Purple v4.1

High level changes

The 2 key changes to PURPLE in v6 are improvements to the somatic fit and the introduction of a biallelic likelihood. There are also a few other minor changes:

Category	Changes
SOMATIC fit improvements	• Additional adjustment on subclonality for somatic peak fitting in T/N WGS mode • Somatic fitting enabled for T/O and targeted modes. Reported • New trigger for somatic fitting for panels (Purity > 92%; 1.8<Ploidy<2.2) • Regions of negative implied copy number excluded from somatic Poisson calculations • TBD: allow VAF peak max to be slightly > 0.5 in T/N WGS???
Biallelic Likelihood	• Biallelic flag replaced by new model to calculate a biallelic Likelihood.
Other	• Tetrasomy 9P and Trisomy 9P added as chromosomal aberrations • Minor improvements to CIRCOS and INPUT.png to support HRD CNN classifier • Only use VAF > 2% for MSI, TMB, TML • No longer exclude INDELS in long repeats from reportability

Somatic fit improvement

Previously, the somatic fit only functioned in tumor normal WGS mode. This is extended to cover targeted panels and tumor only samples, to deal with the issue of samples with low aneuploidy often being fit to tumor fractions close to 100%

Trigger for somatic mode

PURPLE triggers somatic mode if the sample appears highly diploid. The criteria for this differs between WGS and TARGETED samples:

Mode	Trigger
WGS	Trigger somatic mode if maxDiploidProprition >= 0.97
Targeted	Trigger somatic mode if best fitted purity > 92% AND 1.8 < ploidy < 2.2

Behaviour in somatic mode

For Tumor-Normal WGS, the existing behaviour is retained with the addition of a subclonal peak fit adjustment (TO DO: JUNRAN to add details)

If the sample is tumor only AND/OR targeted panel then the tumor purity is simply estimated as 2x the median of HOTSPOT and reportable variants. Note that variants with vaf < 0.04 are ignored for the fit and that non-HOTSPOT variants are filtered if the VAF > 0.35 as potential germline. If no variant satisfies these criteria then the purity is set to 0.08 and the sample is marked as NO_TUMOR. The DELETED_GENES check is also not applied in these modes.

*NB: somatic variants are **NOT** used for tumor only samples in the somatic fit deviation in normal mode*