Supplementary Methods

Driver mutation identification

We previously adapted an approach described by Papaemmanuil et al. to identify high-confidence driver mutations in the bone marrow samples without matched germline control.[1,2] First, only variants that would introduce a protein-coding change were kept for further analysis. Mutations with ANNOVAR annotation of nonsynonymous, stop-gain, stop-loss, splicing, frameshift insertion, frameshift deletion, nonframeshift insertion, or non-frameshift deletion were considered able to introduce a protein-coding change. Second, common polymorphisms were removed to reduce the load of possible germline contamination due to the absence of matched normal control. Specifically, a series of public variant databases including the 1000 Genome Database (http://www.1000genomes.org/), ESP6500 Database (http://evs.gs.washington.edu/EVS/), dbSNP ver.132 (http://www.ncbi.nlm.nih.gov/SNP/), and Exome Aggregation Consortium database (http://exac.broadinstitute.org/), were used. Variants with a population frequency of 0.14% or more in any of those databases were considered possible germline polymorphisms and were therefore removed from further analysis.

Finally, a hierarchical classification system was used to assign a confidence level for each remaining variant to facilitate the identification of putative somatic driver mutations. Variants were considered drivers if they were confirmed somatic mutations based on annotation in the COSMIC database (version 81); were loss-of-function mutations such as splicing, stopgain, stop-loss, and frameshift mutation in well-characterized tumor suppressor genes; or were recurrent variants that reside within three amino acids of a confirmed somatic mutation. Confirmed variants were also divided into whether they were previously associated with hematologic or non-hematologic malignancies. All other variants were considered variants of unknown significance (VUS). Any VUS that was either likely FLT3-ITD variants or had experimental evidence supporting is pathogenicity was subsequently included (Supplementary Figure 2). To ensure inclusion of all potential "ITD" variants, insertions were considered possible ITD sequences if the insertion sequence was at least 3 nucleotides long, based on prior reports of the range of FLT3-ITD sequence length[3].

Excluding variants

Variants were excluded if they were not known to be associated with AML but appeared in most patients in a dataset, suggesting that the variant was an artifact. Several low-level variants were also excluded because they were consistently mutated in a mean of 1-5% of cells per sample, were not called in available bulk sequencing samples, and were not known to be drivers associated with hematologic malignancies based on COSMIC annotation. These were excluded because of the uncertain clinical relevance of novel and consistently low-level variants, and it was possible that these low-level variants could be related to technical artifacts, such as errors in polymerase chain reaction (PCR) during library preparation. For one specific variant in *ASXL1* that was highly recurrent in the MD Anderson dataset[4] and which is known both to be a driver but also subject to technical artifact from PCR slippage[5], the variant was excluded even though its mean percent of cells mutated was greater than 5%. Supplementary Figure 3 shows the total number of mutation events and the number of specific variants that were blacklisted due to each of these criteria.

Creating mutation trees

For each sample, a mutation matrix was created using loom files from the Tapestri Pipeline output. In these matrices, rows were individual variants and columns were individual cells. The presence (mutant), absence (wildtype), and missingness (no call) of a mutation event was noted for each cell, and zygosity was not considered for modeling clonal evolution. When a single patient had multiple samples, mutation matrices were merged so that samples lacking a mutation present in another sample were considered to have wildtype for that mutation in all cells.

Single Cell Inference of Tumor Evolution (SCITE)[6] was then applied to all mutation matrices, with the following parameters: false positive rate 1%, false negative rate inferred with prior probability set to the mean Tapestri Pipeline allele dropout estimate for samples from that patient, doublet rate (rate of multiple cells sequenced as a single cell) inferred by SCITE, and Monte Carlo Markov Chain length of 30,000. Allele dropout estimates were from Tapestri Pipeline v2 using the updated allele dropout calculation compared to v1. In the few cases this allele dropout estimate was unavailable, either a Tapestri Pipeline v1 allele dropout estimate (if available) or mean imputation was used as the prior allele dropout (false negative) estimate. Trees were inferred using only single-cell data rather than also using bulk data in part because of

the limited benefit of bulk data when the number of cells per sample is high[7]. Due to the computational complexity of accounting for mutation loss and the challenges distinguishing mutation loss from allele dropout, mutations were assumed to be gained exactly once and never lost (infinite sites assumption).

Creating high-confidence trees

Although SCITE efficiently finds optimal trees under a variety of conditions[6], it is possible that some links in the tree are not well-supported by the data from the mutation matrix. For example, the optimal tree architecture may include a path such that a later mutation occurs in the same clone as an earlier mutation, despite few of the cells with later mutation also having the earlier mutation. This can be due to violations of model assumptions, such as mutations occurring multiple times in a tree or reverting to wildtype, or due to challenges reconciling false positives, allele dropout, and doublets. An example tree from the Stanford cohort containing a connection that was not well-supported by the data is depicted in Supplementary Figure 5C.

To address this, we defined a "low-support" directed connection between two mutations as one where <50% of cells with a later mutation also contained an inferred earlier mutation (among cells that had calls for both mutations). To ensure that all connections in the trees were well-supported during mutation order inference while also minimizing post-hoc edits of the trees, variants were iteratively removed from the trees if they contributed to the greatest number of low-support connections, and greater distance to the root node of the tree used as a tiebreaker. Removed variants (Supplementary Table 3) were disproportionately low-level signaling mutations (Supplementary Figure 5A-B).

Tree stability

Although the SCITE algorithm involves randomization, the trees were nearly always stable. The posterior distributions of the trees consisted of mostly the same tree in 94% of cases, and even when sampling >20,000 trees from the posterior, 75% of posterior distributions consisted of only one tree (Supplementary Figure 14). For a concrete interpretation, we ran SCITE a second time on all trees considered for analysis using the same SCITE parameters, and only 5% of trees changed architecture, with only 1% changing architecture after excluding variants of unknown significance and low-support variants.

Merging FLT3-ITD variants

Several samples had multiple different insertion sequences in *FLT3* exons 14 or 15, the characteristic exons for *FLT3* internal tandem duplication (FLT3-ITD) mutations. However, even when the sets of cells containing different sequences were disjoint, we suspected that many cases of multiple insertion sequences per sample were part of the same ITD mutation event for several reasons: 1) the number of distinct insertion sequences per sample varied significantly across datasets, suggesting the number of "ITD" events was due to batch effects (Supplementary Figure 4B); 2) FLT3-ITD variants can be unstable and change over time[8]; 3) ITDs can be longer than the sequencing reads[9], which were 150 base pairs in all of the datasets; and 4) errors during library prep, sequencing, or sequence mapping could also result in the appearance of rare but distinct insertion sequences in different cells.

To capture the different evolutionary patterns that could appear because of these issues, we created trees for each patient using three different FLT3-ITD merging strategies:

1. "Conservative" merging: only merge FLT3-ITD insertion sequences if one was a perfect subsequence of another, and if they started at the same genomic locus. This strategy allows for sequences to be shorter due to limitations in read length or clipping but assumes that all insertion sequences that aren't perfect matches represent different ITD events.
2. "Liberal" merging: merge sequences that were at least 90% similar irrespective of genomic locus. Similarity was determined using local sequence alignment, and in the similarity assessment, alignment gaps had the same weight as a change in nucleotide. This strategy assumes that sequences that have a highly similar subsequence are the same ITD event. Mapping ITDs precisely to a genomic locus can be challenging, so this strategy also allows for highly similar sequences to be merged irrespective of the precise genomic position.
3. "All" merging: merge all potential FLT3-ITD sequences, irrespective of sequence similarity and genomic locus. This assumes that any potential ITD sequences were part of the same event and that the sequencing captured different sections of the same insertion.

At baseline, we assumed that all perfect sequence matches from the same locus should be merged, so all trees had at least "Conservative" merging of potential FLT3-ITD variants. After

this merging, the number of ITDs dropped substantially, and this strategy had similar results as the "Liberal" strategy (Supplementary Figure 4B). Otherwise, to choose between these three different strategies, we leveraged the clonal architecture inferred from the single-cell data in two ways:

1. FLT3-ITD events were merged if all potential ITDs were terminal events (leaves), originated from the same parent event in the tree, and could be merged without changing the rest of the tree architecture. This rule allows merging events that reflect the same pattern of clonal evolution.
2. If there were still multiple ITDs, the merging scheme was chosen such that the fewest assumptions were made about the number of ITD events (i.e. "Conservative" merging was prioritized, followed by "Liberal", followed by "All") but only if the tree did not acquire more low-support connections, as defined above. This strategy captures whether merging variants improves the strength of tree's connections overall.

