nature genetics

Supplementary information

https://doi.org/10.1038/s41588-025-02196-0

Genomic landscape of multiple myeloma and its precursor conditions

In the format provided by the authors and unedited

Supplementary Information for "Genomic landscape of multiple myeloma and its precursor conditions"

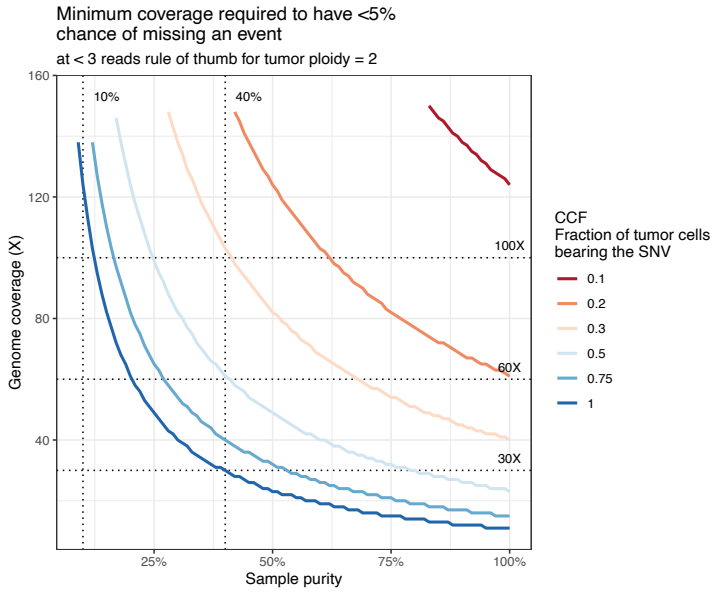
Table of Contents

Table of Supplementary Figures

Supplementary Notes

1 Theoretical sensitivity to detect SNVs from deep WGS

In this study, we decided to sequence samples to a depth of 60X (tumor) and 30X (matched normal). This sequencing depth was chosen based on a binomial model [1] reproduced in the figure below (***Supplementary Figure 1***). We assumed that tumors would be ≥40% pure after enrichment (median effective purity: 81%). In that case, we aimed to remain powered to detect mutations down to a cancer cell fraction (CCF) 50% (diploid tumor).

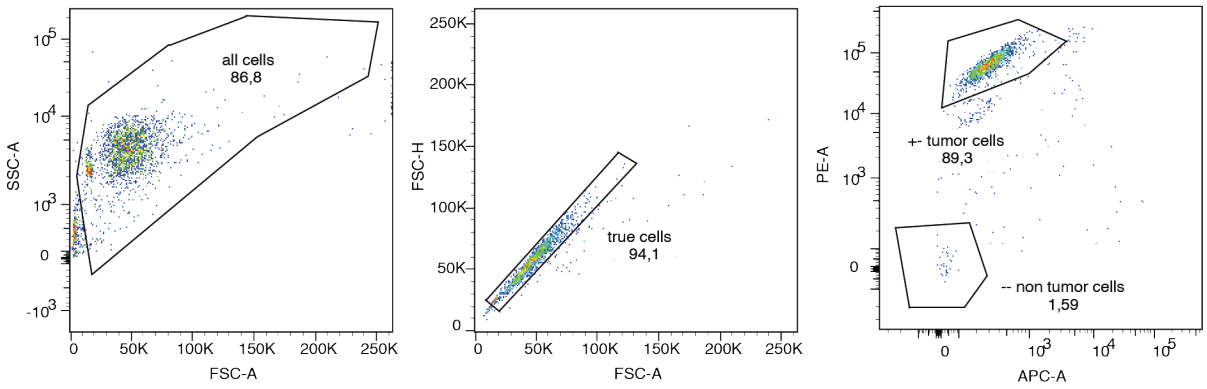


Supplementary Figure 1 *Binomial model (simulation) representing the minimum coverage required for a 95% chance of capturing at least three reads, based on tumor purity in the sample (x-axis), cancer cell fraction (CCF; colored lines), and genomic coverage (genome units; y-axis).*

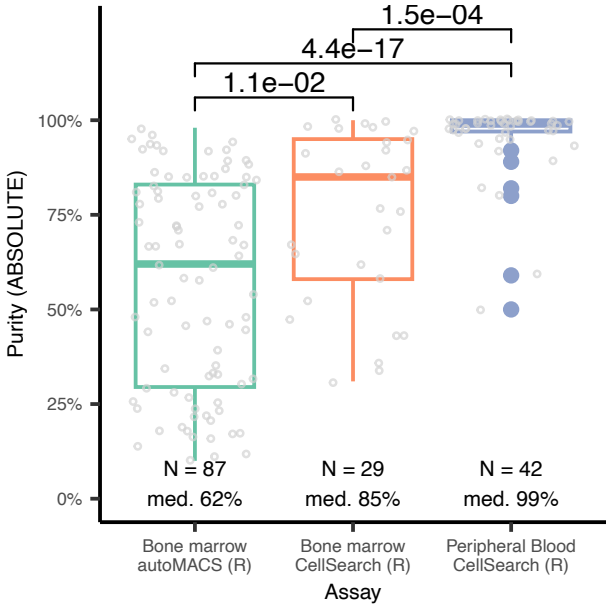
2 Sorting of tumor plasma cells with the Cellsearch platform

The CellSearch® platform can be used to enumerate circulating tumor cells (CTCs) from multiple myeloma (MM). We previously showed that this system, which isolates CTCs in a background of leukocytes and few normal plasma cells, can be coupled with sorting on a standard flow-activated cell sorter (FACS) to produce high purity mini-pools of CTCs. Then, these CTCs pools can be used to generate standard whole-genome sequencing (WGS) data. Here, we extended this approach to bone marrow (including marrows from clinical pathology laboratories). First, approximately 250 uL of bone marrow aspirate are taken into a CellRescue® tube with 3.75 mL of PBS (to reach a total of 4 mL). Then, 4mL of diluted bone marrow are split over 3 to 4 channels

of the CellSearch® system, which are then filled up with sample buffer (14mL). Following CTC enrichment, the same methods as previously published under the MinimuMM-seq protocol are applied [2] (***Supplementary Figure 2***). Of note, this method (CellSearch+FACS) yielded a significant increase in tumor purity compared to autoMACS CD138+ when assessed with the ABSOLUTE algorithm (***Supplementary Figure 3***), however, to a lower extent than with CTCs.



Supplementary Figure 2 *Example of fluorescence-activated cell sorting strategy for isolating tumor plasma cells with an abnormal phenotype following plasma cell enrichment by CellSearch ®. First, candidate events are gated based on Forward Scatter (FSC) and Side Scatter (SSC) (left), then FSC gating is refined based on area (FSC-A) over height (FSC-H) ratio (middle). Finally, tumor plasma cells from the bone marrow are isolated based on a CD19-CD45- (APC negative) and CD38+CD138+ (PE positive) phenotype.*



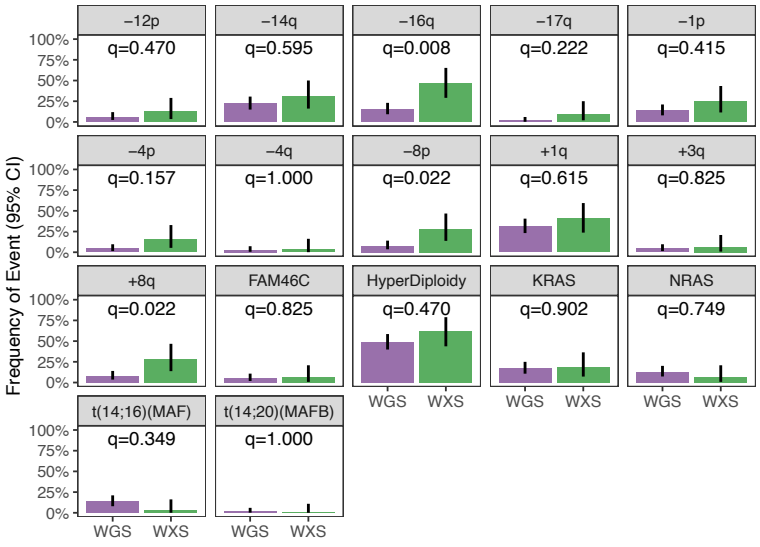
Supplementary Figure 3 *Comparison of tumor purity assessed by the ABSOLUTE algorithm between autoMACS CD138+ bone marrow processing and CellSearch®+FACS (bone marrow or circulating tumor cells from peripheral blood). Purity is given after manual review. Bonferroni-adjusted P values from the Dunn post hoc tests after Kruskal-Wallis (P=1.4E-16). From left to right: bone marrow autoMACS® N=87, median 62%, median absolute deviation (mad) 36%; bone marrow CellSearch® N=29, median 85% mad 21%; peripheral blood CellSearch® N=42, median 99%, mad 1%. In boxplots, the center line represents the median, the box limits represent the upper and lower quartiles, and the whiskers indicate 1.5 times the interquartile range. Values beyond whiskers appear individually as plain circles.*

3 Hazard Ratios of Features from the MM-like score taken individually

We calculated the hazard ratios for each component of the MM-like score and adjusted their Wald p-value for the false discovery rate (q-value, **Supplementary Table 13**). At a q<0.1, 7 variables of the score are significantly associated with a higher risk of progression from smoldering multiple myeloma (SMM) to multiple myeloma (MM) (*KRAS*, *NRAS*, -14q, -16q, -1p, -4p, and -8p). All variables with a sign value "+1" happen more frequently in MM than in monoclonal gammopathy of undetermined significance (MGUS)/SMM. Variables with a "-1" (*MAF* and *MAFB* translocations) are more common in MGUS/SMM than in MM.

4 Frequency of mutations in the MM-like score between whole genomes and exomes

In this study, we combined the mutations from whole genomes and exomes [3]. In the exomes, a sequencing depth of ~109X was reported, which was higher than the genomes (~50X). When we looked specifically at mutations from the MM-like score, we found most of them at comparable frequencies (q>0.05, n=14/17; ***Supplementary Figure 4***), even though the studies were run >4 years apart. In particular, *KRAS* and *NRAS* were found at comparable frequencies despite the difference in coverage. In addition, in the MM-like score, we combined mutations that are redundant to avoid double counting of events and to allow detectability by both genome and exome sequencing (e.g., detection of a focal +8q24 at *MYC* and of a t(*MYC*) translocation).



Supplementary Figure 4 Comparison of the frequency of events from the MM-like score between whole genome sequencing (WGS, N=118) and whole exome sequencing (WXS, N=132) of unmatched patients. $CI_{95\%}$ are binomial confidence intervals (error bars), and q values are false-discovery rates adjusted from Fisher's exact P values.

5 Analysis of phylogenetic trees from two patients with non-progressive disease

*This paragraph mentions **Main Figure 2** found in the main manuscript.*

To better understand the evolution of the tumor in the two patients (Pts. SCI054 and SCI116) that increased their score without progression, we analyzed the subclonal structure of the tumors and their phylogenetic trees. Patient SCI054 (**Fig. 2B**) represented an interesting case of light chain MGUS / low-risk SMM with a follow-up sample collected after three years. At baseline, we detected a canonical t(11;14) translocation, a clonal del(17p)(*TP53*) and two subclonal missense mutations in *TP53* (p.S241F, CCF: 24%; p.Y205D, CCF: 28%). After three years, we observed an increase in both tumor burden (free light chain ratio: 4.6 to 9.4) and CCF of the *TP53* mutations (CCF: 24 to 50% and 28 to 36%, respectively), as well as a newly detected subclonal hyperdiploidy (gains of chr3, 9, 11, and 15; **Fig. 2B**). We analyzed the phylogenetic structure of these samples and identified hyperdiploidy in the follow-up sample in cluster 5 (**Fig. 2B**, orange), yielding an increasing MM-like score from 1 to 2. In contrast, patient SCI116, was initially characterized by hyperdiploidy with no other MM driver event (MM-like score of 1; **Fig. 2C**), but in the follow-up sample, we identified del(1p), increasing the MM-like score to 2. Phylogenetic analysis (using PhylogicNDT), identified an increasing subclone with the del(1p) (C5 in **Fig. 2C**) and other subclones in a separate part of the tree that either shrank or were stable over time (C2, C3, and C4 in **Fig. 2C**).

6 Bradley-Terry Models for clonality-based timing competition of subclones

6.1 Significant hotspots of Copy-number alterations (GISTIC2) for the Bradley-Terry models

In order to identify recurrent focal regions of CNAs in the tumor's genomic landscape we first applied GISTIC2.0 [4] on the whole cohort of Copy Number profiles, with the following parameters: Join_segments_size = 50, Focal_length_cutoff = 0.25, Q-value = 0.01, arm_peel = "YES", sample_centering = "NO". All the other input parameters were selected as default. In order to further correct the false positives generated by the germline CNVs, a list of regions to filter out in the analysis was created by using the DGV (Database of Genomic Variants) version 107 database (available at http://dgv.tcag.ca/dgv/app/home) and selecting for all the region reported in the database in at least 100 samples, across at least 2 different studies which show a reported MAF > 5% in the human population. This input will be used for the Clonality League Comparison (CLC, see below).

6.2 The relative timing of copy-number alterations based on their clonality

The Bradley-Terry model (BTM) is a statistical model used to analyze paired comparison data [5]. The model is implemented in the *BradleyTerry2* R package [6]. This model assumes that in a "match" between any two "players", say player *i* and player *j* (*i*, *j* ∈ (1, . . . , *K*)), the odds that *i* beats *j* are α_i/α_j , where α_i and α_j are positive-valued parameters which represents the "ability" to win. Applications are many, ranging from the analysis of sports tournaments to the field of genomic alterations timing in cancers [7]. In this context, "players" are the different genomic alterations, which can be compared in "matches". In our model, a genomic alteration "wins" if it is found to be less clonal than another one. Therefore, the Bradley-Terry model (BTM) output "ability" score can be interpreted as a "subclonality score" (BT score). The BT score in our model reflects how much an alteration is likely to be observed at a subclonal level in the analyzed tumor type compared to all the other alterations included in the BTM. The BTM runs in two main steps:

- **Input matrix generation**: The number of "win" points for each alteration is computed by performing pairwise clonality "matches" and summed cumulatively across a "league" containing all the aggregated tumor samples. This procedure generates a matrix of wins and losses per alteration.
- **BT score computation**: This matrix is used as input to a function that fits a model by employing maximum likelihood estimation. This step generates relative "abilities" (BT scores) for each player/alteration. The "BTm" function from the "BradleyTerry2" R package, version 1.1-2, performs this step.

BT score interpretation: Generally, by following the clonal evolution Darwinian model, it's possible to interpret that clonal alterations happen earlier in time than subclonal alterations. Consequently, a high BT score implies that the alteration is subclonal and may happen late in time during oncogenesis. On the contrary, a low BT score implies that the alteration is frequently observed as clonal and may happen early during oncogenesis. A second interpretation is also possible: a high BT score could reflect that an event remains subclonal due to negative selection, while a low BT score can reflect a rapid positive selection (fitness advantage).

Modifications to the BTM input matrix

- **Statistical assessment of matches**

To account for the noise and variation in CN data, we implemented a statistical test into the match procedure to check if the difference in clonality between two events is significant before we decide the result of a match (win or loss). This test is a summarized t-test derived from the means and standard deviations of the copy number (CN) values of the two alterations. Only when the

difference is statistically significant ($p<0.05$) a win is awarded. Otherwise, the match is considered a draw, with both alterations receiving zero points.

- **Quantitative matches outcomes**

In our BT models, the "win" points assigned to events during pairwise clonality matches are calculated to enhance the model's resolution by maximizing the available clonality information. Specifically, the number of points gained in a win is made proportional to the clonality difference between the two competing alterations, ranging from 1 (small difference <20% clonality) to 4 (large difference >80% clonality). This differs from previous methods that used a constant number of "win" points (usually 1), regardless of the actual difference in clonality levels in matches. This adjustment leverages the quantitative magnitude of the difference in clonality, with the goal of providing a more detailed view of the alteration's clonality relationships.

- **Threshold to call alterations based on sample purity.**

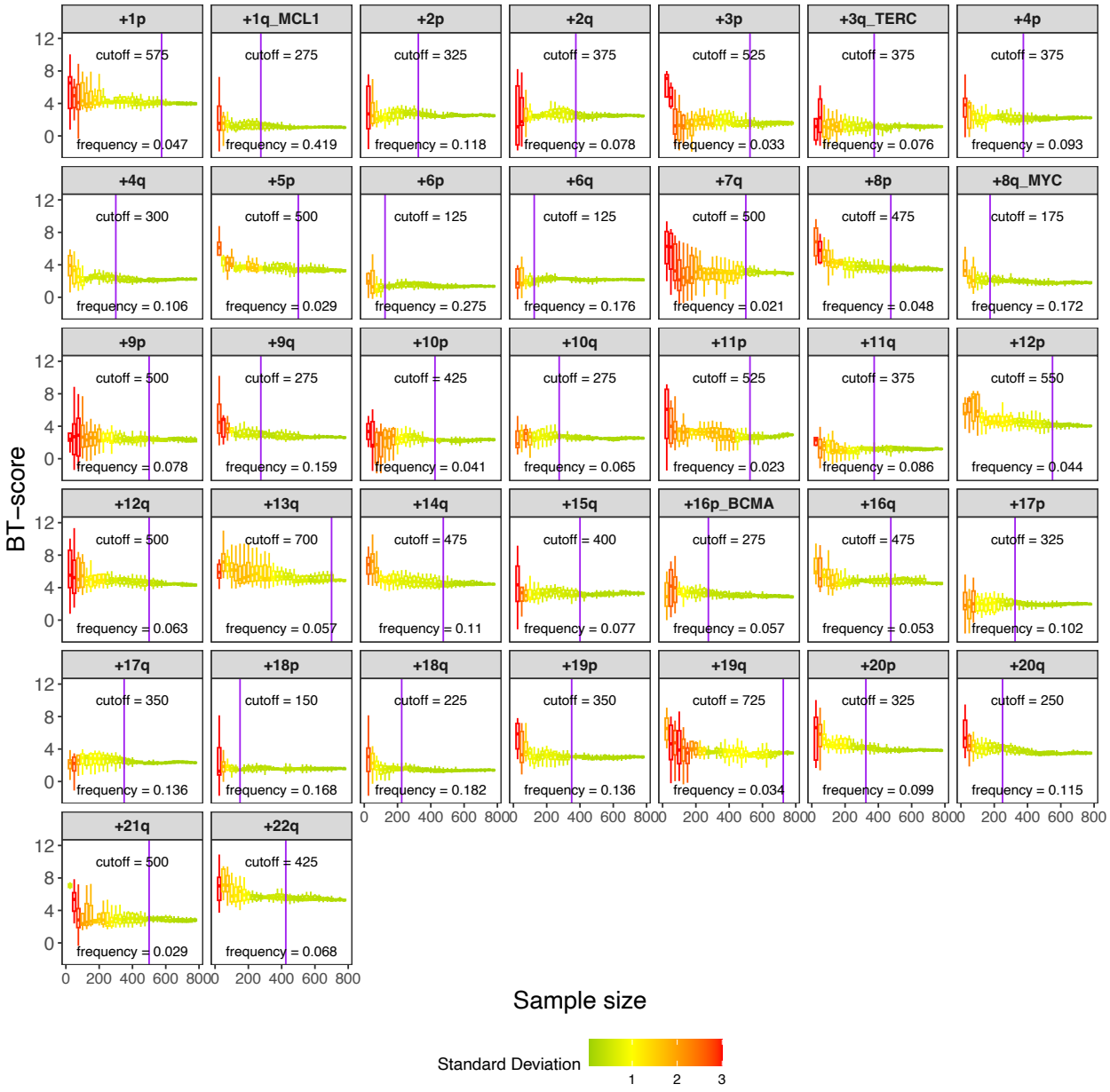
We recognized that 1) high precision in clonality estimation is crucial for obtaining optimal BTM results, and 2) low sample purity is detrimental to the detection of subclonal alterations. Consequently, in performing alteration calls to use in the BTM, we chose to scale the minimum clonality threshold for considering the presence of an event based on tumor purity. For example, while the default clonality threshold to detect an event in this analysis is set at 5%, this threshold is adjusted to 10% in samples with 50% tumor purity, and to 20% in samples with 25% tumor purity. This scaling ensures that our analysis accounts for the dilution effect of non-tumor cells in the sample, improving the accuracy of alteration detection.

6.3 Stability of BT scores and relationship with the frequency of alterations

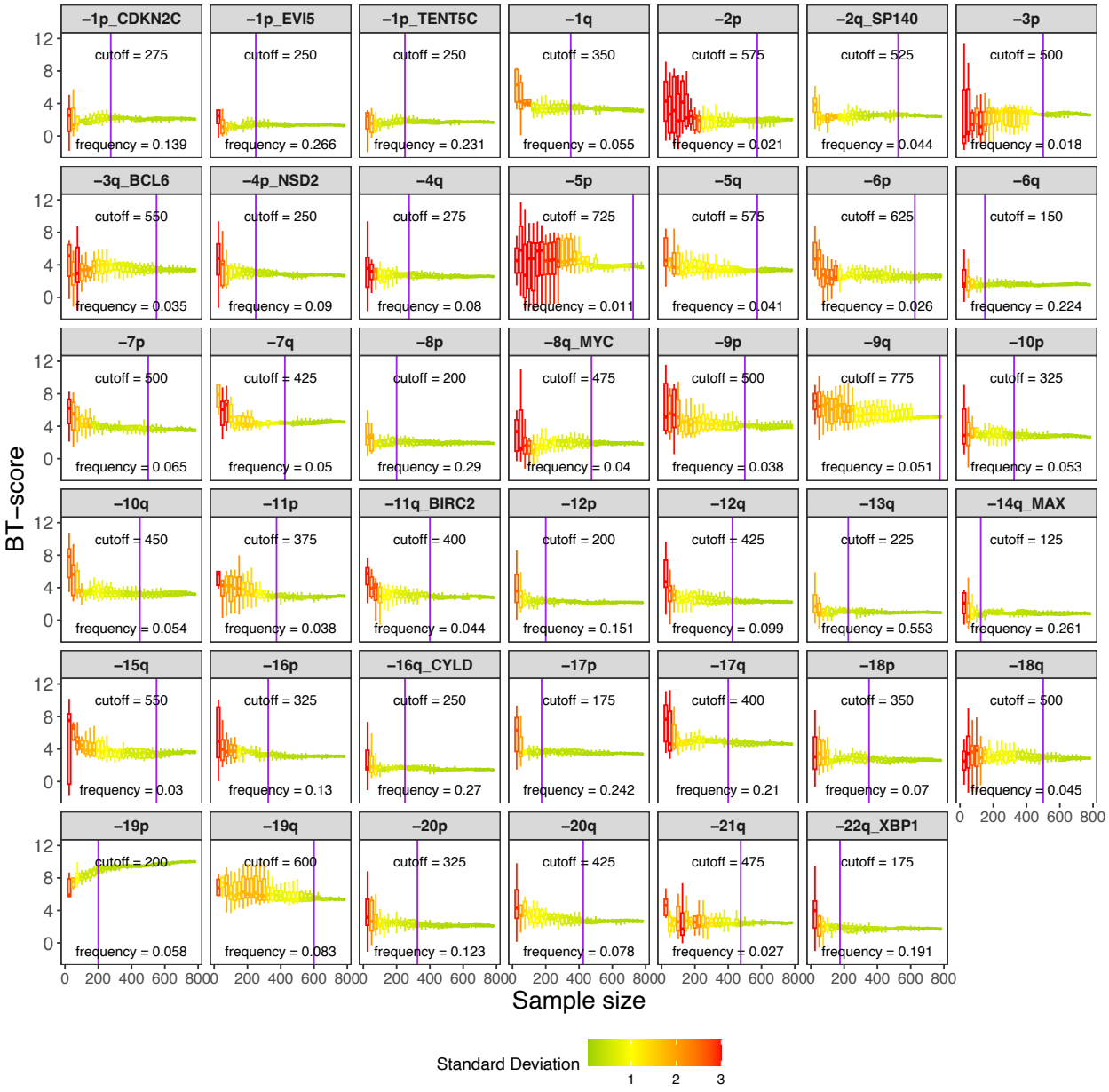
Our study also aimed to identify the minimum number of samples required to achieve stable timing estimates for each genomic alteration by utilizing a bootstrap approach. In detail, by leveraging the CoMMpass cohort (n=847 samples for this analysis), we generated a series of sub-cohorts of increasing size. Next, within each sub-cohort, we performed 20 independent BTMs to generate timing estimates, thereby assessing the stability of BT scores of all alterations across varying sample sizes. We quantitatively defined stability as the cutoff point where the standard deviation of timing estimates from the bootstrapped analyses consistently fell below 0.5 across subsequent sub-cohorts, indicating that additional samples did not significantly enhance the stability of the timing scores. This stability cutoff was crucial to determine the optimal sample size

required to obtain reliable and consistent timing estimates, enhancing the accuracy and reliability of timing analysis interpretation in our study (***Supplementary Figure 5***).

Starting with samples from the CoMMpass cohort, each genomic alteration included in the CLM model was bootstrapped 20 times in sub-cohorts with increasingly more samples to find the minimum number of samples required to obtain a stable BT score for each alteration. Stability here is quantitatively defined as achieving a SD of less than 0.5 among the 20 BT scores from the bootstrapped analyses within a specific sub-cohort size (top panel: gains; bottom panel: losses).



