Supplemental information

Classification

F94Y, originating from a Glioma tumor, marked the sole mutation in POLE or POLD1 within that sample, boasting 16 SNV variants and a PPD + MMRd ratio of 0.33. E978G surfaced in two GENIE Glioma samples, both registering as positive cases. In one instance, it stood as the solitary POLE or POLD1 clonal mutation, while in the other, it contended with a less frequent mutation. Notably, both samples exhibited an SNV count exceeding 400.

Mis-classified tumors

During the analysis we identified 21 tumors that harbor a clonal driver mutation, but have not shown the PPD phenotype. Of those 0 MMRp and 9 MMRd tumors had more than 10 SNVs but did not present high activity of the PPD signatures. In addition we found 12 MMRp cases with less than 10 SNVs that harbored driver variants including 4 of the most frequent POLE variants (5 P286R, 5 V411L,1 A456P and 1 S297F). Of those, 9 were from a specific sequencing assay (NKI-CHPV2-SOCV2-NGS) that had a very small sequencing target which can explain their misclassification. Of the remaining 3, 1 had very low purity, while the issue with the other two is not clear. .

We searched for tumors that present PPD phenotype, but had no POLE/D1 mutation. To achieve that we applied the same SVM tool to the tumors without any POLE/D1 mutation (Fig. S13), and found 7 tumors that were classified as PPD (Another one is in lung cancer). In addition we found 2 PPD tumors that had a mutation in POLE/D1, but did not have a clonal driver mutation. The explanation for these 9 tumors could be that they harbor a germline mutation that was not detected in the somatic sequencing data in GENIE. An alternative explanation could be a technical artifact during the sequencing or mutation calling.

In our analysis, we identified 612 samples with a subclonal POLE/D1 mutation but without a clonal mutation. Among these, we found 4 MMRp cases that could be classified as potential PPD samples. 3 of these cases included a known driver mutation, while 1 exhibited an unfamiliar driver (*POLE* p.F367C). In addition we found 3 MMRd, we identified 3 tumors, one with a known driver mutation and two with unfamiliar *POLE* variants.

TCGA dataset

While validating with the TCGA dataset, we initially identified two variants in the ExoD of *POLE*. One of these variants was the R705W variant. The other mutation, L1235I, was found in three samples within the TCGA dataset. Among these samples, two exhibited a known driver mutation (P286R or V411L), while in the third sample, L1235I was the only *POLE/D1* mutation present. After careful consideration, we determined that the evidence was insufficient to classify L1235I as a driver mutation, and therefore it was classified as non-driver.