In most cases, there was only one ITD remaining after "Conservative" merging, and among cases with multiple ITDs, usually all ITDs were merged (Supplementary Figure 4C-D).

Despite exploration of different merging strategies, it remains unclear how many FLT3-ITD events truly occurred for the reasons described above. To ensure that our merging strategy did not substantially affect mutation order inference, we also explored whether *NPM1* and *FLT3* mutation ordering changed between "Conservative" merging and our resulting merging scheme. Of the 44 patients with both *NPM1* and *FLT3* mutations, only one (AML-88 from the MD Anderson cohort, Supplementary Figure 15) had a change in the mutation order, although the FLT-ITD variant in this graph was not used for subsequent order inference because of low support connections with downstream *FLT3* variants. Among cases that did not have "Conservative" merging, the merging strategy described largely reduced the number of distal (far from root node) FLT3-ITD mutations (Supplementary Figure 16).

Defining early signaling mutations

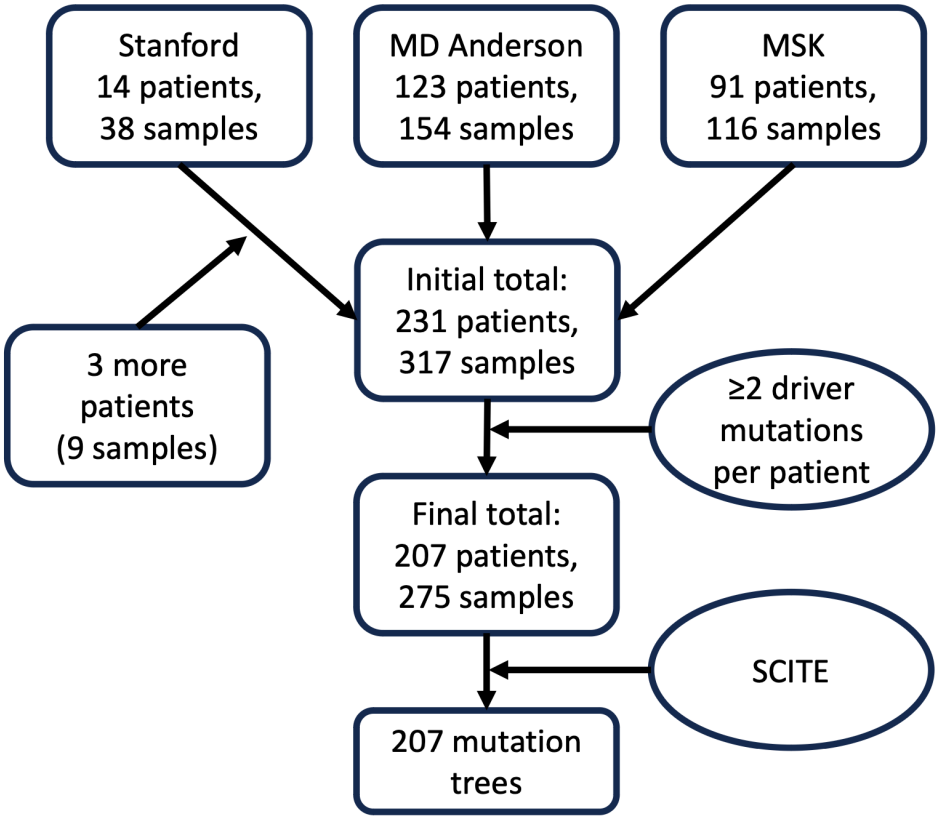
In select trees, a signaling mutation, such as in *FLT3* or *NRAS*, arose from the root node without a preceding mutation, and sometimes, it was the only mutation in a clone. We considered mutations that arose without a preceding mutation to be "early," but given the limited coverage of the targeted sequencing panels, we could not exclude the possibility that another mutation

arose before the signaling mutation but was missed by our panels. We leveraged available bulk sequencing data either from published whole exome sequencing or extended targeted panels that were used for routine clinical care to explore this possibility. Of the 39 cases with early signaling mutations, 2 had corresponding whole exome sequencing[10], and an additional 15 had extended targeted sequencing. Review of the mutations from these bulk sequencing data for DNA methylation, transcription factors, and chromatin/cohesin mutations that may have met our criteria as "driver" mutations are shown in Supplementary Table 7. Only 4 of the 17 patients had potential early driver mutations that were unaccounted for. Furthermore, careful review of these mutations suggests they are rare in AML and/or that their functional consequences are unclear because their pathogenicity is computationally derived and/or the was rarely associated with AML. Given the uncertainty in placing these mutations in the trees and their uncertain pathogenicity, these mutations were not considered in cases with early signaling mutations.

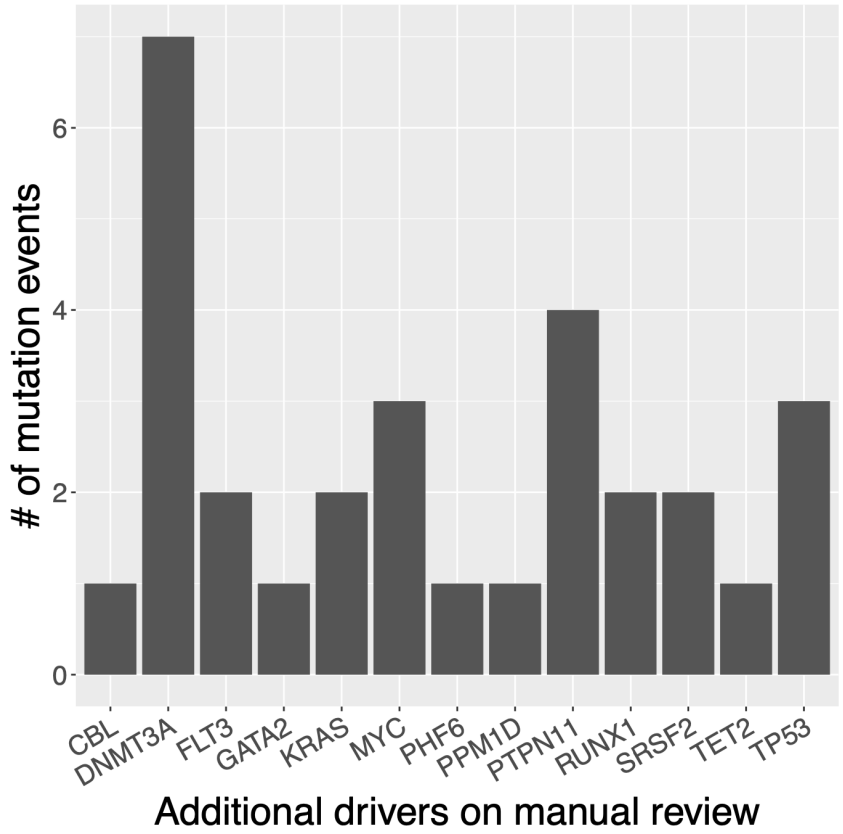
1. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627; quiz 3699 (2013).
2. Morita, K. *et al.* Clearance of Somatic Mutations at Remission and the Risk of Relapse in Acute Myeloid Leukemia. *J Clin Oncol* **36**, 1788–1797 (2018).
3. Patnaik, M. M. The importance of FLT3 mutational analysis in acute myeloid leukemia. *Leuk Lymphoma* **59**, 2273–2286 (2018).
4. Morita, K. *et al.* Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat Commun* **11**, 5327 (2020).
5. Alberti, M. O. *et al.* Discriminating a common somatic ASXL1 mutation (c.1934dup; p.G646Wfs*12) from artifact in myeloid malignancies using NGS. *Leukemia* **32**, 1874–1878 (2018).
6. Jahn, K., Kuipers, J. & Beerenwinkel, N. Tree inference for single-cell data. *Genome Biology* **17**, 86 (2016).

7. Malikic, S., Jahn, K., Kuipers, J., Sahinalp, S. C. & Beerenwinkel, N. Integrative inference of subclonal tumour evolution from single-cell and bulk sequencing data. *Nature Communications* **10**, 2750 (2019).
8. Kottaridis, P. D. *et al.* Studies of FLT3 mutations in paired presentation and relapse samples from patients with acute myeloid leukemia: implications for the role of FLT3 mutations in leukemogenesis, minimal residual disease detection, and possible therapy with FLT3 inhibitors. *Blood* **100**, 2393–2398 (2002).
9. Kayser, S. *et al.* Insertion of FLT3 internal tandem duplication in the tyrosine kinase domain-1 is associated with resistance to chemotherapy and inferior outcome. *Blood* **114**, 2386–2392 (2009).
10. Corces-Zimmerman, M. R., Hong, W.-J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2548–2553 (2014).
11. Bottomly, D. *et al.* Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell* **40**, 850-864.e9 (2022).
12. The Cancer Genome Atlas Research Network. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *https://doi.org/10.1056/NEJMoa1301689* (2013) doi:10.1056/NEJMoa1301689.
13. Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* **374**, 2209–2221 (2016).
14. Metzeler, K. H. *et al.* Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. *Blood* **128**, 686–698 (2016).

