Somatic variant calling and identification of de novo mutations

MuTect and DeNovoGear

Edinburgh Genomics
Edinburgh, UK
23rd October 2015

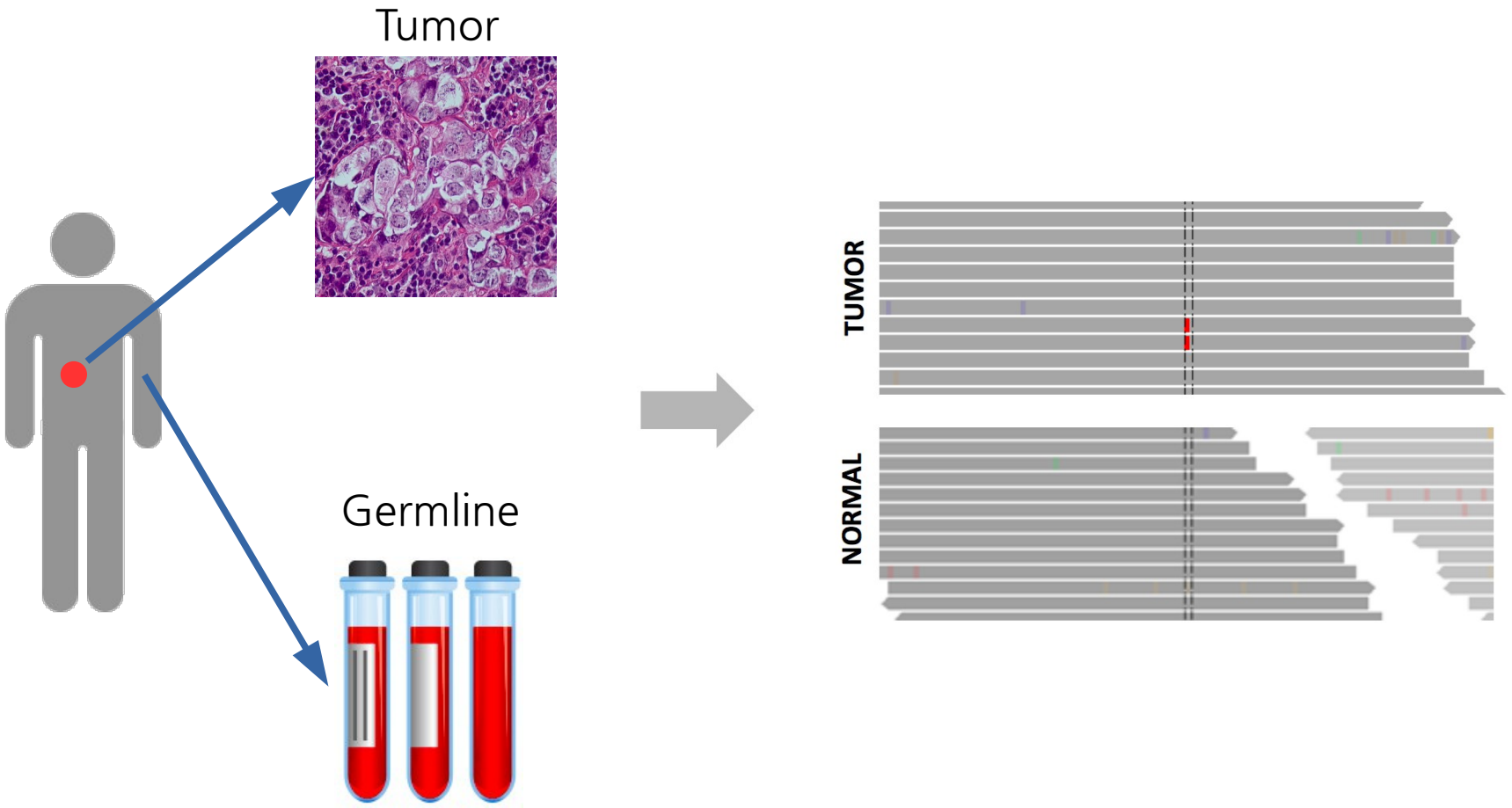
Marta Bleda Latorre
mb2033@cam.ac.uk
Research Assistant at the Department of Medicine
University of Cambridge
Cambridge, UK

UNIVERSITY OF CAMBRIDGE

* Slides based on GATK presentations (link)