Supplementary Figure 5 *Stability analysis of timing estimates from a Bradley-Terry (BT) model. The cohort was bootstrapped 20 times at various sample sizes (downsampled from the MMRF CoMMpass study), and variation of bootstraps was measured for each arm-level copy-number gain. The x-axis represents the number of samples in bootstraps. The y-axis represents the BT score estimate distribution. We considered a standard deviation of 0.5 to be stable and calculated the number of samples required to achieve such stability (which depends on the frequency of those events). In boxplots, the center line represents the median, the box limits represent the upper and lower quartiles, and the whiskers indicate 1.5 times the interquartile range. Values beyond whiskers appear individually as plain circles.*

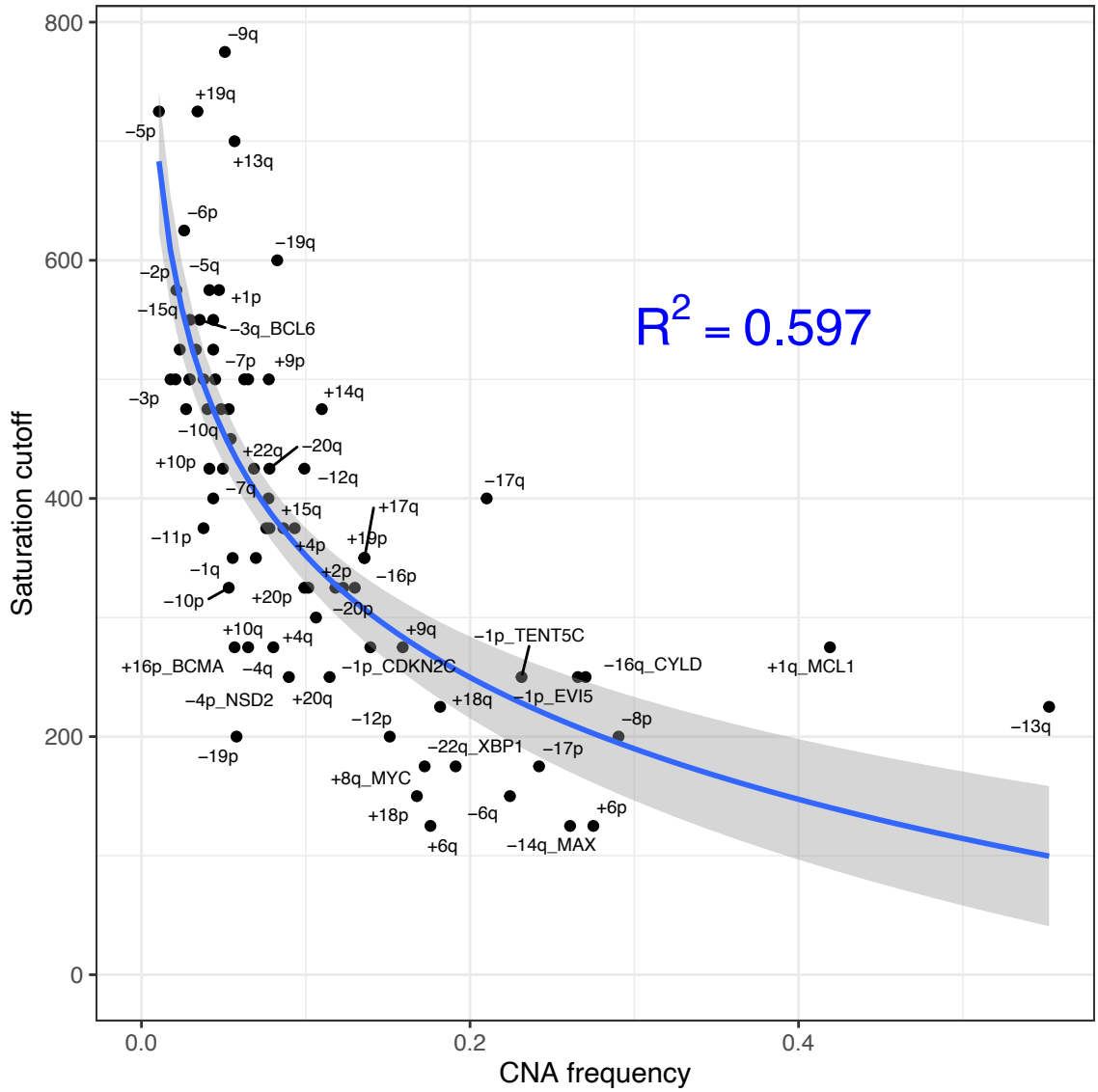


Supplementary Figure 5 (bis) *Stability analysis of timing estimates from a Bradley-Terry (BT) model. The cohort was bootstrapped 20 times at various sample sizes (downsampled from the MMRF CoMMpass study), and variation of bootstraps was measured for each arm-level copy-number loss. The x-axis represents the number of samples in bootstraps. The y-axis represents the BT score estimate distribution. We considered a standard deviation of 0.5 to be stable and calculated the number of samples required to achieve such stability (which depends on the frequency of those events). In boxplots, the center line represents the median, the box limits represent the upper and lower quartiles, and the whiskers indicate 1.5 times the interquartile range. Values beyond whiskers appear individually as plain circles*

These critical cutoff points, where further sample inclusion does not substantially improve the timing score's stability, are marked in the plot facets by purple vertical bars for each individual genomic alteration (and annotated with "sat" labels in the plot facets). Each alteration's frequency,

as computed in the total cohort, is annotated in the bottom part of the corresponding facet ("freq" labels).

Then, we wondered how the number of samples required to stabilize the Bradley-Terry model evolved with the frequency of copy-number events (***Supplementary Figure 6***).



Supplementary Figure 6 *Relationship between stability cutoffs and the alteration's frequencies. Scatter plot showing a negative logarithmic regression model (blue line, formula: y= - log(x)), fitted to study the relationship between stability cutoffs (y-axis) and frequencies of copy number alterations (CNA; x-axis). The model correlation (R^2 = 0.567) indicates that a significant portion of the variation in the stability cutoff variable is explained by the alteration frequency variable. The grey band around the regression line represents the 95% confidence interval of the model.*

When we explored the relationship between the stability cutoffs and the frequencies of each alteration. Our analysis, employing a negative logarithmic regression model, revealed a significant relationship ($R^2 = 0.567$), indicating that the frequencies of the alterations could explain a notable portion of the variability in stability cutoffs.

6.4 Comparison of Bradley-Terry models (Clonality Leagues Comparison)

The Clonality League Comparison (CLC) differentiates between two independent BTMs generated with samples from different individuals diagnosed with diverse stages of the disease. In this study, we developed the CLC method to investigate the temporal dynamics of copy-number alterations (CNAs) in MM progression by comparing the BT score of genomic alterations between MGUS/SMM and newly diagnosed MM (NDMM) stages of the disease. To achieve this, we constructed two separate BTMs, one for each disease stage, using the aggregated dataset of genomic profiles from patients diagnosed with SMM and NDMM, respectively. The primary objective is to elucidate the progression dynamics of genomic alterations by analyzing and comparing the alterations in BT scores derived from these two models.

Each BT model was constructed using pairwise matches of CNAs computed at each disease stage. The models assign a BT score to each CNA, indicating its relative clonality level within the disease stage. To facilitate a direct comparison between the MGUS/SMM and the NDMM models, we centered the BT scores on “Hyperdiploidy,” which was found to be the most clonal alteration in both models. This alteration was assigned a BT score of zero in both models, serving as a reference point for comparing other alterations.

We compared the BT scores between the two models without a specific adjustment because a Blend-Altman analysis showed only a small mean difference (0.21 units) between the MGUS/SMM and NDMM BT scores and because most alterations were detected within the agreement interval (±1.96 SD) (***Supplementary Figure 7***).

To compare the BT scores, we computed the difference in scores for each CNA between the NDMM and MGUS/SMM models. A positive difference indicates that the alteration is observed at a lower level of clonality in NDMM when compared to MGUS/SMM, while a negative difference indicates a greater level of clonality in NDMM when compared to MGUS/SMM (***Supplementary Figure 8***). Importantly, this analysis can provide key insights into the evolutive dynamics of alterations as MM progresses from MGUS/SMM to NDMM disease phase, as described below:

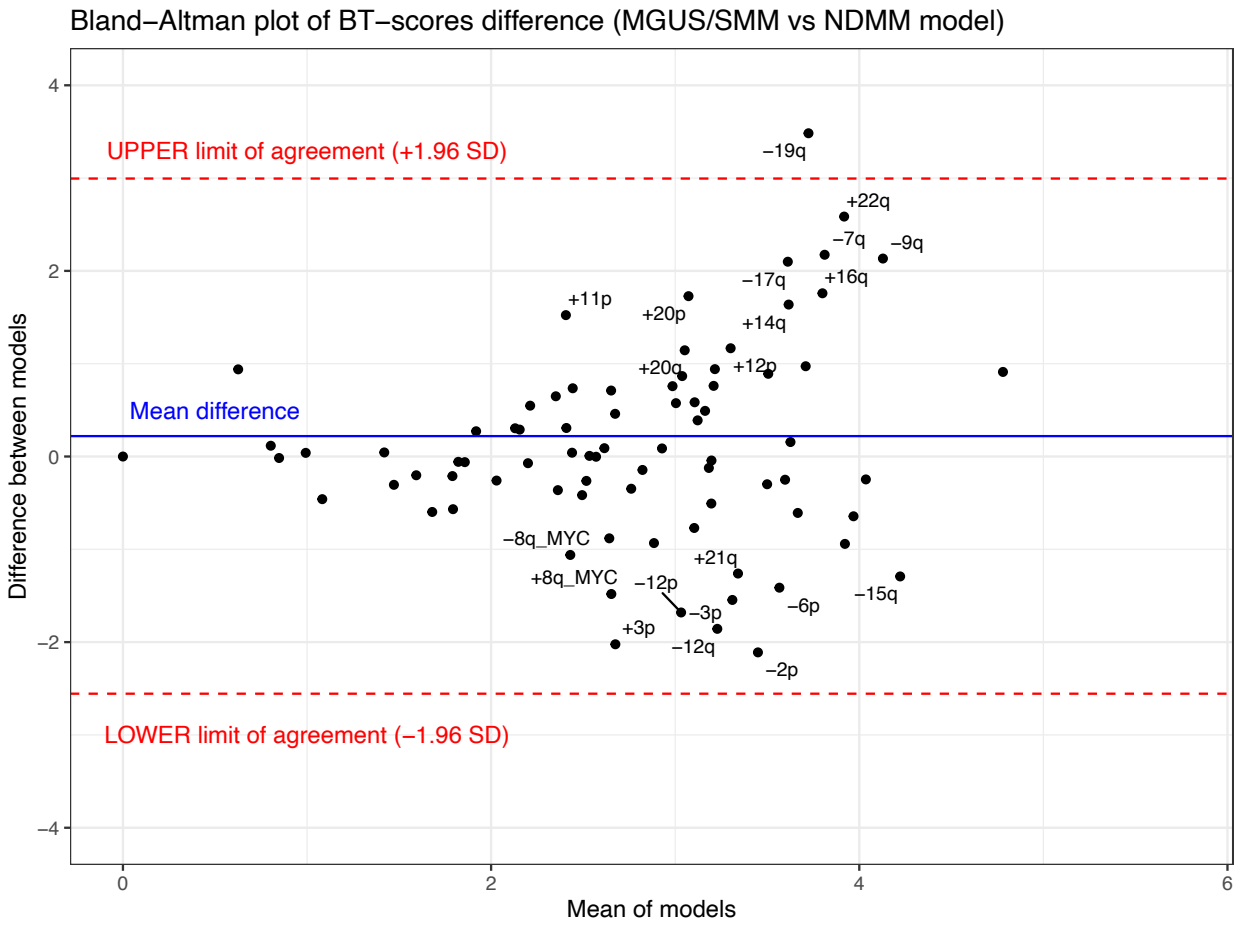
- subclonal SMM and clonal NDMM alterations represent alterations that increase in clonality between the two disease phases; consequently, they may be subject to positive selection (they may be associated with a fitness advantage that implies a clonal growth).
- clonal SMM and subclonal NDMM alterations: represent alterations that decrease in clonality between the two disease phases. Consequently, they may be affected by negative selection (they may be associated with a fitness disadvantage that implies a clonal extinction).

Overall, we found a positive correlation of BT scores between MGUS/SMM and MM (R=0.61, $P=1.1x10^{-5}$), suggesting the overall sequence of acquisition of genomic events is comparable across stages (**Fig. 3E, Extended Data Figure 5**). Indeed, in both league competitions, events that are known to happen early had the lowest scores (hyperdiploidy, del(14q), gain(11q), del(13q), and gain(1q); BT scores < 1.4: **Fig. 3E; Supplementary Table 7**). In contrast, deletion of 17p (encompassing *TP53*), which is a secondary event in models of progression and remains subclonal during disease development[8,9], had an elevated score in both leagues (MGUS/SMM: 3.4, MM: 2.6; $P=3.3x10^{-4}$), validating that our model captures accurate patterns of MM oncogenesis.

Statistical Analysis

Statistical analyses were conducted to ascertain the significance of the observed differences in BT scores between the two disease stages. We employed a bootstrap approach, generating 30 BT score distributions per alteration in each disease stage, to ensure robust BT score estimations and estimate measures of uncertainty (95% CI and SD). Following this, we computed the difference between NDMM and SMM bootstrap-derived median BT scores for each alteration (by using the Wilcoxon signed-rank test). Finally, to address multiple comparisons across genomic alterations, we applied the Benjamini-Hochberg procedure to control the false discovery rate (q-value). $q<0.05$ were considered significant.

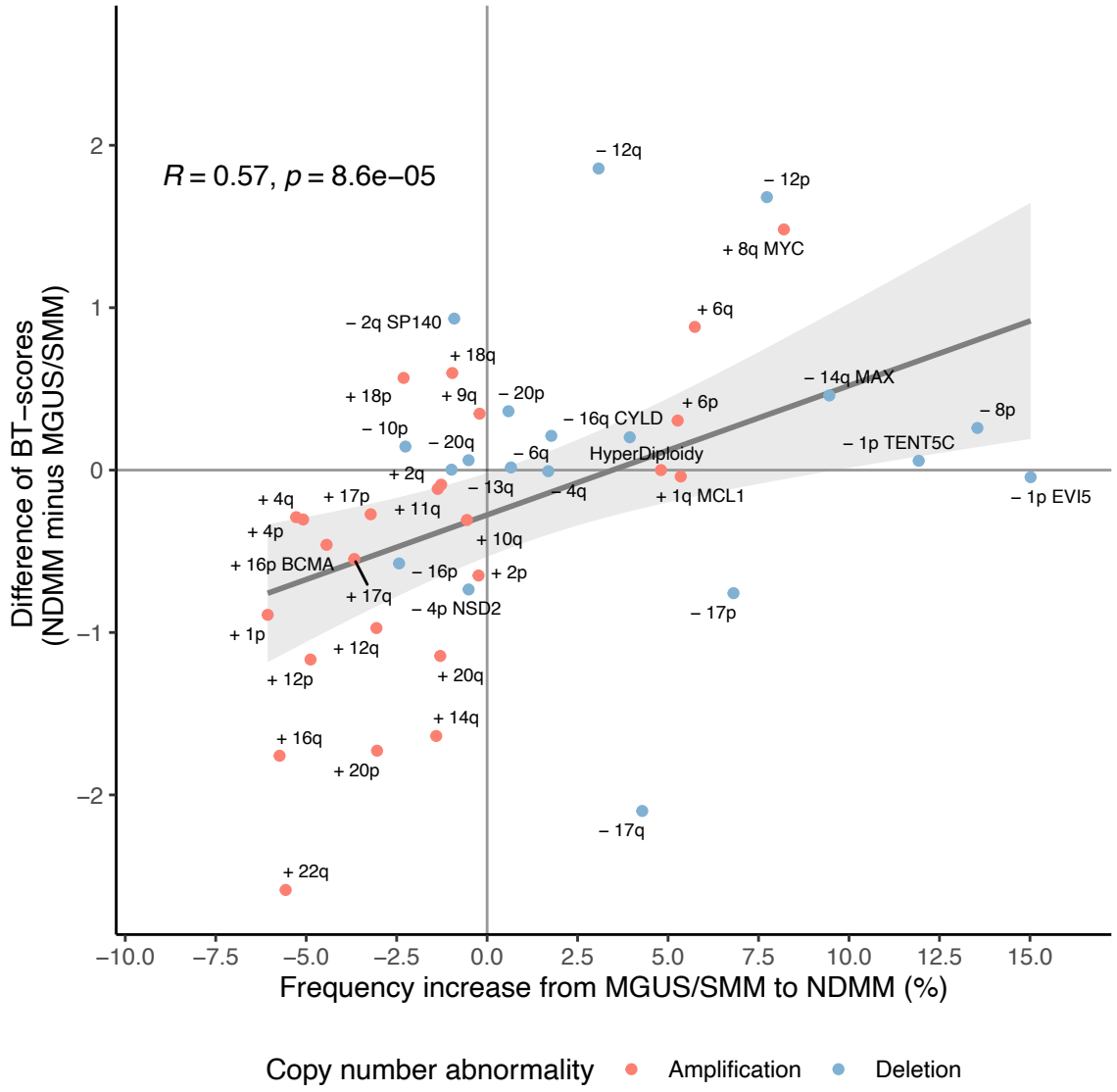
Next, we wanted to validate that the point estimates from Bradley-Terry models could be compared. We used the Bland-Altman method, representing the mean difference between models, as a heuristic to validate that both models (MGUS/SMM and NDMM) agree overall.



Supplementary Figure 7 *Bland-Altman plot of Bradley-Terry (BT) scores obtained from two BT models. Bland-Altman plot of BT scores obtained from two BT models in MGUS/SMM versus the BT scores obtained from the BT model in NDMM. Each point represents a different genomic alteration. The solid blue line represents the mean difference between BT scores, while the two red dashed lines represent the upper and lower limits of agreement, respectively (defined as ± 1.96 SD of the BT scores difference distribution).*

Next, we wanted to compare the frequency of events in MGUS/SMM versus their relative ordering (i.e., are events more common in MGUS/SMM also earlier). By subtracting the NDMM BT scores from SMM BT scores, we obtained a measure of the Clonality Difference (CD). Genomic alterations showing a positive CD (top part of the plot) were sub-clonal in MGUS/SMM and more clonal in NDMM (i.e., these events increased their clonality during progression to active disease). On the contrary, alterations showing a negative CD (bottom part of the plot) were more clonal in MGUS/SMM and sub-clonal in NDMM (i.e., these events decreased their clonality during the progression to active disease (bottom part of the plot). A linear regression model (dark grey line) showed a non-zero correlation (R^2=0.57, p<0.001) between the CD metric and the difference in the frequency of alterations. This suggests that alterations that increase their clonality also tend to increase their frequency, and vice-versa, between SMM and NDMM disease phases. The grey

band around the linear regression line represents the 95% confidence interval of the linear regression model.

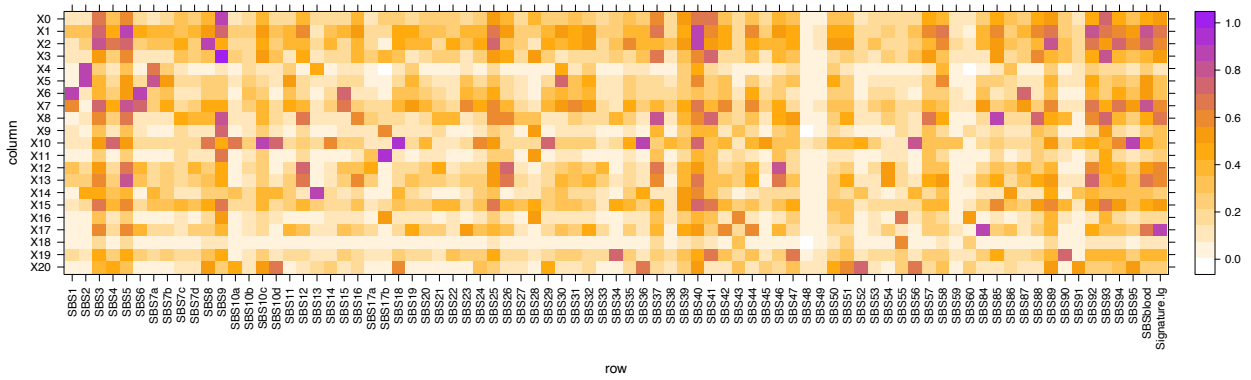


Supplementary Figure 8 *Clonality difference and frequency variation between SMM and NDMM. Correlation graph of the difference in frequency of events between MGUS/SMM and NDMM (x-axis), and the difference in BT scores between MGUS/SMM league and the NDMM league (y-axis). The resulting linear correlation coefficient was non-null (R=0.57, P=8.6E-5).*

7 Mutational signatures extracted with the *hdp* method

(*The following paragraph is copied from the Methods section of the manuscript to help with the understanding of the following Supplementary Figures*.) Mutational signatures were extracted from SNVs across all donors using a Bayesian hierarchical Dirichlet process (hdp package: https://github.com/nicolaroberts/hdp). If a given donor had multiple different tumor samples sequenced, SNVs were grouped into pseudo-samples based on their presence or absence across the different samples, such that a given SNV was only counted once. SNVs were further divided into three categories, each forming a further pseudo-sample: SNVs within immunoglobulin genes, clustered SNVs with an inter-mutational distance less than 1,000 bases, and other SNVs. SNVs were collapsed into their 96 counts of base change within their trinucleotide context. The hierarchy used by hdp consisted of (pseudo-)samples grouped by donor. The hdp algorithm was run in 20 different chains, for 40,000 iterations with a burn-in of 20,000 iterations.

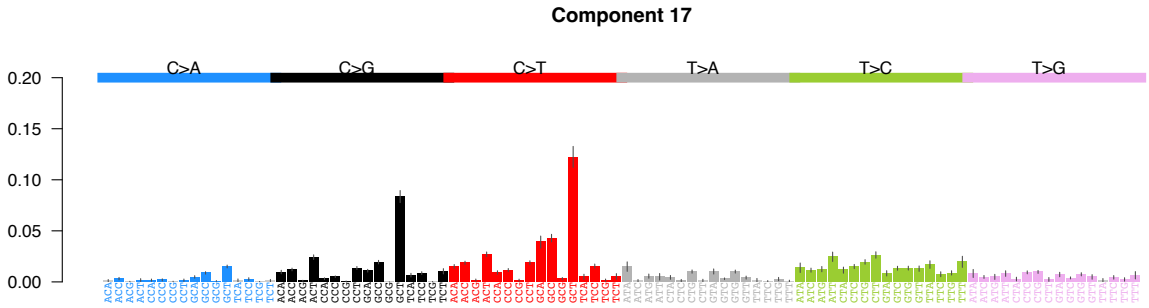
The resulting extracted signatures were further compared to the reference COSMIC signatures (v3.3).



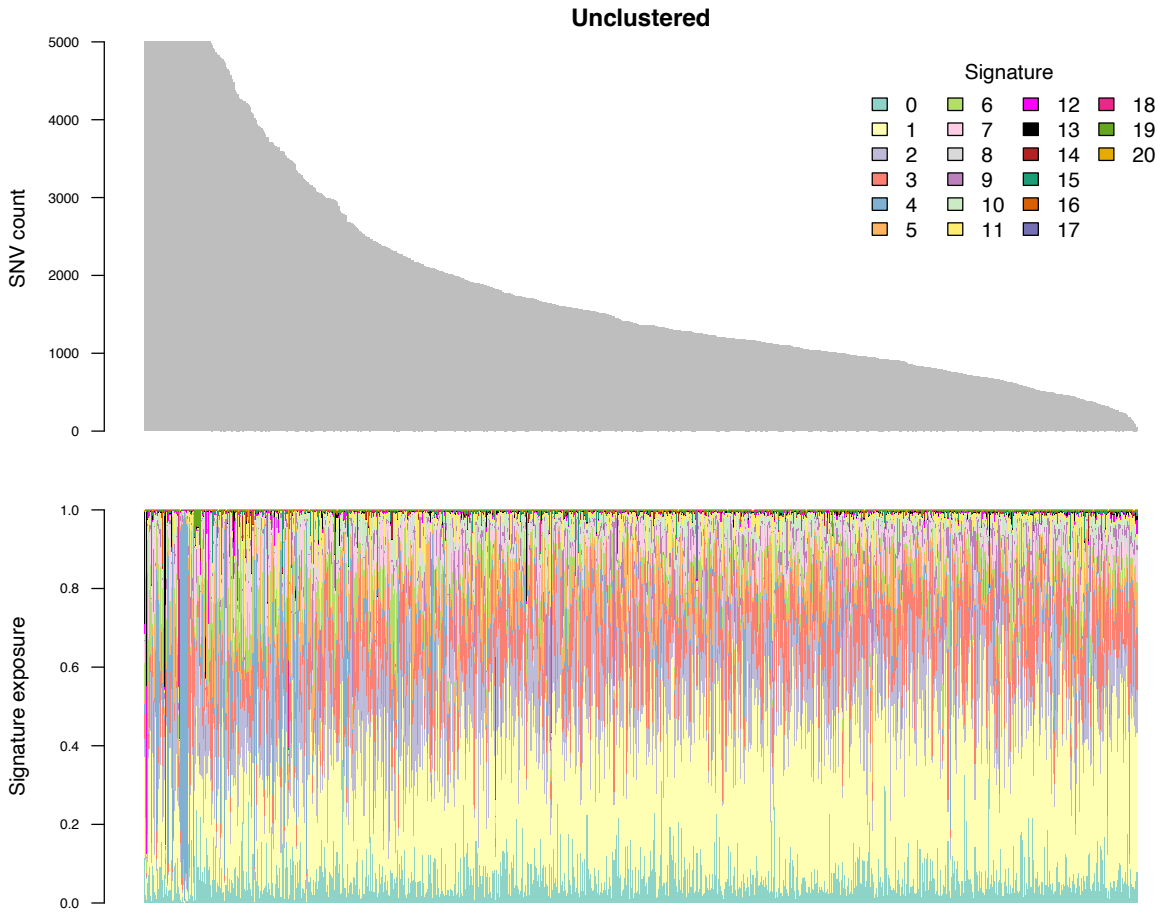
Supplementary Figure 9 *Cosine similarities (in a colored scale) between mutational signatures from the hdp method run on the entire cohort (MGUS/SMM and NDMM; on the y-axis) and the COSMIC (v3.3) catalog of mutations on the x-axis.*

Further, SBS16 and SBS17 were unexpected in MM. The spectrum for SBS16 was recently updated, which may explain why it remained undetected in previous work. SBS17 was initially linked to chemotherapy (5'-Fluoroacil) in solid cancers[10,11], and therefore excluded from untreated MM. However, it was recently described in B-cell lymphomas[12] and normal memory B cells[13], and our finding further supports a role for SBS17 in shaping the mutational landscape of MM as early as in precursor stages. We also detected eight signatures that were not in the COSMIC catalog, of which five possibly are technical artefacts which all accounted for less than <5% of mutations, were mostly subclonal, and could be associated to specific techniques used to sequence their WGS (**Supplementary Table 8**). Finally, three signatures could not be matched to the existing

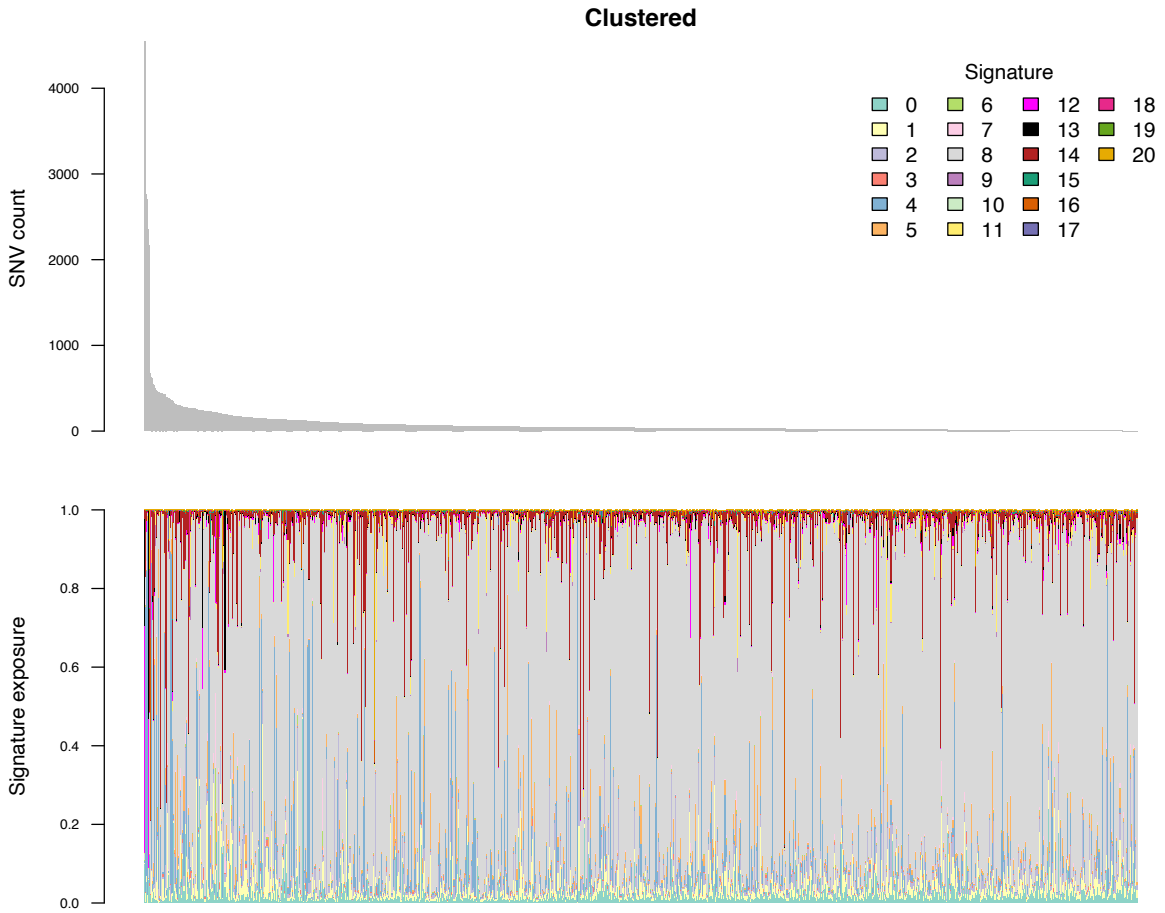
COSMIC catalogue, and we could not yet identify a preferential disease stage or subgroup. The unmatched signature N9 showed similarity with SBS9 and SBS17b with additional T>C and T>G components; N13 was marked mostly with C>T and T>C transitions resembling clock-like signatures SBS5/40 with an additional T>C component, while N15 showed limited similarity with SBS9 and SBS85 (AID) and SBS5/40 (clock-like; **Supplementary Notes**).