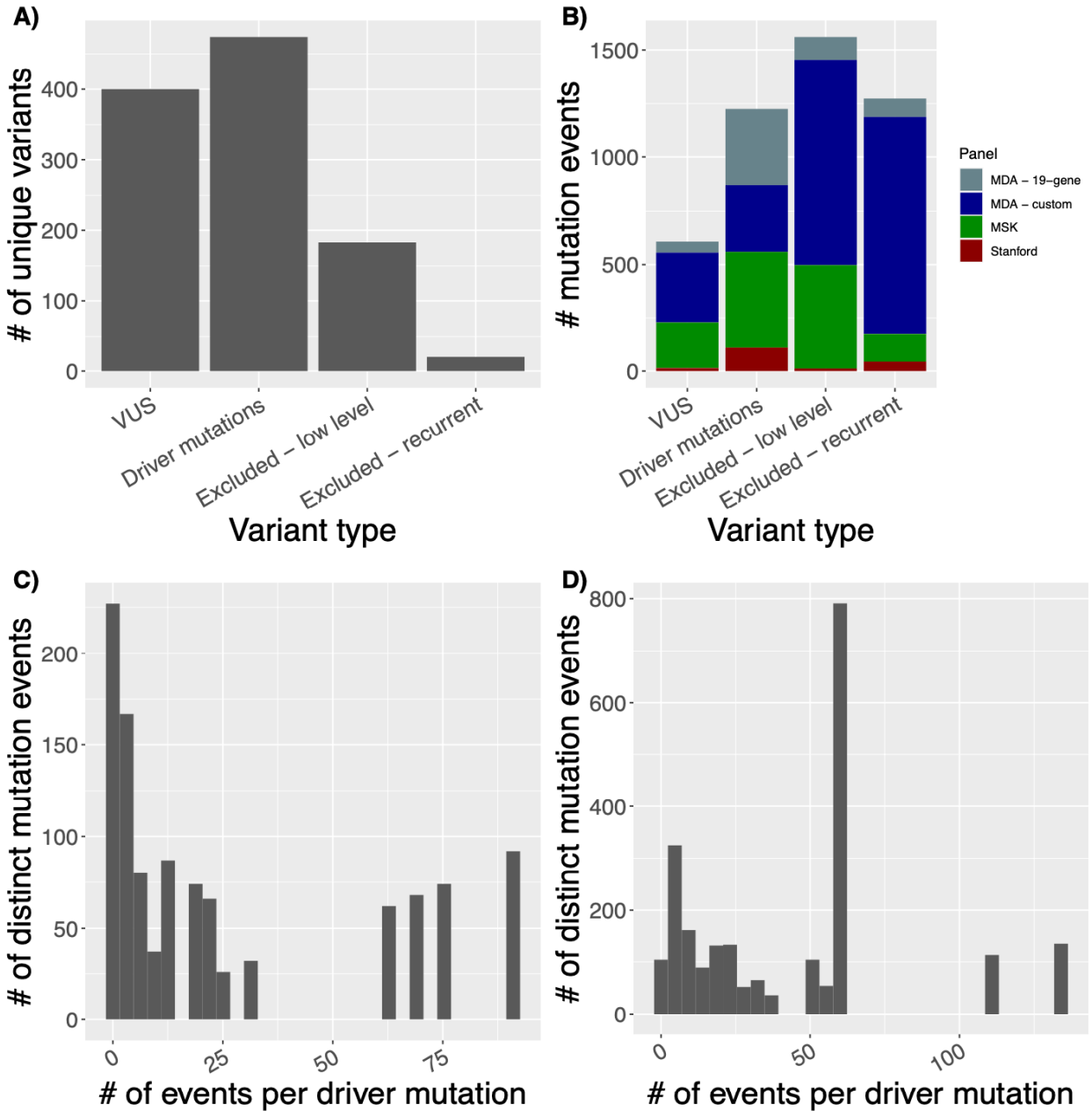
Supplementary Figures



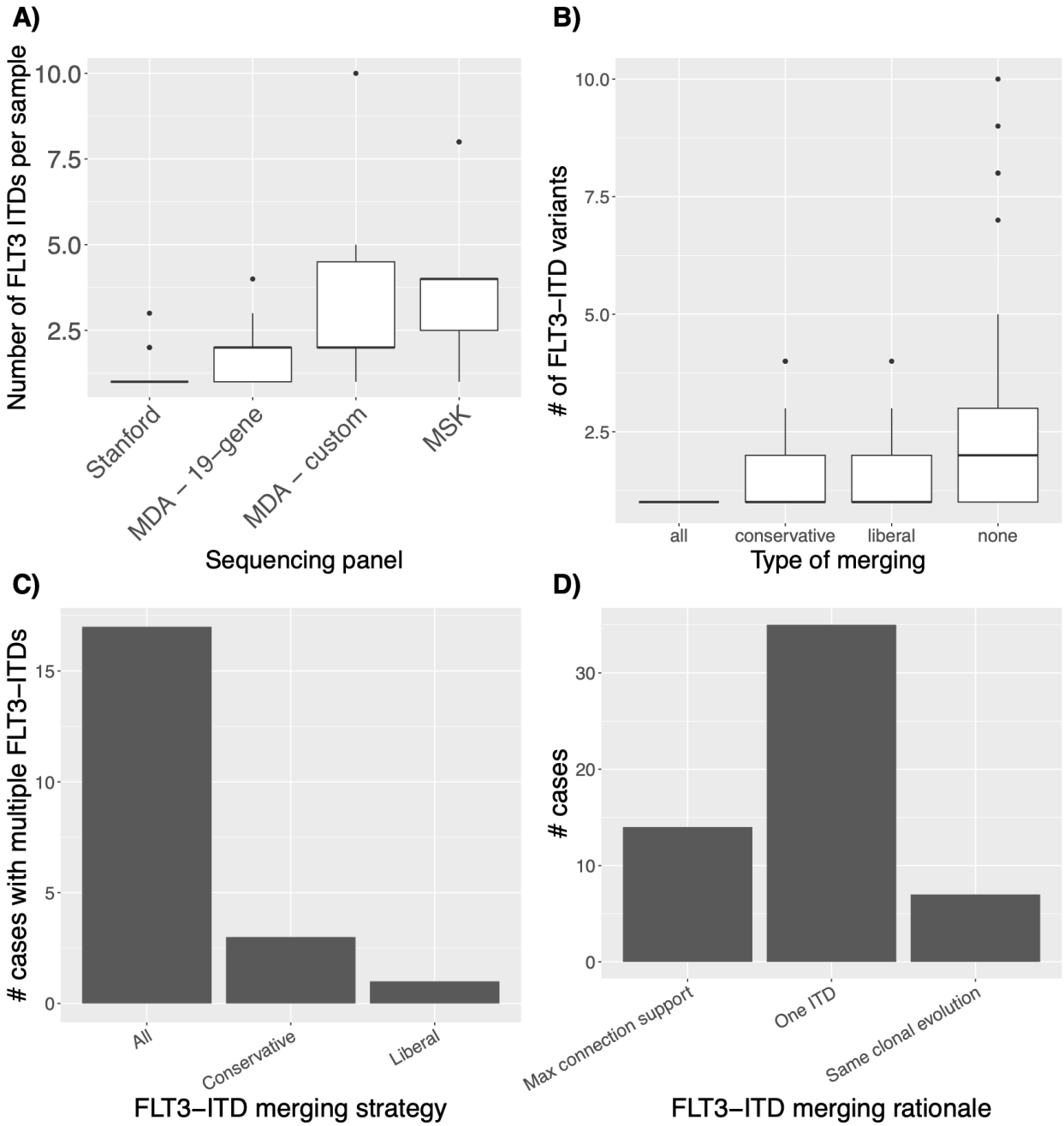
Supplementary Figure 1: Diagram showing studies included in the analysis, including number of patients and samples from Stanford, MD Anderson, and Memorial Sloan Kettering (MSK).