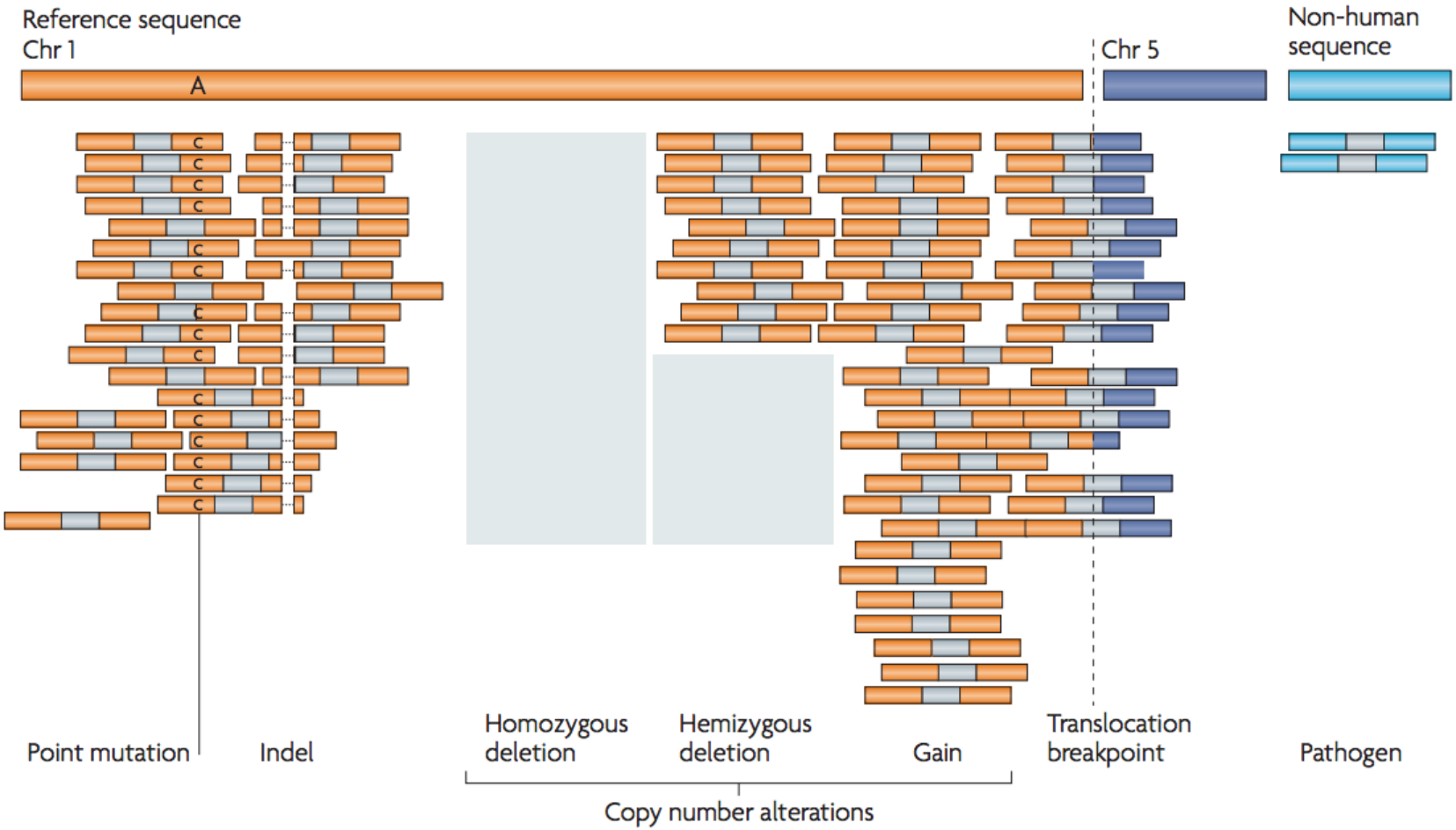
Somatic calling

Tumor vs normal



Somatic calling

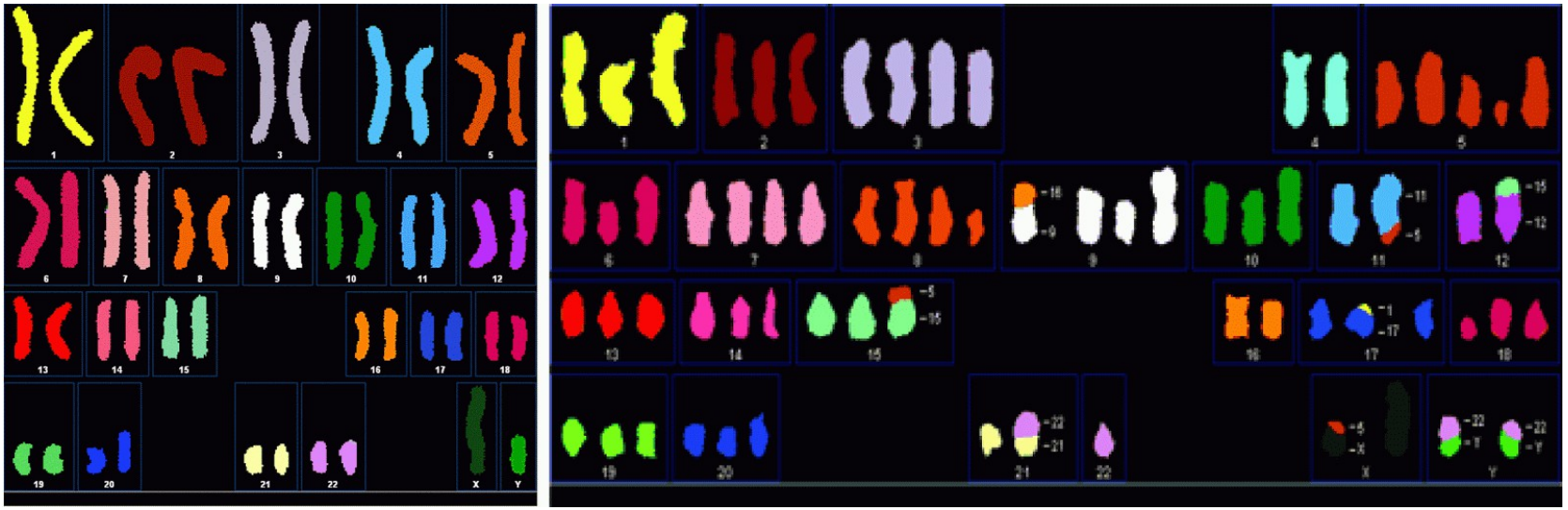
Type of variants



Meyerson, Gabriel and Getz. Nat. Rev. Genet. (2010)