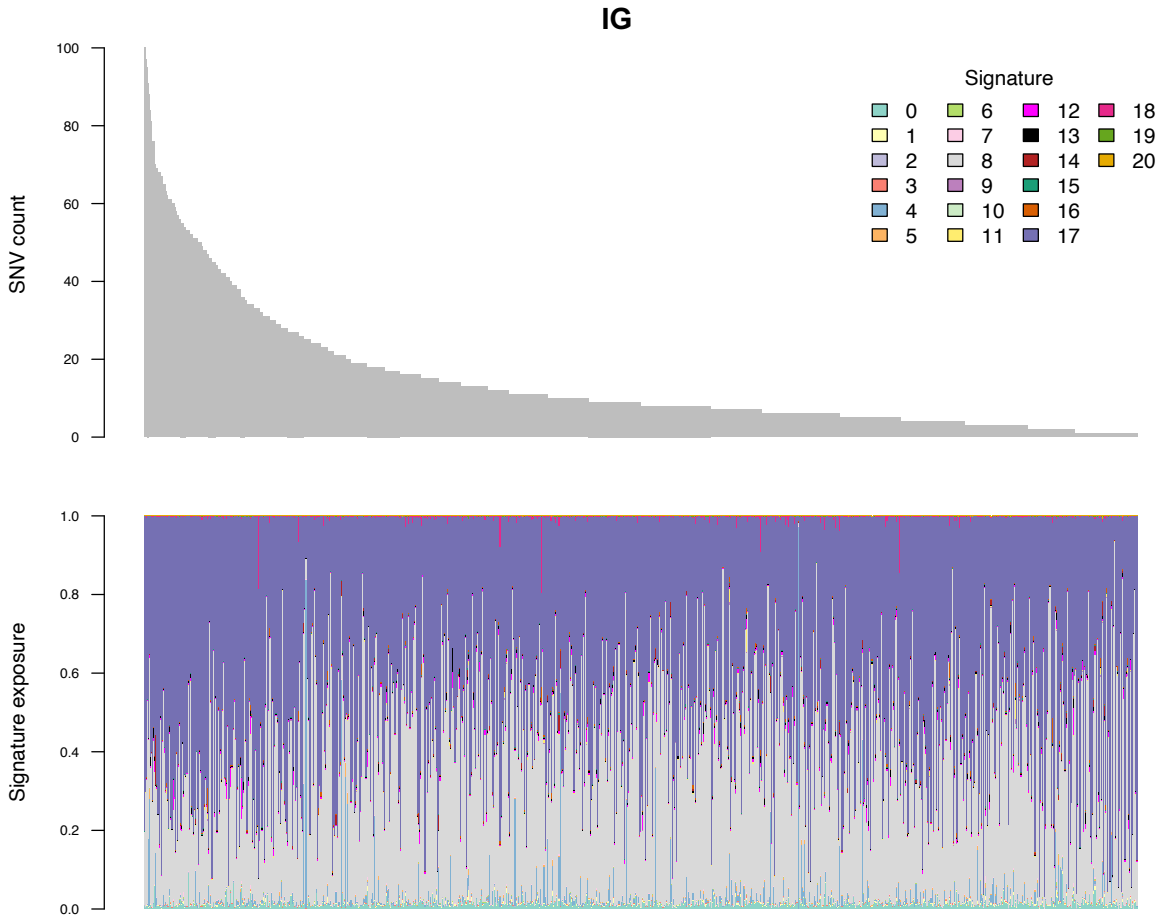
Supplementary Figure 10 *Weights of the signature X17 (SBS84-like) in their trinucleotide context estimated with the hierarchical Dirichlet process (hdp; Methods). Bars represent the mean value and 95% confidence intervals obtained from hdp.*



Supplementary Figure 11 *Bar plots representing the mutational burden per patient (N=958) top row, in total number of SNV) and the relative contribution of mutational signatures (bottom row) obtained from the hdp algorithm (cutoff: 5,000 SNVs).*



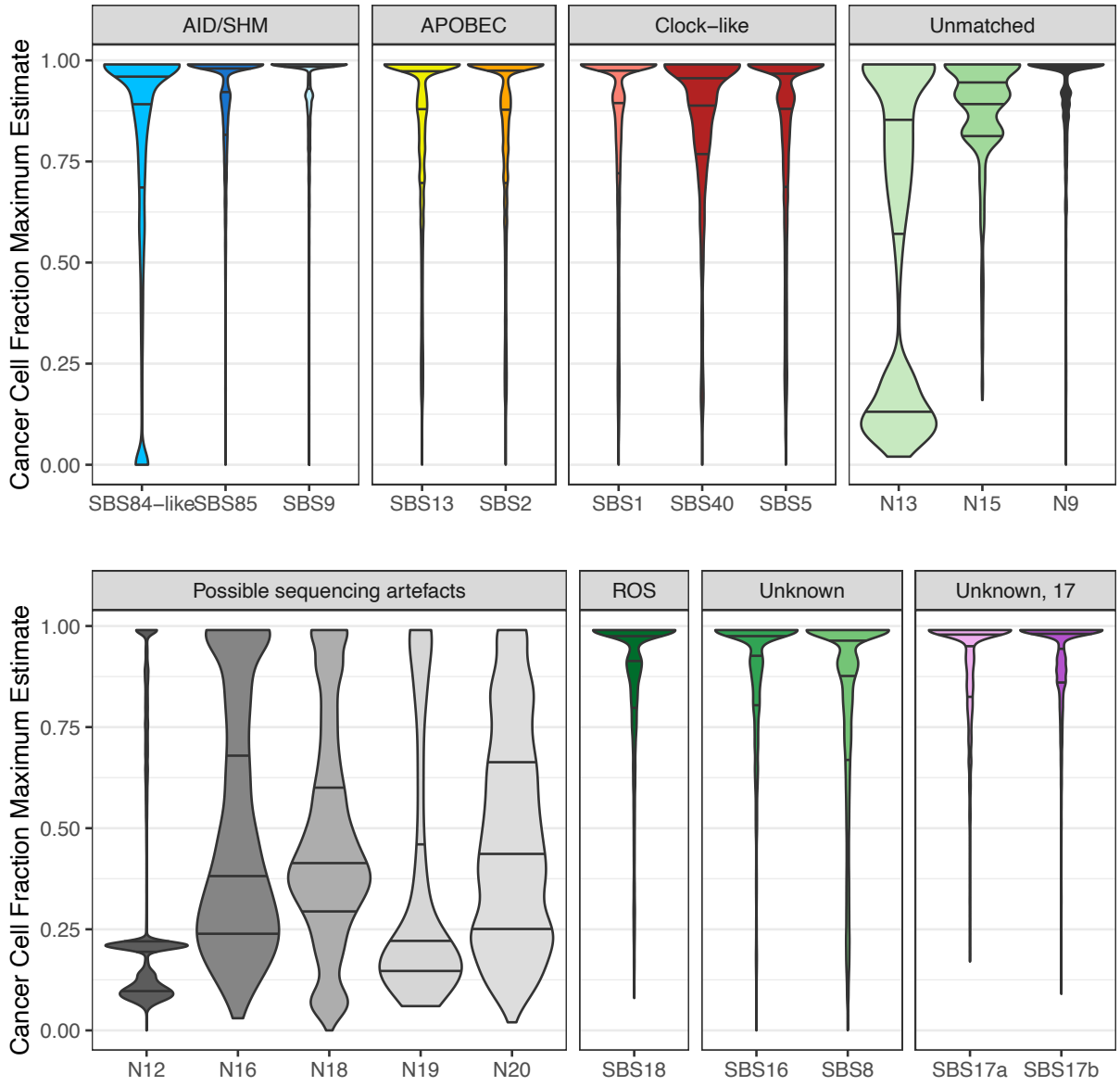
Supplementary Figure 12 *Mutational burden (top row, N=860) and relative contribution of mutational signatures (bottom row) obtained from the hdp algorithm for clustered mutations only (i.e., mutations that fall within 1kb of each other).*



Supplementary Figure 13 *Mutational burden (top row, N=842) and relative contribution of mutational signatures (bottom row) obtained from the hdp algorithm for mutations found in the immunoglobulin loci only: IGH (hg38: chr14:105,586,437-106,879,844), IGK (hg38: chr2:88,857,361-90,235,368), and IGL (hg38: chr22:22,026,076-22,922,913).*

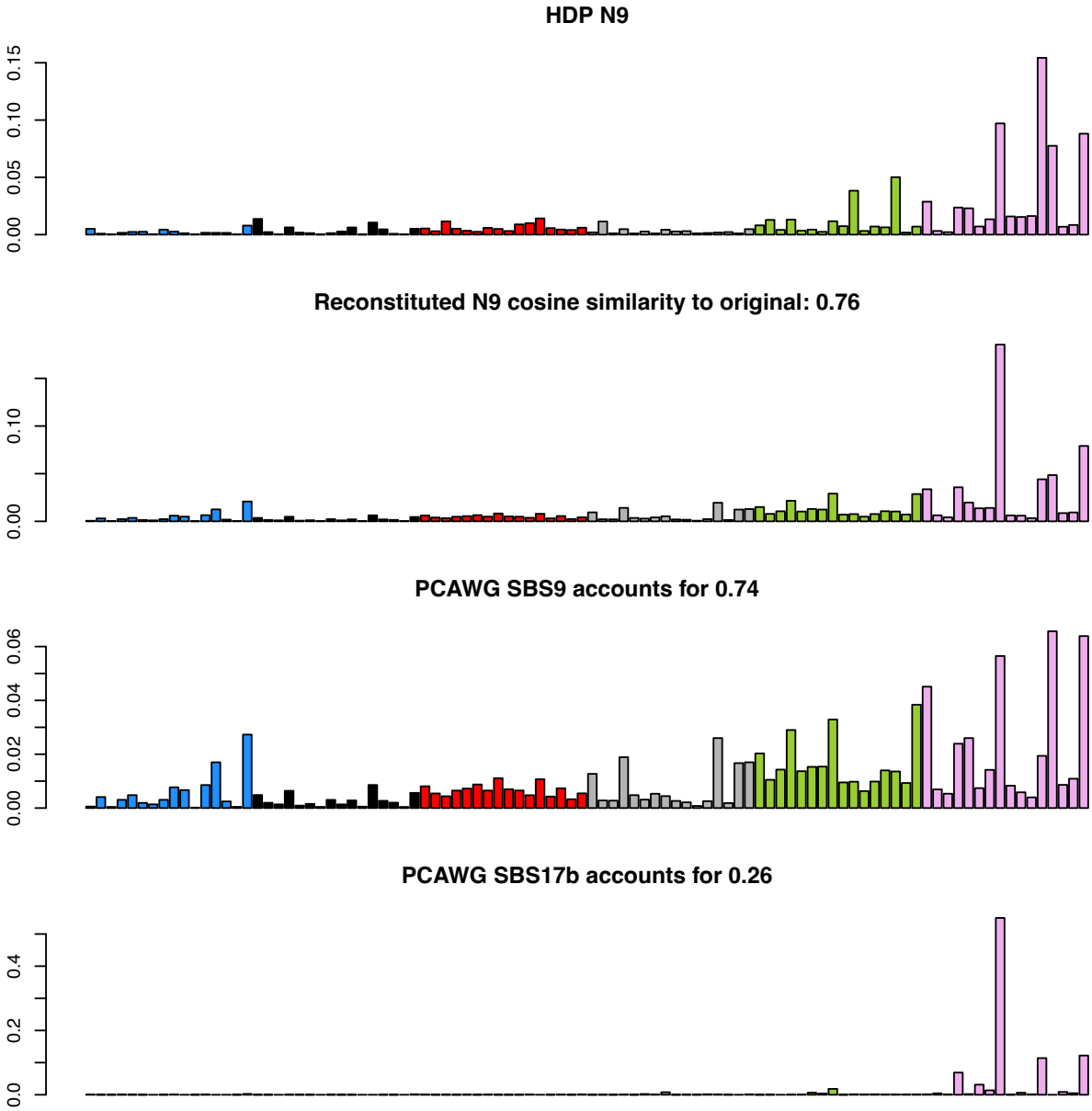
In order to identify mutational signatures caused by potential technical artifacts that were not filtered during WGS data analysis, we extracted all mutations with a CCF hat (maximum likelihood) > 0.5. This represents a total of 907,940 point mutations out of the original 2,097,410 unique point mutations in the cohort. We then represented the distribution of this CCF hat value across mutational signatures and show that potential technical artifacts indeed have low CCF estimates (median < 50%; ***Supplementary Figure 14***). This is consistent with known COSMIC signatures present in the WGS data, which mostly have high CCF (median > 50%). In addition, we showed that potential technical artifacts were enriched in certain library construction modalities: N12 was found enriched in samples processed with the Menarini Silicon Biosystems CellSearch CTC isolation machine, which will require a larger panel of normals with more data

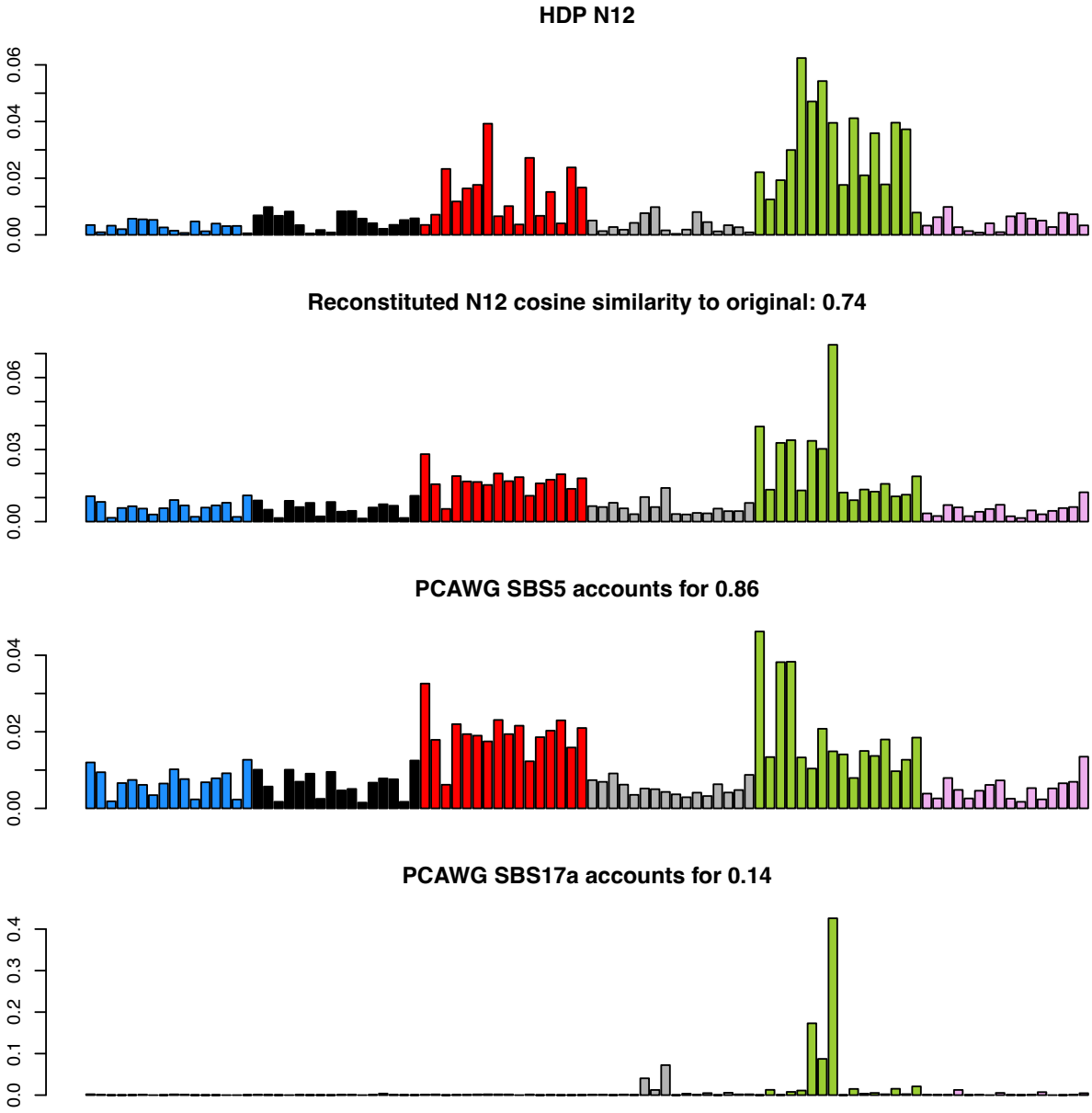
available. N16 was found in DNA libraries constructed with the NEB Ultra II but neither in CoMMpass nor in other available WGS data used for this study, and this indicates that low-input DNA protocols generate unfiltered mutations (<5%). Signatures N18 and N20 were found only in 5 samples processed in 2012-2014 and may reflect degradation of the DNA over time under suboptimal storage conditions. Finally, signature N19 was found only in samples sequenced at the Sanger Institute with a characteristic T[T>A]A component.

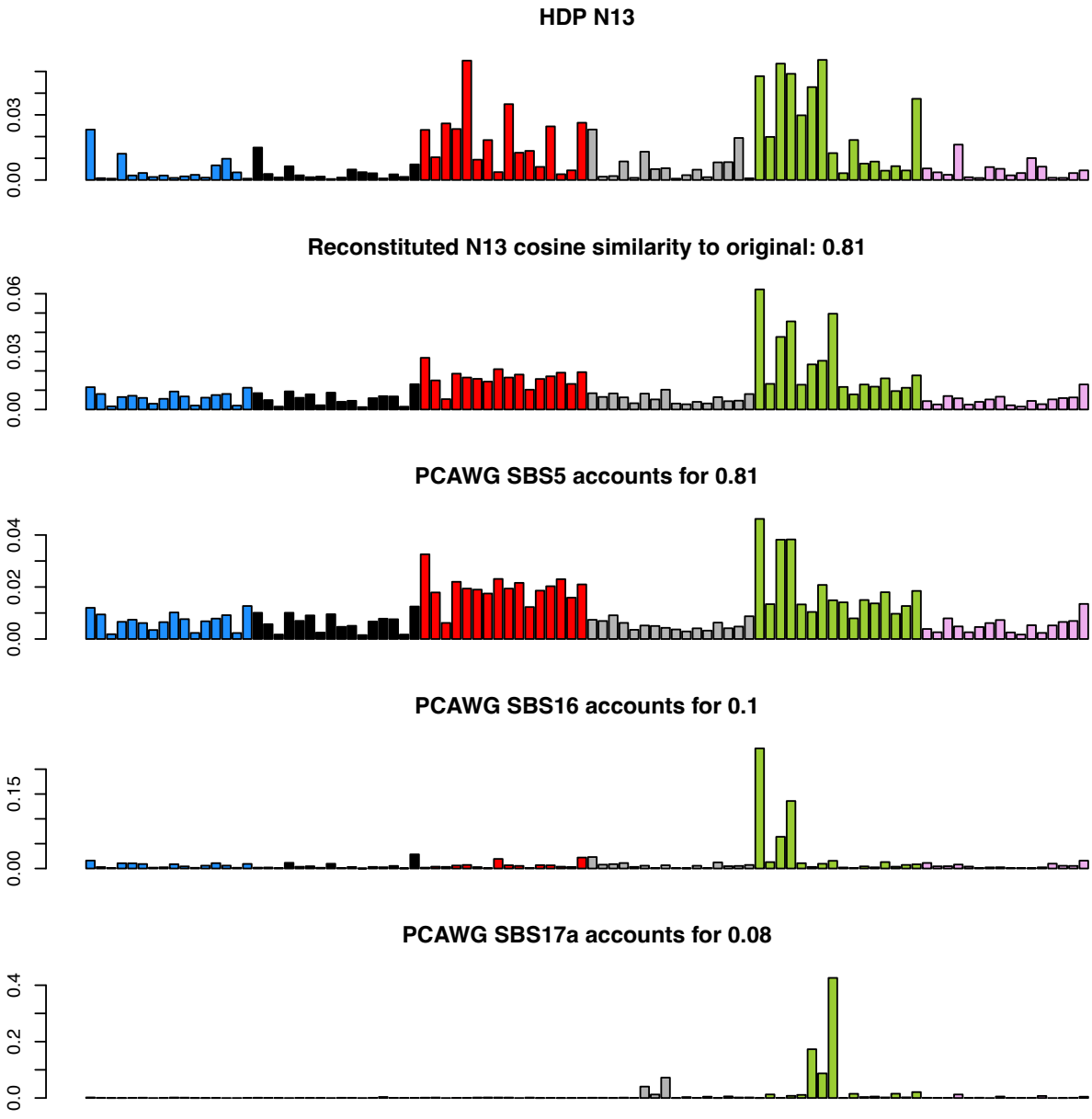


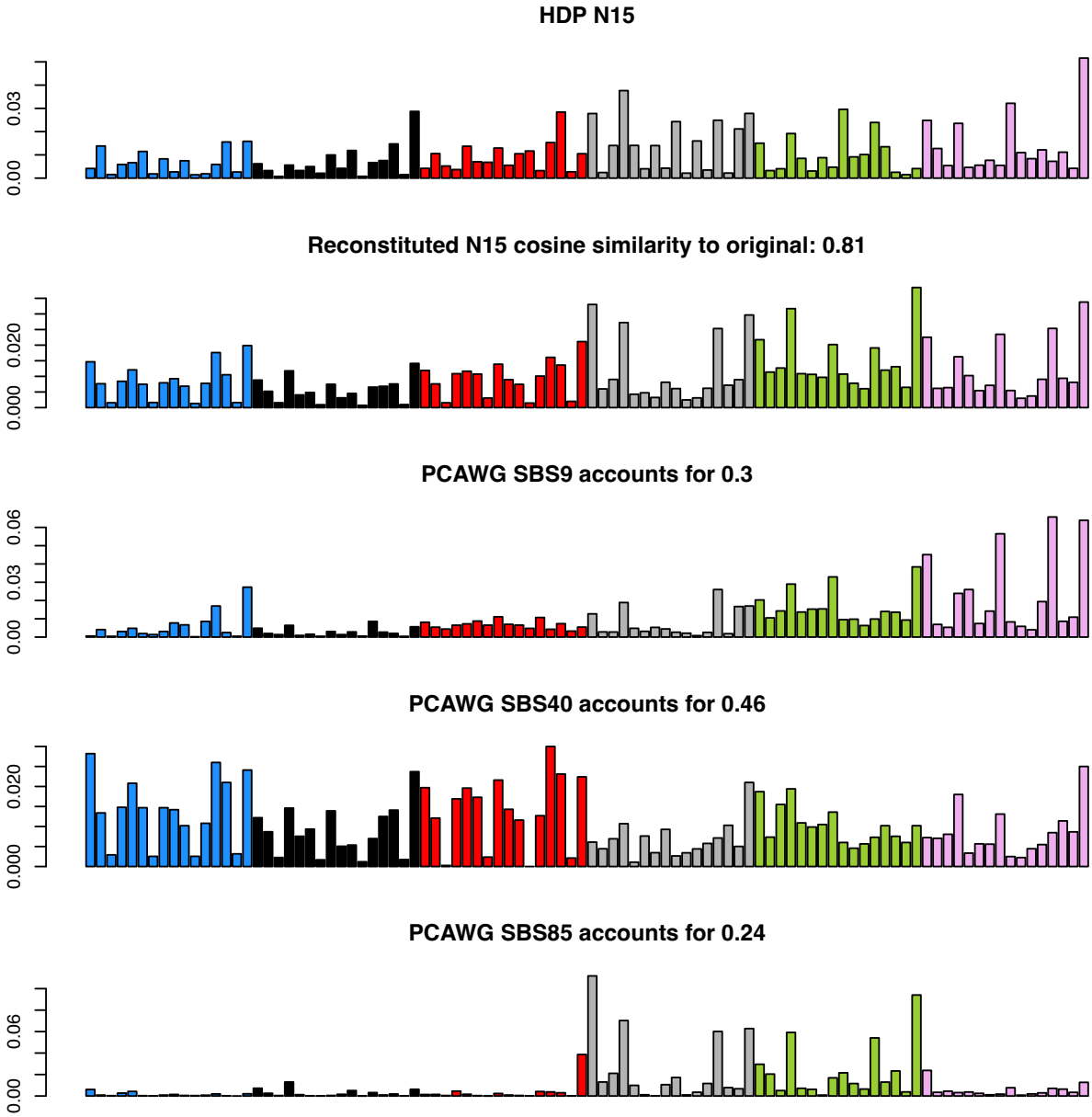
Supplementary Figure 14 *Distribution of CCF estimates per mutational signatures, for COSMIC-matched signatures and possible sequencing artifacts. Horizontal bars represent the 25, 50, and 75% quantiles of the distribution.*

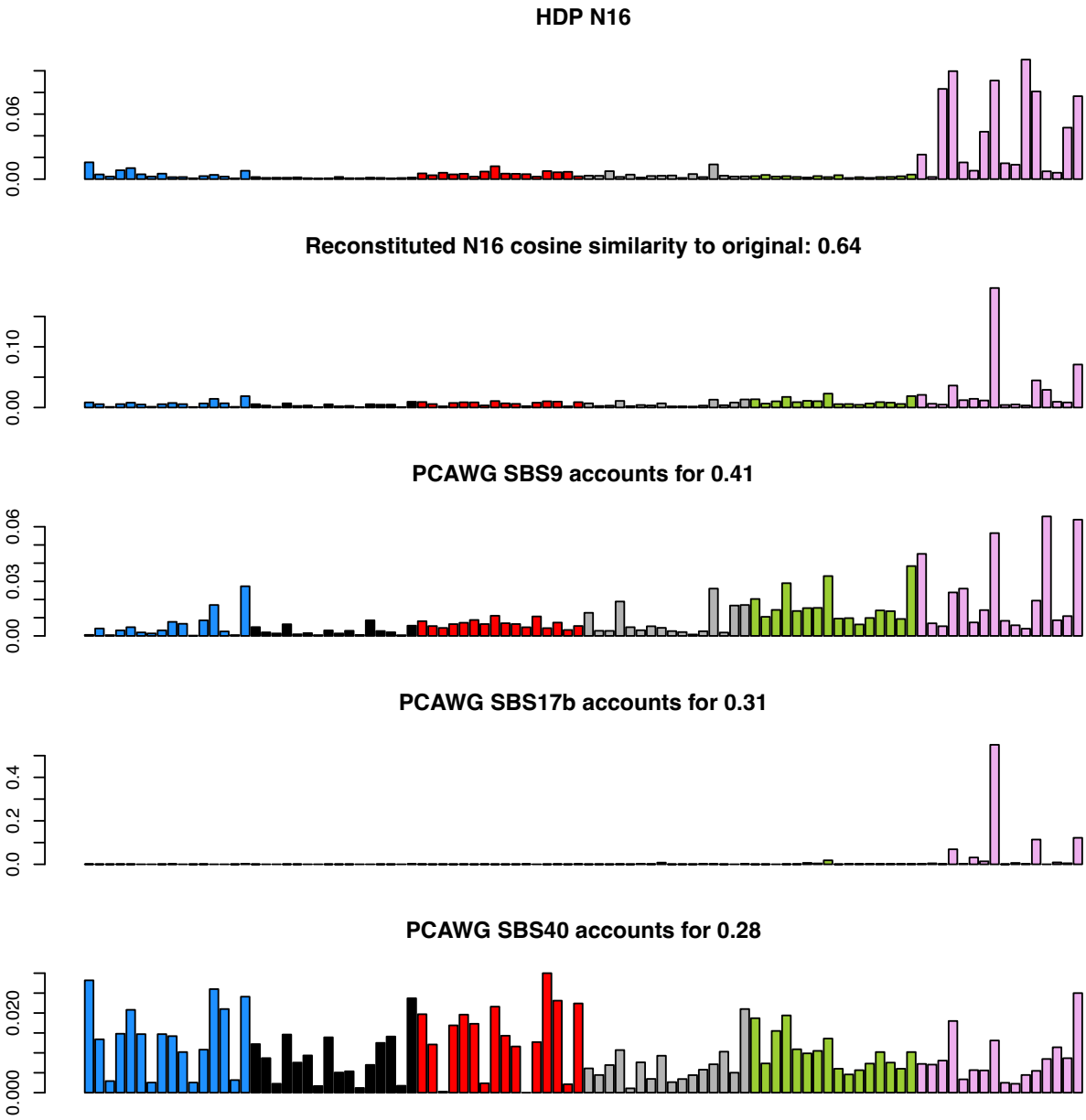
Unmatched mutational signatures obtained with the hdp algorithms did not reach the predetermined threshold of 0.85 at the reconstruction step (**Supplementary Figure 15**).

