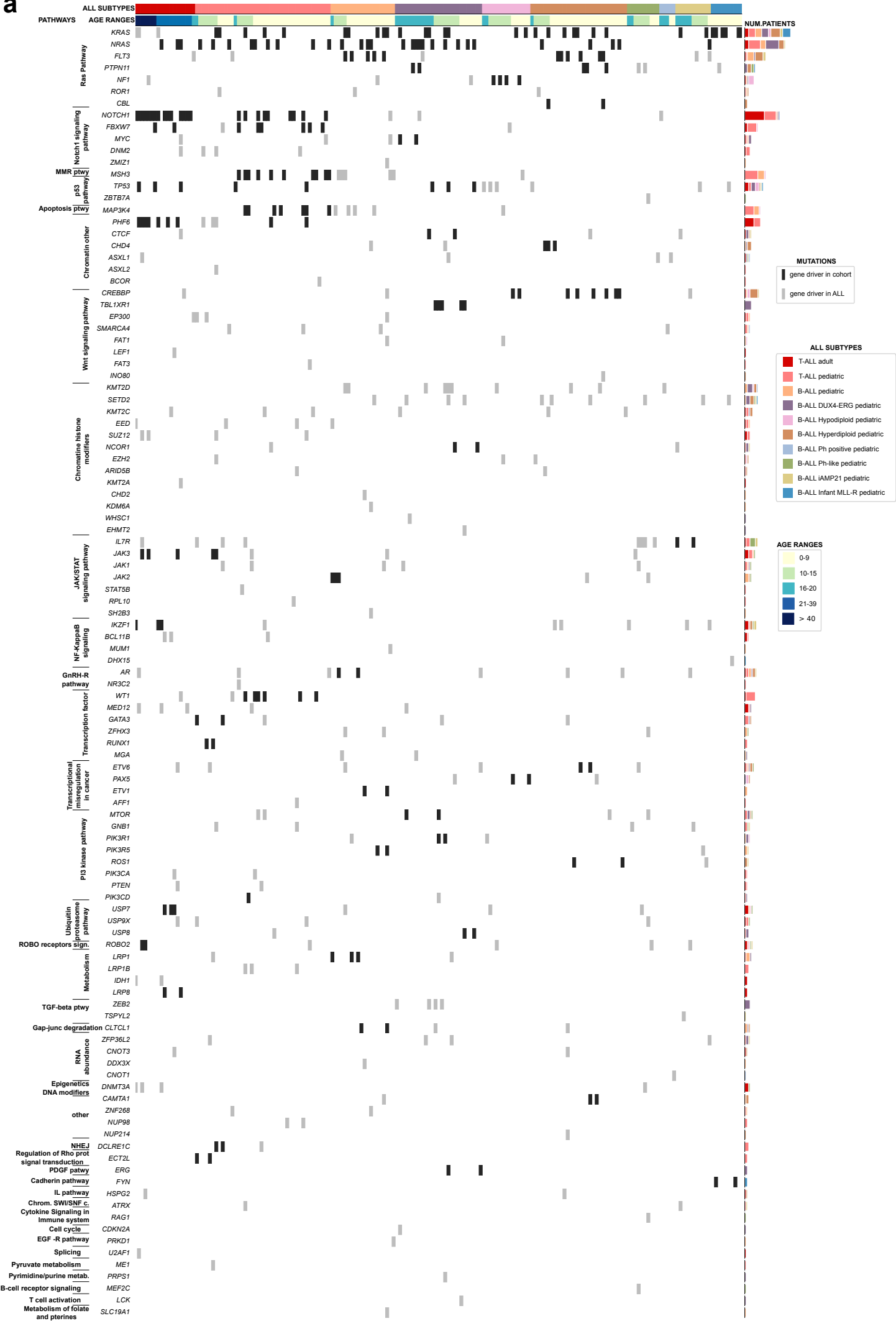


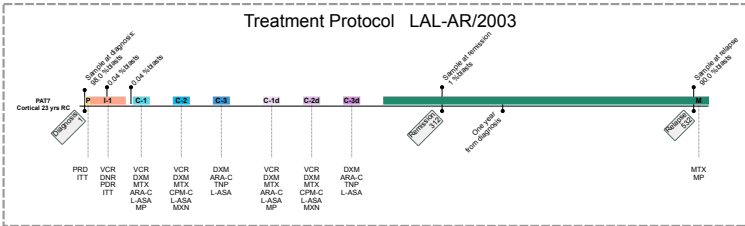
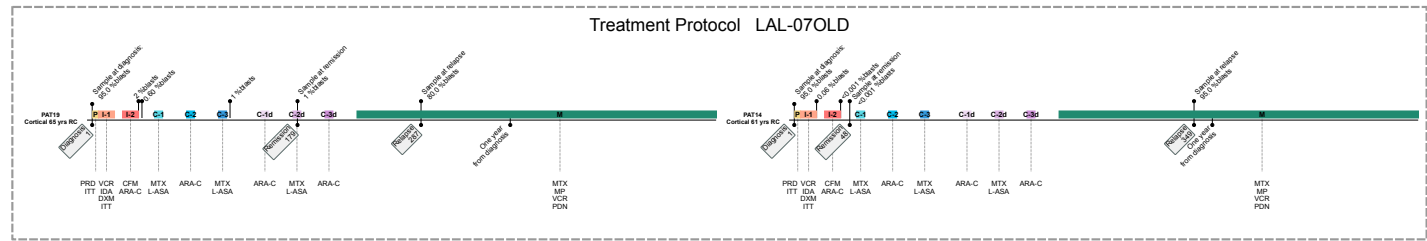
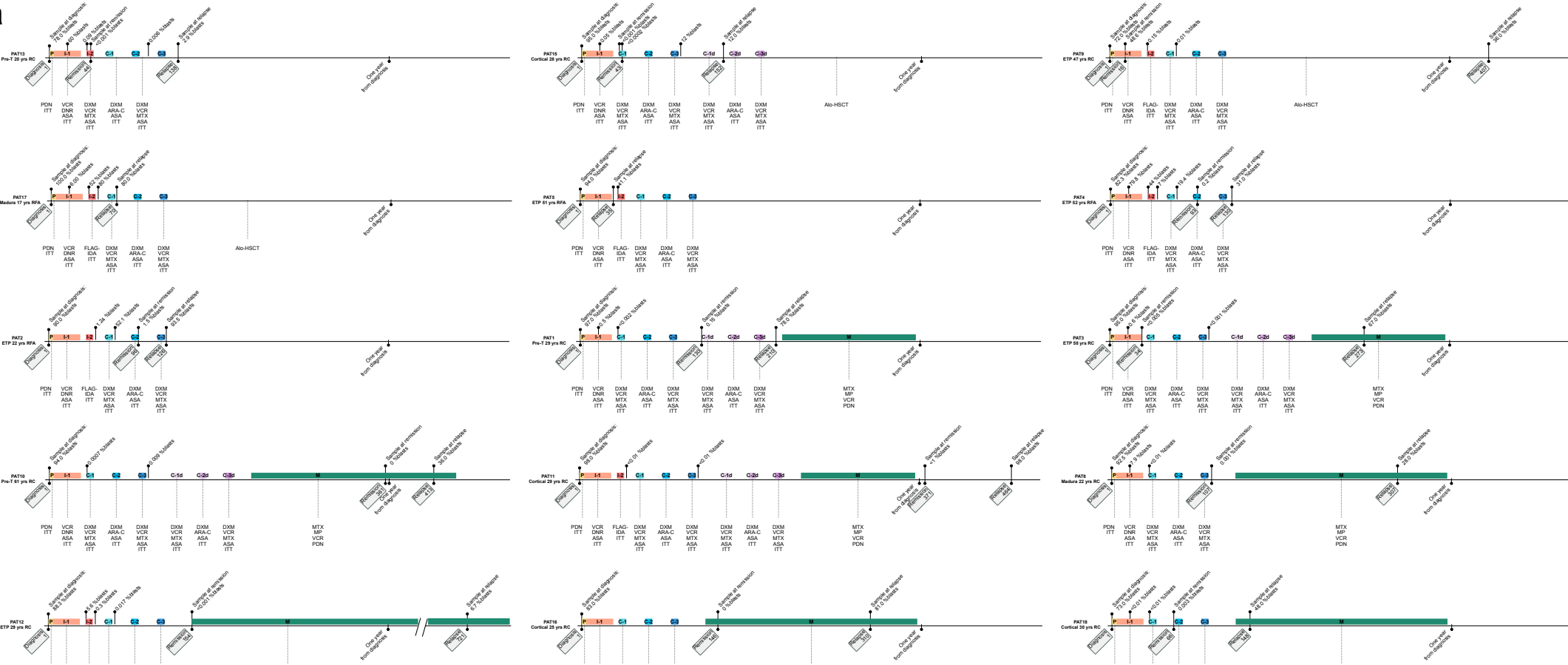
Fig. S2. Mutations in driver genes in primary ALLs

Rows are driver genes in ALL (collected from the literature or identified across these cohorts using IntOGen; see Supp. Methods) grouped by protein family, biological process or pathway. Columns are ALL samples grouped by cohort, and sorted by age. Each full rectangular cell represents a protein-affecting mutation in a driver gene annotated from the literature (grey) or directly detected as driver in that cohort through the IntOGen pipeline (black). The bars on the left represent the total number of patients in each ALL cohort with mutations of the gene.

Fig. S3

a

Treatment Protocol LAL-AR/2011



DXM: Dexamethasone
VCR: Vincristine
MTX: Methotrexate
PDN: Prednisone
ARA-C: Cytarabine
MP: Mercaptopurine
MXN: Mitoxantrone
TNP: Teniposide
ASA: L-asparaginase
CPM: Cyclophosphamide
DNR: Daunorubicin

ITT Intrathecal treatment:
- Methotrexate
- Cytarabine
- Hydrocortisone

Alo-HSCT: Allogeneic hematopoietic
stem cell transplantation

FLAG-IDA treatment:
- Idarubicin
- Fludarabine
- Cytarabine
- Granulocyte-colony
stimulating factor

P Prephase
I-1 Induction 1
I-2 Induction 2
C-1 Consolidation 1 **C-1d** Delay consolidation 1
C-2 Consolidation 2 **C-2d** Delay consolidation 2
C-3 Consolidation 3 **C-3d** Delay consolidation 3
M Maintenance

Fig. S3. Clinical course of patients in the in-house T-ALL cohort.

Each figure represents the clinical course of a patient. Day 1 represents the sample taken at diagnosis and the total extension of the line comprises one and a half years. Lymphoblast counts at regular checks and sample extractions performed at the hospital are represented above the line. Colored boxes represent treatment cycles according to clinical protocols (see Supp. Methods), and rectangular grey labels below correspond to sample extraction timepoints. The width of the treatment box is approximately scaled to the time that corresponds to protocols guidelines. Treatments received appear below the timelines. PAT16 most likely received treatment based on protocol LAL-AR/2011, however, this information has never been confirmed.

a



Fig. S4. Mutations in driver genes in primary and relapse T-ALL tumors of three cohorts

Rows are driver genes in T-ALL (collected from the literature or identified across these cohorts using IntOGen; see Supp. Methods) grouped by protein family, biological process or pathway. Columns are ALL samples grouped by cohort, and sorted by age. We added immunophenotypic information for the in-house T-ALL cohort. Primary-private and relapse-private mutations are represented as blue and yellow semicircles, respectively. Shared mutations are represented as green circles. The total number of patients affected by mutations of each gene across the three cohorts are indicated by stacked bars at the right-side of the graph. Calls from the X chromosome in the pediatric cohorts are not included (only in adults).