Somatic calling

Number of somatic mutations

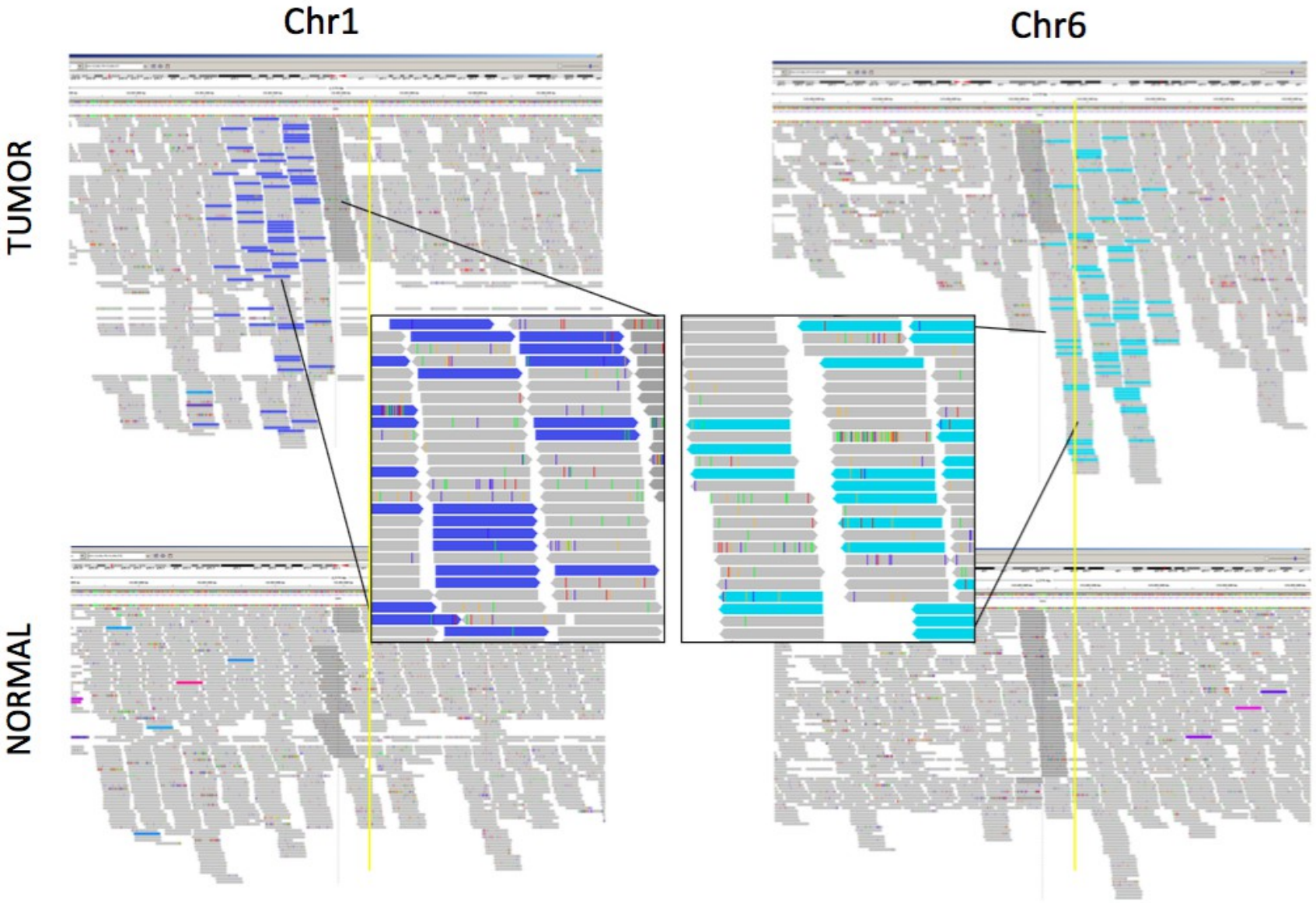


In a genome of 3 x 109 bases:

- 1000s to 10,000s somatic single nucleotide variations (sSNVs)
- 100s to 1,000s somatic small insertions and deletions (sINDELs)
- 100s to 1,000s somatic structural variations (sSVs)
- 100s to 1,000s somatic copy number alterations (sCNAs)