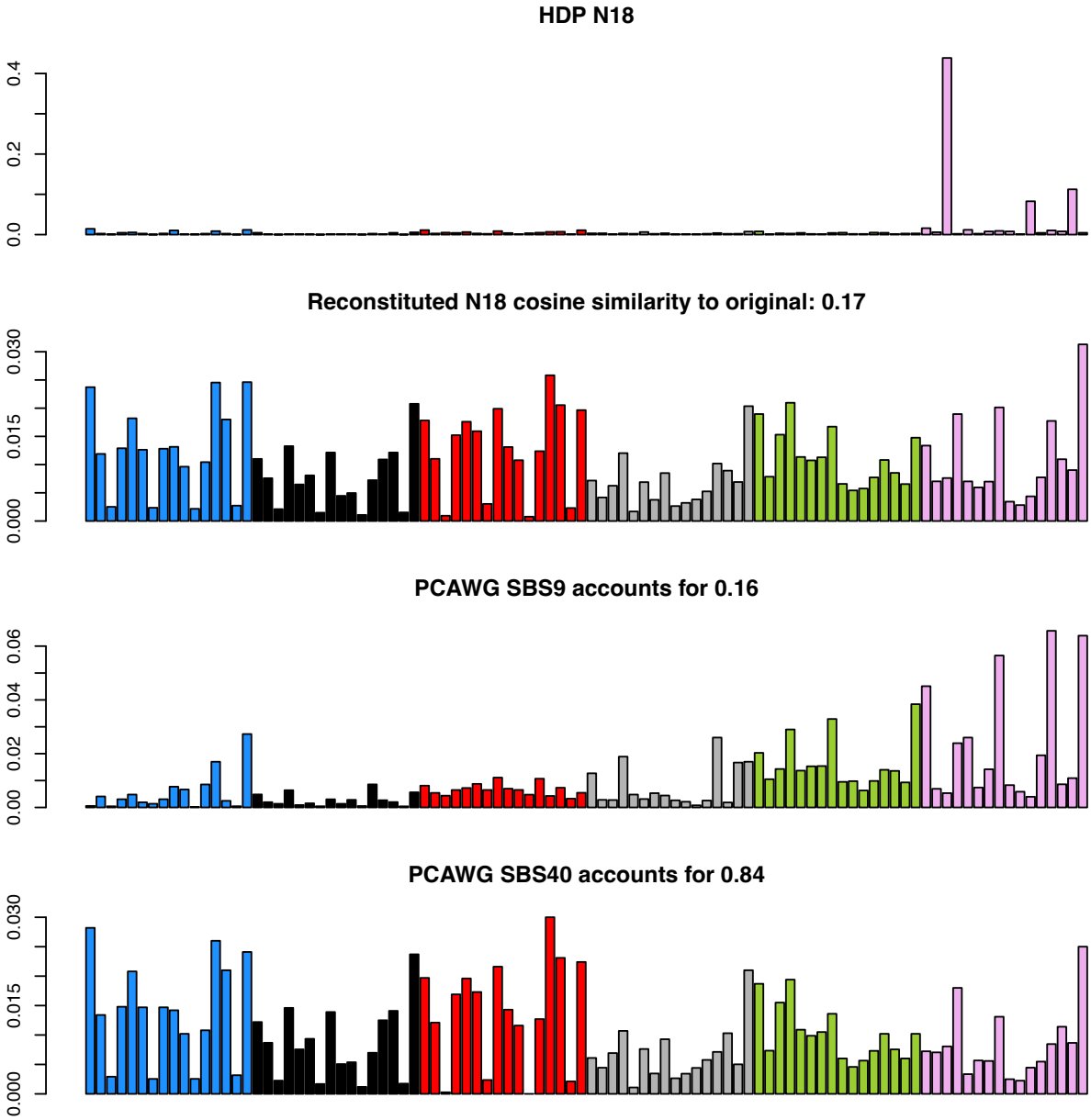


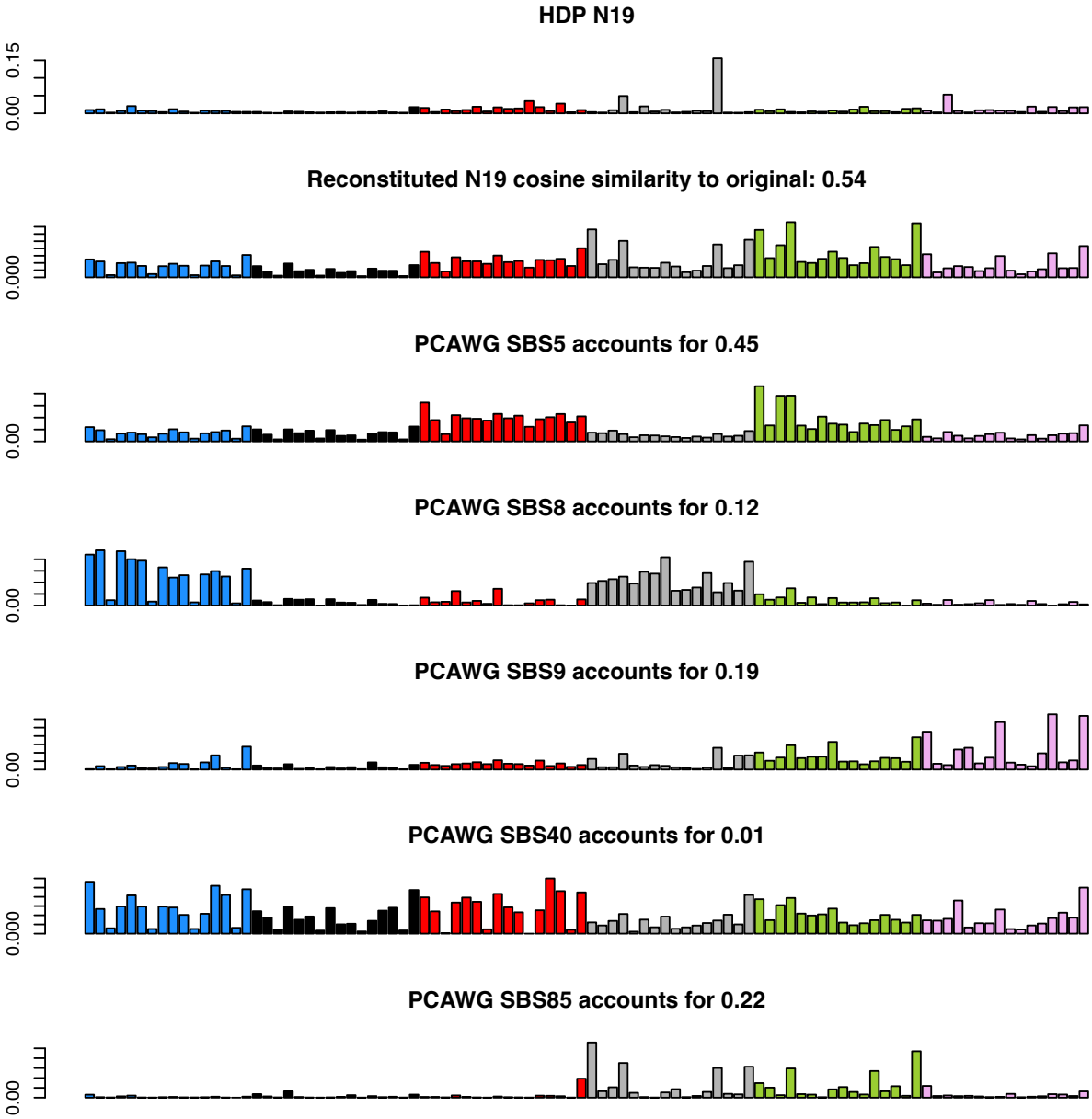

HDP N12
0.00
0.02
0.04
0.06
Reconstituted N12 cosine similarity to original: 0.74
0.00
0.03
0.06
PCAWG SBS5 accounts for 0.86
0.00
0.02
0.04
PCAWG SBS17a accounts for 0.14
0.0
0.1
0.2
0.3
0.4

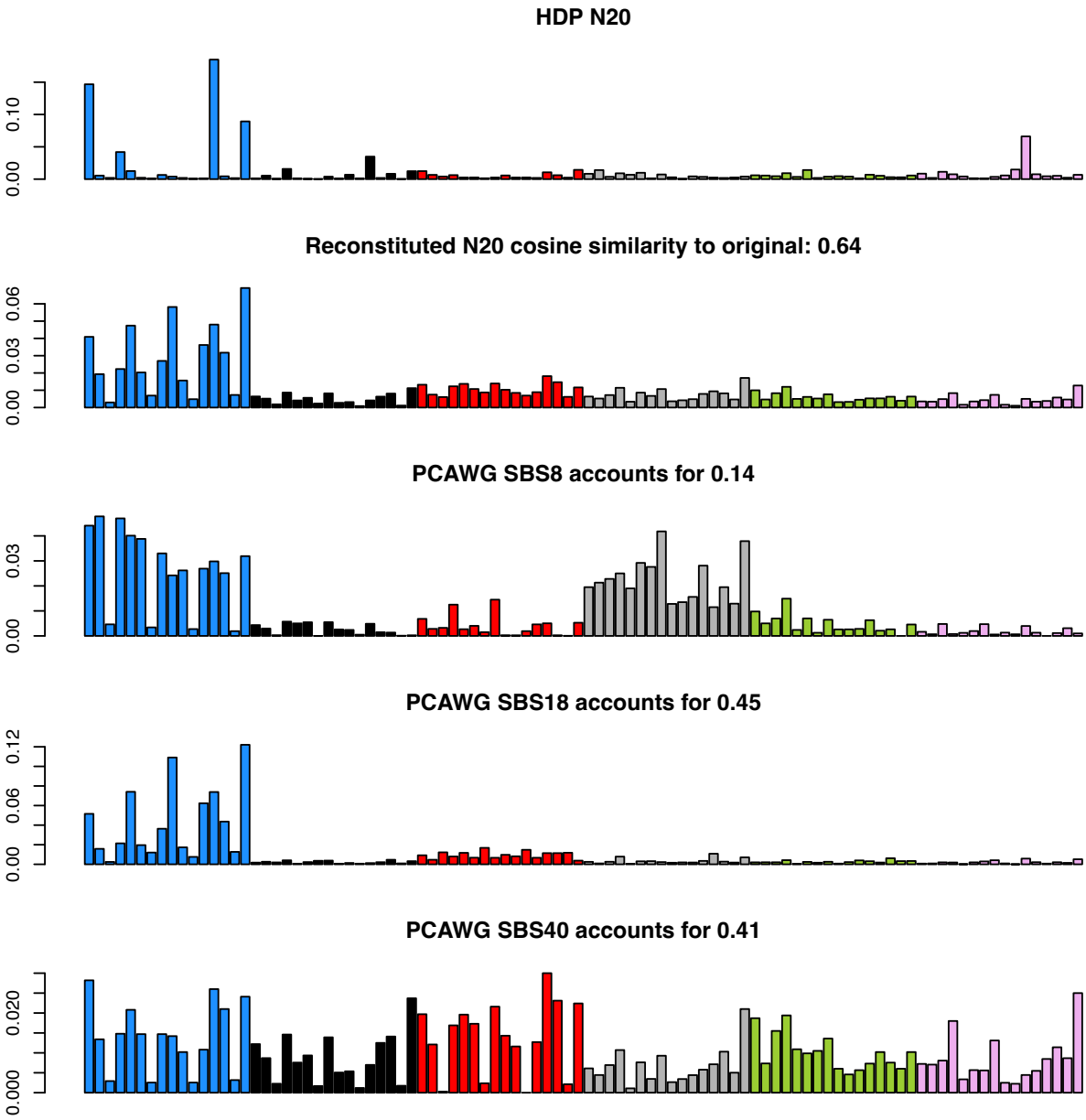

HDP N13
0.00
0.03
Reconstituted N13 cosine similarity to original: 0.81
0.00
0.03
0.06
PCAWG SBS5 accounts for 0.81
0.00
0.02
0.04
PCAWG SBS16 accounts for 0.1
0.00
0.15
PCAWG SBS17a accounts for 0.08
0.0
0.2
0.4


HDP N15
0.00
0.03
Reconstituted N15 cosine similarity to original: 0.81
0.000
0.020
PCAWG SBS9 accounts for 0.3
0.00
0.03
0.06
PCAWG SBS40 accounts for 0.46
0.000
0.020
PCAWG SBS85 accounts for 0.24
0.00
0.06


HDP N16
0.00
0.06
Reconstituted N16 cosine similarity to original: 0.64
0.00
0.10
PCAWG SBS9 accounts for 0.41
0.00
0.03
0.06
PCAWG SBS17b accounts for 0.31
0.0
0.2
0.4
PCAWG SBS40 accounts for 0.28
0.000
0.020


HDP N18
0.0
0.2
0.4
Reconstituted N18 cosine similarity to original: 0.17
0.000
0.015
0.030
PCAWG SBS9 accounts for 0.16
0.00
0.02
0.04
0.06
PCAWG SBS40 accounts for 0.84
0.000
0.015
0.030


HDP N19
0.00
0.15
Reconstituted N19 cosine similarity to original: 0.54
0.000
PCAWG SBS5 accounts for 0.45
0.00
PCAWG SBS8 accounts for 0.12
0.00
PCAWG SBS9 accounts for 0.19
0.00
PCAWG SBS40 accounts for 0.01
0.000
PCAWG SBS85 accounts for 0.22
0.00



Supplementary Figure 15 *Fitting of new signatures to COSMIC (unmatched signatures and potential technical artifacts). For each new hdp signature, we provide the optimal deconvolution against the COSMIC catalog of mutational signatures. Nucleotide contexts in the same order as in Suplementary Figure 10.*

8 Relation between MYC SVs, MM-like score, and detection of focal gains and losses at chr8q24

MYC was selected as one of the features in our MM-like score, which includes focal gains and losses around chr8q24.21 in 7/9 (78%) precursor patients with a *MYC* translocation.

9 Background on the detection of driver candidates among non-coding elements of the genome

Approximately 98-99% of the human genome is "non-coding"[14]. There are different types of annotated elements in the non-coding part of the genome, including non-coding parts of genes (3' and 5' untranslated regions (UTRs), introns, non-coding exons), regulatory elements (promoters, enhancers, insulators), and other elements (e.g. CpG islands not associated to a gene, transposable elements, non-coding RNAs, etc). Few non-coding elements, such as *TERT*[15,16] and *FOXA1*[17] promoter regions, have been shown to drive solid cancers[18]; however, the existence and role of non-coding mutation hotspots in MM was yet unknown. This motivated us to leverage the DIG algorithm[19] to discover non-coding MM candidate drivers. (See Methods and Results)

Because the MM mutational landscape is likely to resemble that of normal B cell lineage (affected by SHM, and off-target AID activity), we compared our findings with mutations from naïve and memory B cells obtained from Machado et al. [13] publicly available (https://github.com/machadoheather/lymphocyte_somatic_mutation/tree/main/data).

9.1 Note on ILF2 promoter mutation in the MAF subgroup

The MAF subgroup is defined as having an *IgH* translocation with one of the MAF genes, i.e., *MAF* (on chr. 16), *MAFA* (chr. 8), or *MAFB* (chr. 20). However, as opposed to the *CCND1* mutations, *ILF2* is not translocated in the MAF subgroup and hence the mutations are not related to the translocation process. More specifically, APOBEC-induced TpC>T and TpC>G mutations were localized to two nucleotide positions on chr1q21.2 (chr1:153,671,247 and chr1:153,671,254; hg38), which are not in an APOBEC-preferred hairpin loop[20]. Interestingly, these hotspots are within a 200-300 bp CTCF binding peak region in ChIP-seq data in human myeloma (NCI-H929, MM.1S, and KMS-11) and in immortalized lymphoblastoid B-cell (GM12878) cell lines; however, we did not find a canonical CTCF binding motif in this peak region (**Extended Data Figure 8**). Interestingly, in addition to *ILF2*, 3 of the 22 other genes within the same topologically associated domain (TAD; S*NAPIN*, *S100A4,* and *S100A6*) were also significantly overexpressed in the mutant cases (q<0.1 for all genes), suggesting that these mutations are associated with changes in the TAD or activation of enhancers that regulate multiple genes within the TAD (**Fig. 6E**).

9.2 Note on ILF2 expression in the CoMMpass study

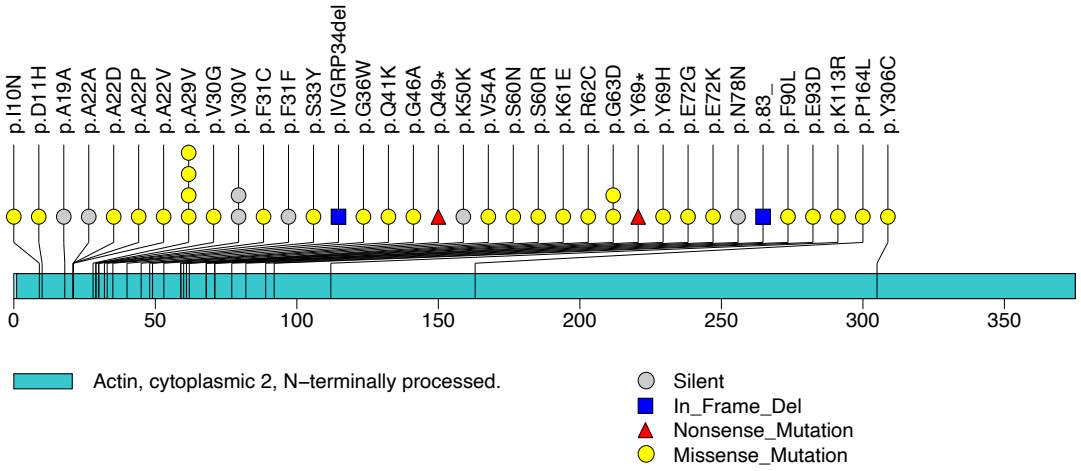
To further validate these hotspots in MM, we examined the CoMMpass low-pass WGS data and searched for evidence of these mutations (even if present in a single read). We found these

hotspot mutations in 26 patients among the 769 CoMMpass samples with at least two reads in at least one hotspot locus (**Supplementary Table 11**). Consistent with the deep WGS data, the MAF subgroup was overrepresented in CoMMpass with an *ILF2* promoter hotspot mutation (13/26; Fisher's Exact $p=4.7x10^{-11}$; **Supplementary Table 11**).

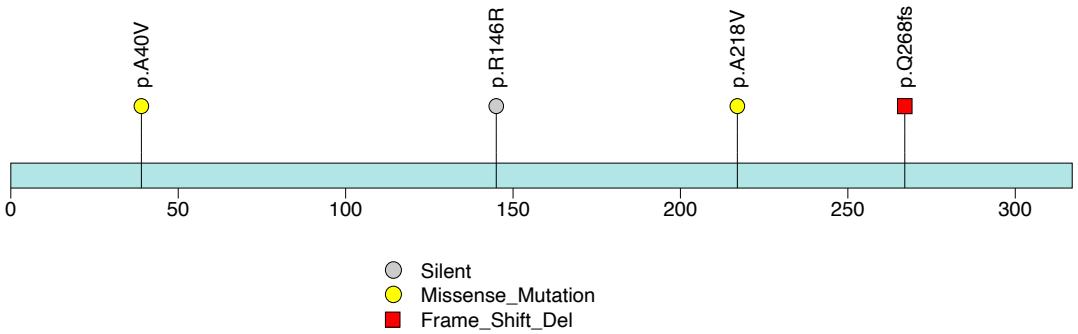
10 Distribution of mutations from candidate drivers on protein sequences

Significant hits from MutSig2CV (including rare drivers <1%) are represented below (***Supplementary Figure 16***). For each candidate driver, we list the mutations found in their coding sequence. For a given protein change, a maximum of 10 samples are shown below. For exact mutations found in >10 samples, the number of samples found mutated is indicated between parenthesis (e.g., *NRAS* p.Q61R (66)).