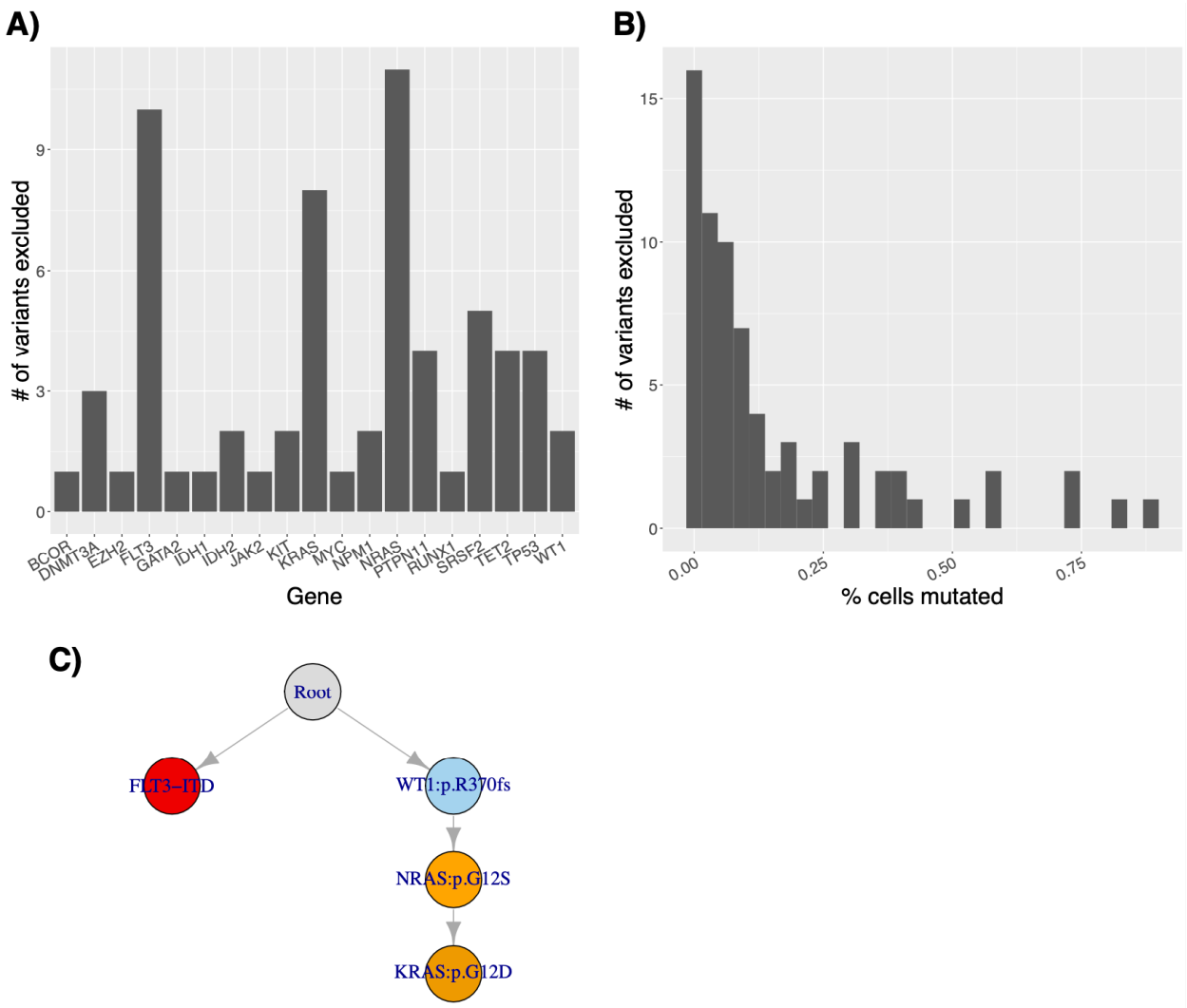
Supplementary Figure 2: Number of additional driver mutations discovered on manual review of variants that were initially of unknown significance, stratified by gene.



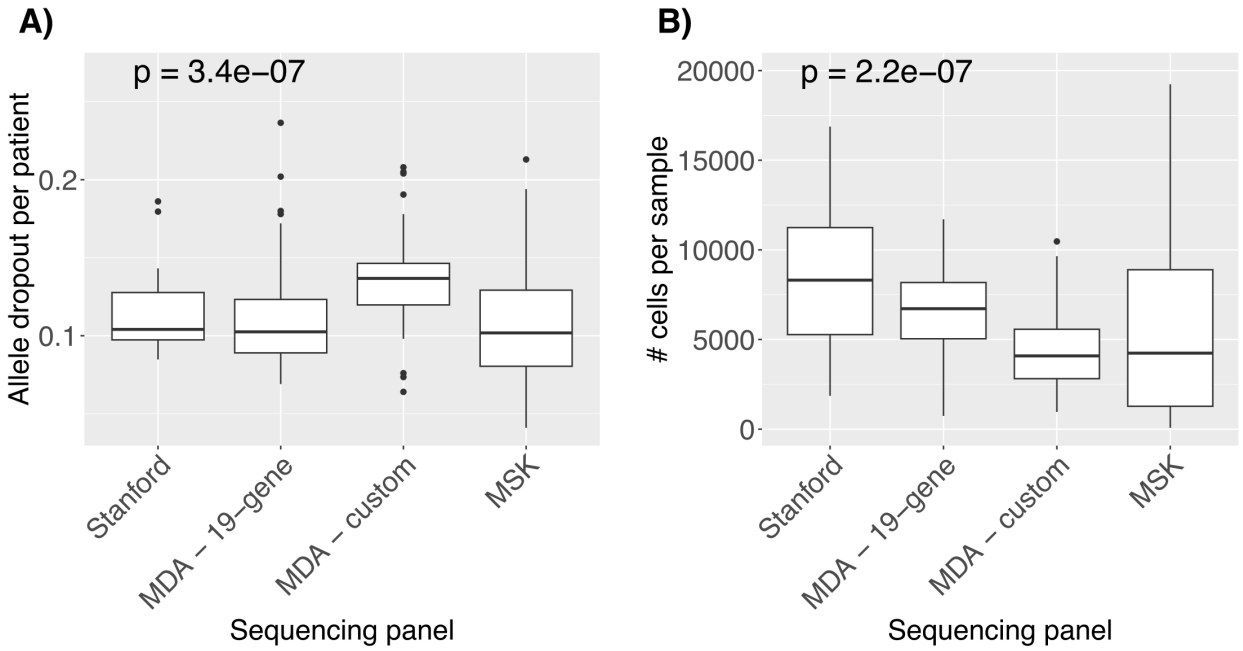
Supplementary Figure 3: Plots showing statistics about variants. A) Number of unique driver mutations, variants of unknown significance (VUS), and variants that were excluded (blacklisted) because they were not known to be associated with AML and either 1. occurred in most patients (Excluded – recurrent) or 2. occurred repeatedly in less than 5% of cells (Excluded – low level). B) Source of different types of variants broken down by dataset. C) Distribution of the number of events per driver mutation (where FLT3-ITD is considered a single type of driver mutation), or D) per blacklisted variant.