a

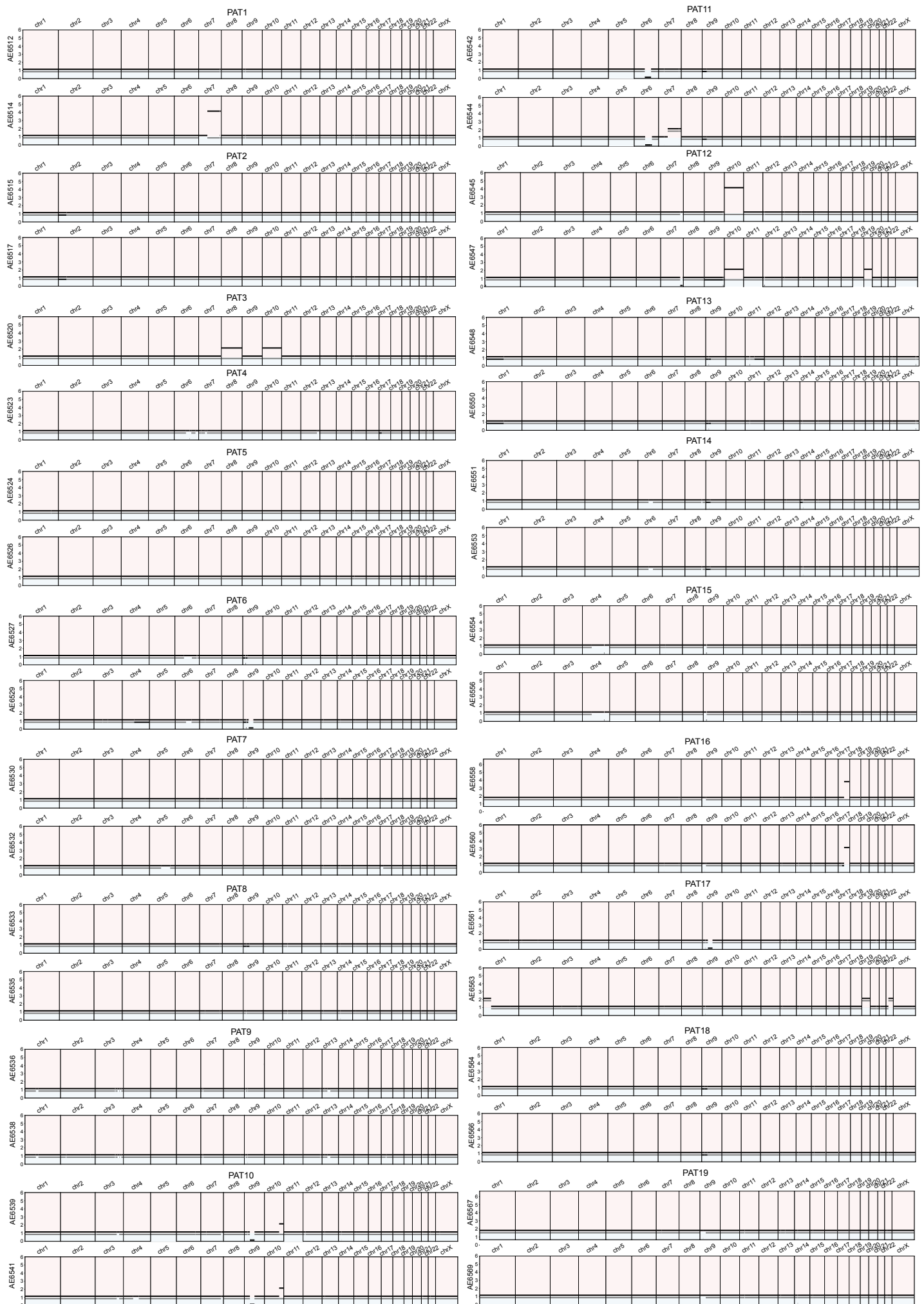


Fig. S5. Copy Number Variants detected in primary and relapse T-ALLs in the in-house cohort

In each panel, that corresponds to one T-ALL sample in the cohort, chromosomes are represented in the x-axis, with their copy number in the y-axis. Copy number of the major allele is represented as a black line and that of the minor allele as a grey line. In diploid segments, both the black and grey lines appear close to 1 (total sum of 2) but not overlapping for visual purposes. Red and blue shaded backgrounds represent amplifications and deletions, respectively. For all patients (indicated in the graph title), the top and bottom plots correspond to the primary and relapse samples, respectively.