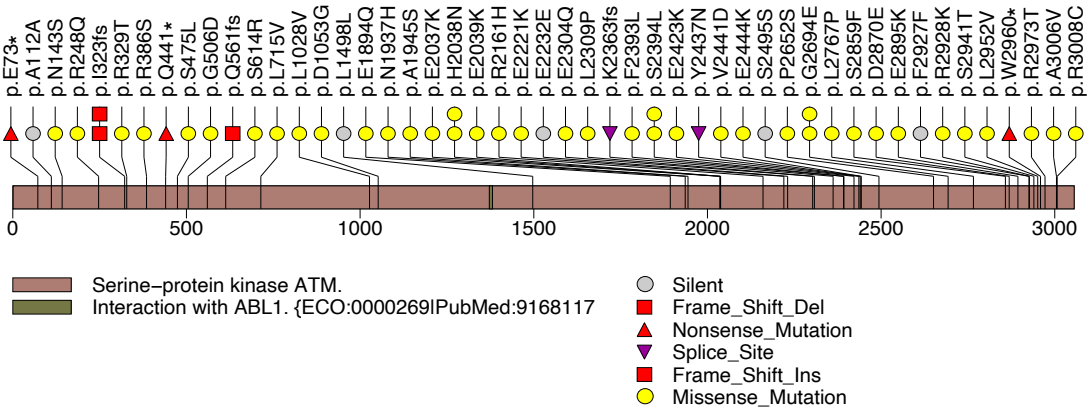
ACTG1



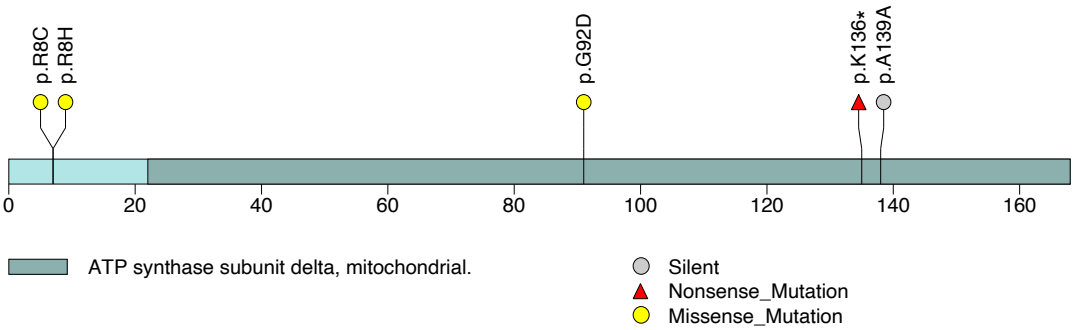
ANKRD9


p.A40V
p.R146R
p.A218V
p.Q268fs
0
50
100
150
200
250
300
Silent
Missense_Mutation
Frame_Shift_Del

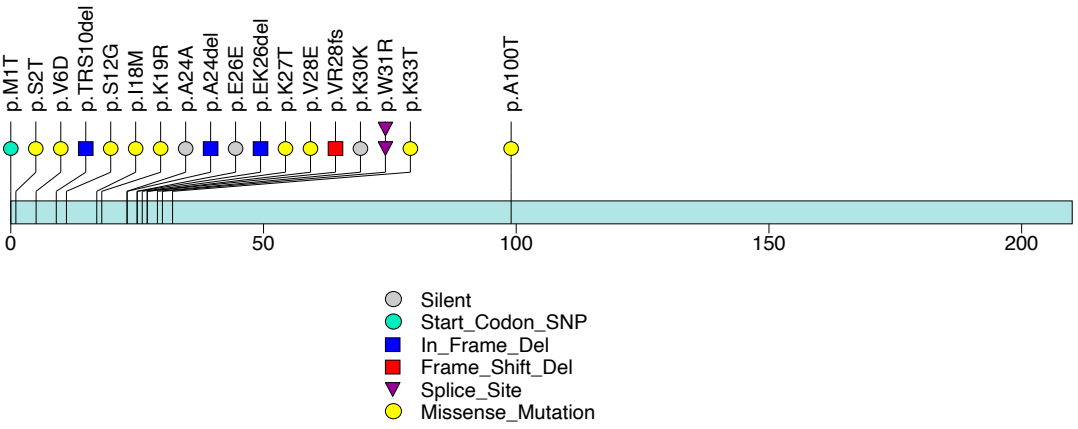
ATM


p.E73*
p.A112A
p.N143S
p.R248Q
p.I323fs
p.R329T
p.R386S
p.Q441*
p.S475L
p.G506D
p.Q561fs
p.S614R
p.L715V
p.L1028V
p.D1053G
p.L1498L
p.E1894Q
p.N1937H
p.A1945S
p.E2037K
p.H2038N
p.E2039K
p.R2161H
p.E2221K
p.E2232E
p.E2304Q
p.L2309P
p.K2363fs
p.F2393L
p.S2394L
p.E2423K
p.Y2437N
p.V2441D
p.E2444K
p.S2495S
p.P2652S
p.G2694E
p.L2767P
p.S2859F
p.D2870E
p.E2895K
p.F2927F
p.R2928K
p.S2941T
p.L2952V
p.W2960*
p.R2973T
p.A3006V
p.R3008C
0
500
1000
1500
2000
2500
3000
Serine−protein kinase ATM.
Interaction with ABL1. {ECO:0000269|PubMed:9168117
Silent
Frame_Shift_Del
Nonsense_Mutation
Splice_Site
Frame_Shift_Ins
Missense_Mutation

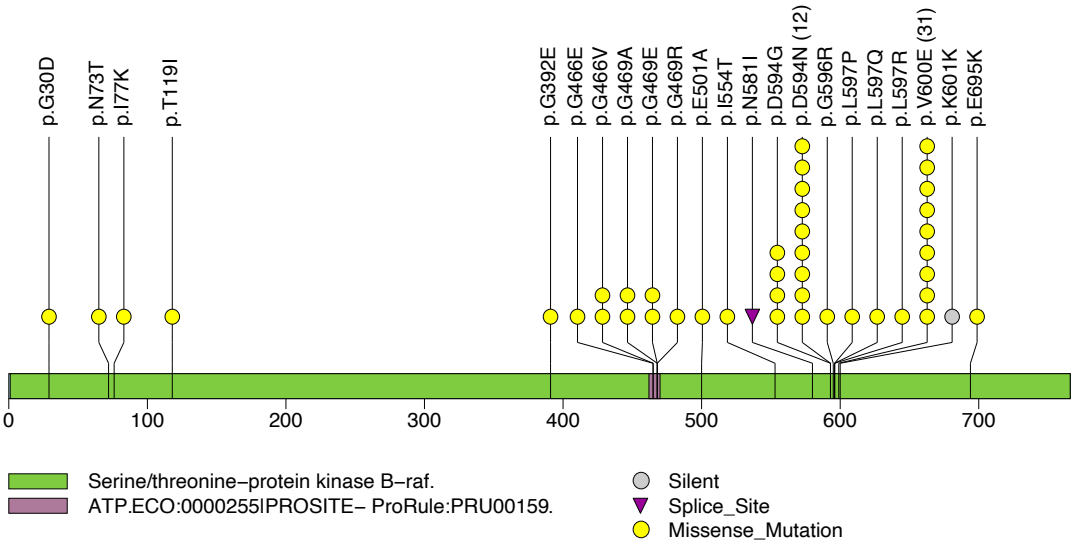
ATP5D


p.R8C
p.R8H
p.G92D
p.K136*
p.A139A
0
20
40
60
80
100
120
140
160
ATP synthase subunit delta, mitochondrial.
Silent
Nonsense_Mutation
Missense_Mutation

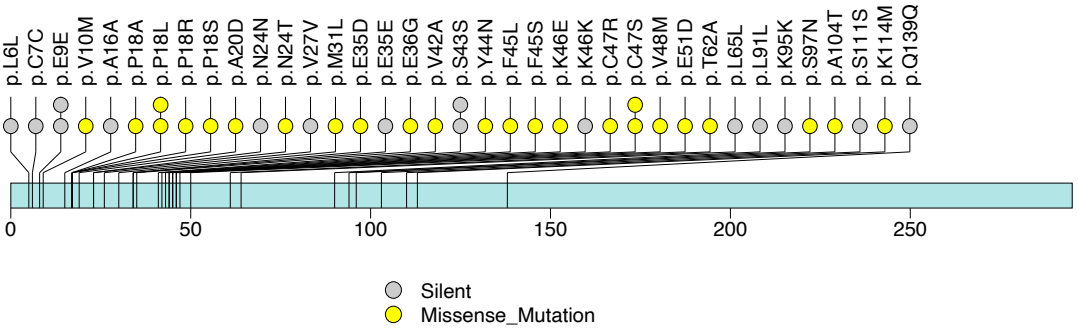
BCL7A


p.M1T
p.S2T
p.V6D
p.TRS10del
p.S12G
p.I18M
p.K19R
p.A24A
p.A24del
p.E26E
p.EK26del
p.K27T
p.V28E
p.VR28fs
p.K30K
p.W31R
p.K33T
p.A100T
0
50
100
150
200
Silent
Start_Codon_SNP
In_Frame_Del
Frame_Shift_Del
Splice_Site
Missense_Mutation

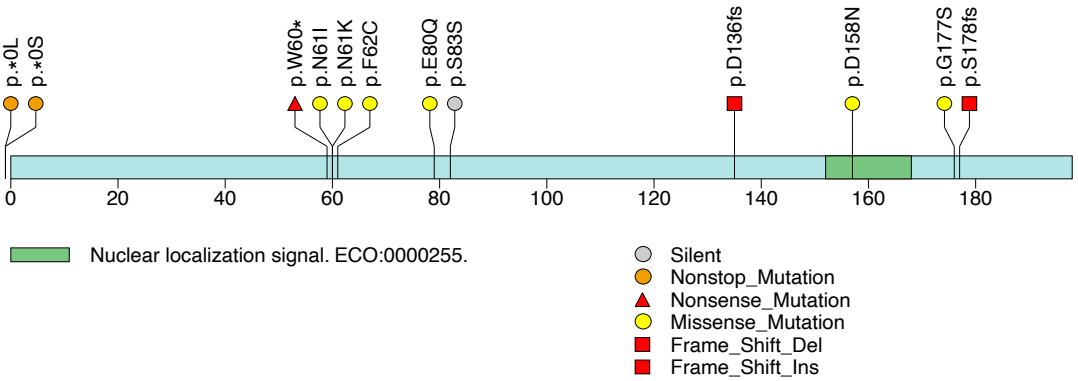
BRAF


p.G30D
p.N73T
p.I77K
p.T119I
p.G392E
p.G466E
p.G466V
p.G469A
p.G469E
p.G469R
p.E501A
p.I554T
p.N581I
p.D594G
p.D594N (12)
p.G596R
p.L597P
p.L597Q
p.L597R
p.V600E (31)
p.K601K
p.E695K
0
100
200
300
400
500
600
700
Serine/threonine−protein kinase B−raf.
ATP.ECO:0000255|PROSITE− ProRule:PRU00159.
Silent
Splice_Site
Missense_Mutation

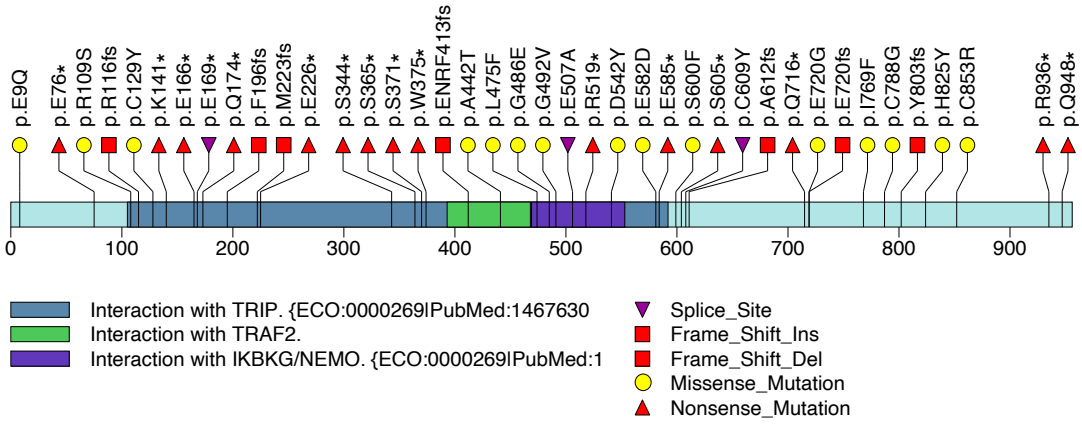
CCND1


p.L6L
p.C7C
p.E9E
p.V10M
p.A16A
p.P18A
p.P18L
p.P18R
p.P18S
p.A20D
p.N24N
p.N24T
p.V27V
p.M31L
p.E35D
p.E35E
p.E36G
p.V42A
p.S43S
p.Y44N
p.F45L
p.F45S
p.K46E
p.K46K
p.C47R
p.C47S
p.V48M
p.E51D
p.T62A
p.L65L
p.L91L
p.K95K
p.S97N
p.A104T
p.S111S
p.K114M
p.Q139Q
0
50
100
150
200
250
Silent
Missense_Mutation

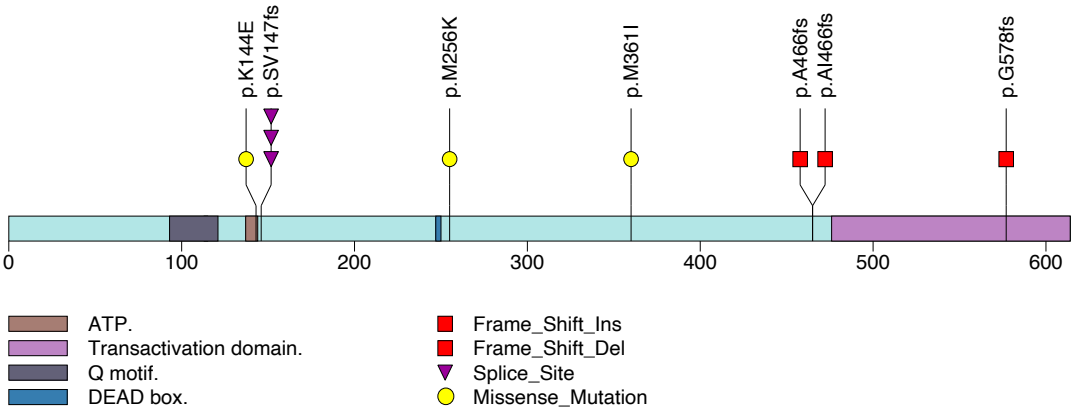
CDKN1B


p.*0L
p.*0S
p.W60*
p.N61I
p.N61K
p.F62C
p.E80Q
p.S83S
p.D136fs
p.D158N
p.G177S
p.S178fs
0
20
40
60
80
100
120
140
160
180
Nuclear localization signal. ECO:0000255.
Silent
Nonstop_Mutation
Nonsense_Mutation
Missense_Mutation
Frame_Shift_Del
Frame_Shift_Ins

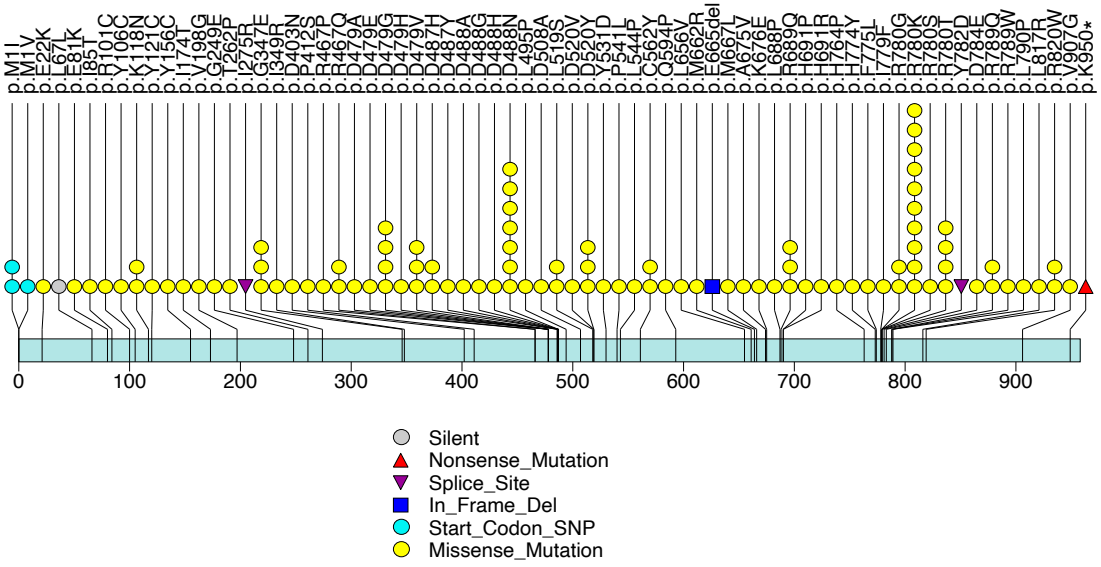
CYLD


p.E9Q
p.E76*
p.R109S
p.R116fs
p.C129Y
p.K141*
p.E166*
p.E169*
p.Q174*
p.F196fs
p.M223fs
p.E226*
p.S344*
p.S365*
p.S371*
p.W375*
p.ENRF413fs
p.A442T
p.L475F
p.G486E
p.G492V
p.E507A
p.R519*
p.D542Y
p.E582D
p.E585*
p.S600F
p.S605*
p.C609Y
p.A612fs
p.Q716*
p.E720G
p.E720fs
p.I769F
p.C788G
p.Y803fs
p.H825Y
p.C853R
p.R936*
p.Q948*
0
100
200
300
400
500
600
700
800
900
Interaction with TRIP. {ECO:0000269|PubMed:1467630
Interaction with TRAF2.
Interaction with IKBKG/NEMO. {ECO:0000269|PubMed:1
Splice_Site
Frame_Shift_Ins
Frame_Shift_Del
Missense_Mutation
Nonsense_Mutation

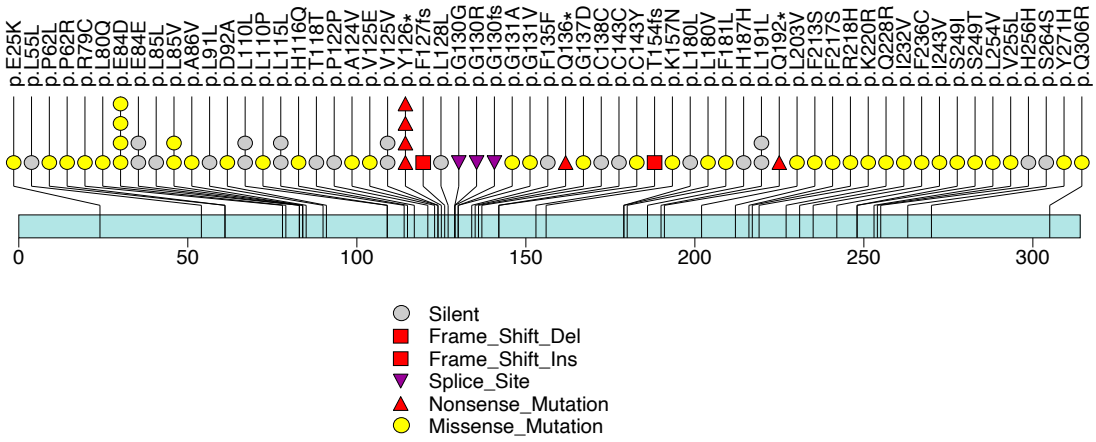
DDX5


p.K144E
p.SV147fs
p.M256K
p.M361I
p.A466fs
p.AI466fs
p.G578fs
0
100
200
300
400
500
600
ATP.
Transactivation domain.
Q motif.
DEAD box.
Frame_Shift_Ins
Frame_Shift_Del
Splice_Site
Missense_Mutation

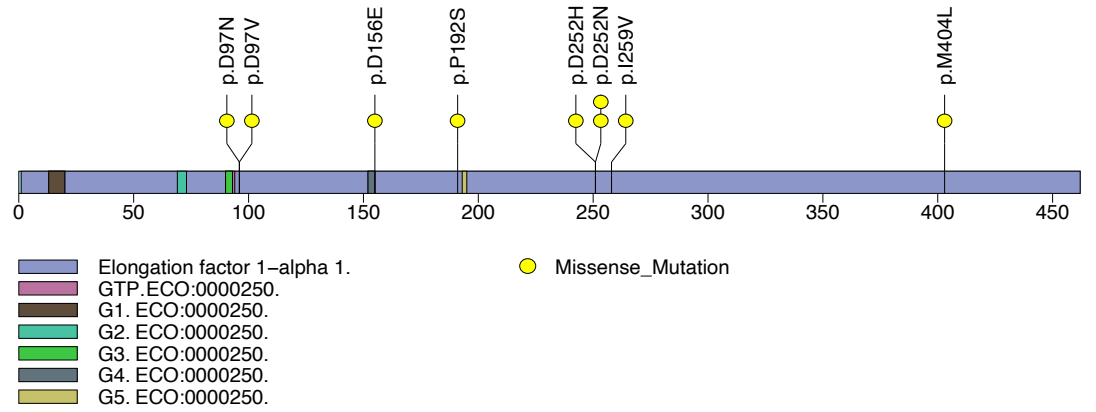
DIS3


p.M1I
p.M1V
p.E22K
p.L67I
p.E81K
p.I85T
p.R101C
p.Y106C
p.K118N
p.Y121C
p.Y156C
p.I174T
p.V198G
p.G249E
p.T262P
p.I275R
p.G347E
p.I349R
p.D403N
p.P412S
p.R467P
p.R467Q
p.D479A
p.D479E
p.D479G
p.D479H
p.D479N
p.D479V
p.D487H
p.D487Y
p.D488A
p.D488G
p.D488H
p.D488N
p.L495P
p.D506A
p.L519S
p.D520V
p.D520Y
p.Y531D
p.L544P
p.C562Y
p.Q594P
p.L656V
p.M662R
p.E665del
p.M667L
p.A675V
p.K676E
p.L688P
p.R689Q
p.H691R
p.H691P
p.H764Y
p.F775L
p.I779F
p.R780G
p.R780K
p.R780S
p.R780T
p.Y782D
p.D784E
p.R789Q
p.R789W
p.L790P
p.L817R
p.R820W
p.V907G
p.K950*
0
100
200
300
400
500
600
700
800
900
Silent
Nonsense_Mutation
Splice_Site
In_Frame_Del
Start_Codon_SNP
Missense_Mutation

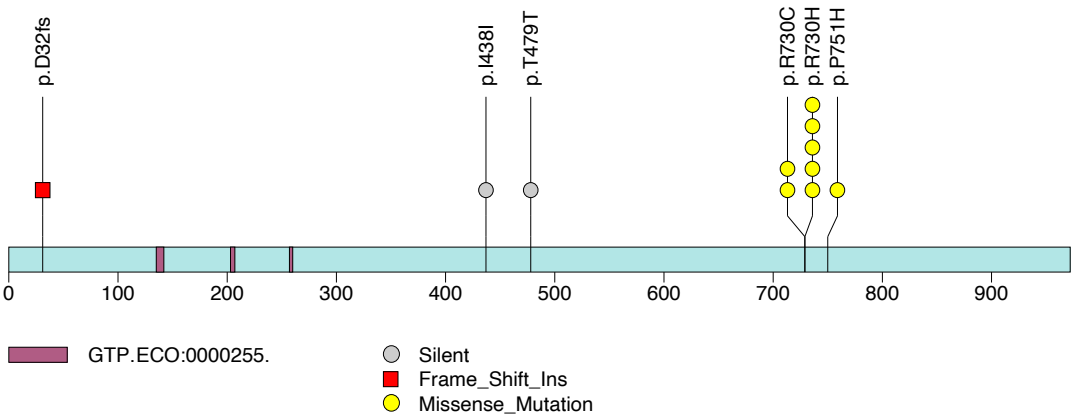
DUSP2



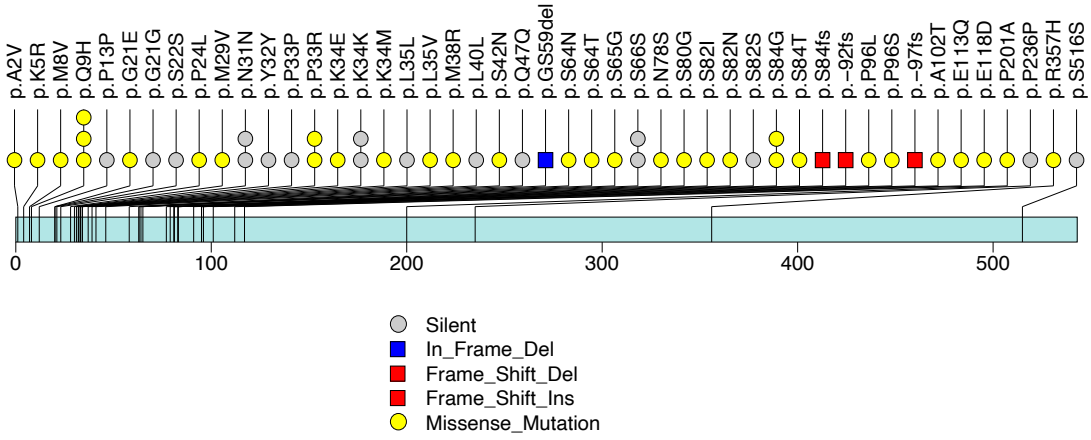
EEF1A1



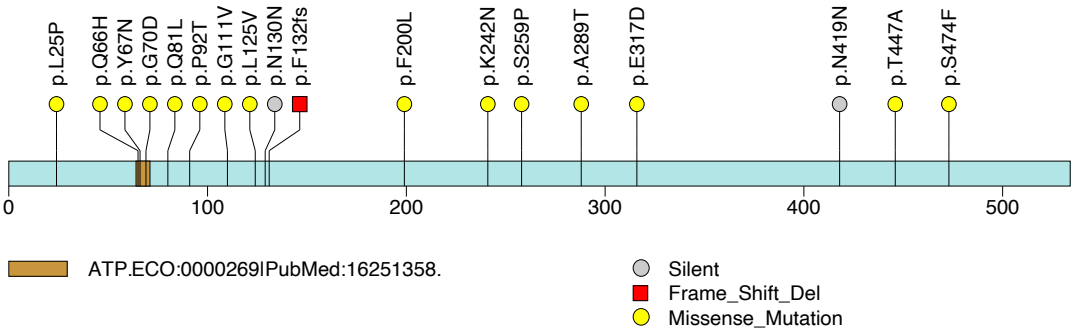
EFTUD2


p.D32fs
p.I438I
p.T479T
p.R730C
p.R730H
p.P751H
0
100
200
300
400
500
600
700
800
900
GTP.ECO:0000255.
Silent
Frame_Shift_Ins
Missense_Mutation

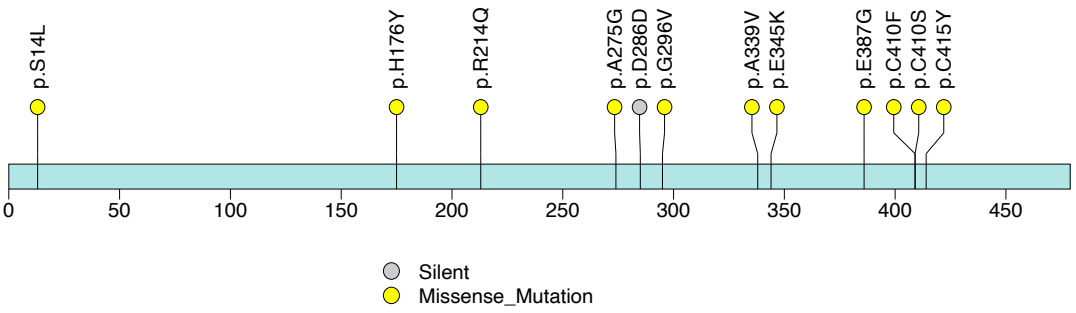
EGR1


p.A2V
p.K5R
p.M8V
p.Q9H
p.P13P
p.G21E
p.G21G
p.S22S
p.P24L
p.M29V
p.N31N
p.Y32Y
p.P33P
p.P33R
p.K34E
p.K34K
p.K34M
p.L35L
p.L35V
p.M38R
p.L40L
p.S42N
p.Q47Q
p.GS59del
p.S64N
p.S64T
p.S65G
p.S66S
p.N78S
p.S80G
p.S82I
p.S82N
p.S82S
p.S84G
p.S84T
p.S84fs
p.-92fs
p.P96L
p.P96S
p.-97fs
p.A102T
p.E113Q
p.E118D
p.P201A
p.P236P
p.R357H
p.S516S
0
100
200
300
400
500
Silent
In_Frame_Del
Frame_Shift_Del
Frame_Shift_Ins
Missense_Mutation