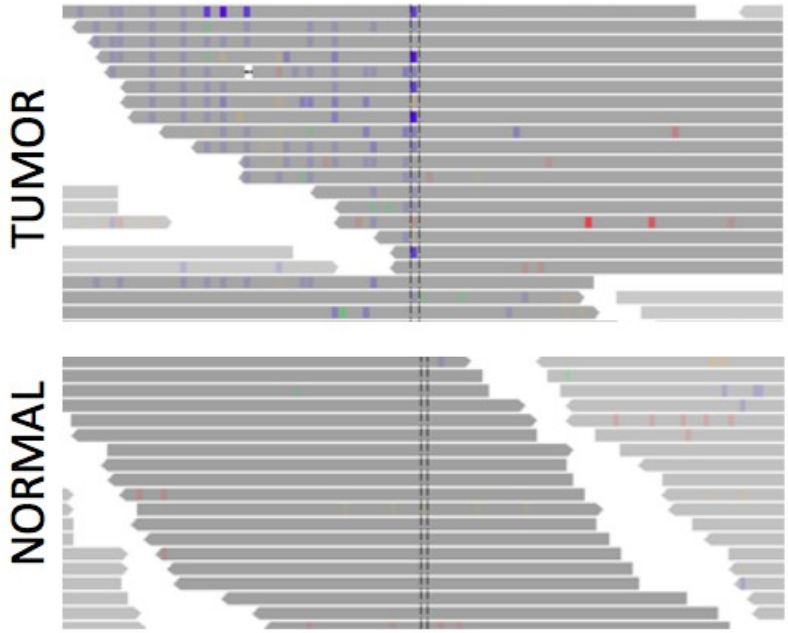
Somatic calling

Visualization of rearrangements

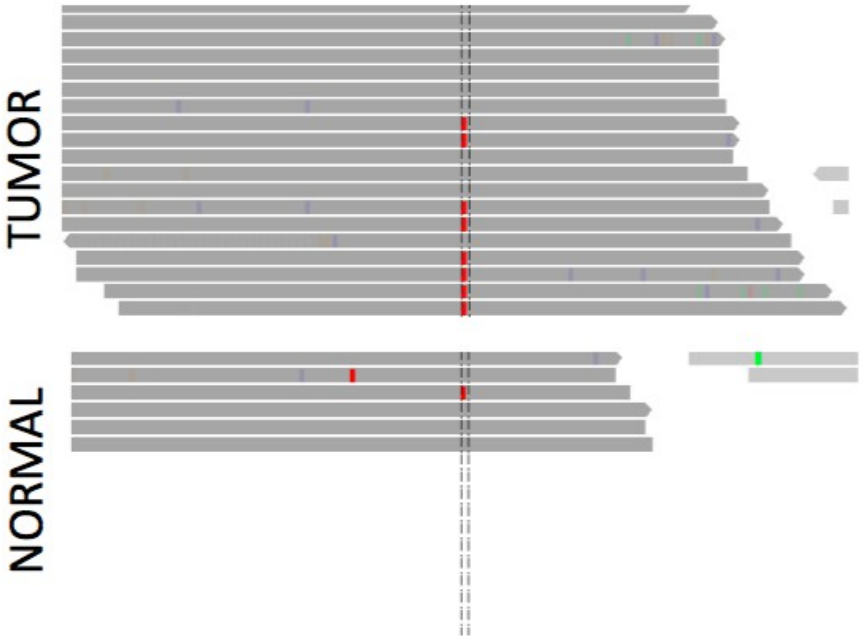


Somatic calling
False positives

Noise