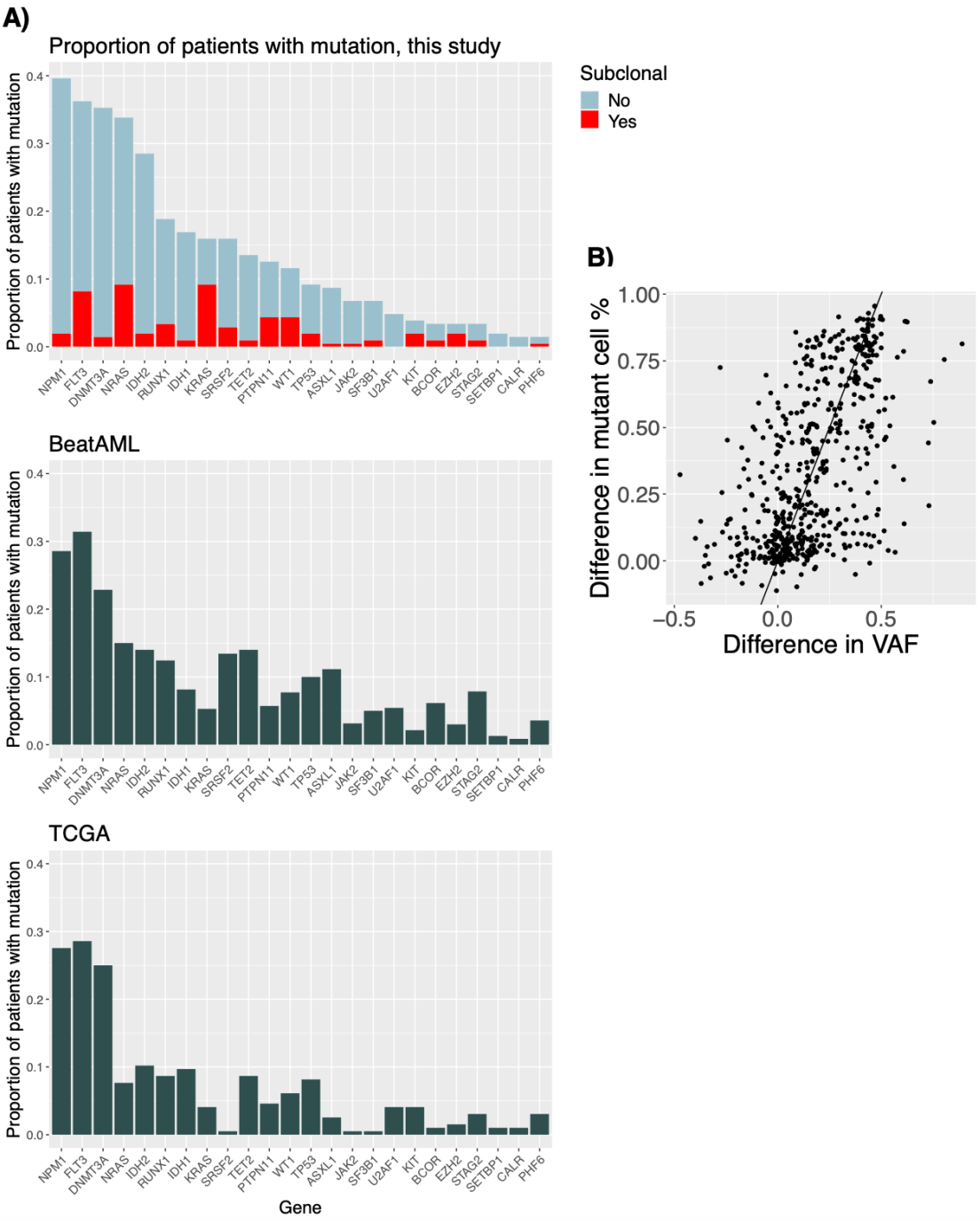
Supplementary Figure 4: A) Number of FLT3-ITDs per sample across each dataset. P-value was calculated with a Kruskal-Wallis test. Total ITDs = 151, and total patients = 58. B) Number of FLT3-ITD variants that result with different types of merging strategies (see Supplementary Methods). C) The number of cases that underwent different merging strategies based on our algorithm for choosing a merging strategy. D) The reasons for merging across all cases, where "Max connection support" means that the tree minimized low-support connections, "Same clonal evolution" means that all ITDs were terminal events in the tree and had the same parent event, and "One ITD" means either there was only one ITD or that the sequence of all ITDs were subsequences of another ITD.



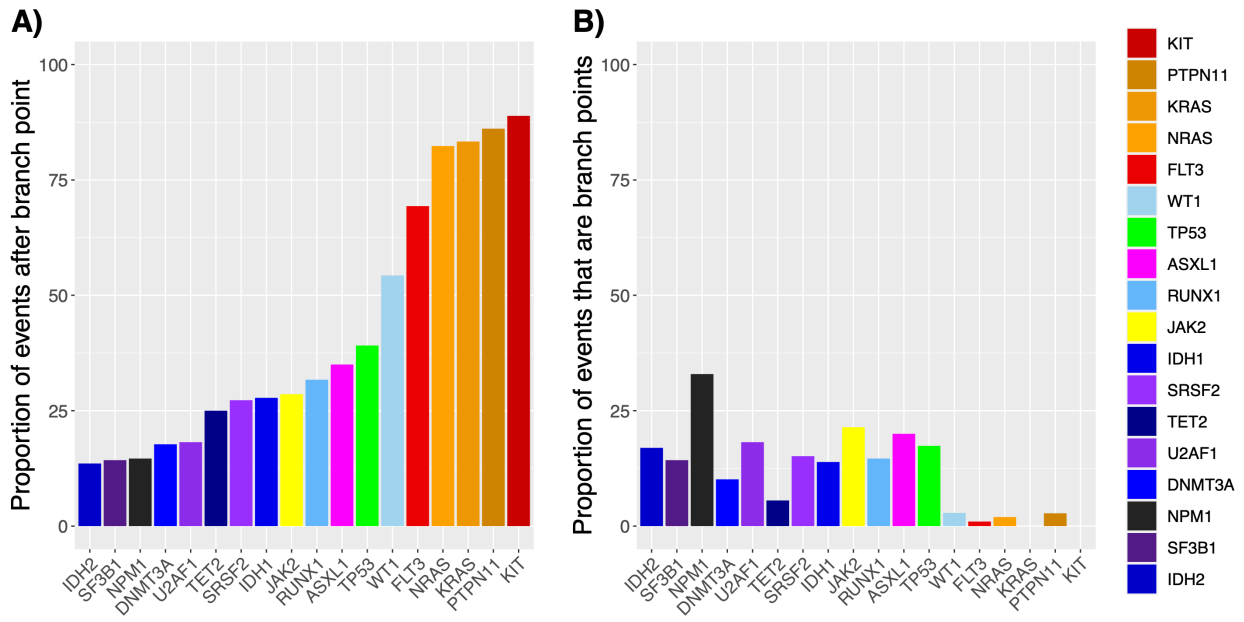
Supplementary Figure 5: After low-support connections were identified in a tree (<50% cells with the later mutation also contained the earlier mutation), mutations were excluded either because they contributed to the most low support connections or were more distal in the tree (Supplementary Methods). A) Bar plot of the number of variants excluded per gene because of low support, across the entire dataset. B) Distribution of proportion of cells mutated among those excluded variants. C) An example tree with a low-support connection (NRAS → KRAS).



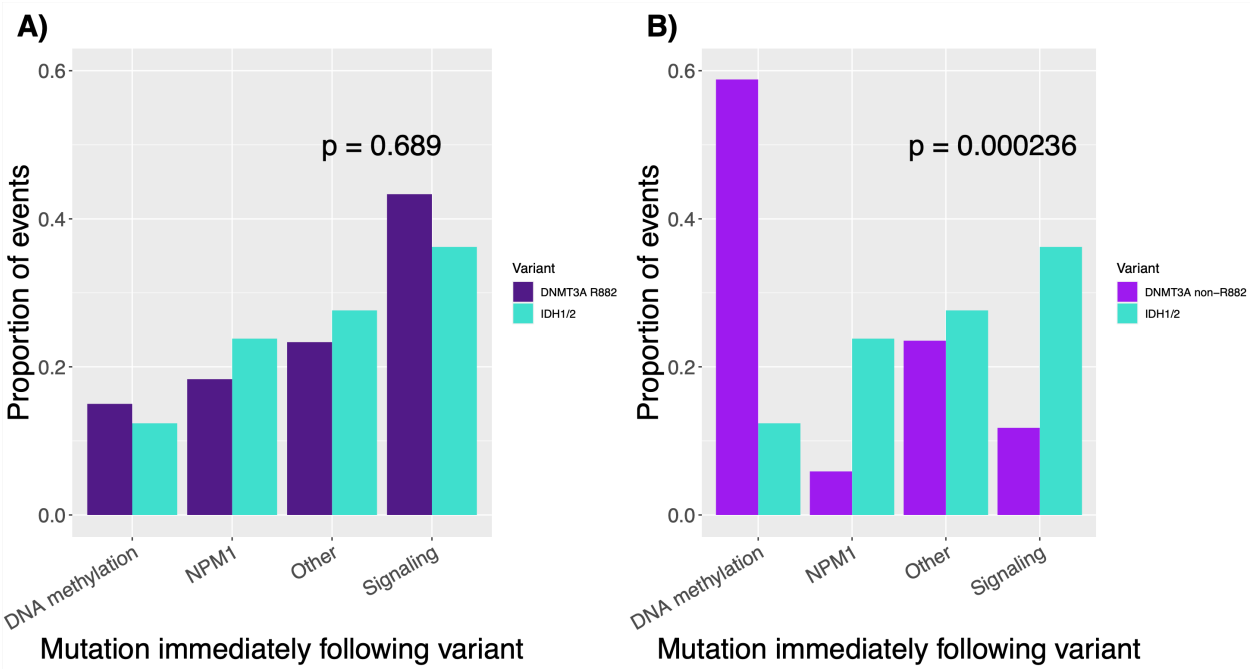
Supplementary Figure 6: A) Allele dropout estimate and B) number of cells per sample stratified by dataset and sequencing panel. Stanford and the "MDA 19-gene panel" are the same Mission Bio sequencing panels at different institutions. The "MDA custom panel" is a 37-gene panel created by collaborators at MD Anderson, and "MSK" refers to the 31-gene panel created by collaborators at Memorial Sloan Kettering. P-values were calculated with the Kruskal-Wallis test.