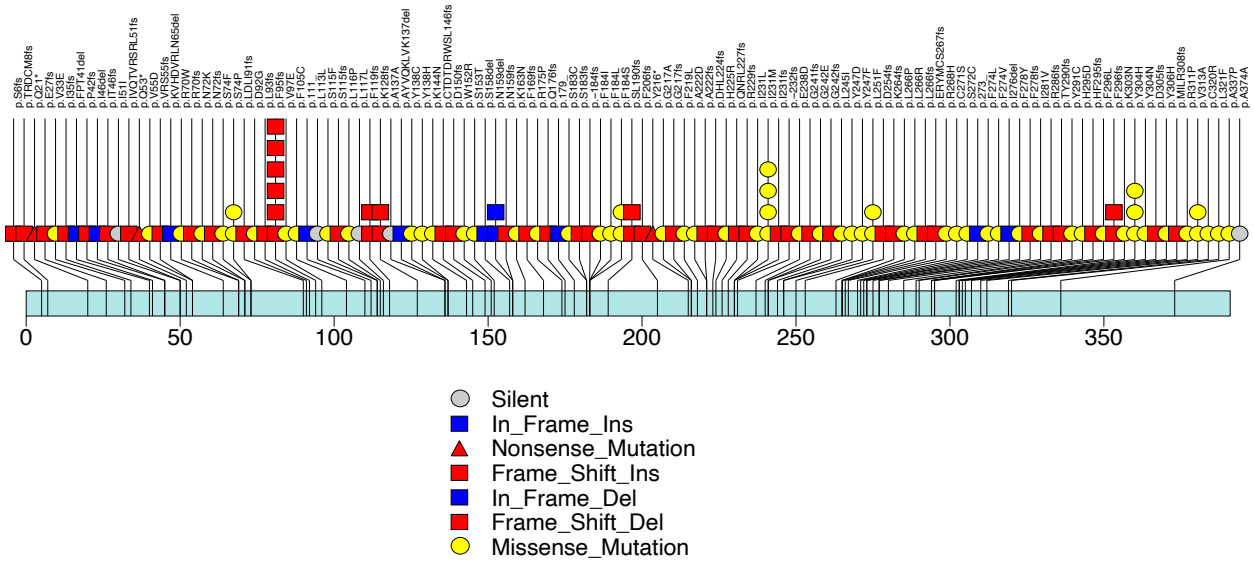
EHD1



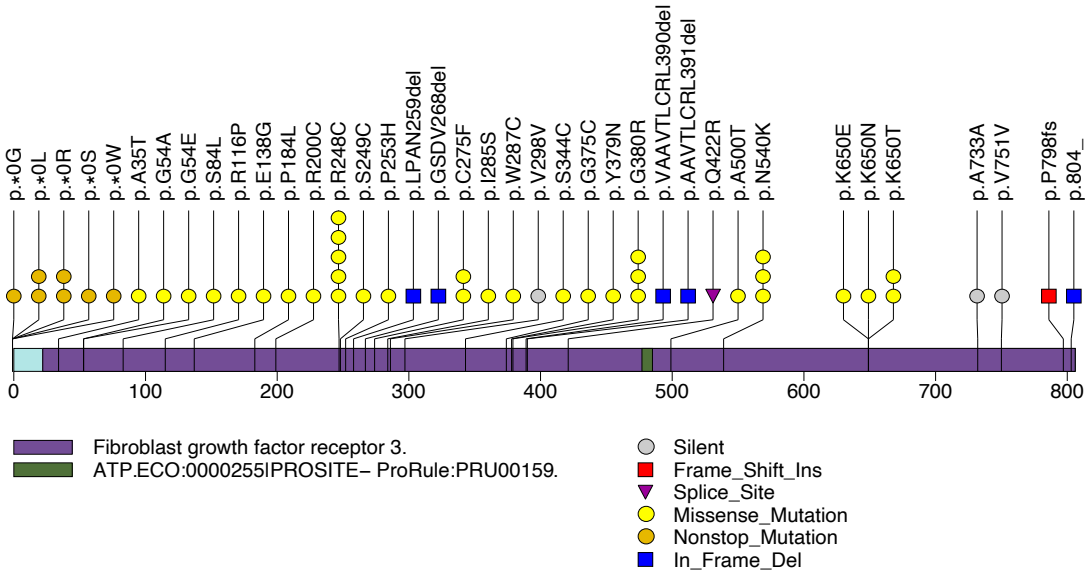
FAM196A



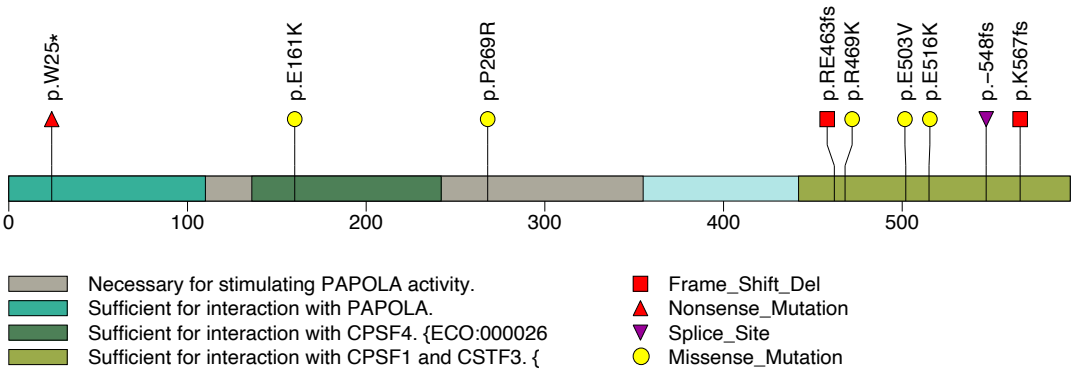
FAM46C



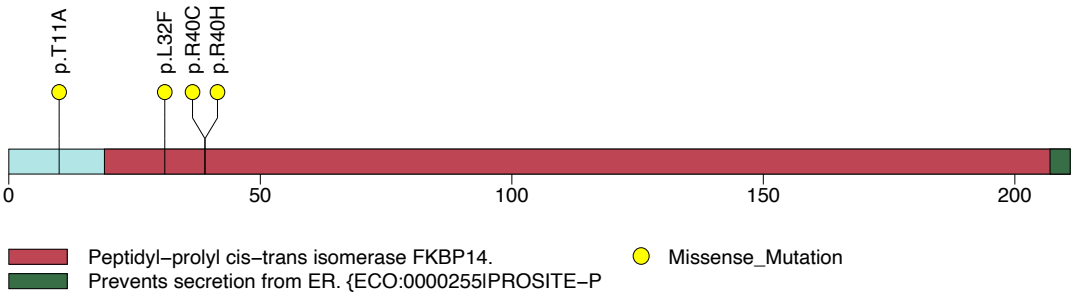
FGFR3



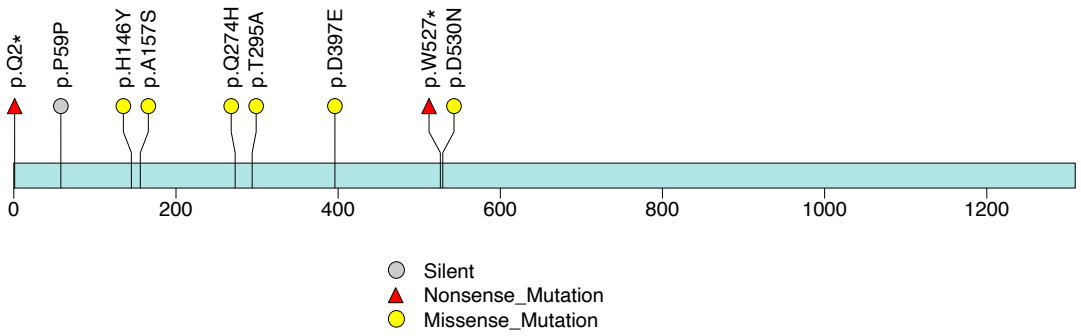
FIP1L1



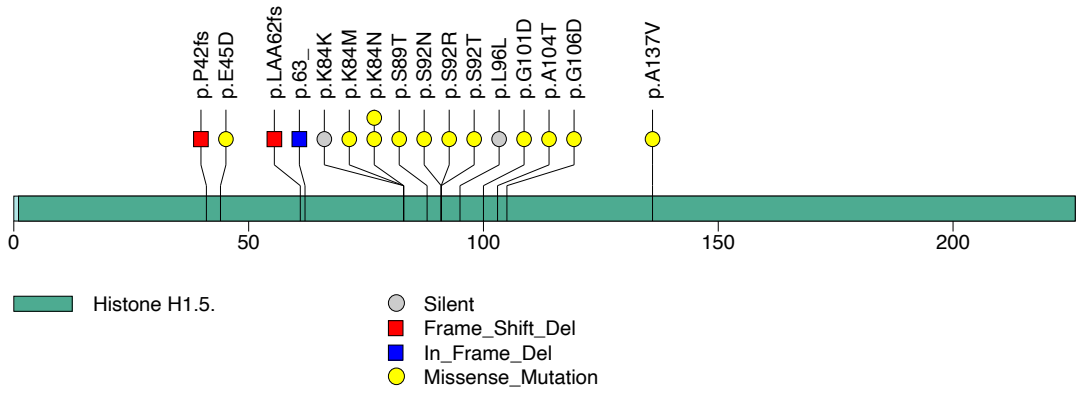
FKBP14



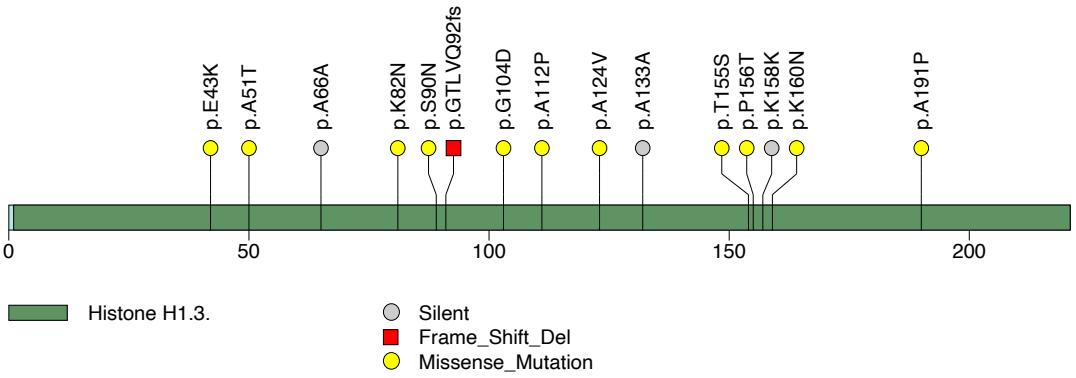
FRMPD2



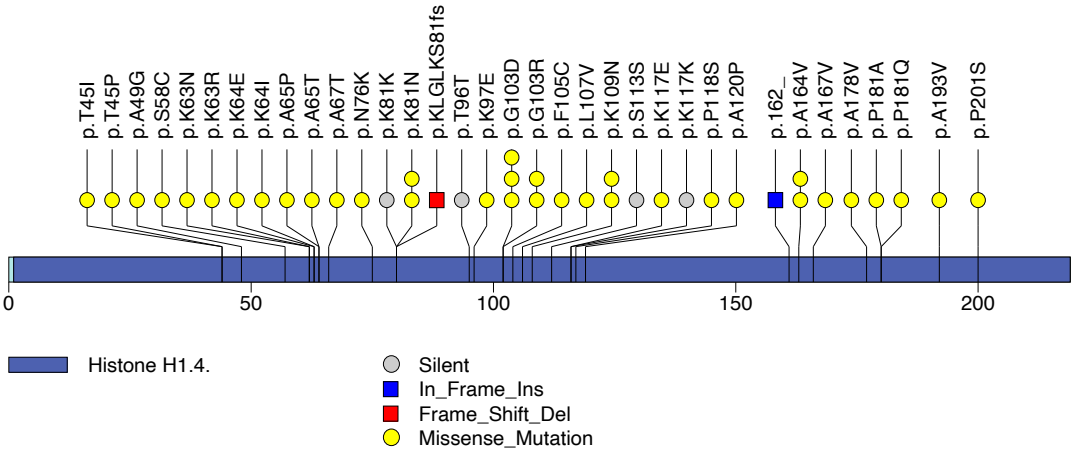
HIST1H1B



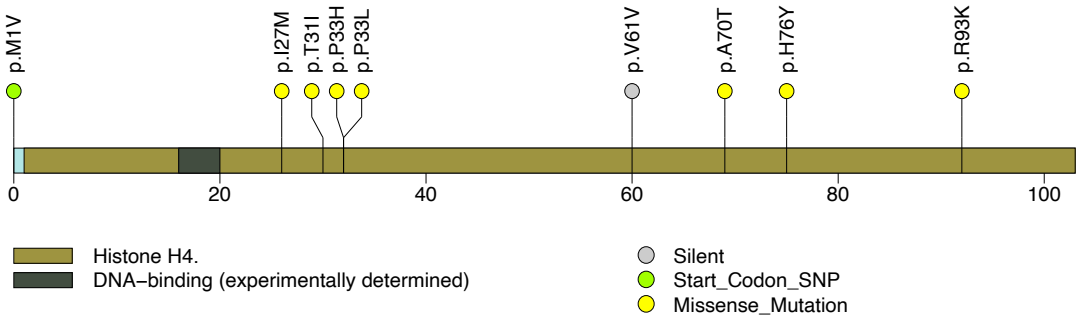
HIST1H1D



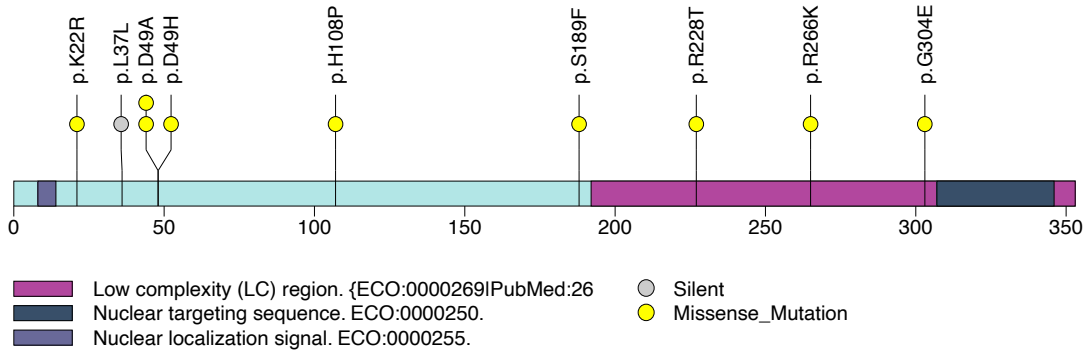
HIST1H1E



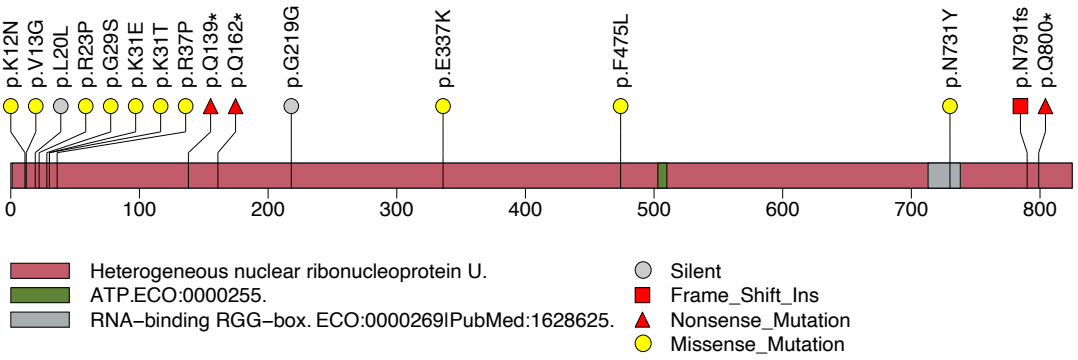
HIST1H4E



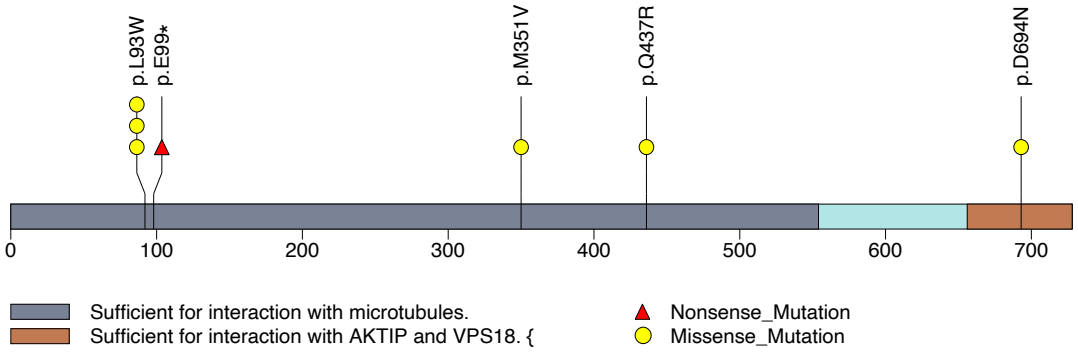
HNRNPA2B1



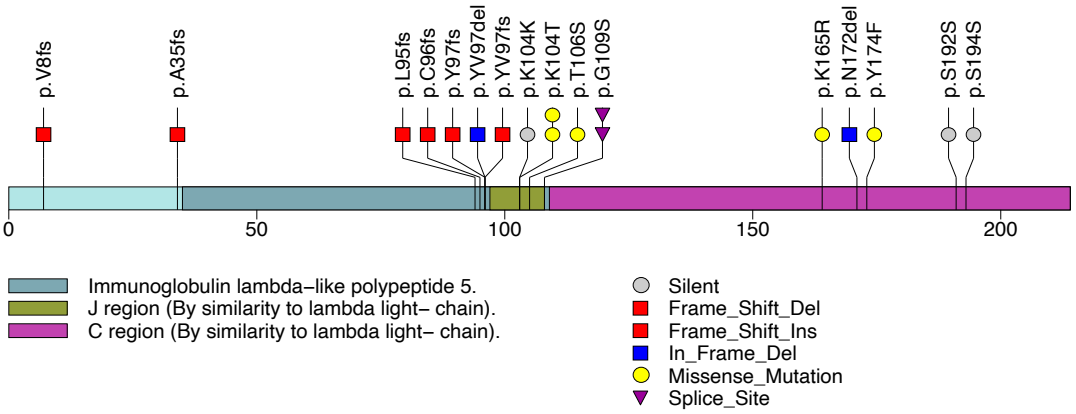
HNRNPU



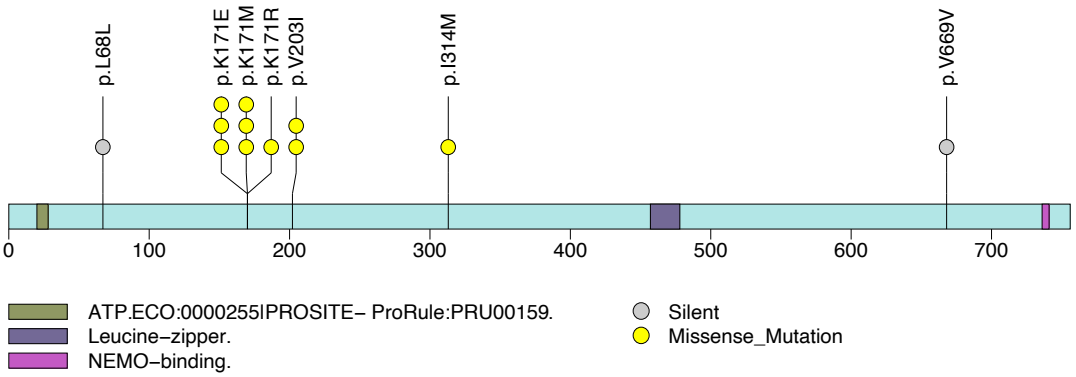
HOOK1



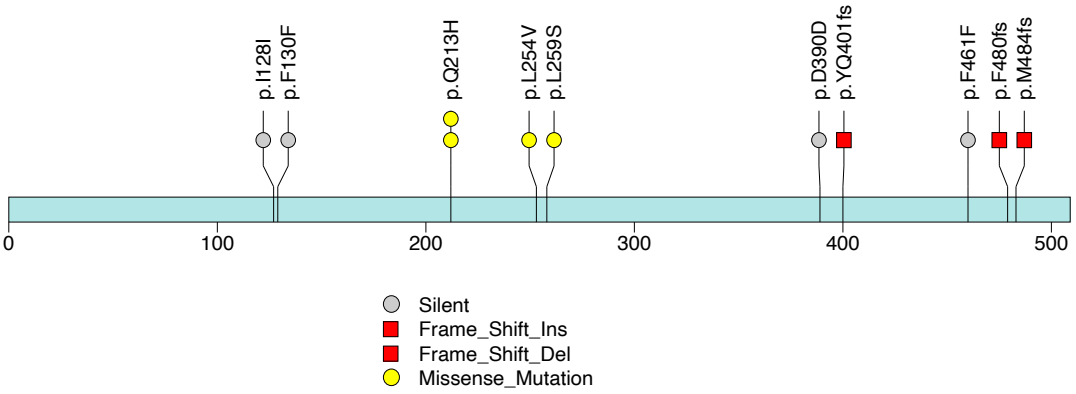
IGLL5


p.V8fs
p.A35fs
p.L95fs
p.C96fs
p.Y97fs
p.YV97del
p.YV97fs
p.K104K
p.K104T
p.T106S
p.G109S
p.K165R
p.N172del
p.Y174F
p.S192S
p.S194S
0
50
100
150
200
Immunoglobulin lambda−like polypeptide 5.
J region (By similarity to lambda light− chain).
C region (By similarity to lambda light− chain).
Silent
Frame_Shift_Del
Frame_Shift_Ins
In_Frame_Del
Missense_Mutation
Splice_Site

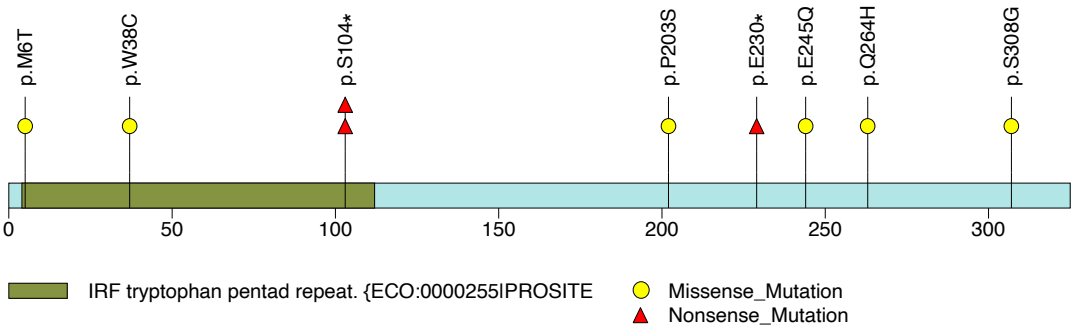
IKBKB


p.L68L
p.K171E
p.K171M
p.K171R
p.V203I
p.I314M
p.V669V
0
100
200
300
400
500
600
700
ATP.ECO:0000255|PROSITE− ProRule:PRU00159.
Leucine−zipper.
NEMO−binding.
Silent
Missense_Mutation

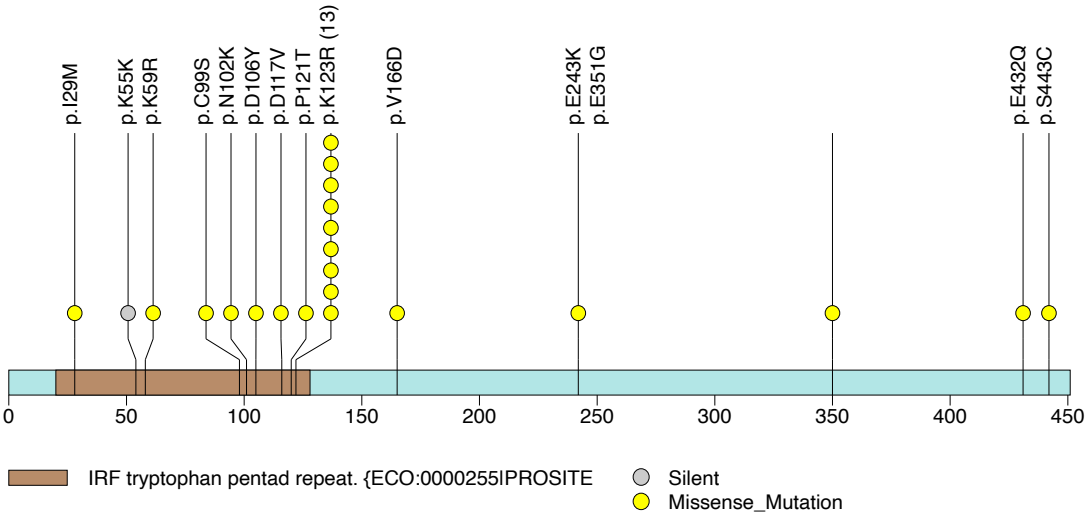
IKZF3


p.I128I
p.F130F
p.Q213H
p.L254V
p.L259S
p.D390D
p.YQ401fs
p.F461F
p.F480fs
p.M484fs
0
100
200
300
400
500
Silent
Frame_Shift_Ins
Frame_Shift_Del
Missense_Mutation

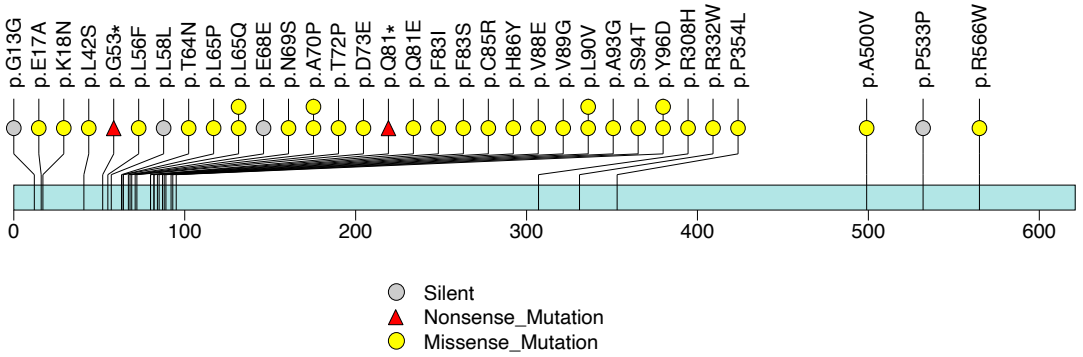
IRF1


p.M6T
p.W38C
p.S104*
p.P203S
p.E230*
p.E245Q
p.Q264H
p.S308G
0
50
100
150
200
250
300
IRF tryptophan pentad repeat. {ECO:0000255|PROSITE
Missense_Mutation
Nonsense_Mutation

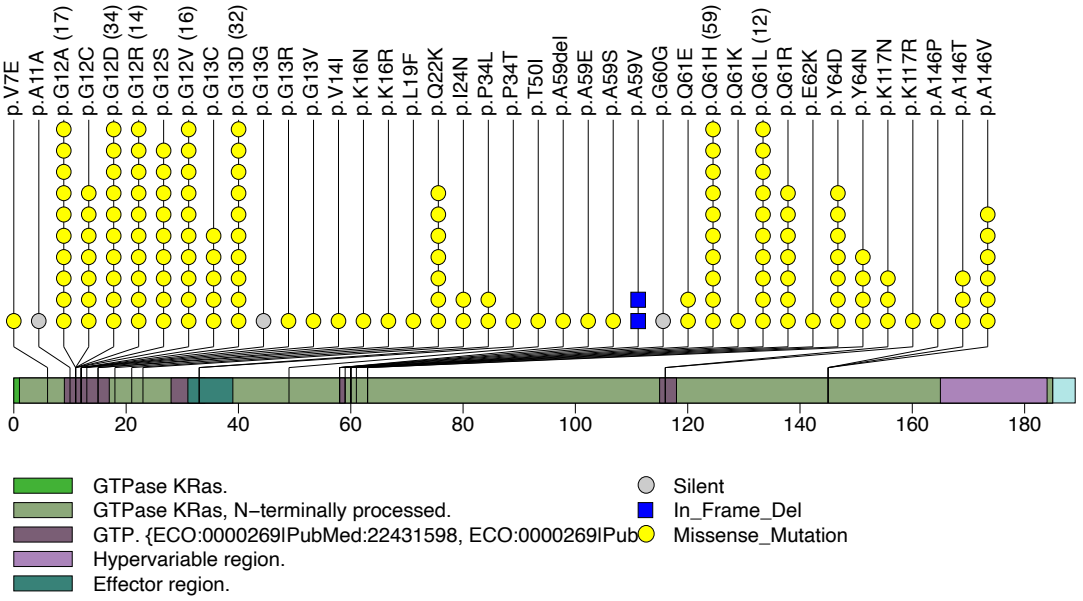
IRF4


p.I29M
p.K55K
p.K59R
p.C99S
p.N102K
p.D106Y
p.D117V
p.P121T
p.K123R (13)
p.V166D
p.E243K
p.E351G
p.E432Q
p.S443C
0
50
100
150
200
250
300
350
400
450
IRF tryptophan pentad repeat. {ECO:0000255|PROSITE
Silent
Missense_Mutation

