



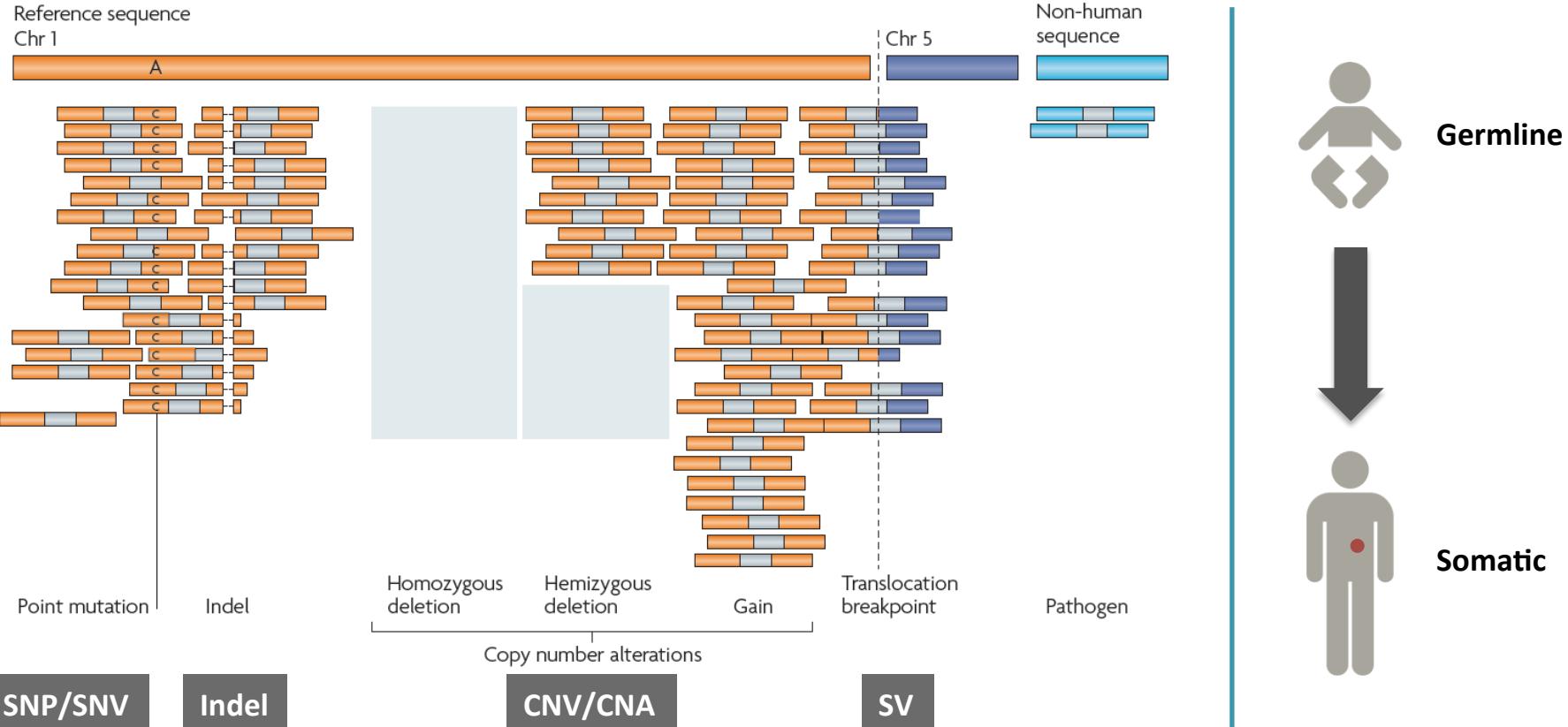
GATK Best Practices for Variant Discovery