Fig. S6

a

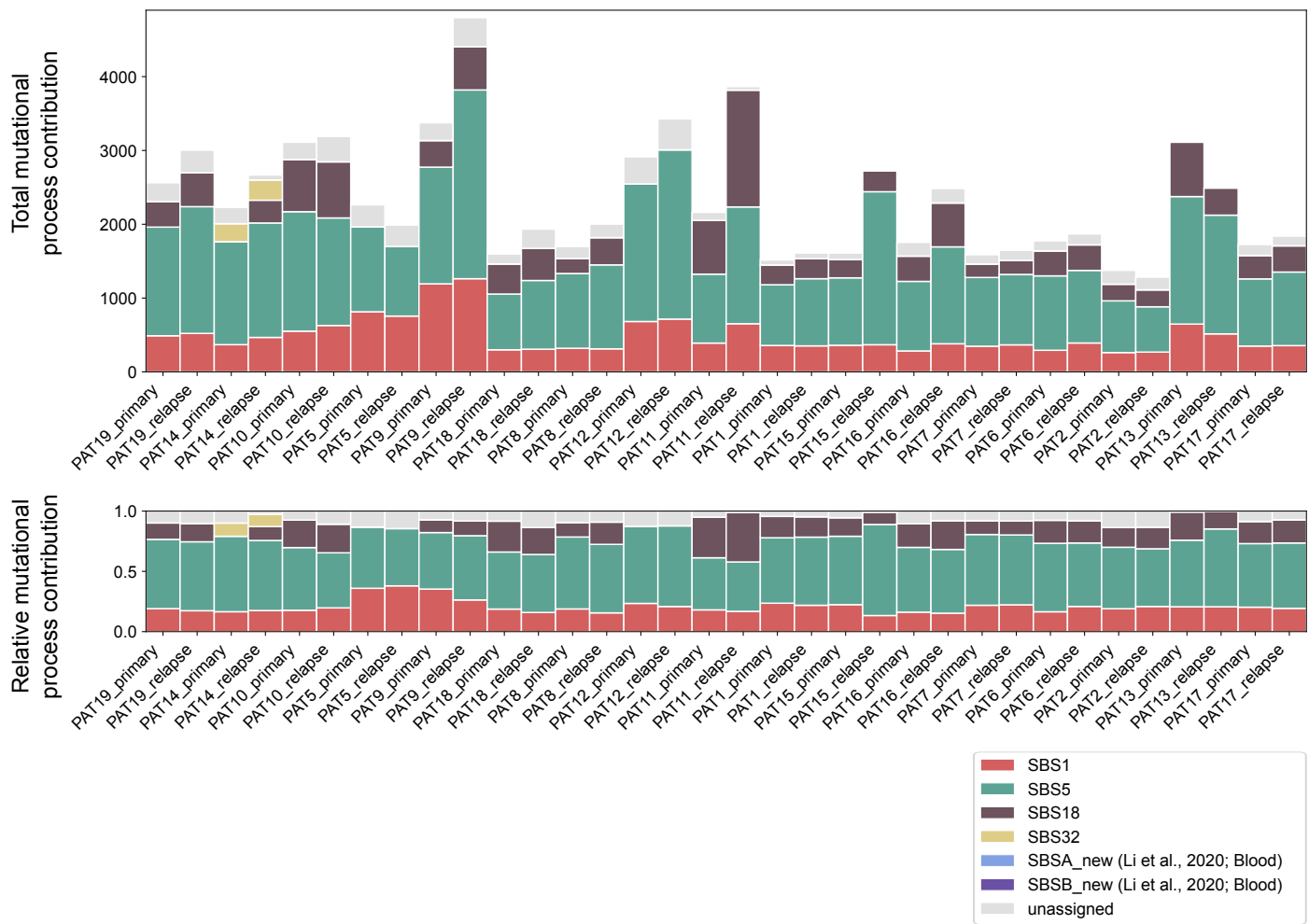


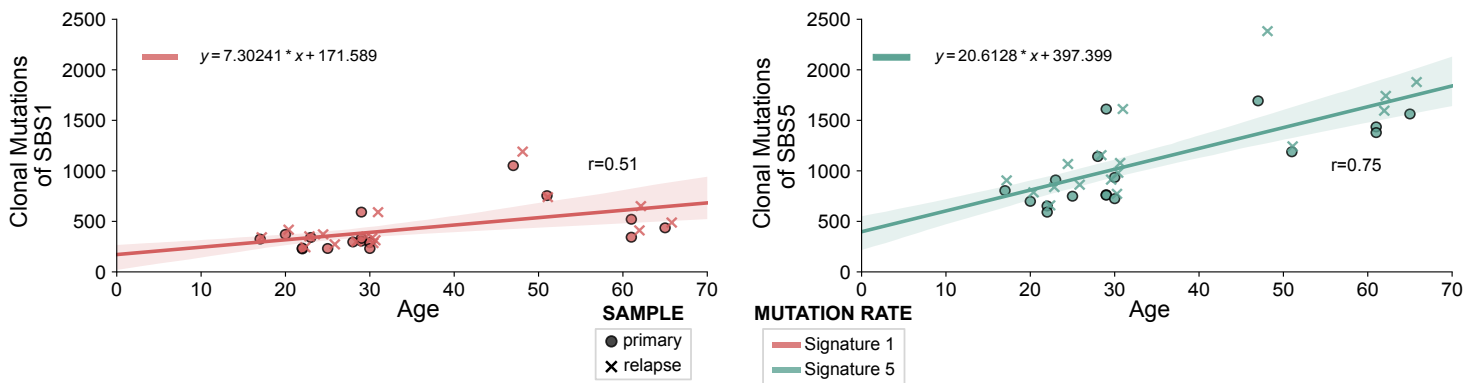
Fig. S6. No mutational footprints attributable to treatments are detected

Panels represent the absolute (top) and relative (bottom) contribution of mutational processes (signatures listed in legend) to the mutation burden of the primary and relapse malignancy of each patient. As indicated in the legend, colors of the bars indicate the signatures used in the fitting process. Note that, although included in the fitting, the mutational signatures (or footprints) of drugs used in ALL treatment (recently identified in pediatric relapse tumors; [57]) are not detected in relapse T-ALLs in the in-house cohort.

Fig. S7

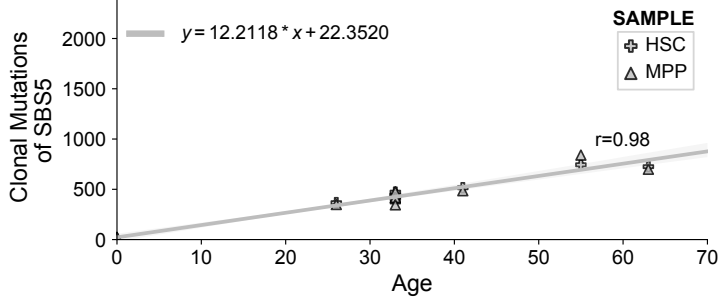
a

Adult T-ALL (in-house) cohort



b

SBS5 from Hematopoietic Cells
(Osorio et al., 2018; Cell Reports)