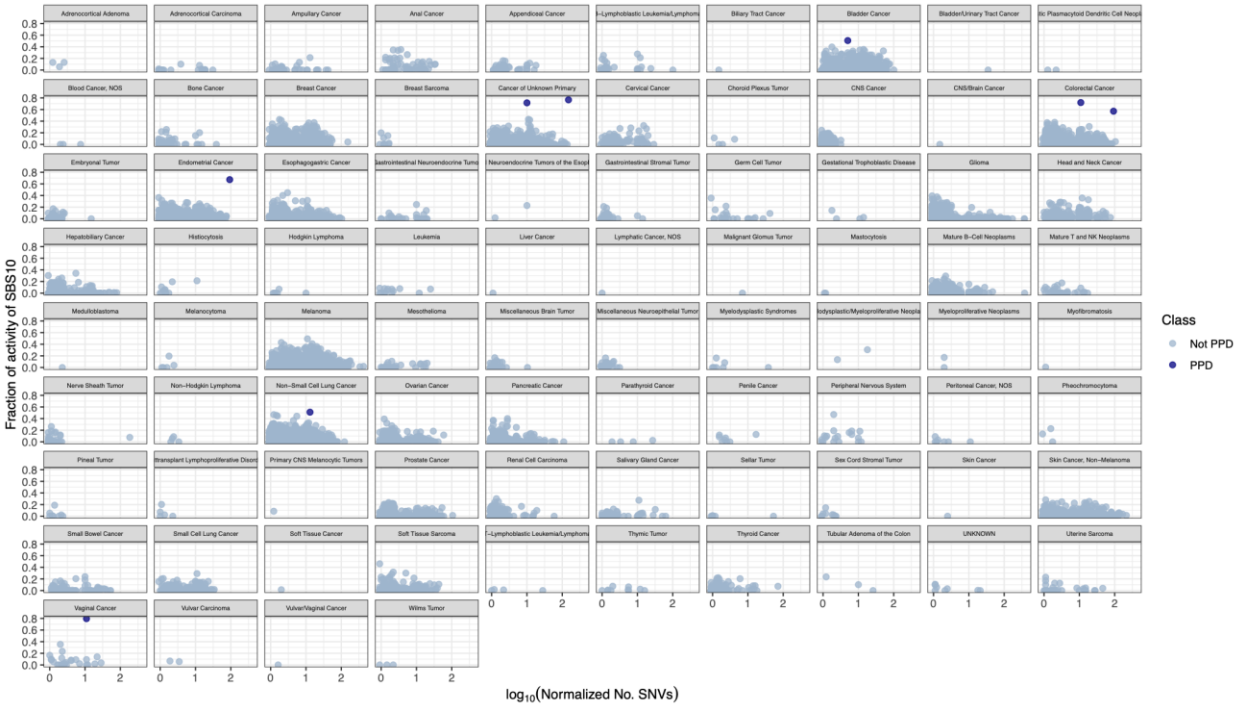


Figure S13
The clustering of the PPD status of the not POLED1 tumors separated per cancer type (for all cancer types)

Interesting Cases

MMRp

GENIE-PROV-a54179a210-41cc2d3220

This sample was a PPD candidate and met all conditions except for the P286R *POLE* mutation, as a single mutation in the ExoD is subclonal. In this sample, there are 5 variants in *POLE*, with 2 of them occurring at the same position (286) - one missense and one deletion. The remaining 3 variants are outside the ExoD.
Please help with the theory

GENIE-MSK-P-0056725-T01-IM6

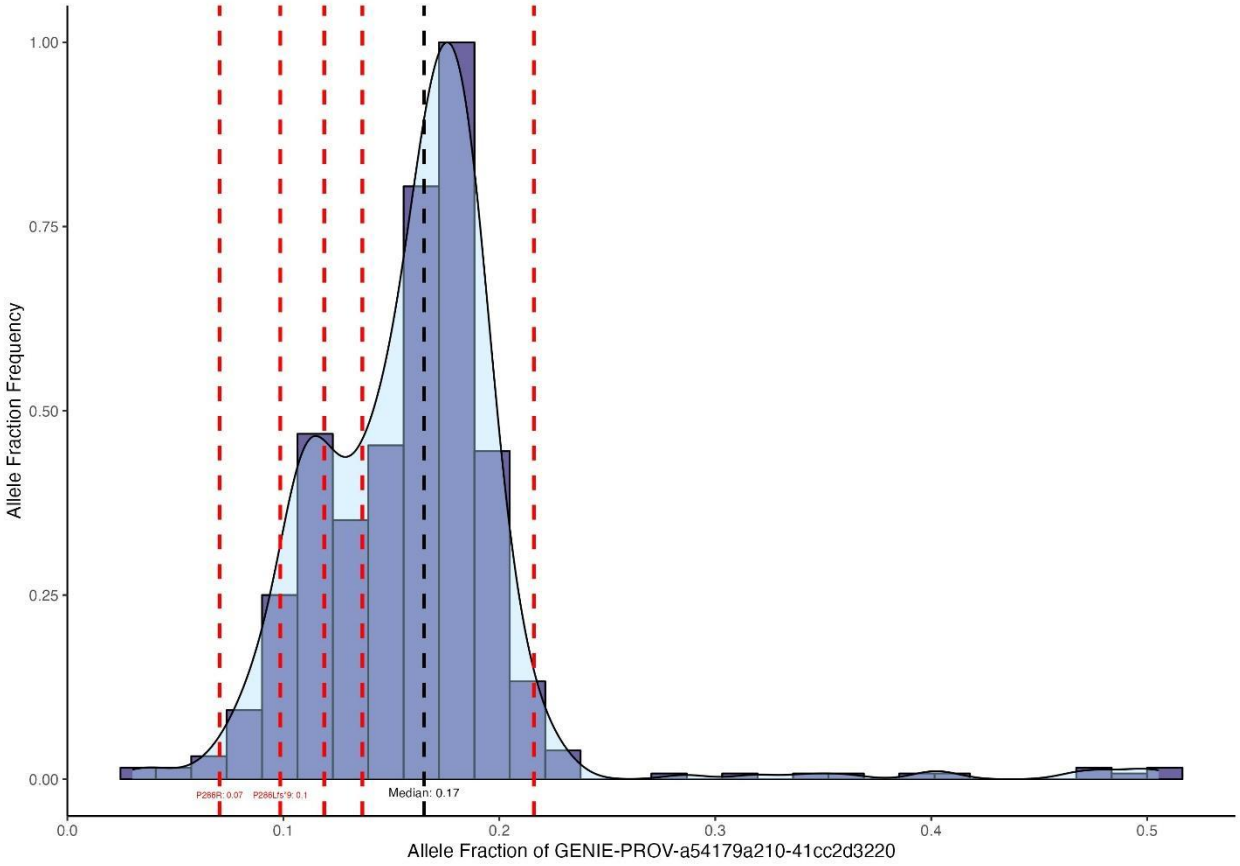
The sample is a PPD candidate with a single SNV in *POLE*, outside of ExoD, at position 1870. Additionally, the sample had a DNP in POLE at position 456, which is a driver position. The median allele fraction of this sample is 0.121118, indicating low purity. This may compromise the reliability of the sequencing. Despite the SNV standing in all conditions, we will not consider it as a driver for this sample.