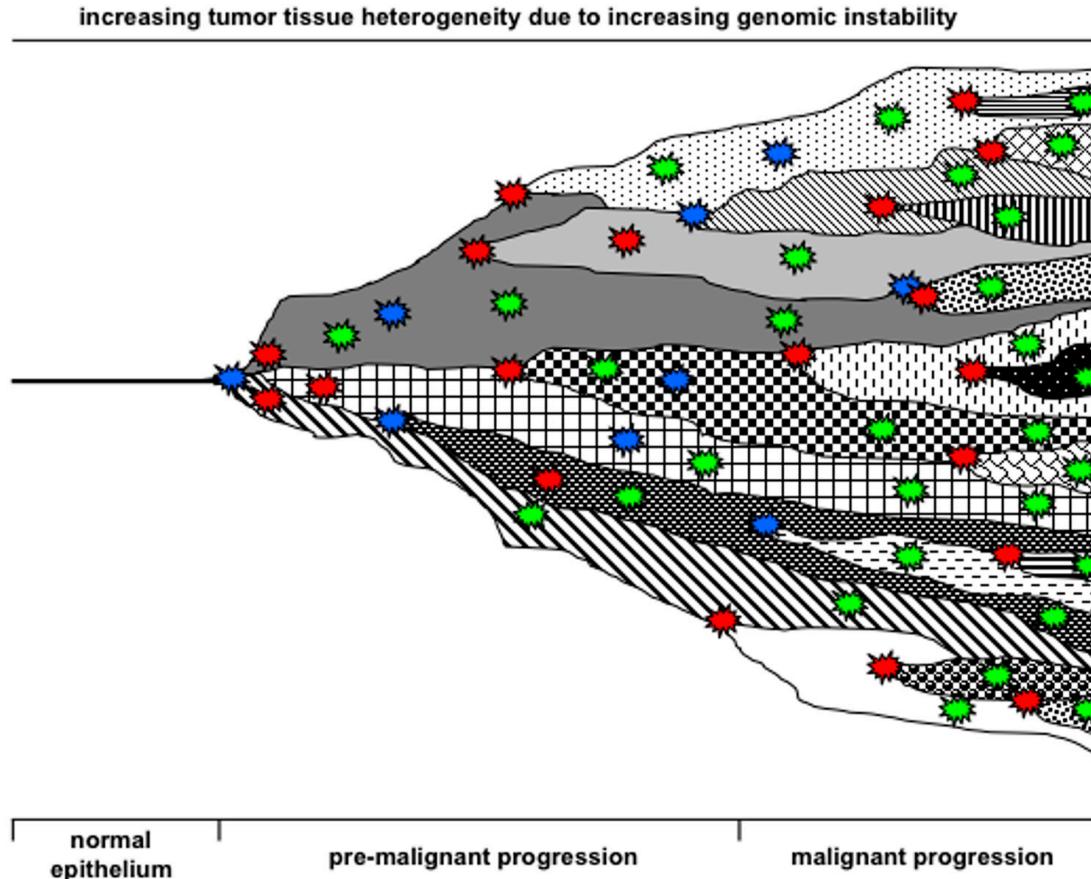
Introduction to Somatic Variant Discovery