c

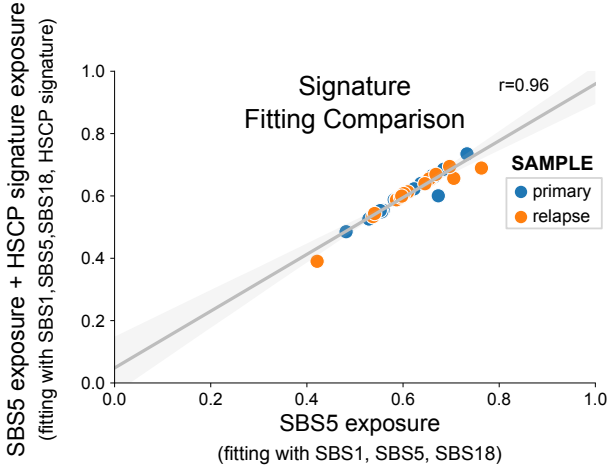
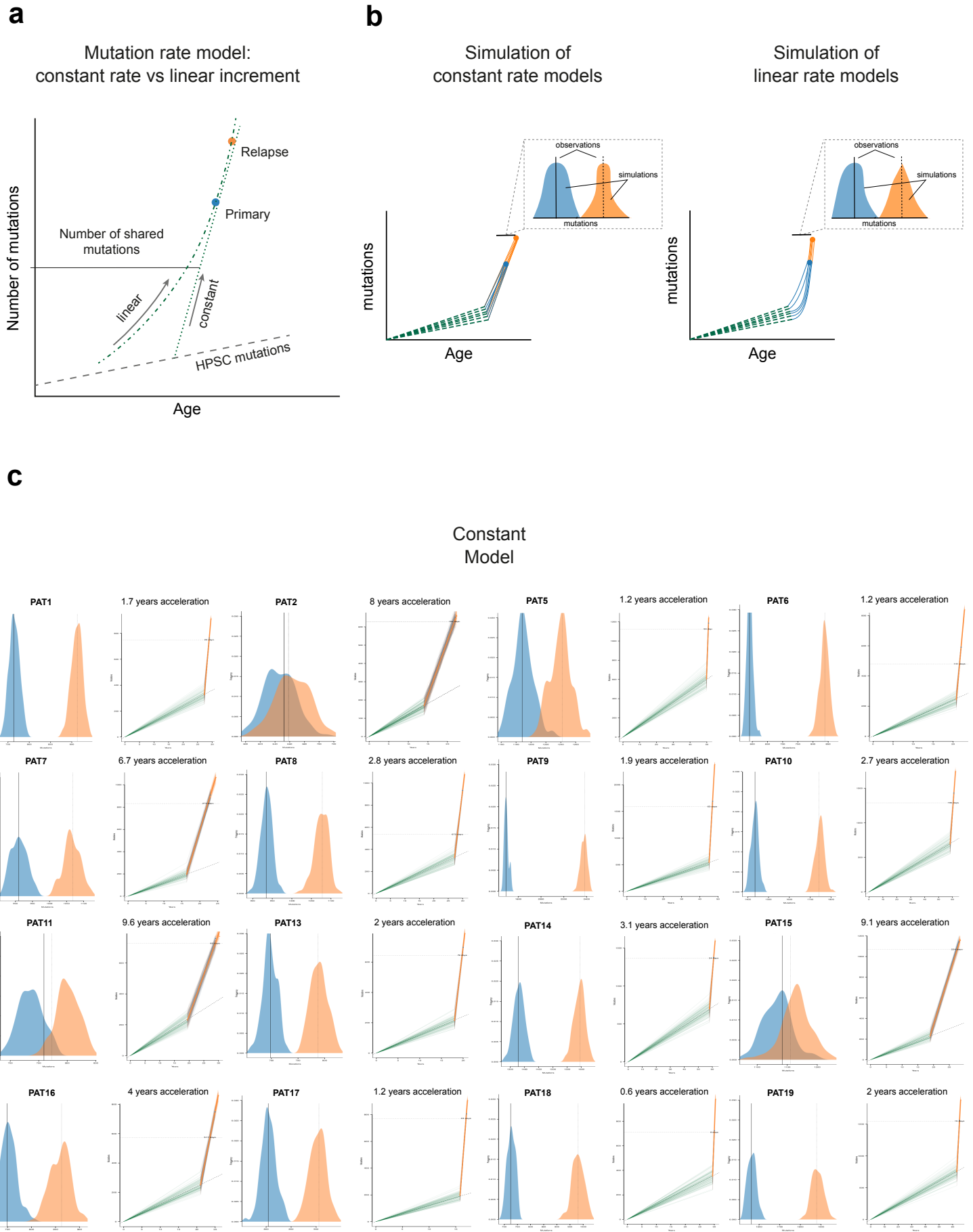


Fig. S7. The contribution of clock-like Signature 5 to the mutation burden of T-ALLs

a) Known clock-like signatures 1 (left panel) and 5 (right panel) contribute clonal mutations at a steady rate across the lives of T-ALL patients in the in-house cohort. Although the number of clonal mutations contributed by each process significantly correlates with the patients' age (signature 1 $p=2.30 \times 10^{-3}$, signature 5 $p=3.64 \times 10^{-7}$), the correlation is stronger in the case of signature 5, which had been previously observed [46]. Moreover, signature 5 contributes more age-related mutations than signature 1. Dots are primary and cross relapse samples. Trendlines following the regression are added and their equations are indicated at the top right of each panel.

b,c) Signature 5 also fits well the steady accumulation of mutations in healthy hematopoietic stem cells and multipotent progenitors (HSC and MPP; Osorio et al., 2018 [47]) with aging. Signature 5 contributed mutations fit very well ($r=0.98$) the mutational burden of clonally expanded HSC and MPP (b), implying that it is probably the main mutational process active in these cells and that signature 5 mutations accumulate steadily over time. When a *de novo* signature extraction (rather than the signature deconstruction presented so far in this article; see Supp. Methods) is carried out on the mutational profile of HSC and MPP a specific HSC population (HSCP) signature is extracted [47,48]. The activity (fraction of contributed mutations) of this signature across HSC and MPP cells correlates very well ($r=0.96$) with that obtained for the fitted signature 5. This supports the idea that the HSCP signature and signature 5 represent the same underlying mutational process active in HSC and MPP.