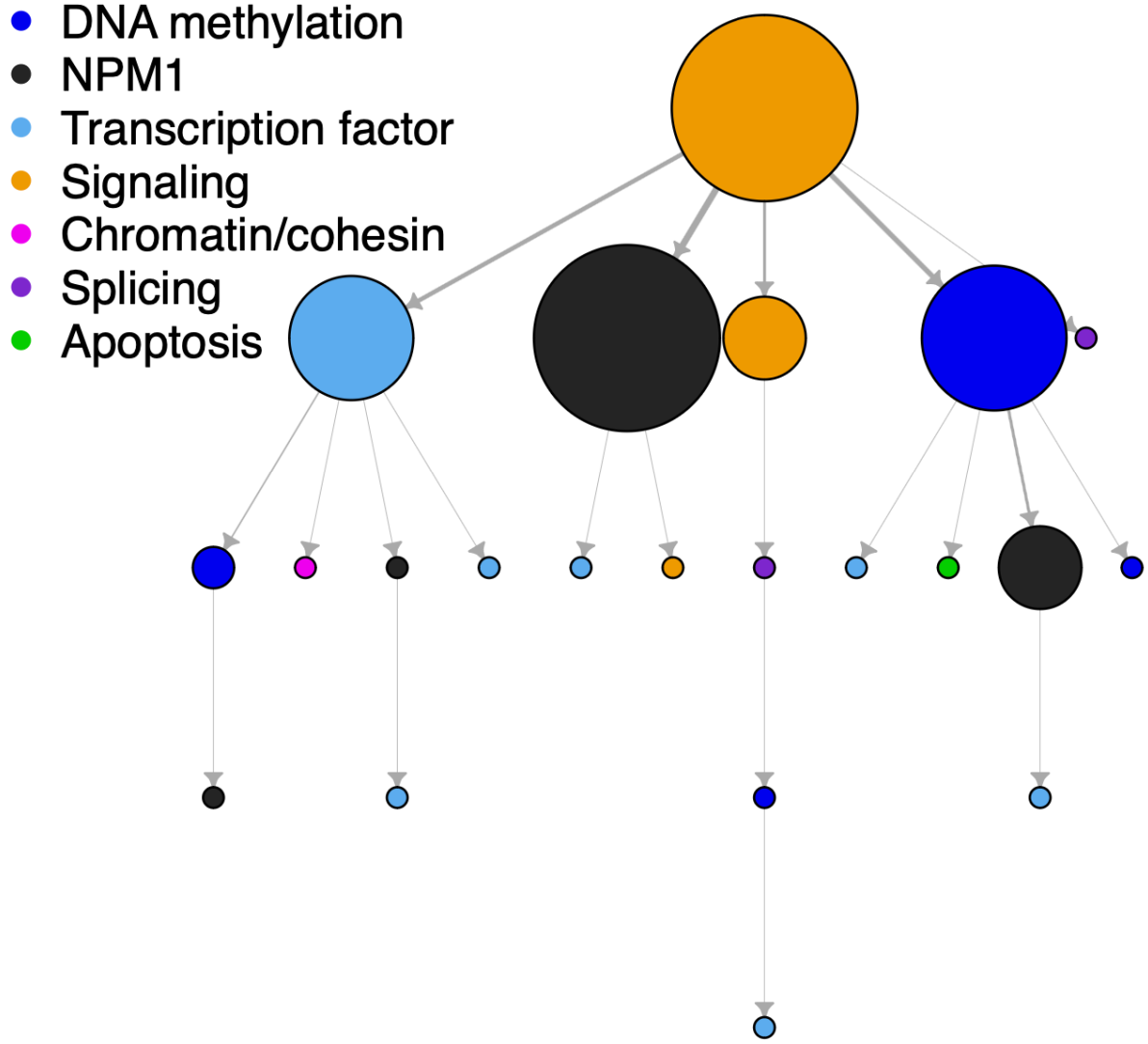
Supplementary Figure 7: A) Distribution of mutations across different datasets. The top plot is from the current study, second plot from the most recent BeatAML study [11], and third plot from The Cancer Genome Atlas [12] study. "Subclonal" means that the mutation was present in < 10% of cells. B) Comparison of the difference in percentage of cells mutated in single-cell data and the difference in variant allele frequency (VAF), which is a proxy for the number of cells mutated, in bulk sequencing data. The line represents the predicted association between these values if all variants were heterozygous. Plot B) was created using all available bulk sequencing data from the samples and variants in the single cell data, a total of 577 pairwise comparisons, 377 variants, and 139 patients.



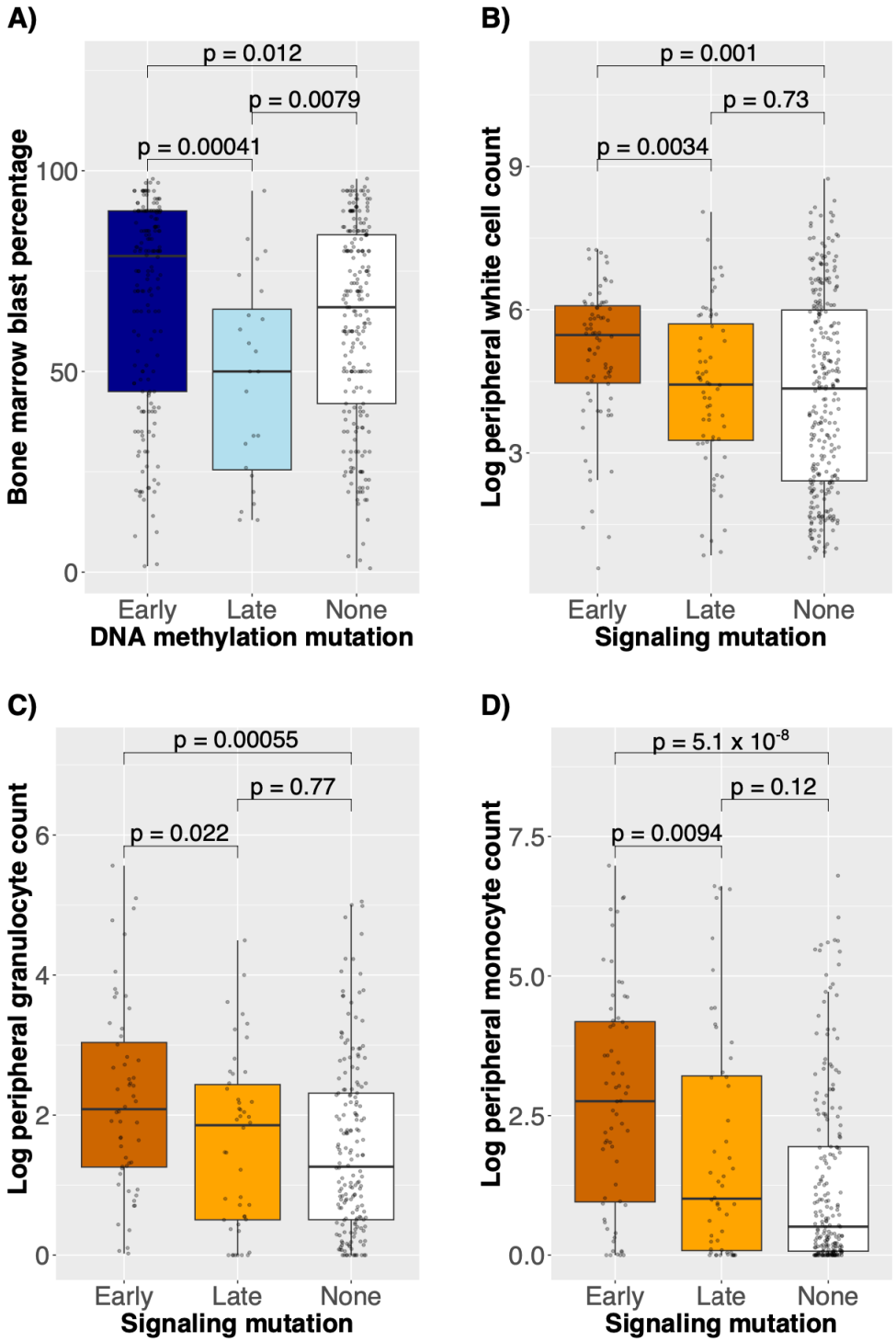
Supplementary Figure 8: A) Percent of mutation events for that gene that immediately follow a branch point, ordered by this percentage. Signaling mutations mostly follow branch points while others generally do not. B) Percentage of times a gene's mutations serve as a branching point. NPM1 mutations most commonly serve as branching points in evolution, largely because they often immediately precede signaling mutations.