Key considerations and workflow logic

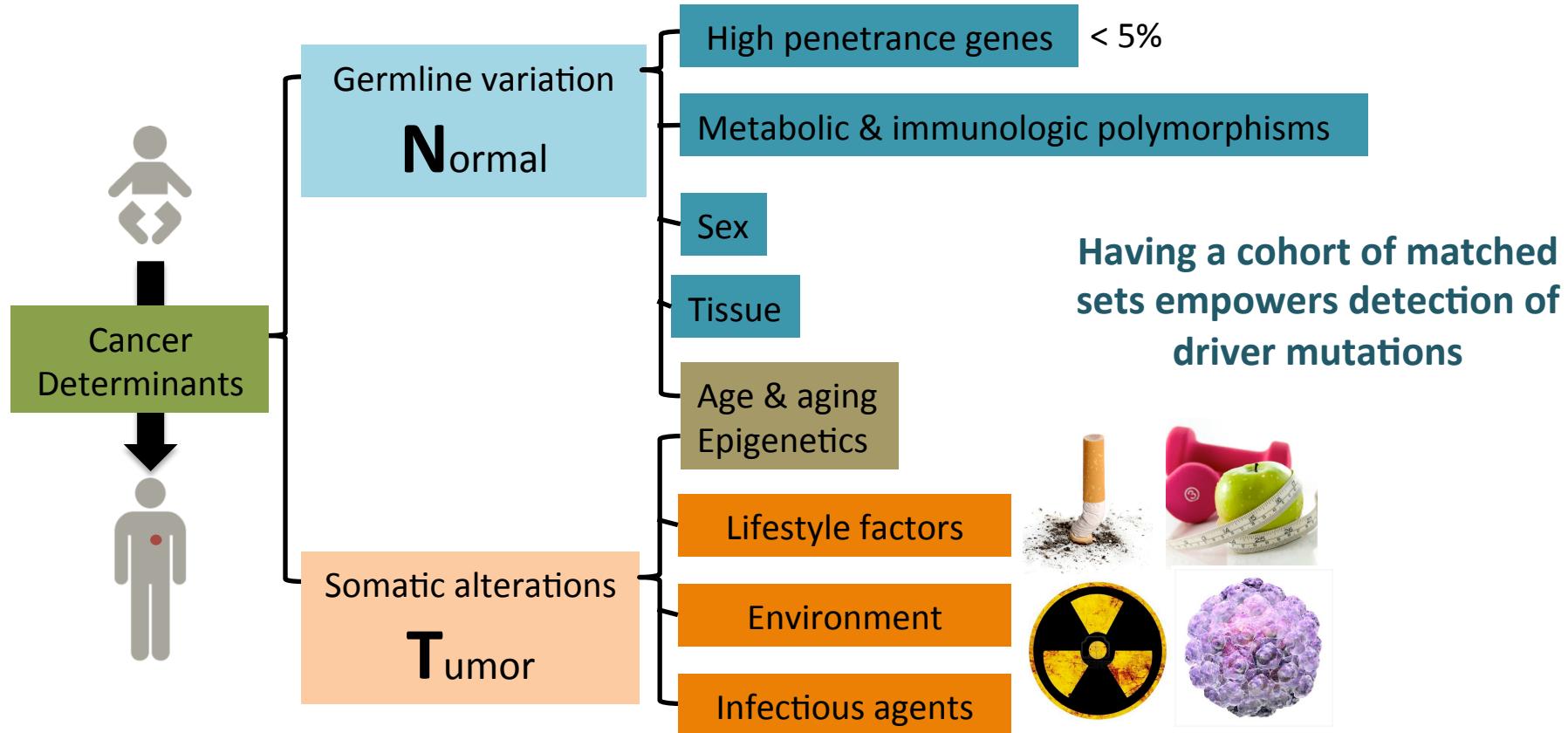
Different types of variants



Role of mutation events in tumor progression



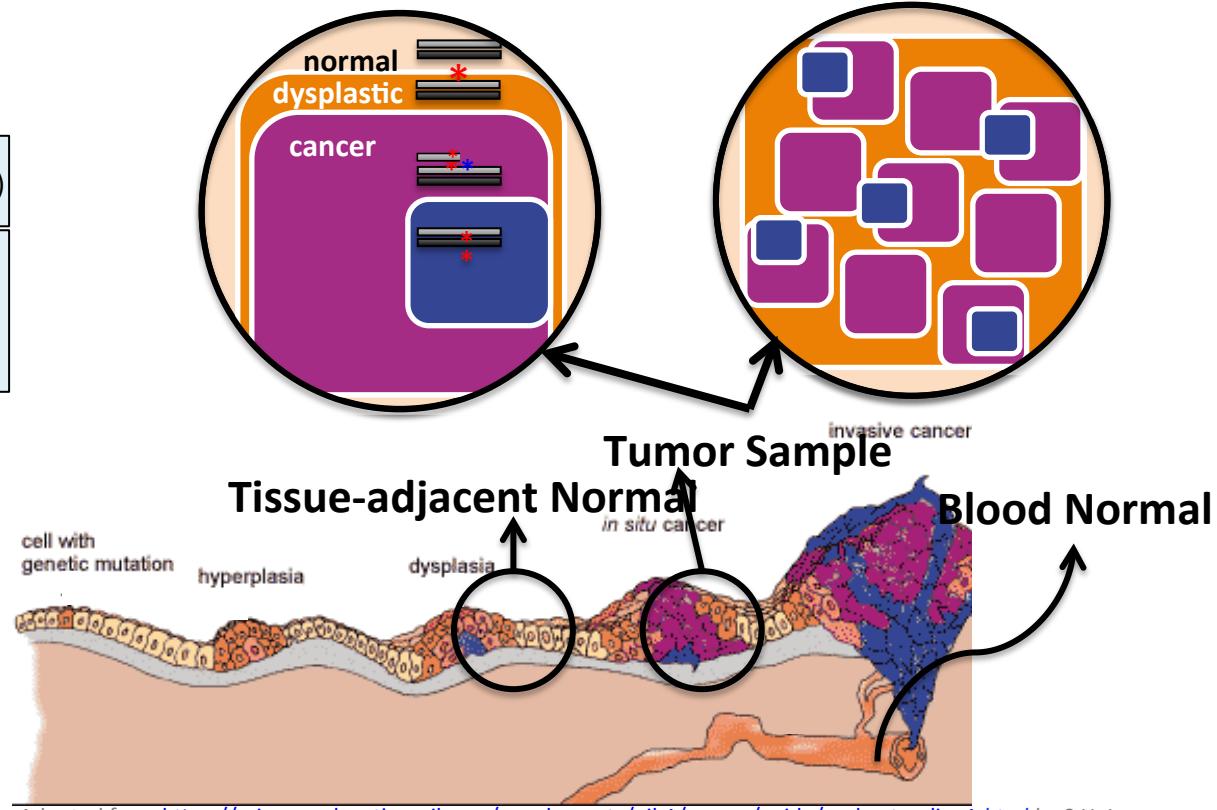
Cohorts of paired T-N data to detect driver mutations



Tumor and normal contamination and heterogeneity

$$\text{Tumor purity} = \frac{(\text{tumor cells})}{(\text{normal} + \text{tumor cells})}$$

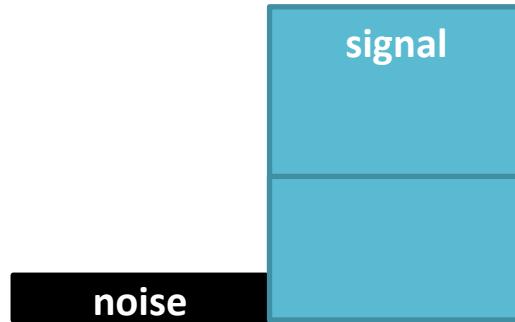
Tumor heterogeneity is based on polygenomic populations, segregated or intermixed, due to ongoing subclonal evolution.