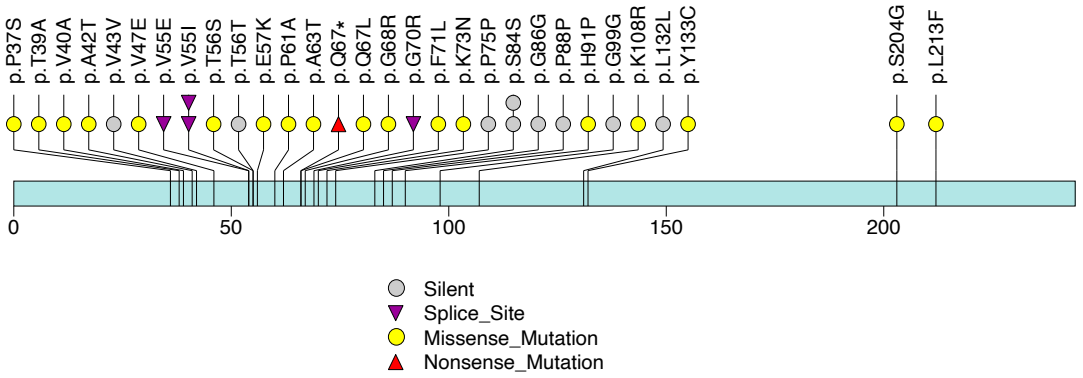
KLHL6


p.G13G
p.E17A
p.K18N
p.L42S
p.G53*
p.L56F
p.L58L
p.T64N
p.L65P
p.L65Q
p.E68E
p.N69S
p.A70P
p.T72P
p.D73E
p.Q81*
p.Q81E
p.F83I
p.F83S
p.C85R
p.H86Y
p.V88E
p.V89G
p.L90V
p.A93G
p.S94T
p.Y96D
p.R308H
p.R332W
p.P354L
p.A500V
p.P533P
p.R566W
0
100
200
300
400
500
600
Silent
Nonsense_Mutation
Missense_Mutation

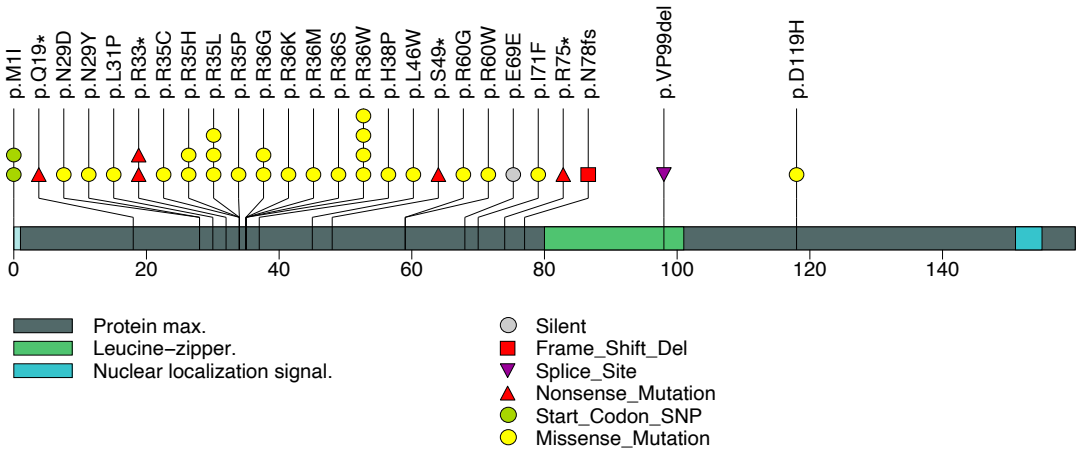
KRAS


p.V7E
p.A11A
p.G12A (17)
p.G12C
p.G12D (34)
p.G12R (14)
p.G12S
p.G12V (16)
p.G13C
p.G13D (32)
p.G13G
p.G13R
p.G13V
p.V14I
p.K16N
p.K16R
p.L19F
p.Q22K
p.I24N
p.P34L
p.P34T
p.T50I
p.A59del
p.A59E
p.A59S
p.A59V
p.G60G
p.Q61E
p.Q61H (59)
p.Q61K
p.Q61L (12)
p.Q61R
p.E62K
p.Y64D
p.Y64N
p.K117N
p.K117R
p.A146P
p.A146T
p.A146V
0
20
40
60
80
100
120
140
160
180
GTPase KRas.
GTPase KRas, N−terminally processed.
GTP. {ECO:0000269|PubMed:22431598, ECO:0000269|Pub
Hypervariable region.
Effector region.
Silent
In_Frame_Del
Missense_Mutation

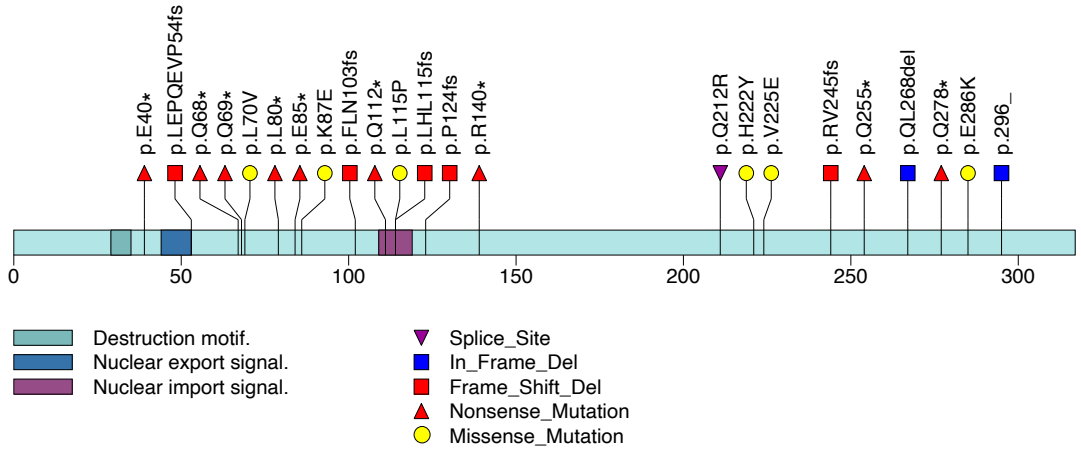
LTB


p.P37S
p.T39A
p.V40A
p.A42T
p.V43V
p.V47E
p.V55E
p.V55I
p.T56S
p.T56T
p.E57K
p.P61A
p.A63T
p.Q67*
p.Q67L
p.G68R
p.G70R
p.F71L
p.K73N
p.P75P
p.S84S
p.G86G
p.P88P
p.H91P
p.G99G
p.K108R
p.L132L
p.Y133C
p.S204G
p.L213F
0
50
100
150
200
Silent
Splice_Site
Missense_Mutation
Nonsense_Mutation

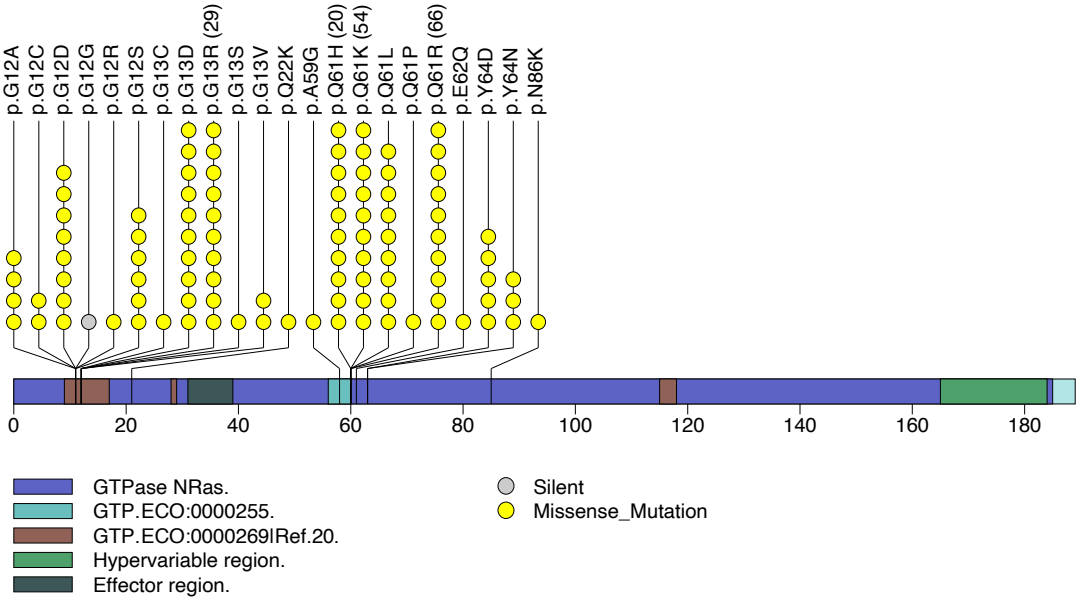
MAX


p.M1I
p.Q19*
p.N29D
p.N29Y
p.L31P
p.R33*
p.R35C
p.R35H
p.R35L
p.R35P
p.R36G
p.R36K
p.R36M
p.R36S
p.R36W
p.H38P
p.L46W
p.S49*
p.R60G
p.R60W
p.E69E
p.I71F
p.R75*
p.N78fs
p.VP99del
p.D119H
0
20
40
60
80
100
120
140
Protein max.
Leucine−zipper.
Nuclear localization signal.
Silent
Frame_Shift_Del
Splice_Site
Nonsense_Mutation
Start_Codon_SNP
Missensе_Mutation

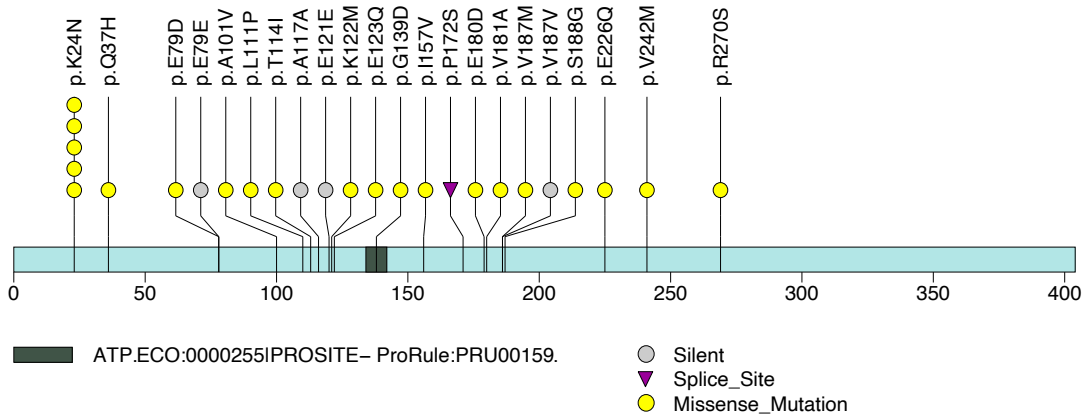
NFKBIA


p.E40*
p.LEPQEVP54fs
p.Q68*
p.Q69*
p.L70V
p.L80*
p.E85*
p.K87E
p.FLN103fs
p.Q112*
p.L115P
p.LHL115fs
p.P124fs
p.R140*
p.Q212R
p.H222Y
p.V225E
p.RV245fs
p.Q255*
p.QL268del
p.Q278*
p.E286K
p.296_
0
50
100
150
200
250
300
Destruction motif.
Nuclear export signal.
Nuclear import signal.
Splice_Site
In_Frame_Del
Frame_Shift_Del
Nonsense_Mutation
Missensе_Mutation

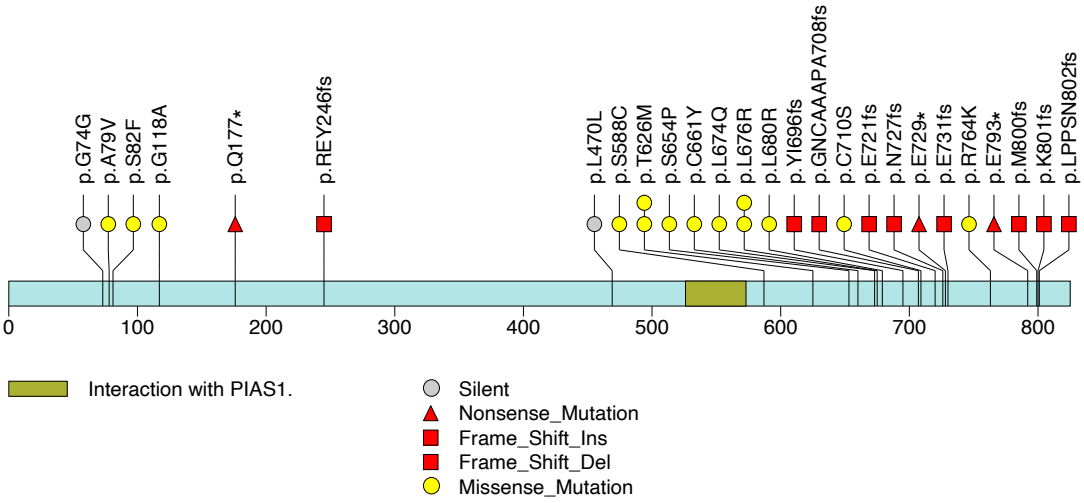
NRAS


p.G12A
p.G12C
p.G12D
p.G12G
p.G12R
p.G12S
p.G13C
p.G13D
p.G13R (29)
p.G13S
p.G13V
p.Q22K
p.A59G
p.Q61H (20)
p.Q61K (54)
p.Q61L
p.Q61P
p.Q61R (66)
p.E62Q
p.Y64D
p.Y64N
p.N86K
0
20
40
60
80
100
120
140
160
180
GTPase NRas.
GTP.ECO:0000255.
GTP.ECO:0000269IRef.20.
Hypervariable region.
Effector region.
Silent
Missense_Mutation

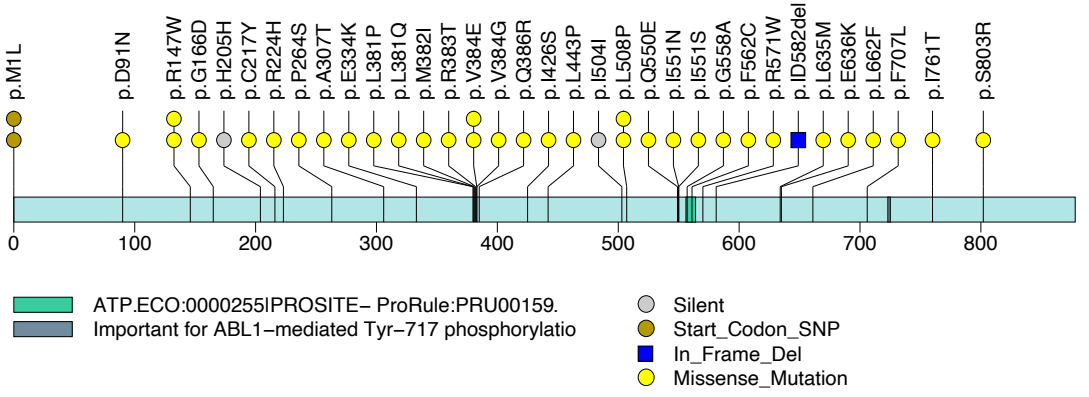
PIM1


p.K24N
p.Q37H
p.E79D
p.E79E
p.A101V
p.L111P
p.T114I
p.A117A
p.E121E
p.K122M
p.E123Q
p.G139D
p.I157V
p.P172S
p.E180D
p.V181A
p.V187M
p.V187V
p.S188G
p.E226Q
p.V242M
p.R270S
0
50
100
150
200
250
300
350
400
ATP.ECO:0000255IPROSITE– ProRule:PRU00159.
Silent
Splice_Site
Missense_Mutation

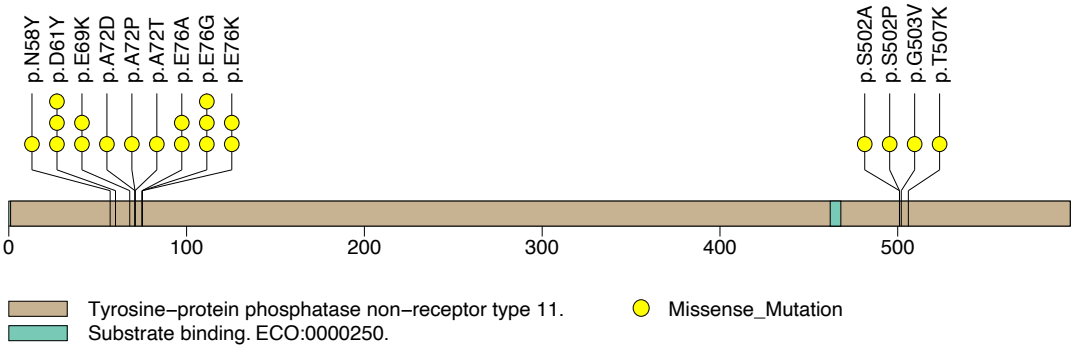
PRDM1


p.G74G
p.A79V
p.S82F
p.G118A
p.Q177*
p.REY246fs
p.L470L
p.S588C
p.T626M
p.S654P
p.C661Y
p.L674Q
p.L676R
p.L680R
p.YI696fs
p.GNCAAAPA708fs
p.C710S
p.E721fs
p.N727fs
p.E729*
p.E731fs
p.R764K
p.E793*
p.M800fs
p.K801fs
p.LPPSN802fs
0
100
200
300
400
500
600
700
800
Interaction with PIAS1.
Silent
Nonsense_Mutation
Frame_Shift_Ins
Frame_Shift_Del
Missense_Mutation

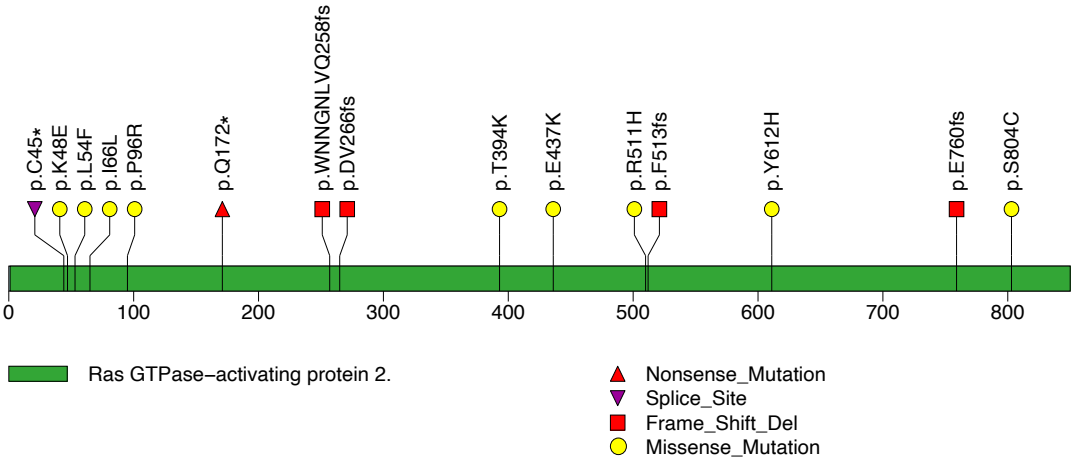
PRKD2


p.M1L
p.D91N
p.R147W
p.G166D
p.H205H
p.C217Y
p.R224H
p.P264S
p.A307T
p.E334K
p.L381P
p.L381Q
p.M382I
p.R383T
p.V384E
p.V384G
p.Q386R
p.I426S
p.L443P
p.I504I
p.L508P
p.Q550E
p.I551N
p.I551S
p.G558A
p.F562C
p.R571W
p.ID582del
p.L635M
p.E636K
p.L662F
p.F707L
p.I761T
p.S803R
0
100
200
300
400
500
600
700
800
ATP.ECO:0000255|PROSITE−ProRule:PRU00159.
Important for ABL1−mediated Tyr−717 phosphorylatio
Silent
Start_Codon_SNP
In_Frame_Del
Missense_Mutation

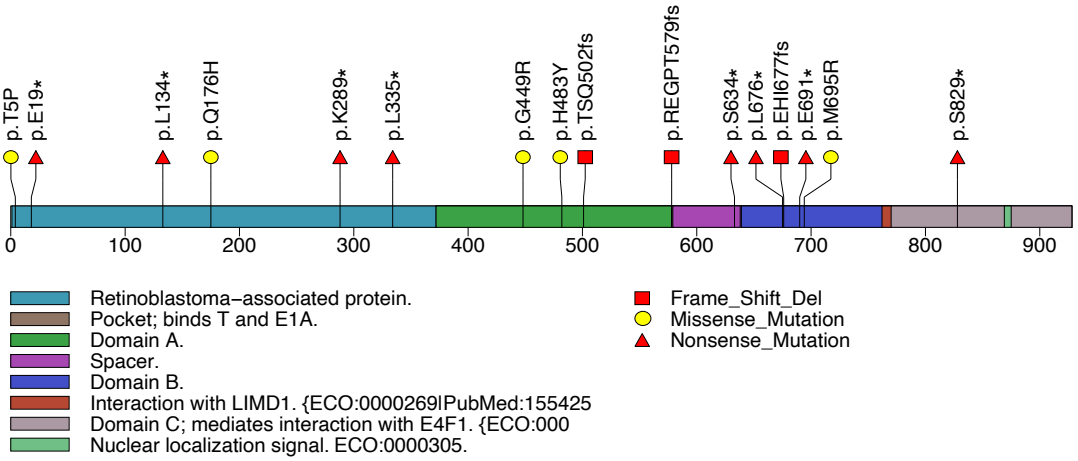
PTPN11



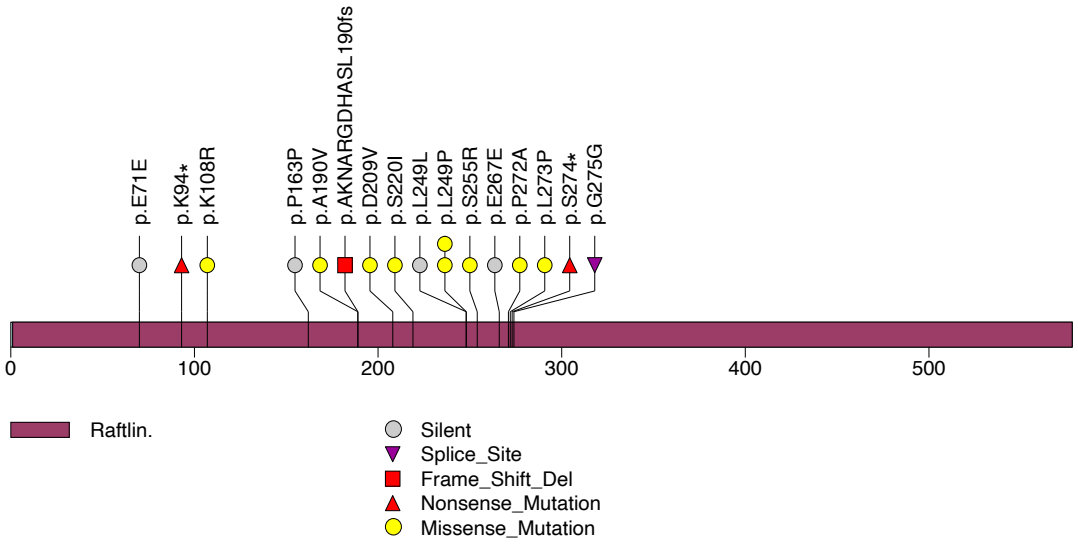
RASA2



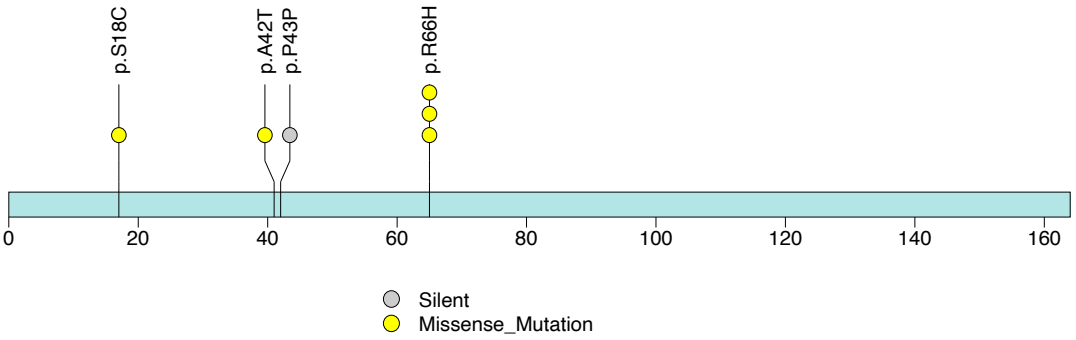
RB1


p.T5P
p.E19*
p.L134*
p.Q176H
p.K289*
p.L335*
p.G449R
p.H483Y
p.TSQ502fs
p.REGPT579fs
p.S634*
p.L676*
p.EHI677fs
p.E691*
p.M695R
p.S829*
0
100
200
300
400
500
600
700
800
900
Retinoblastoma–associated protein.
Pocket; binds T and E1A.
Domain A.
Spacer.
Domain B.
Interaction with LIMD1. {ECO:0000269|PubMed:155425
Domain C; mediates interaction with E4F1. {ECO:000
Nuclear localization signal. ECO:0000305.
Frame_Shift_Del
Missense_Mutation
Nonsense_Mutation

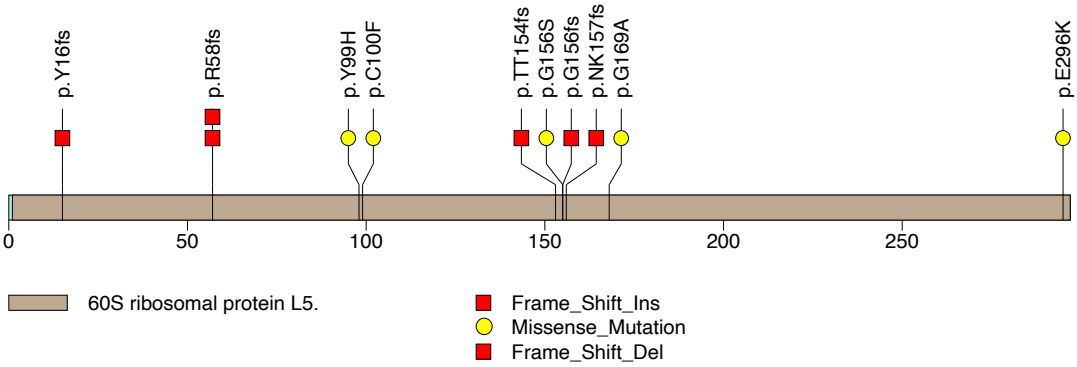
RFTN1


p.E71E
p.K94*
p.K108R
p.P163P
p.A190V
p.AKNARGDHASL190fs
p.D209V
p.S220I
p.L249L
p.L249P
p.S255R
p.E267E
p.P272A
p.L273P
p.S274*
p.G275G
0
100
200
300
400
500
Raftlin.
Silent
Splice_Site
Frame_Shift_Del
Nonsense_Mutation
Missense_Mutation