Fig. S8



Linear Model

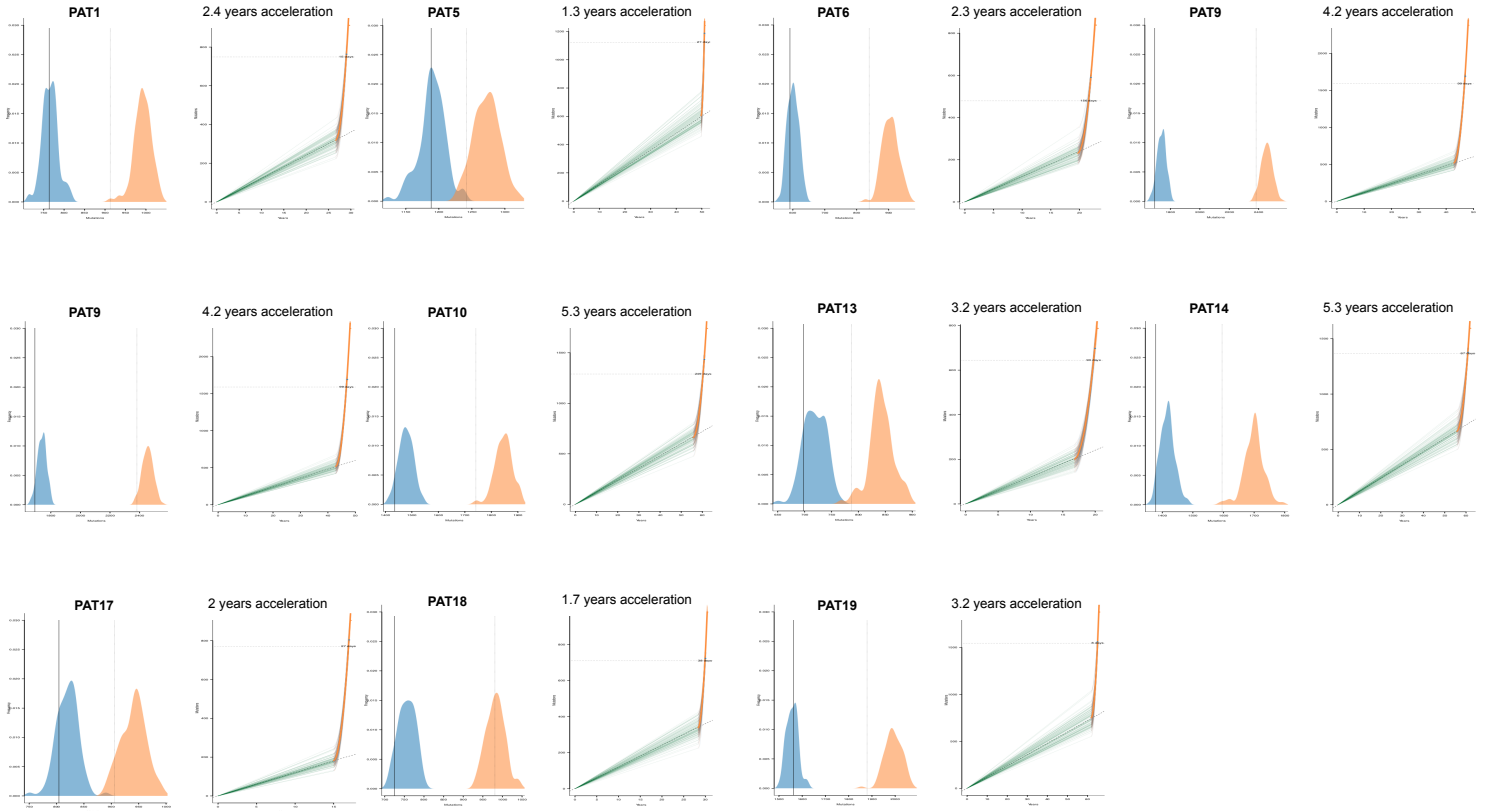


Fig. S8. Models of accelerated mutation rate in T-ALL

a) Schematic representation of two extreme models of the accelerated mutation rate in T-ALL.

The baseline mutation burden increase represents the steady accumulation of mutations of HSC and MPP shown in Supplementary Figure 7. The linear model consists in a steady acceleration of the mutation rate throughout all the evolutionary history of the T-ALL. On the other hand, in the constant model the acceleration occurs only once in the evolution of the T-ALL, which after this point maintains a steady increase of the mutation burden.

b) For both, the constant (left plot) and linear (right plot) models, a number of simulations of accelerated mutation rate are carried out, represented in these schematic graphs by dotted lines. The likelihood of each explaining the observed mutation burden of primary and relapse samples is then computed, as explained in the main manuscript.

c) Real examples of the more likely models given the observed data for each patient of both types of simulations (constant and linear). The years when the mutation rate accelerated and incremented are written above each pair of graphs per patient.