Amount of signal may be comparable to noise



Expectation for germline variants

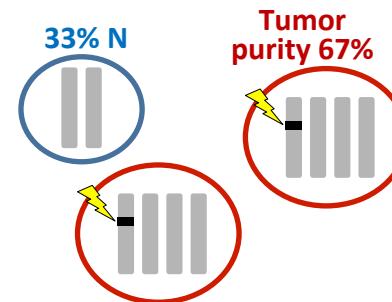


+ AF expected to follow ploidy

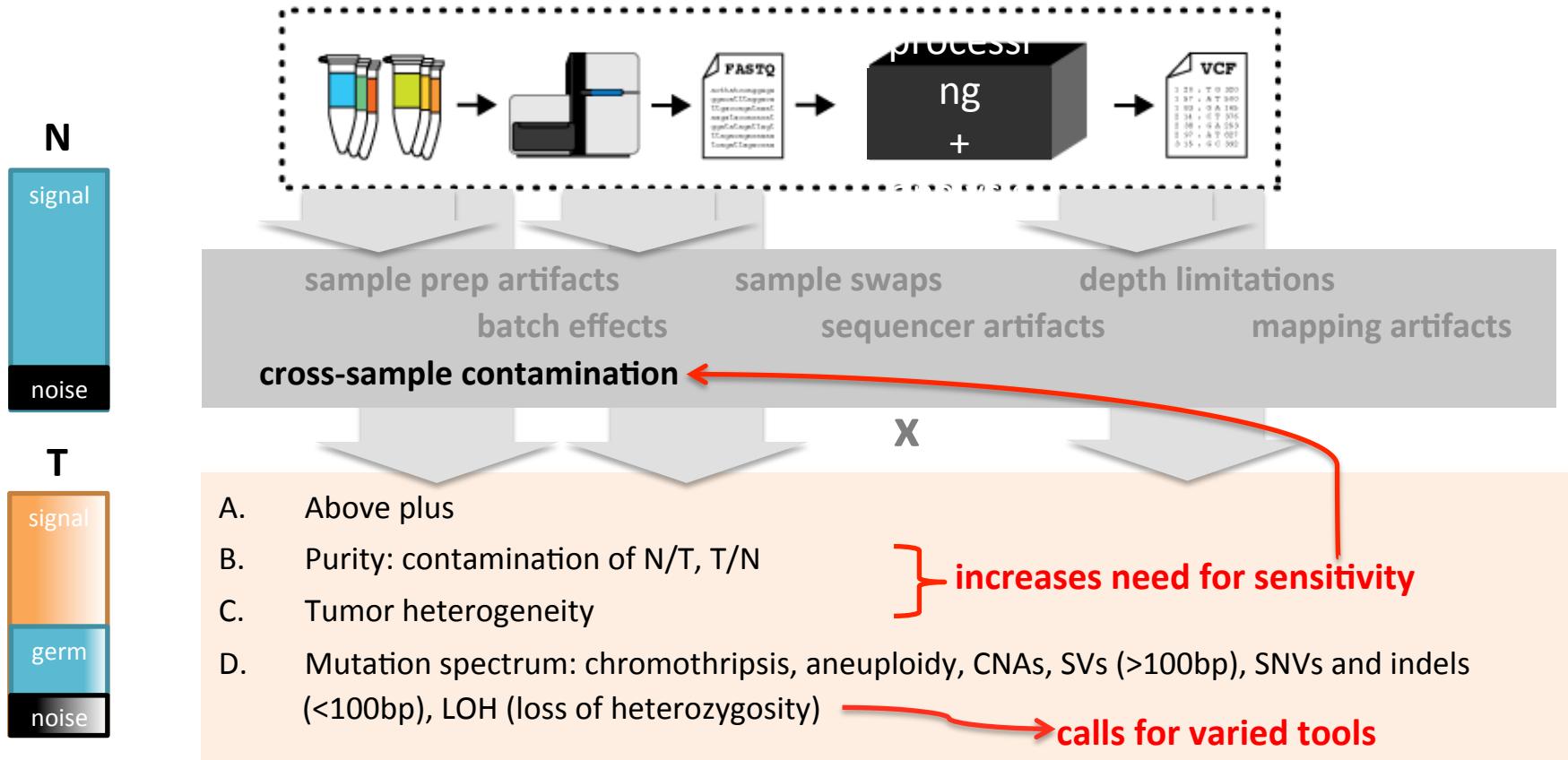
Expectation for somatic variants



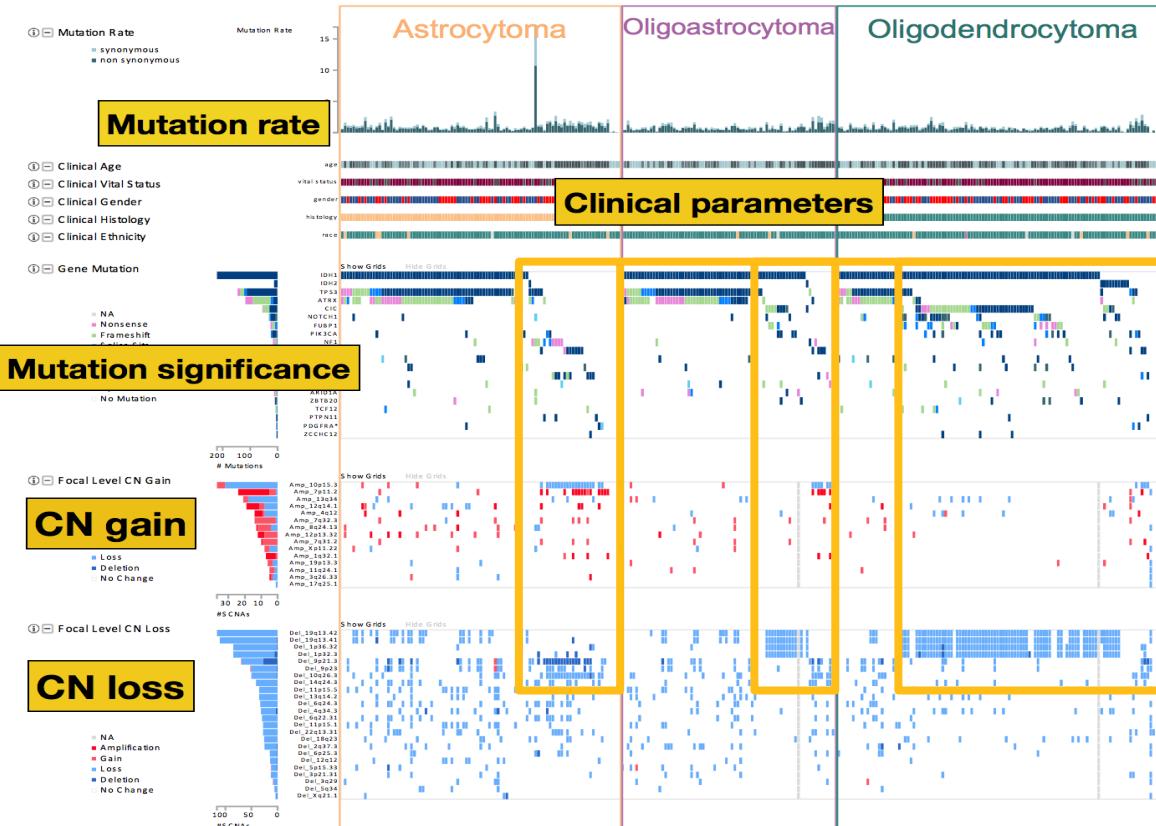
+ no reliance on ploidy for AF



Cancer-specific challenges confound analyses



A tumor's genomic alterations are multilayered

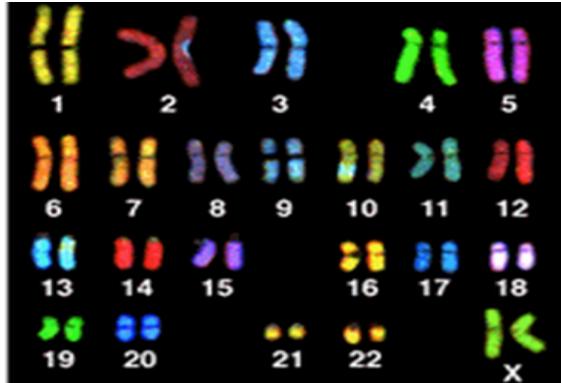


- Brain lower grade glioma (LGG) iCoMut plot from Firebrowse
- Patients lacking characteristic mutations in IDH1/2, TP53 and ATRX have increased focal copy number alterations
- Cancer analysis must take into account a wide spectrum of alterations