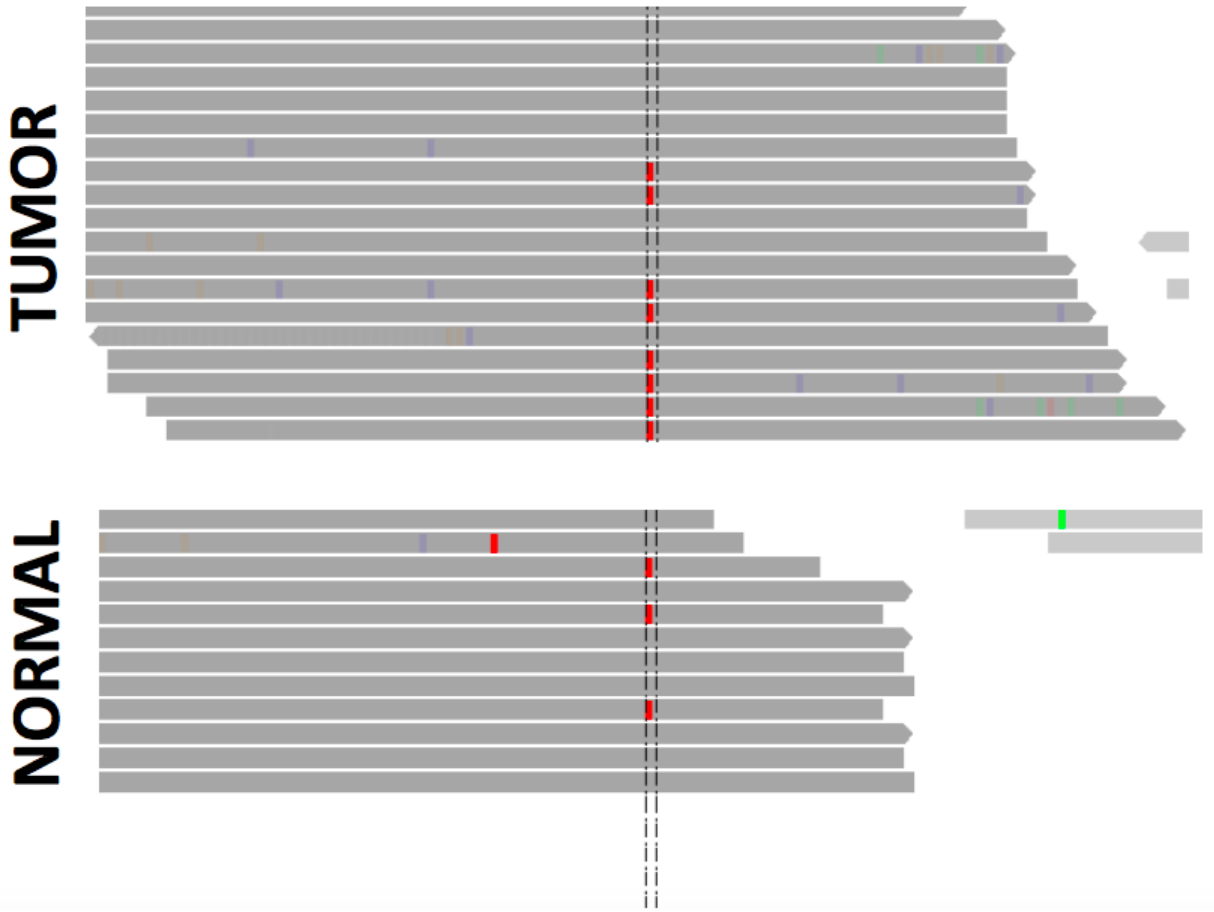


Germline mutation

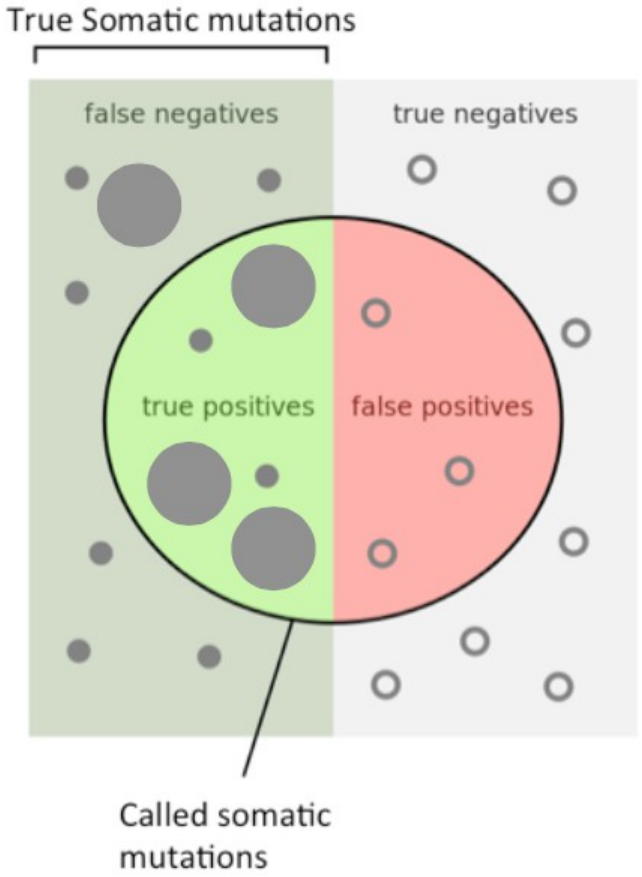


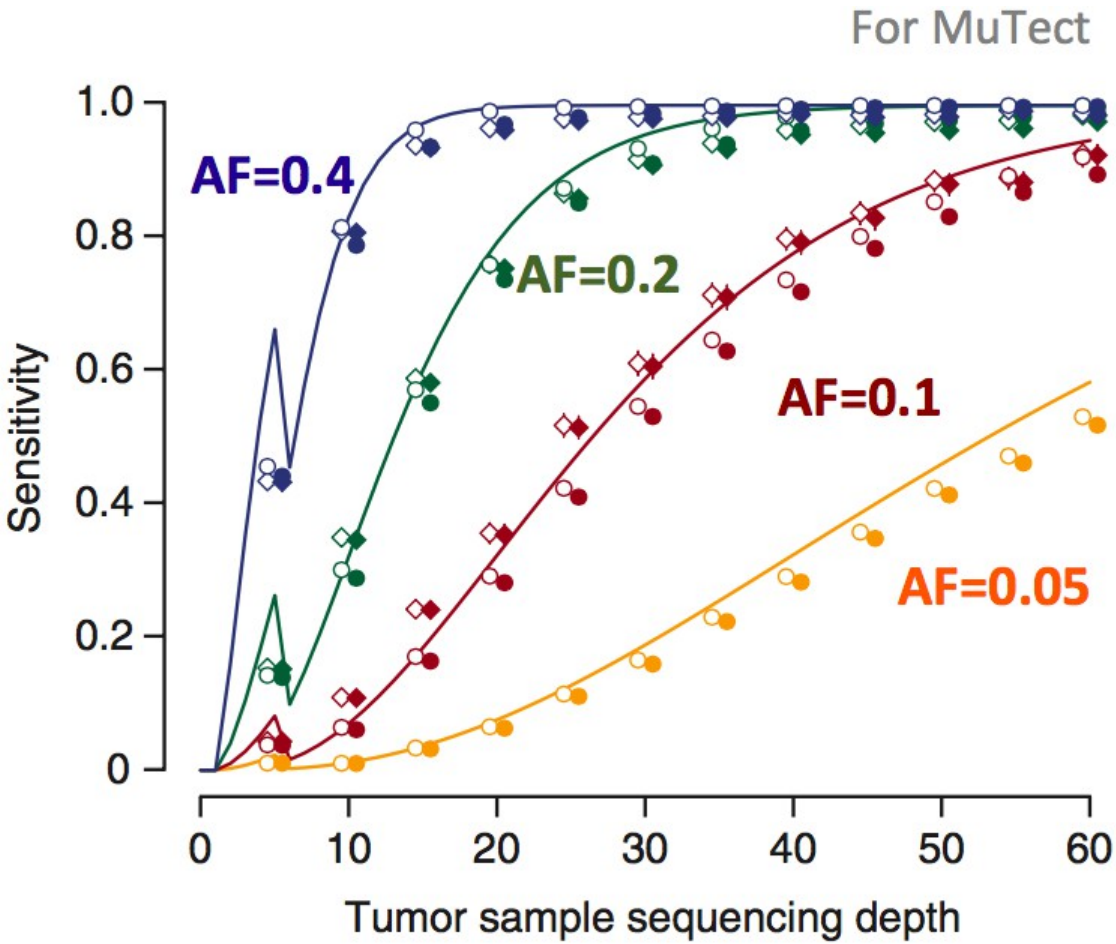
Somatic calling
False positives

Contamination of the normal sample