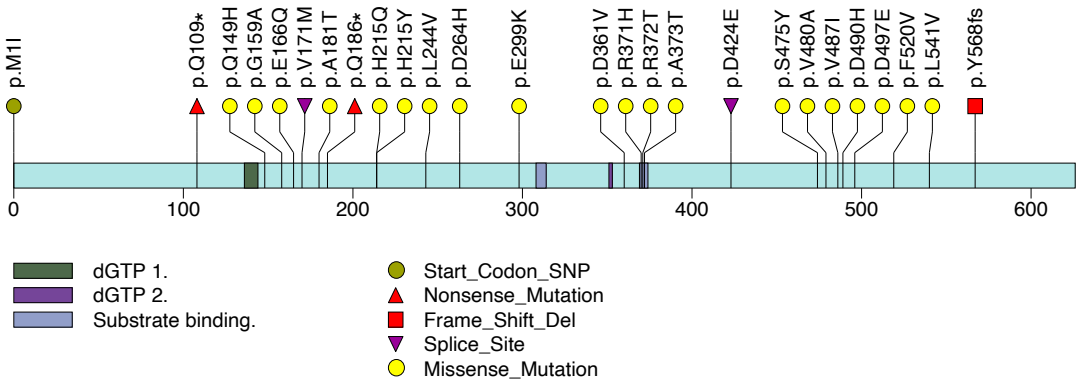
RNASEH2C



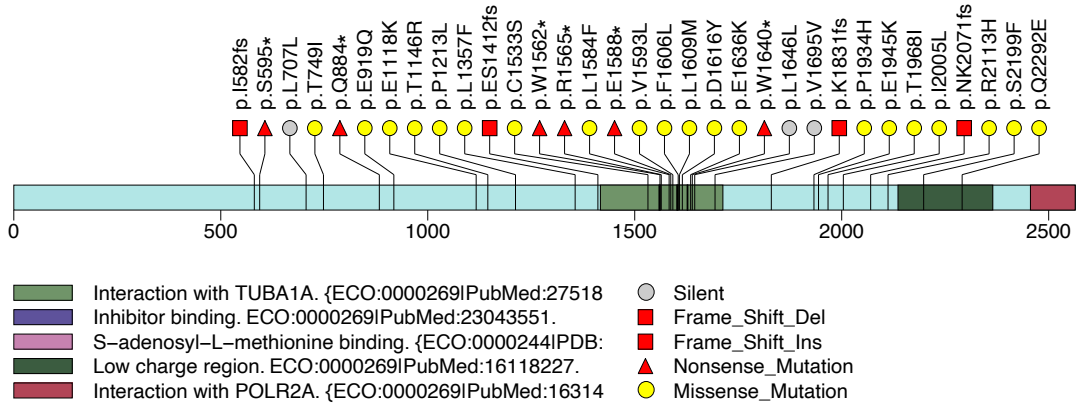
RPL5



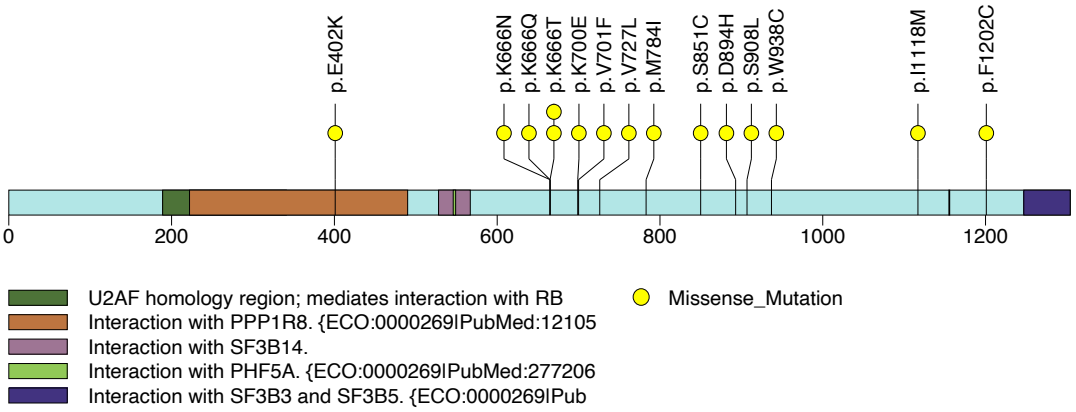
SAMHD1



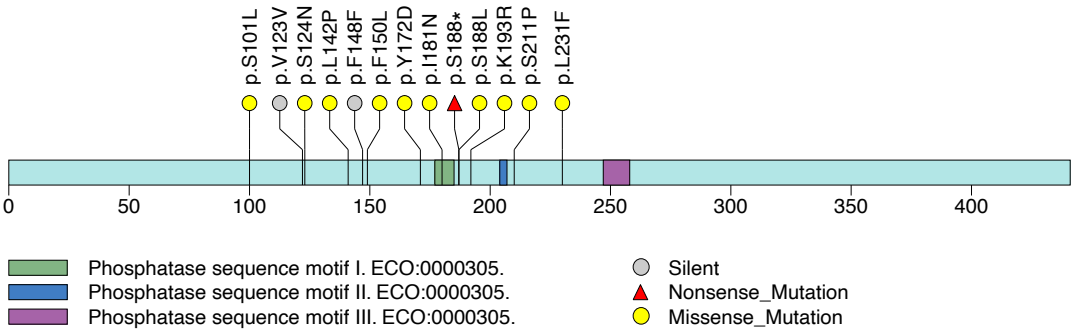
SETD2



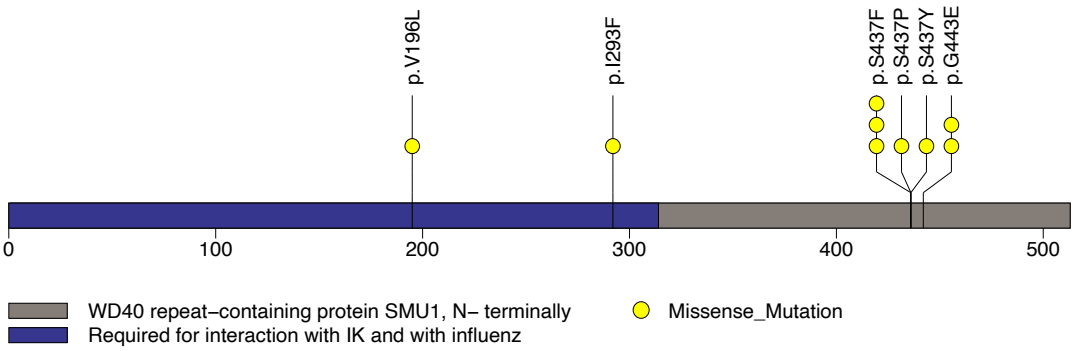
SF3B1


p.E402K
p.K666N
p.K666Q
p.K666T
p.K700E
p.V701F
p.V727L
p.M784I
p.S851C
p.D894H
p.S908L
p.W938C
p.I1118M
p.F1202C
0
200
400
600
800
1000
1200
U2AF homology region; mediates interaction with RB
Interaction with PPP1R8. {ECO:0000269IPubMed:12105
Interaction with SF3B14.
Interaction with PHF5A. {ECO:0000269IPubMed:277206
Interaction with SF3B3 and SF3B5. {ECO:0000269IPub
Missense_Mutation

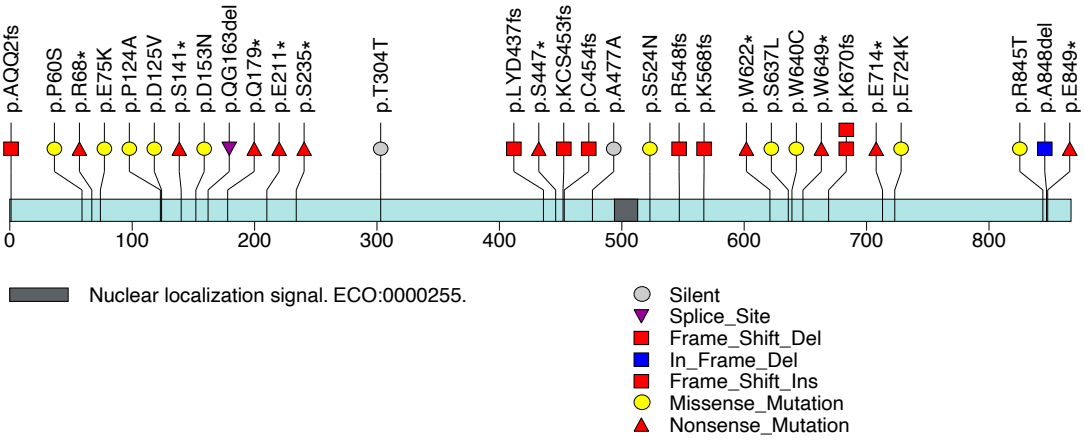
SGPP1


p.S101L
p.V123V
p.S124N
p.L142P
p.F148F
p.F150L
p.Y172D
p.I181N
p.S188*
p.S188L
p.K193R
p.S211P
p.L231F
0
50
100
150
200
250
300
350
400
Phosphatase sequence motif I. ECO:0000305.
Phosphatase sequence motif II. ECO:0000305.
Phosphatase sequence motif III. ECO:0000305.
Silent
Nonsense_Mutation
Missense_Mutation

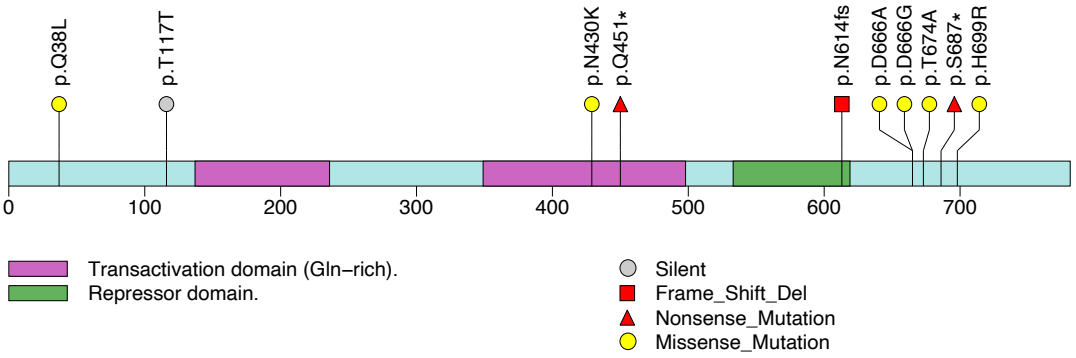
SMU1


p.V196L
p.I293F
p.S437F
p.S437P
p.S437Y
p.G443E
0
100
200
300
400
500
WD40 repeat–containing protein SMU1, N– terminally
Required for interaction with IK and with influenz
Missense_Mutation

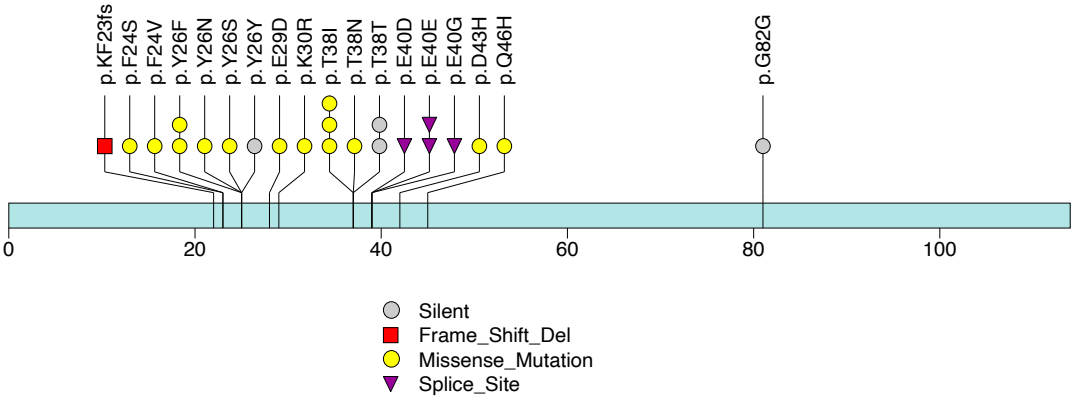
SP140


p.AQQ2fs
p.P60S
p.R68*
p.E75K
p.P124A
p.D125V
p.S141*
p.D153N
p.QG163del
p.Q179*
p.E211*
p.S235*
p.T304T
p.LYD437fs
p.S447*
p.KCS453fs
p.C454fs
p.A477A
p.S524N
p.R548fs
p.K568fs
p.W622*
p.S637L
p.W640C
p.W649*
p.K670fs
p.E714*
p.E724K
p.R845T
p.A848del
p.E849*
0
100
200
300
400
500
600
700
800
Nuclear localization signal. ECO:0000255.
Silent
Splice_Site
Frame_Shift_Del
In_Frame_Del
Frame_Shift_Ins
Missense_Mutation
Nonsense_Mutation

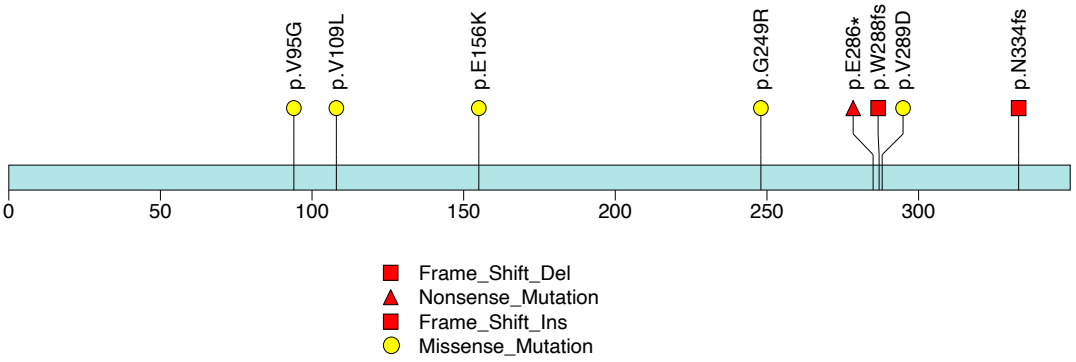
SP3


p.Q38L
p.T117T
p.N430K
p.Q451*
p.N614fs
p.D666A
p.D666G
p.T674A
p.S687*
p.H699R
0
100
200
300
400
500
600
700
Transactivation domain (Gln−rich).
Repressor domain.
Silent
Frame_Shift_Del
Nonsense_Mutation
Missense_Mutation

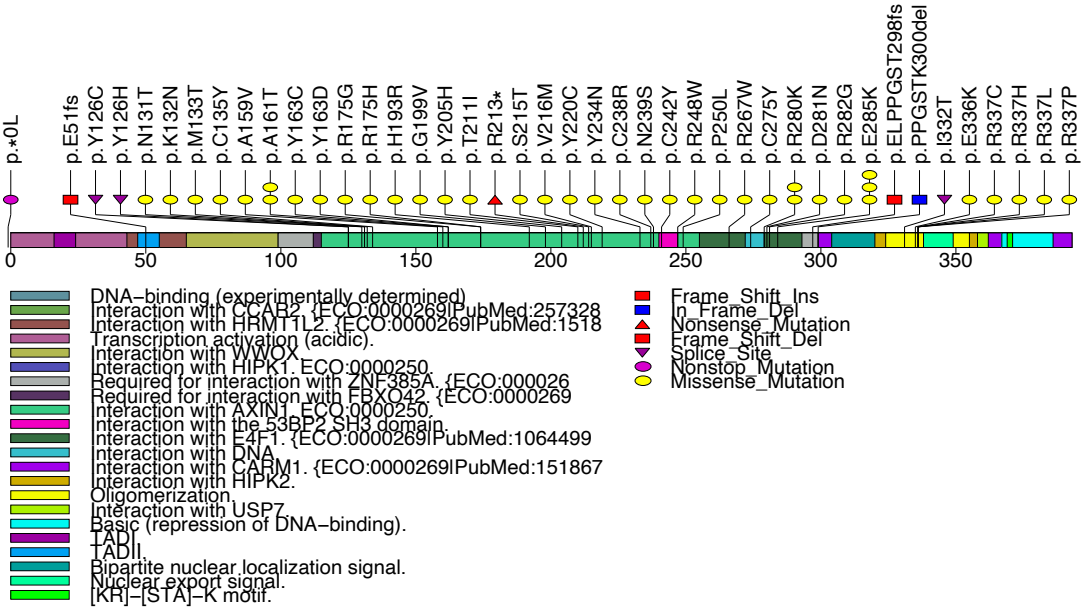
TCL1A


p.KF23fs
p.F24S
p.F24V
p.Y26F
p.Y26N
p.Y26S
p.Y26Y
p.E29D
p.K30R
p.T38I
p.T38N
p.T38T
p.E40D
p.E40E
p.E40G
p.D43H
p.Q46H
p.G82G
0
20
40
60
80
100
Silent
Frame_Shift_Del
Missense_Mutation
Splice_Site

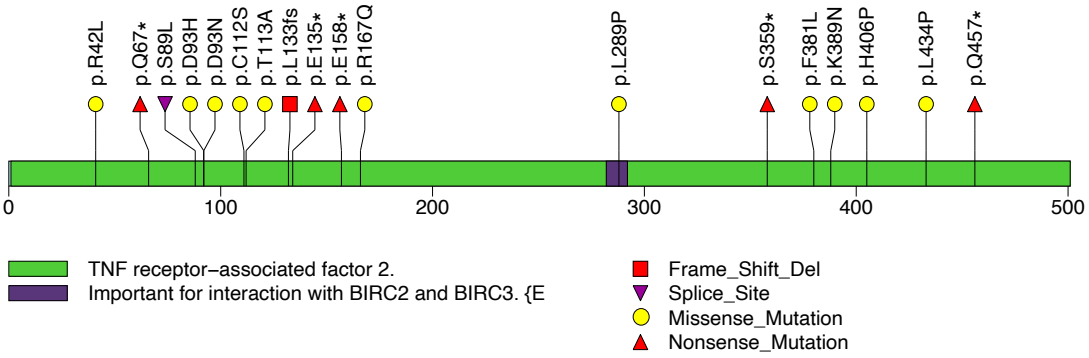
TGDS


p.V95G
p.V109L
p.E156K
p.G249R
p.E286*
p.W288fs
p.V289D
p.N334fs
0
50
100
150
200
250
300
Frame_Shift_Del
Nonsense_Mutation
Frame_Shift_Ins
Missense_Mutation

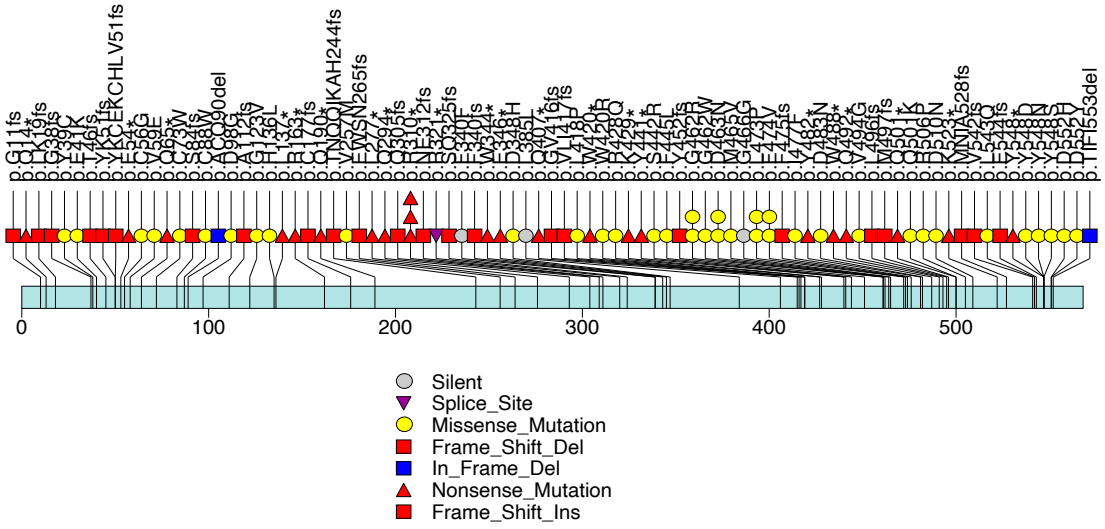
TP53


p.*0L
p.E51fs
p.Y126C
p.Y126H
p.N131T
p.K132N
p.M133T
p.C135Y
p.A159V
p.A161T
p.Y163C
p.Y163D
p.R175G
p.R175H
p.H193R
p.G199V
p.Y205H
p.T211I
p.R213*
p.S215T
p.V216M
p.Y220C
p.Y234N
p.C238R
p.N239S
p.C242Y
p.R248W
p.P250L
p.R267W
p.C275Y
p.R280K
p.D281N
p.R282G
p.E285K
p.ELPPGST298fs
p.PPGSTK300del
p.I332T
p.E336K
p.R337C
p.R337H
p.R337L
p.R337P
0
50
100
150
200
250
300
350
DNA−binding (experimentally determined)
Interaction with CCAR2. {ECO:0000269|PubMed:257328
Interaction with HRMT1L2. {ECO:0000269|PubMed:1518
Transcription activation (acidic).
Interaction with WWOX.
Interaction with HIPK1. ECO:0000250.
Required for interaction with ZNF385A. {ECO:000026
Required for interaction with FBXO42. {ECO:0000269
Interaction with AXIN1. ECO:0000250.
Interaction with the 53BP2 SH3 domain.
Interaction with E4F1. {ECO:0000269|PubMed:1064499
Interaction with DNA.
Interaction with CABIN1. {ECO:0000269|PubMed:151867
Interaction with HIPK2.
Oligomerization.
Interaction with USP7.
Basic (repression of DNA−binding).
TADI.
TADII.
Bipartite nuclear localization signal.
Nuclear export signal.
[KR]−[STA]−K motif.
Frame_Shift_Ins
In_Frame_Del
Nonsense_Mutation
Frame_Shift_Del
Splice_Site
Nonstop_Mutation
Missense_Mutation

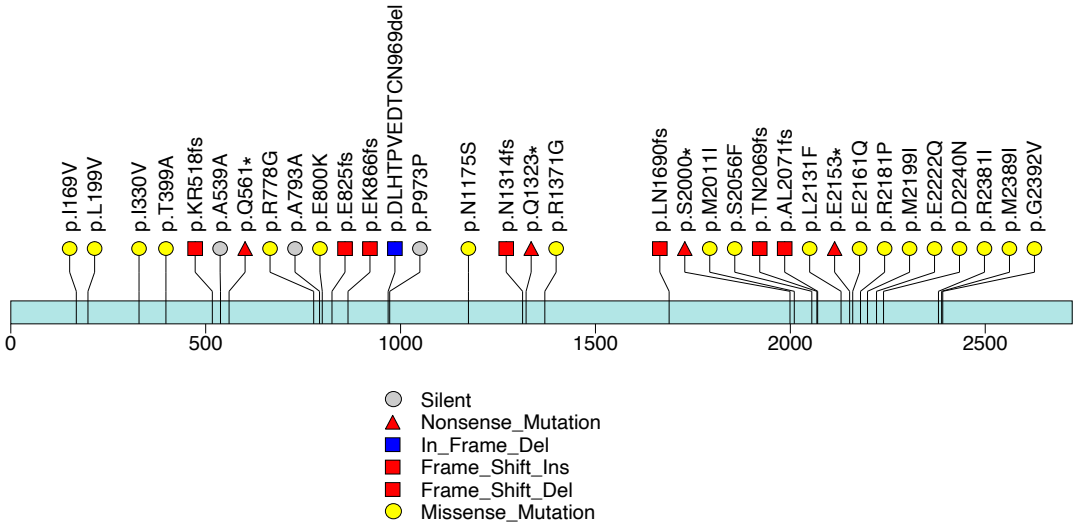
TRAF2


p.R42L
p.Q67*
p.S89L
p.D93H
p.D93N
p.C112S
p.T113A
p.L133fs
p.E135*
p.E158*
p.R167Q
p.L289P
p.S359*
p.F381L
p.K389N
p.H406P
p.L434P
p.Q457*
0
100
200
300
400
500
TNF receptor−associated factor 2.
Important for interaction with BIRC2 and BIRC3. {E
Frame_Shift_Del
Splice_Site
Missensе_Mutation
Nonsense_Mutation

TRAF3


0
100
200
300
400
500
Silent
Splice_Site
Missense_Mutation
Frame_Shift_Del
In_Frame_Del
Nonsense_Mutation
Frame_Shift_Ins

ZNF292



Supplementary Figure 16 *"Lollipop" figures representing protein changes (reference UniProt sequence) found in our study. For each gene, the x-axis represents the position of the amino acid on the reference protein sequence, and the exact same mutations (position and amino acid change) are piled up. For mutations with more than ten occurrences in the cohort, the exact number was added between parenthesis after the protein change annotation (e.g., "NRAS Q61R (66)" was found in 66 unique participants).*

Additional references

1. Tarabichi, M. *et al.* A practical guide to cancer subclonal reconstruction from DNA sequencing. *Nat Methods* **18**, 144-155 (2021).
2. Dutta, A.K. *et al.* MinimuMM-seq: Genome Sequencing of Circulating Tumor Cells for Minimally Invasive Molecular Characterization of Multiple Myeloma Pathology. *Cancer Discov* **13**, 348-363 (2023).
3. Bustoros, M. *et al.* Genomic Profiling of Smoldering Multiple Myeloma Identifies Patients at a High Risk of Disease Progression. *Journal of Clinical Oncology* **38**, 2380-2390 (2020).
4. Mermel, C.H. *et al.* GISTIC2. 0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology* **12**, 1-14 (2011).
5. Bradley, R.A. & Terry, M.E. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* **39**, 324-345 (1952).
6. Turner, H. & Firth, D. Bradley-Terry Models in R: The BradleyTerry2 Package. *Journal of Statistical Software* **48**, 1 - 21 (2012).
7. Jolly, C. & Van Loo, P. Timing somatic events in the evolution of cancer. *Genome Biology* **19**, 95 (2018).
8. Thakurta, A. *et al.* High subclonal fraction of 17p deletion is associated with poor prognosis in multiple myeloma. *Blood, The Journal of the American Society of Hematology* **133**, 1217-1221 (2019).
9. Manier, S. *et al.* Genomic complexity of multiple myeloma and its clinical implications. *Nature Reviews Clinical Oncology* **14**, 100-113 (2017).
10. Christensen, S. *et al.* 5-Fluorouracil treatment induces characteristic T> G mutations in human cancer. *Nature communications* **10**, 4571 (2019).
11. Pich, O. *et al.* The mutational footprints of cancer therapies. *Nature genetics* **51**, 1732-1740 (2019).
12. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
13. Machado, H.E. *et al.* Diverse mutational landscapes in human lymphocytes. *Nature* **608**, 724-732 (2022).
14. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57 (2012).
15. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959-61 (2013).
16. Huang, F.W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-9 (2013).
17. Rheinbay, E. *et al.* Recurrent and functional regulatory mutations in breast cancer. *Nature* **547**, 55-60 (2017).
18. Rheinbay, E. *et al.* Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102-111 (2020).
19. Sherman, M.A. *et al.* Genome-wide mapping of somatic mutation rates uncovers drivers of cancer. *Nat Biotechnol* **40**, 1634-1643 (2022).
20. Buisson, R. *et al.* Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* **364**, eaaw2872 (2019).