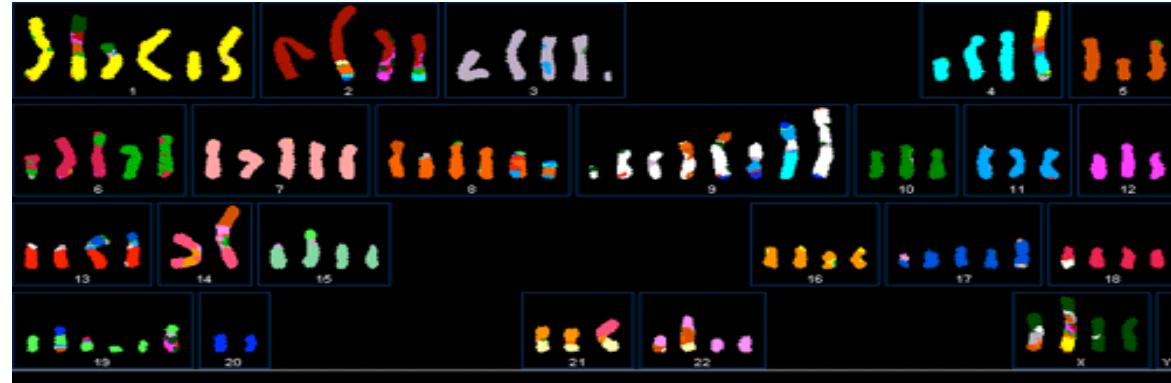
Somatic alterations can be dramatic



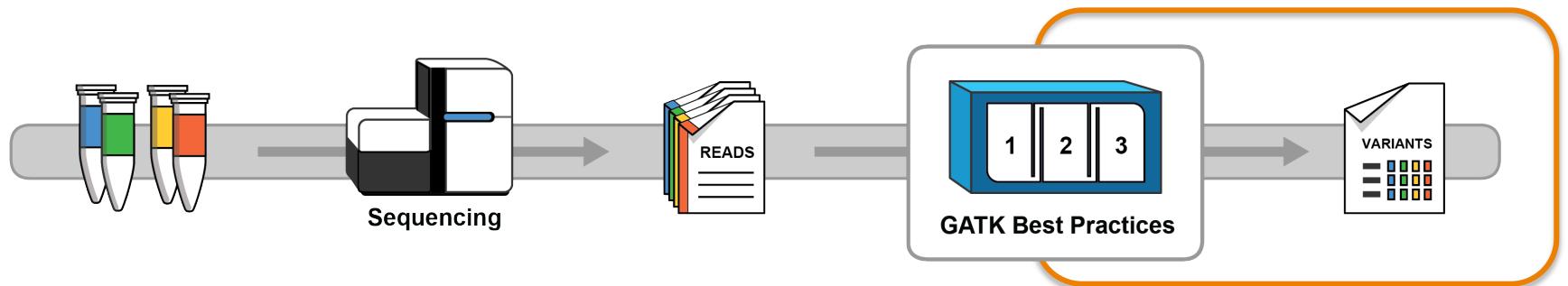
Normal Cell



Cancer Cell Line HCC1954



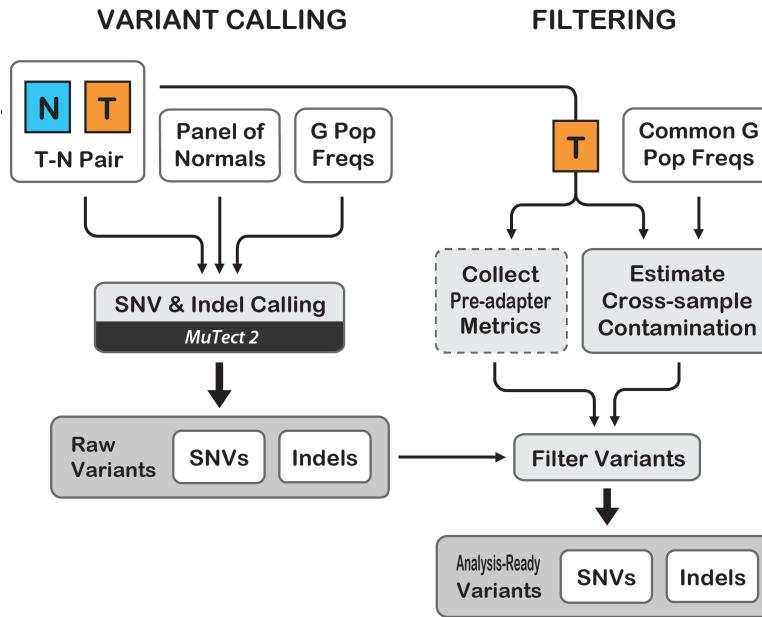
- Spectral karyotyping paints each chromosome pair with a color
- Alterations can vary dramatically between cancers and within cancers