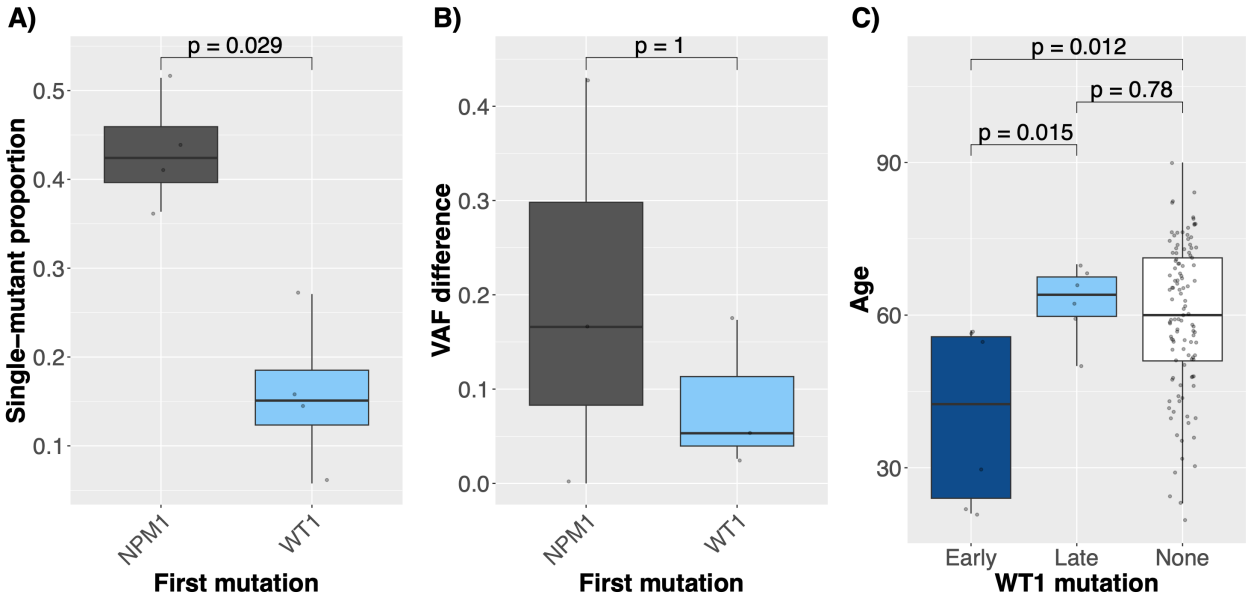
Supplementary Figure 9: A) Percentage of mutations that immediately followed either DNMT3A R882 or IDH1/2 mutations. B) Percentage of mutations that immediately followed non-R882 DNMT3A mutations vs. IDH1/2 mutations. P-values calculated with Fisher's exact test.



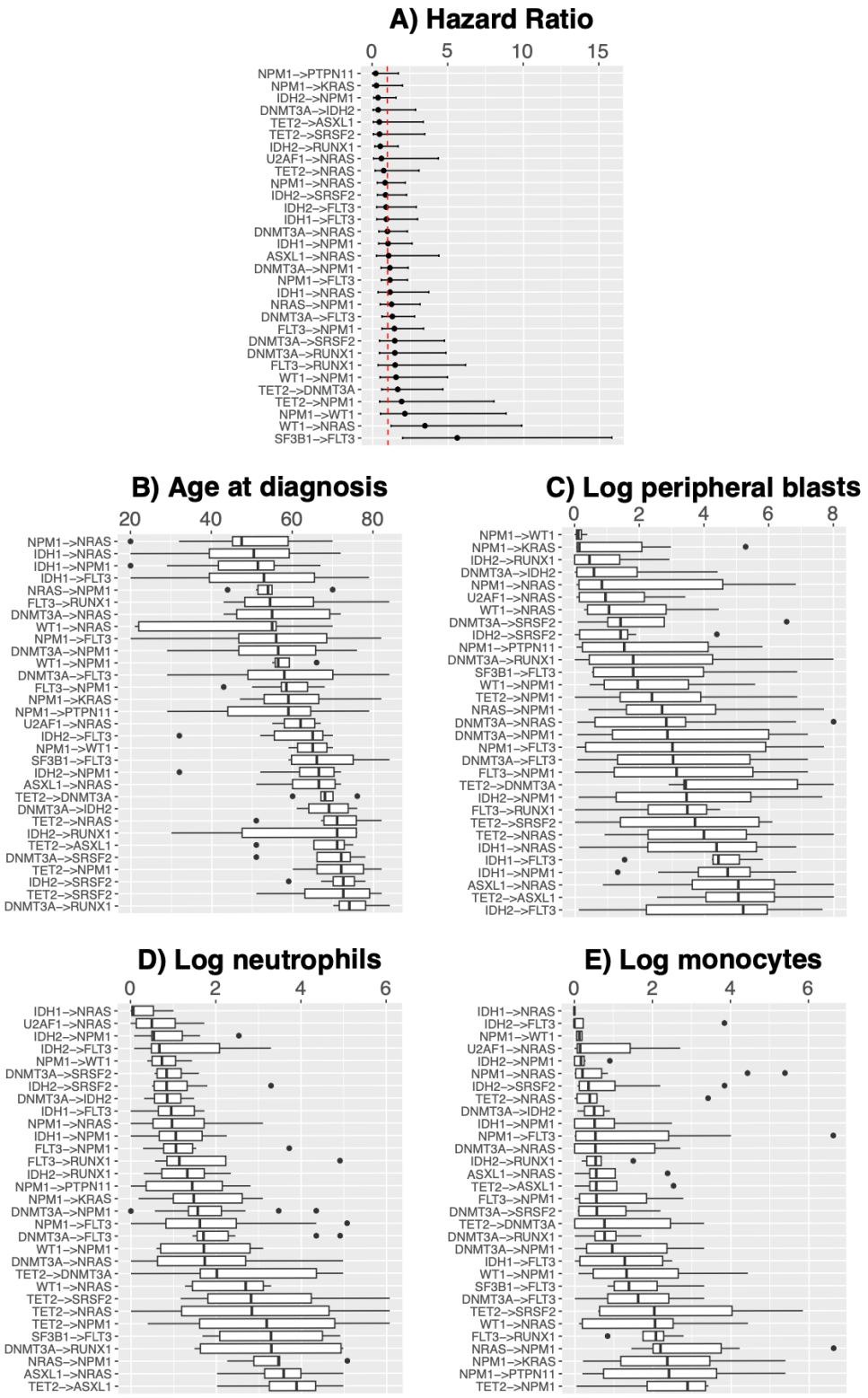
Supplementary Figure 10: Considering all cases where a signaling mutation preceded another mutation (n = 39), sub-trees were created using the signaling mutation as the starting node, and all such sub-trees were merged. This figure shows what mutations tend to follow signaling mutations, and they are predominantly NPM1 and DNA methylation mutations, although many transcription factor mutations (primarily in WT1) also commonly followed different signaling mutations.