Somatic calling
Deep sequencing required



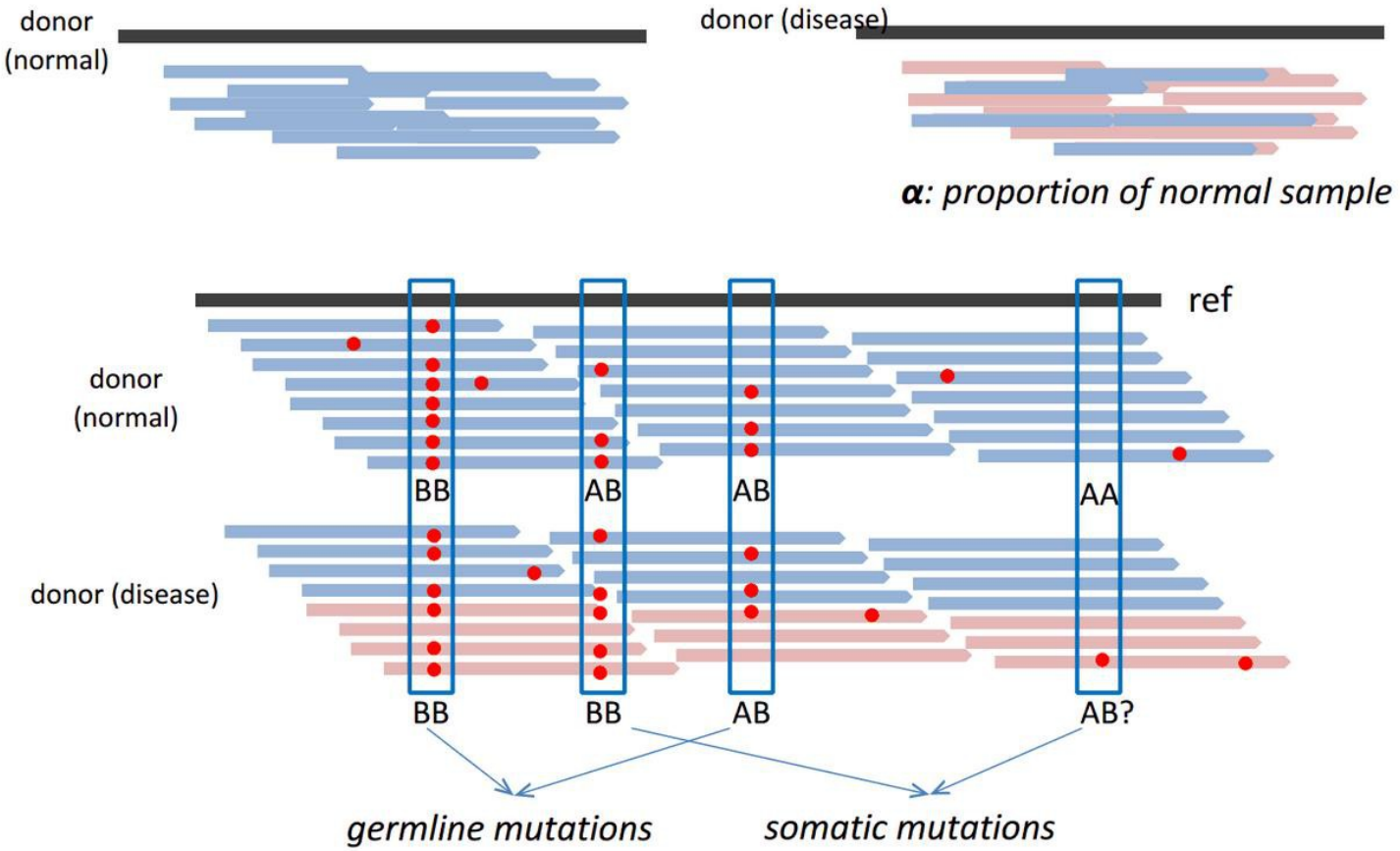


Somatic calling

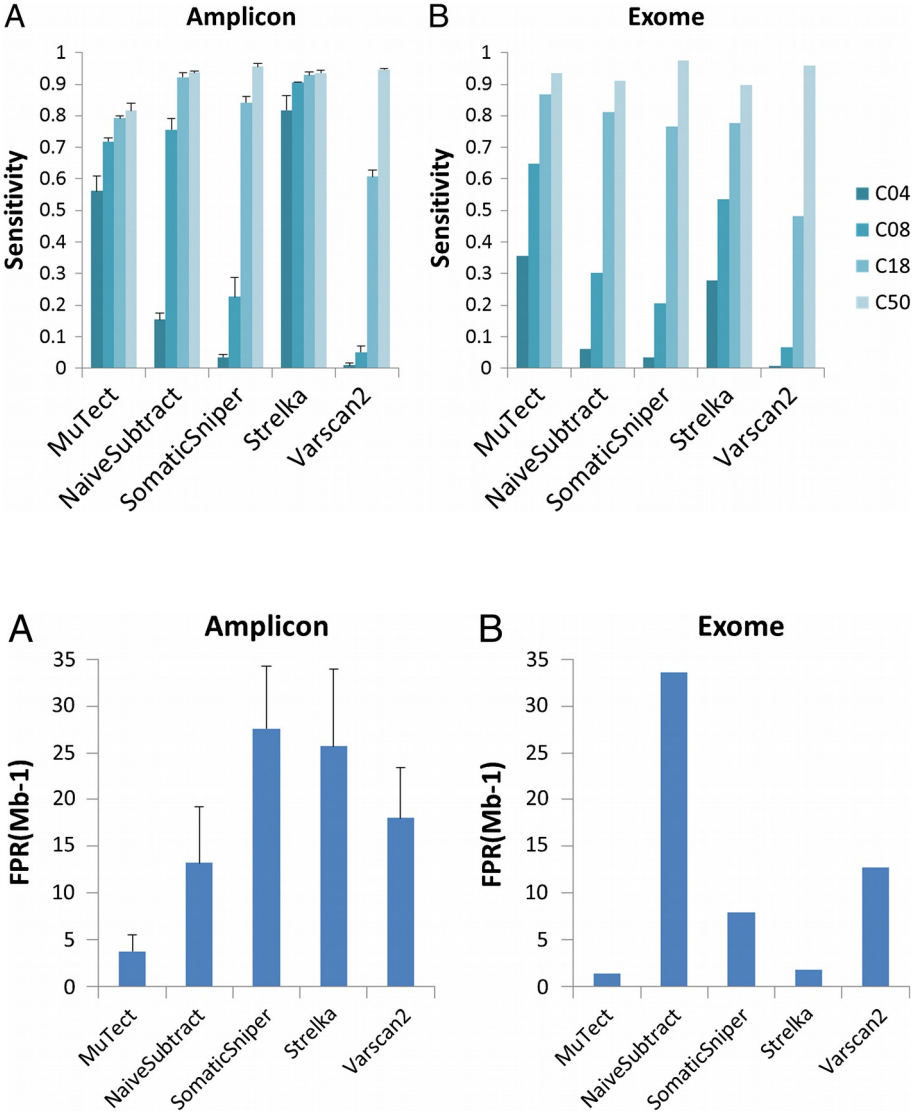
Detecting somatic SNVs in cancer

Challenges:

- Somatic variants occur at low frequency in genome
- Most tumors are impure and heterogeneous

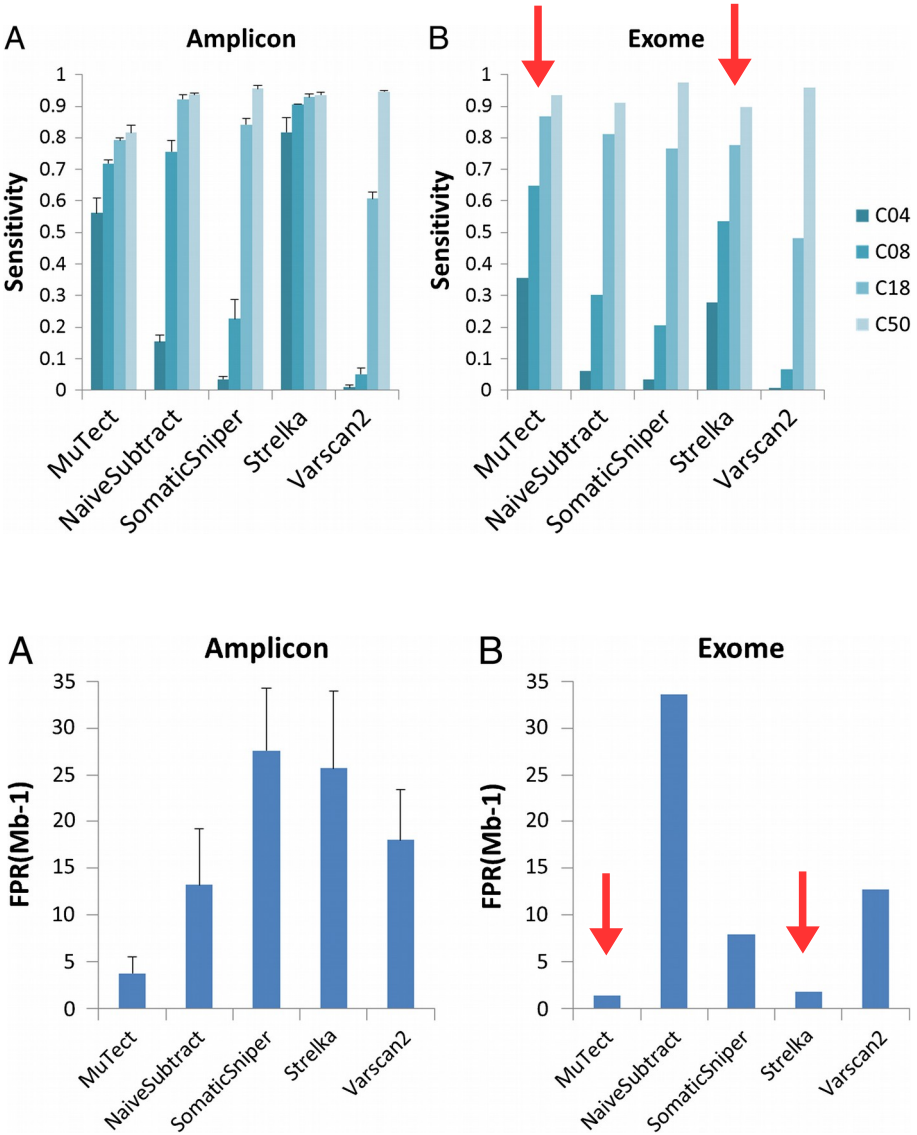


Somatic calling
Software



Xu, Huilei, et al. "**Comparison of somatic mutation calling methods in amplicon and whole exome sequence data.**" *BMC genomics* 15.1 (2014): 244.