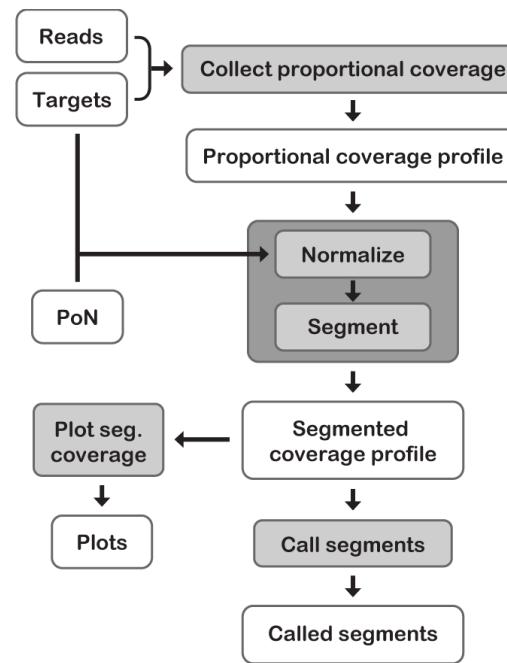
THE WORKFLOWS

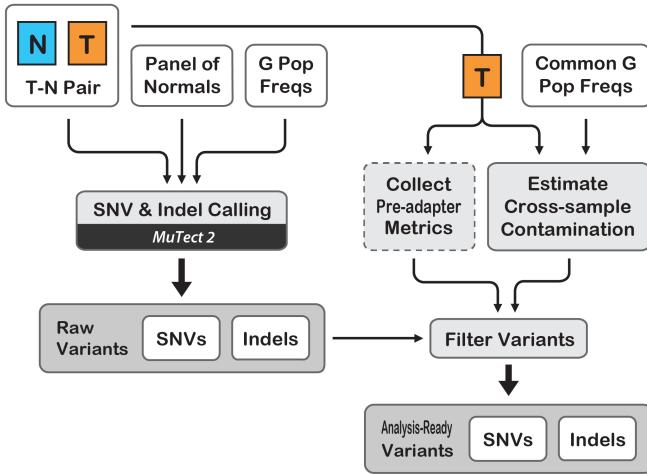
Somatic variant discovery workflows in GATK4

Somatic SNV and Indel Discovery



Somatic CNV Discovery





SOMATIC SNVS & INDELS

Logic of the Tumor-Normal workflow



Comparison to matched normal → subtraction of germline background



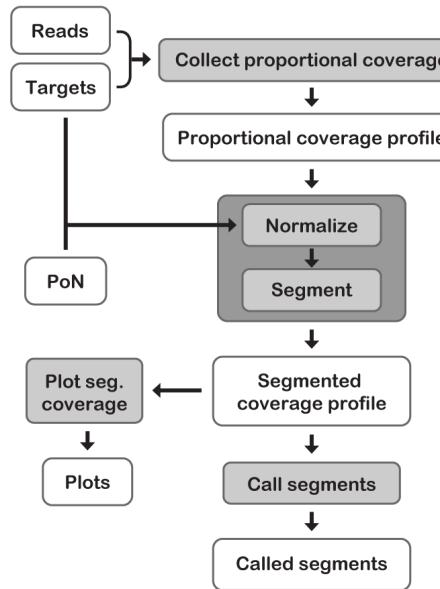
Tumor-only analysis

- It is possible to run the workflow without a matched normal in “tumor-only mode” (normally used for PON creation)
- MUST have a good PON to eliminate common germline variation
- Will still require extra filtering (not described here)

Panel of Normals for SNVs & Indels



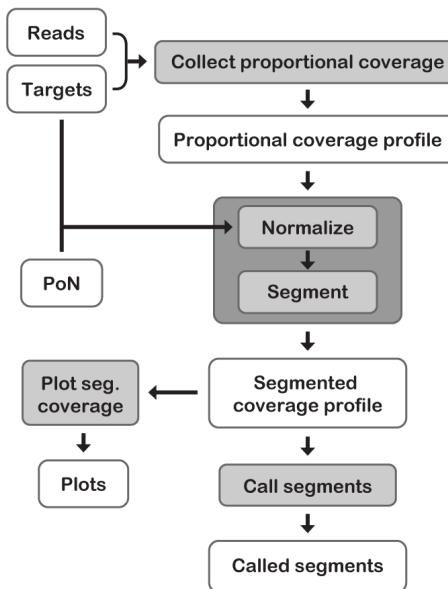
- VCF of calls made from a set of unrelated “normal” samples
- Main purpose:
Eliminate common/recurring technical artifacts
-> should use normals made using the same data generation techniques
(eg same capture kit for exomes, same sequencing platform etc)
- Secondary purpose: also eliminates germline variants not called in the matched normal (or approximates the normal if none is available)