Fig. S9

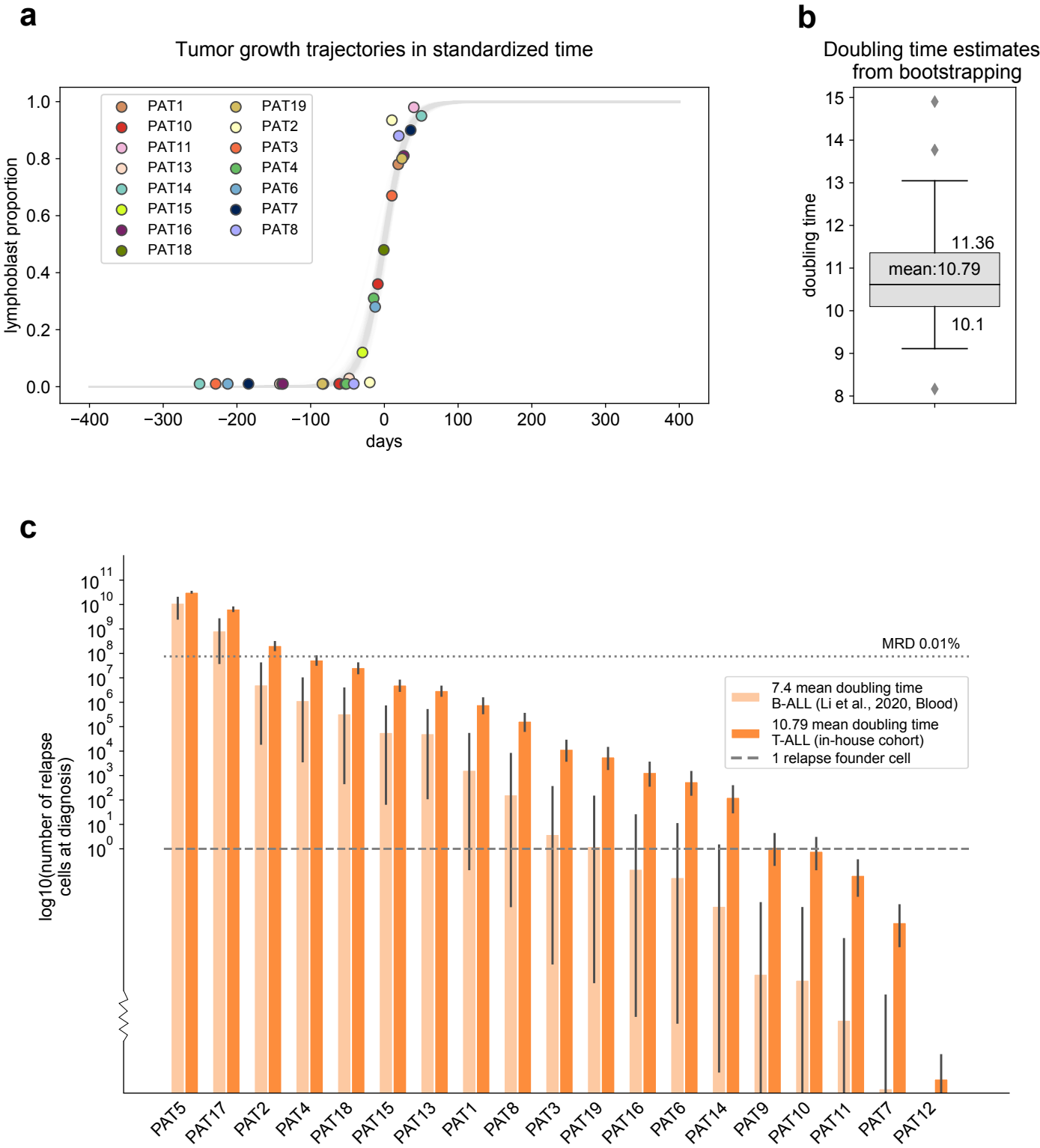


Fig. S9. Estimating the doubling time of T-ALL population from the pathologists' observations

a) Bootstrapping adjustment of growth logistic curves to the counts of lymphoblasts in remission and relapse bone marrow samples carried out by the pathologist. The observations at these two points (dot pair) for each patient are represented with the same color. Paired-dates of the bone marrow sampling are re-scaled to a standardized time where 0.5 blast proportion (y-axis) falls at day 0 (x-axis) of the growth trajectory of the malignancy of each patient.

b) Boxplot with the doubling time estimates resulting from the bootstrapping. The line at the center of the boxplot is the mean. The first and third quartile of the distribution of bootstrapped doubling times are also represented.

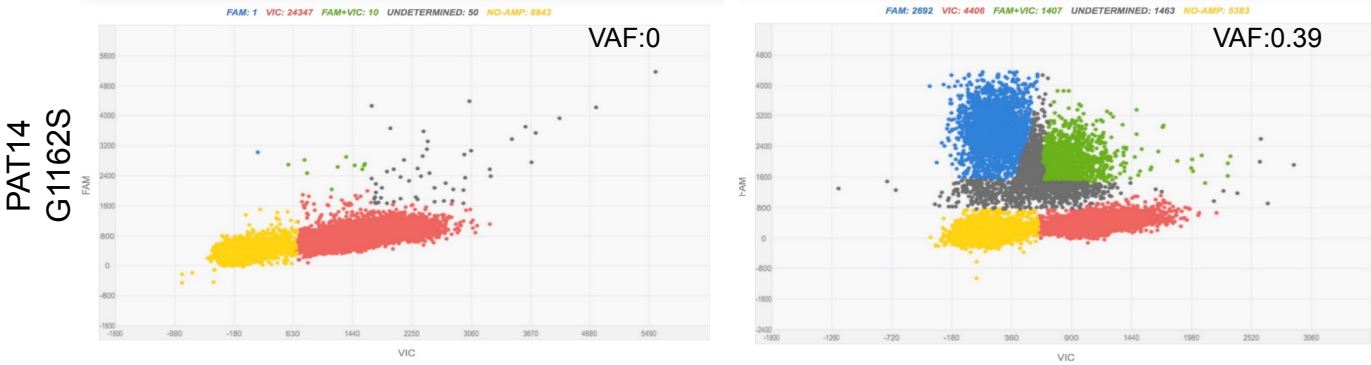
c) Comparison of the number of relapse cells for all patients in the in-house T-ALL cohort computed at time of diagnosis using two different estimates of the doubling time of the lymphoblastic population. Light orange bars represent the size of the T-ALL relapse population computed using the doubling time estimate obtained recently for pediatric B-ALLs (Li et al., 2020) Dark orange bars represent the size of the T-ALL relapse population computed using the doubling time estimate obtained in (b). Error bars are the estimates corresponding to the doubling time in the first and third quartiles of the distribution.