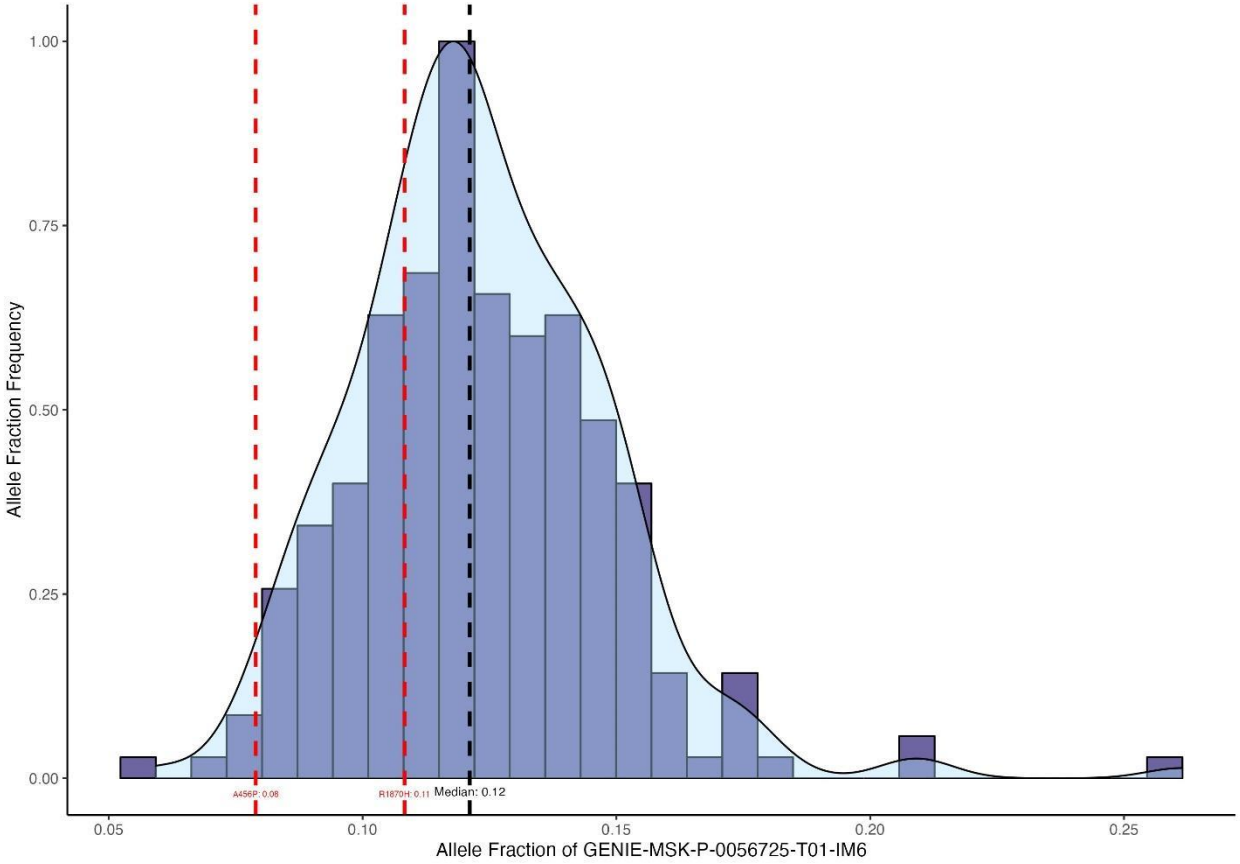
MMRd

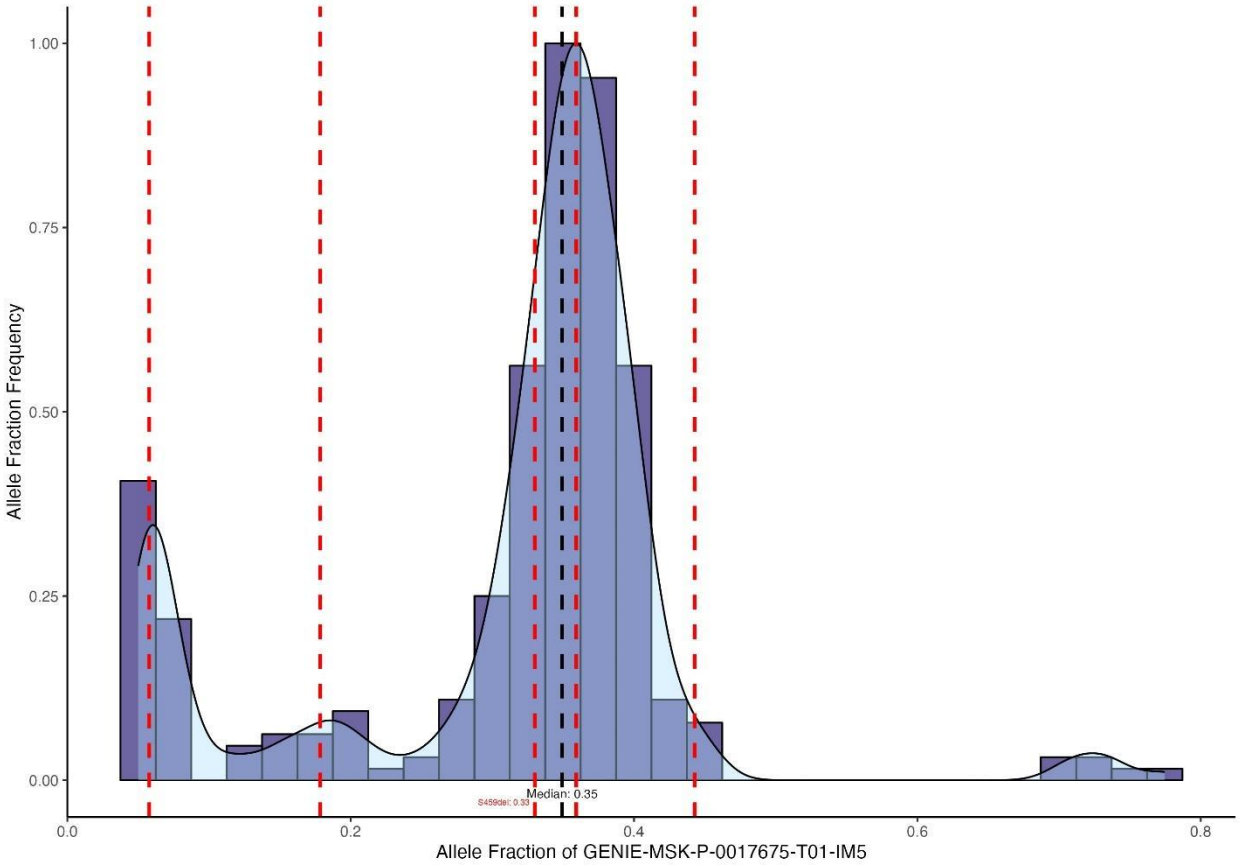
GENIE-MSK-P-0017675-T01-IM5

The sample is PPd + MMRd candidate in Glioma. It has 4 SNV missense *POLE* variants out of ExoD, and a single deletion in *POLE* position 459 which is a driver position. We refer to the deletion as what caused the increase in SNV. The strongest signature here is SBS1.

<u>Interesting cases</u>




1.00
0.75
0.50
0.25
0.00
Allele Fraction Frequency
A456P: 0.08
R1870H: 0.11
Median: 0.12
0.05
0.10
0.15
0.20
0.25
Allele Fraction of GENIE-MSK-P-0056725-T01-IM6



Cases with multiple drivers

Upon review, we identified 5 samples with 2 variants each, classified as drivers: 3 MMRp and 2 MMRd. The first sample, an MMRd case, contained both POLE mutation S461P and POLD1 mutation G205S, both clonal. We counted this as a positive case for S461P because it was the only mutation inside the ExoD.

In the second sample, also an MMRd case, we found two POLD1 variants, L357M and L606M, both clonal. The driver was attributed to L357M for being inside the ExoD of POLD1. It's worth noting that L357M appeared as a driver in only this sample, compared to L606M, which was found in 3 positive samples.

The third sample, an MMRp case, contained POLE mutation V411L and POLD1 mutation D316N, both clonal. We counted this as a positive case for V411L because D316N is a driver found only in MMRd cases.

The fourth sample had two POLE variants, L424I and F367V, but no information on allele fraction, so we couldn't decide. However, we noted that L424I was found only in MMRd-associated cases, while F367V was found in both MMRp and MMRd cases.

The final sample was MMRp and had variants D275G and S459F, with S459F being subclonal. We chose D275G as the clonal ExoD mutation because S459F had a lower allele fraction compared to the other variants in that sample.

Out of the 322 positive samples containing any driver, this represents a relatively small proportion.