Part 4

SOMATIC CNV

Copy number: it's all about coverage and normalization



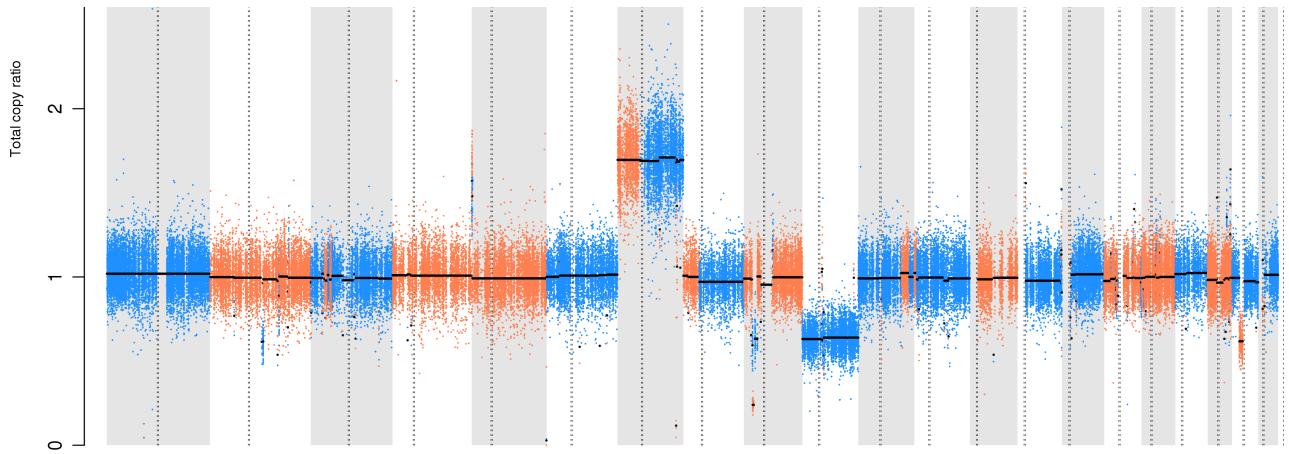
Collect proportional coverage



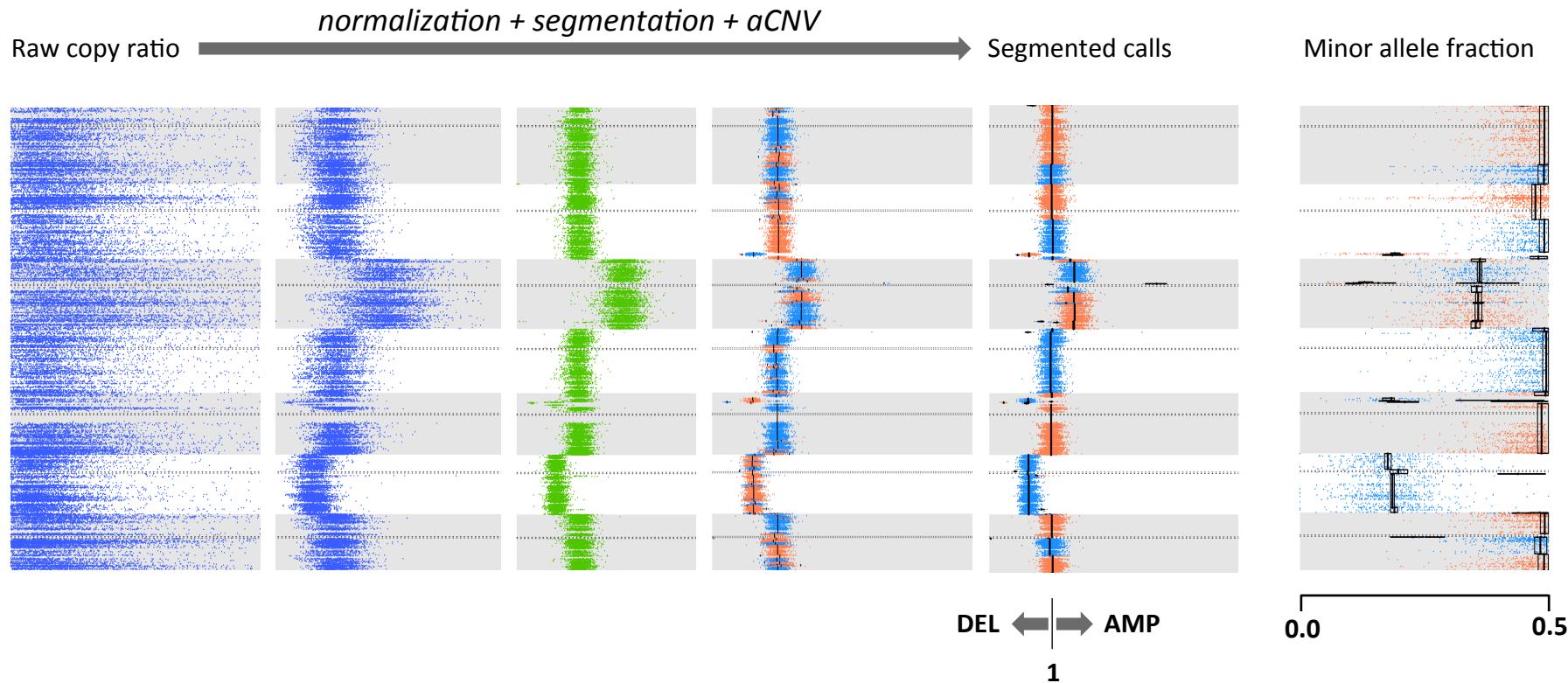
Normalize to remove noise



Identify segment boundaries



“Denoising” normalization process is essential



Panel of Normals for CNVs



- Made from a set of unrelated “normal” samples
BUT not a VCF (unlike SNVs & Indels PON)
- Main purpose:
“Denoising” to compensate for variability in coverage
-> should use normals made using the same data generation techniques
(eg same capture kit for exomes, same sequencing platform etc)