Fig. S10

a

Diagnosis

Relapse



b

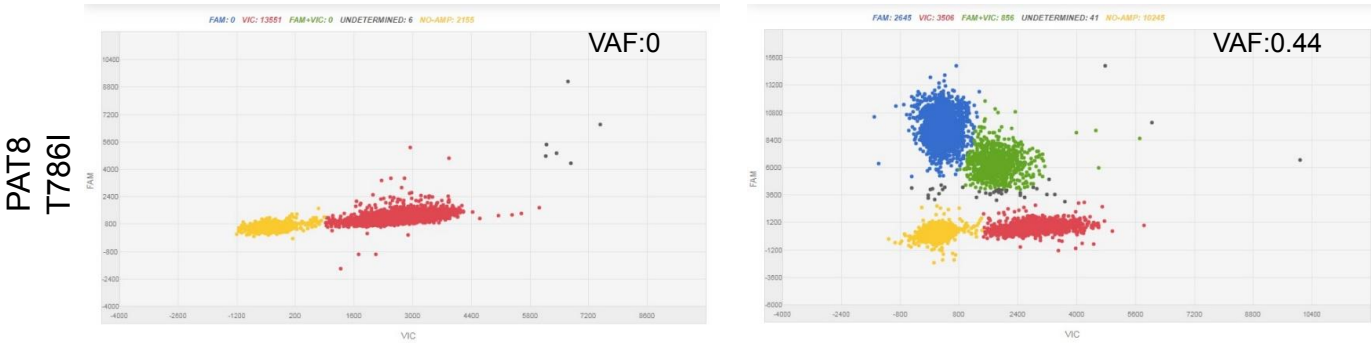


Fig. S10. Results of digital PCR on mutant *SMARCA4* in two primary samples

Detection of mutations a) G1162S and b) T786I in the relapse-enriched *SMARCA4* gene in primary samples of PAT14 and PAT8 respectively was negative in both (VAF= 0). The resolution of the dPCR in PAT14 was 0.089% whereas in PAT8 it was 0.11%. The VAF detected of the mutants in relapse derived from the dPCR is similar to the one detected by NGS which are both close to the expected 0.5 for an heterozygous variant with no normal contamination (0.39 vs 0.403 and 0.44 vs 0.346). Scatter plots showing the distribution of the data points based on the dyes used (VIC and FAM). Blue dots (FAM) represent presence of mutant *SMARCA4* and red dots (VIC) represent wild-type *SMARCA4*; yellow refers to no amplification and green to co-amplified wild-type and mutant species.

Fig. S11

a

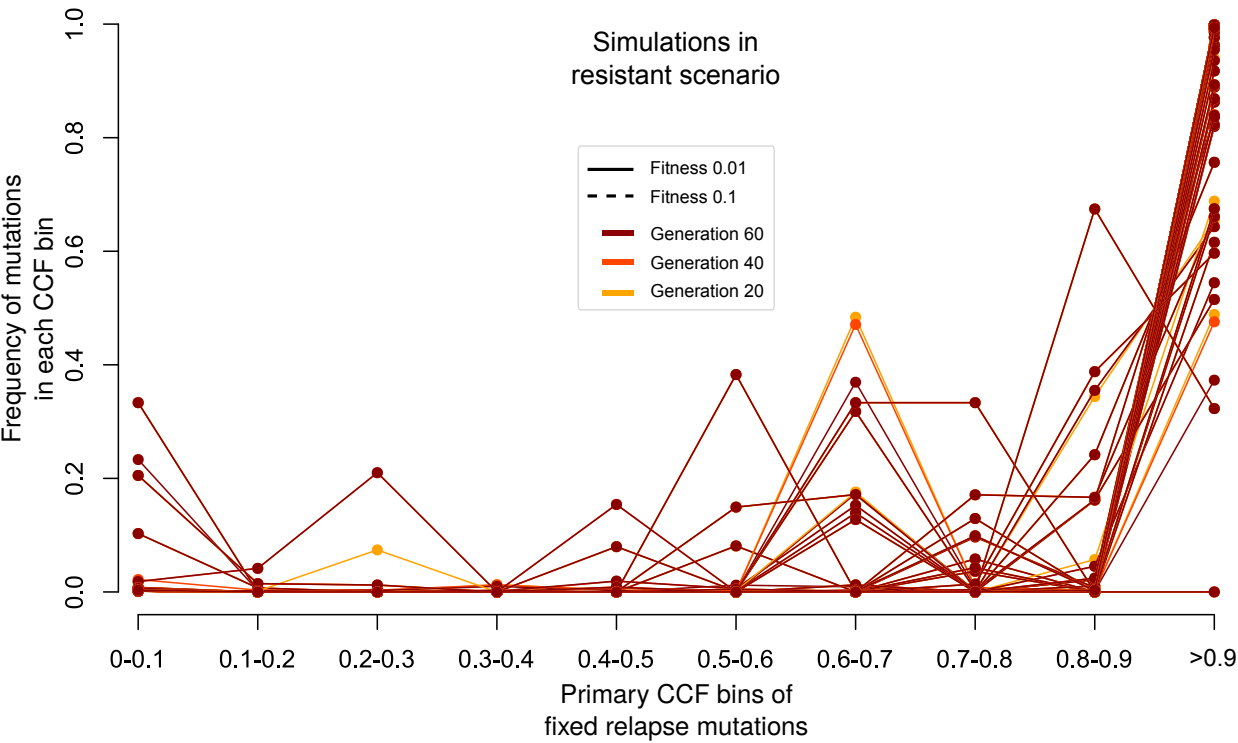


Fig. S11. Distribution of primary CCF of relapse fixed mutations in the simulated resistant scenario

Distribution (frequency) of CCF values of mutations in synthetic primary T-ALL populations in evolutionary simulations following the resistant scenario defined in the toy example of Figure 5b. The dots represent mutations binned at different CCF values with the frequency that each bin represents with respect to all mutations in each synthetic relapse population. All the results of six simulation settings with different values of fitness of driver mutations and number of cell generations are presented.