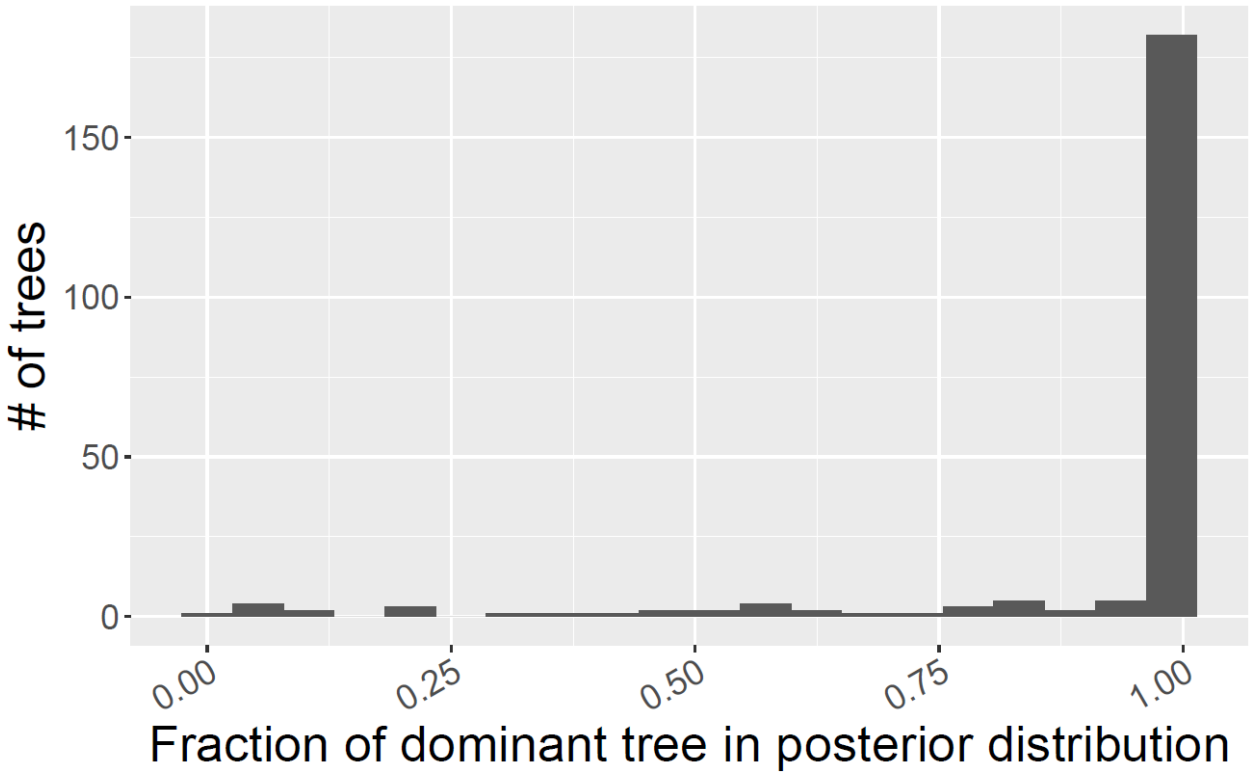
Supplementary Figure 11: Using the BeatAML data [11], A) distribution of bone marrow blast percentage compared to whether DNA methylation mutations were early, late, or absent. B-D) Similar plots comparing signaling mutations to B) log peripheral white blood cell count, C) log peripheral granulocyte counts, and D) log peripheral monocyte counts. Using these bulk sequencing, early and late were defined as VAF (variant allele frequency) ≥ 0.3 or < 0.3, respectively.



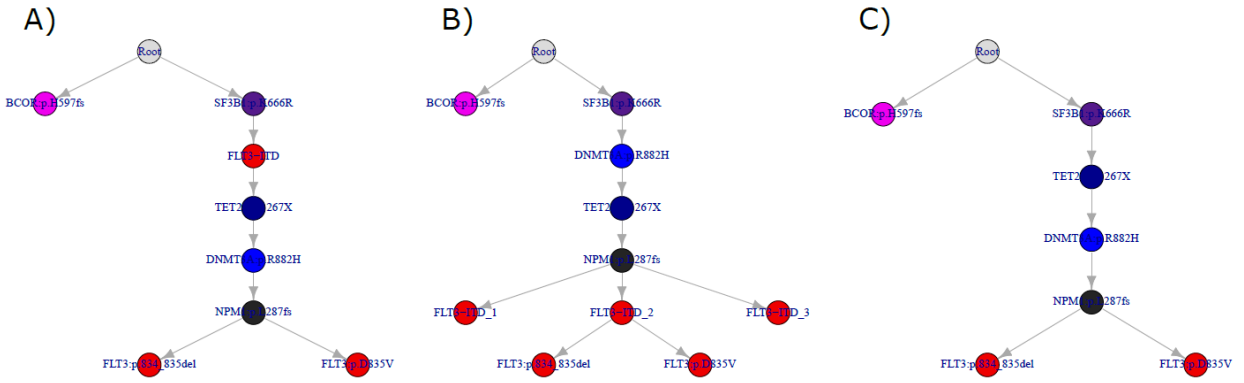
Supplementary Figure 12: A) Single-mutant proportions for WT1-first cases and NPM1-first cases. B) Similar comparison using variant allele frequency (VAF) differences between NPM1 and WT1 from bulk sequencing using the same variants and samples. A) Early, late, or no WT1 mutation at diagnosis compared to age.



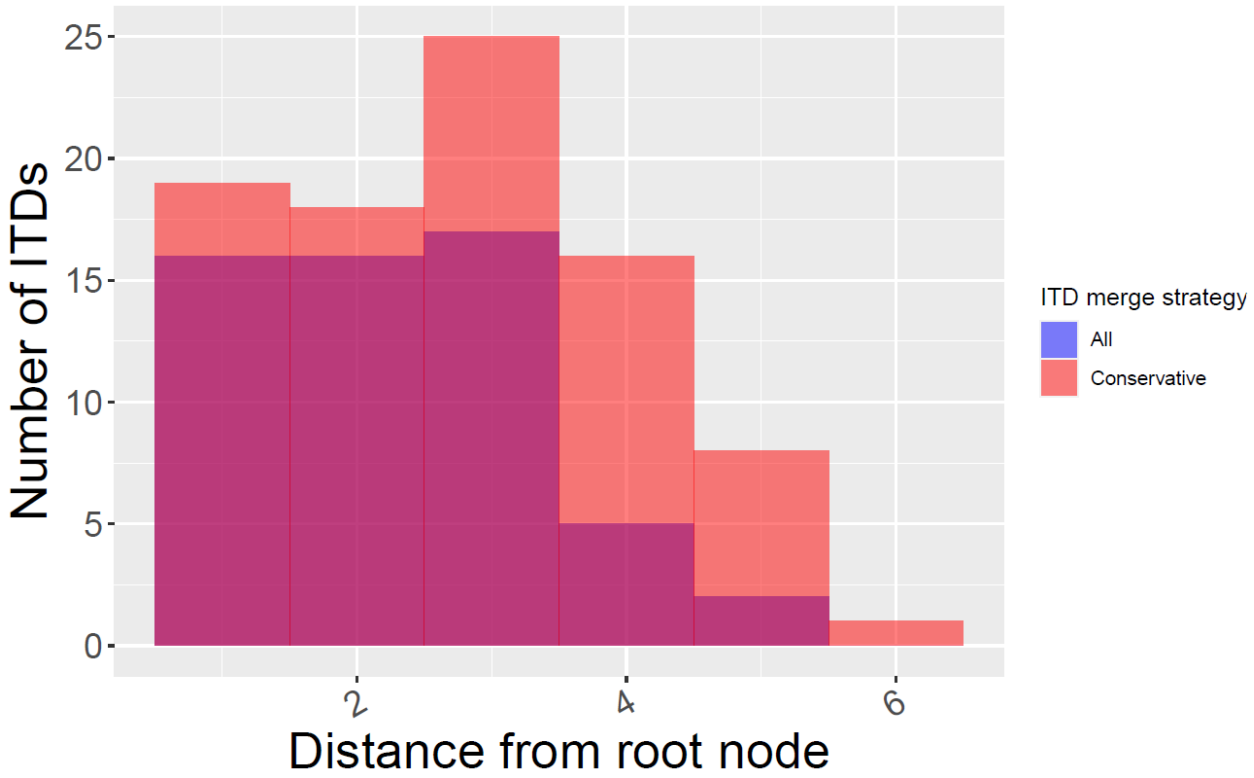
Supplementary Figure 13: Pairwise mutation orderings compared to different distributions of clinical variables, specifically A) hazard ratio of overall survival, B) age, C) peripheral blood log blasts), D) peripheral blood log neutrophils, and E) peripheral blood log monocytes compared to all patients without that pairwise path.



Supplementary Figure 14: Fraction of trees in the posterior distribution that are identical to the final tree used in the analysis. Generally, the posterior distribution was dominated by one tree.



Supplementary Figure 15: Extreme example of the consequences of merging FLT3-ITD variants using case AML-88 from the MD Anderson dataset. In this case, A) merging all variants resulted in the FLT-ITD variant to be higher in the tree than with B) conservative merging. However, the FLT3-ITD variant ultimately could not be used because it contributed to too many low support-connections, result in C) the final tree.



Supplementary Figure 16: Distance of FLT3-ITD variants from root node to the variant when a conservative ITD merging strategy is used (light red) or all ITD variants are merged (blue, becomes purple when mixed with light red in figure). This shows that when merging ITD variants, the more distal ITD variants in the tree are most affected.