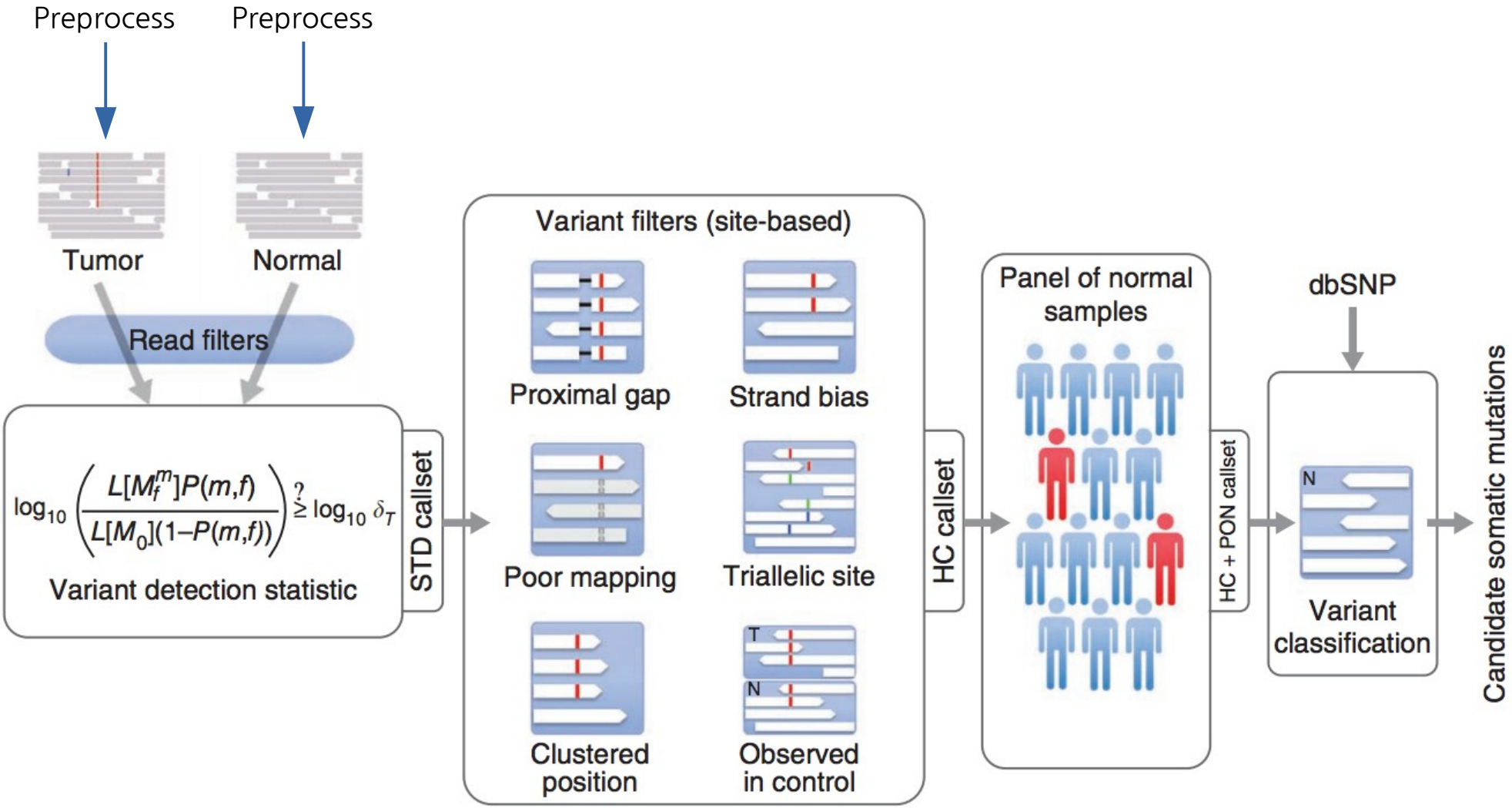
Somatic calling
Software



Xu, Huilei, et al. "**Comparison of somatic mutation calling methods in amplicon and whole exome sequence data.**" *BMC genomics* 15.1 (2014): 244.

Somatic calling

MuTect



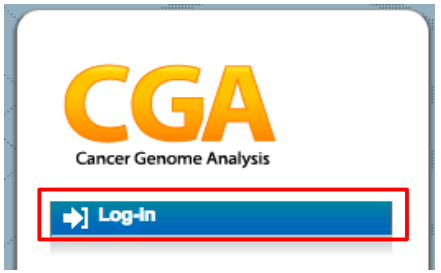
MuTect installation

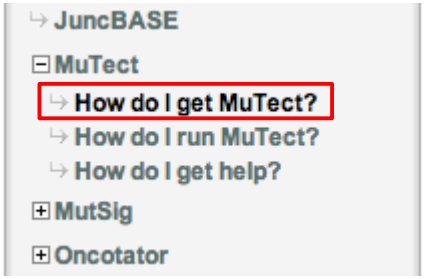
- **MuTect download**

  http://www.broadinstitute.org/cancer/cga/mutect

  - Click ***Log-in*** and go to the ***Create new account*** tab
  - Fill the form
  - Go to ***How do I get mutect*** and accept the license agreement
  - Download the latest version

    muTect-1.1.4-bin.zip
  - Extract the file in the applications folder





- **Check if MuTect is working**

```
java -jar muTect-1.1.4.jar -h
```

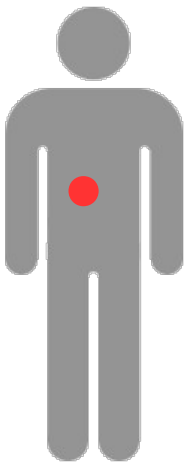
- **Usage**

```
java -jar muTect-1.1.4.jar --analysis_type MuTect [arguments]
```

De novo mutations

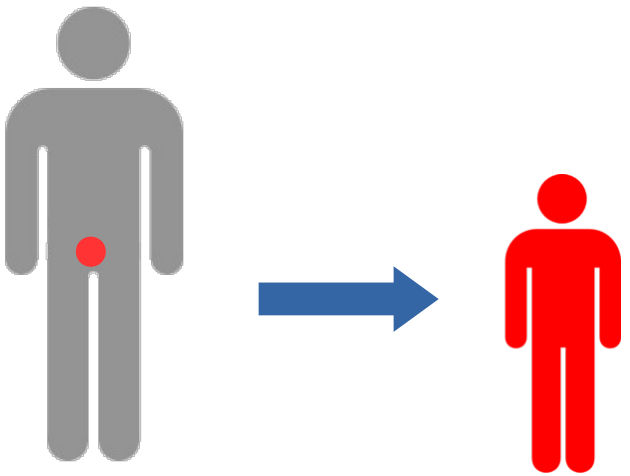
De novo mutation, An alteration in a gene that is present for the **first time** in one family member as a result of a **mutation in a germ cell** (egg or sperm) of one of the parents or in the **fertilized egg** itself.

Somatic variants



- Mutation in tumor only
- Non inheritable

***De novo* mutations**



- Can be inherited
- All cells affected in offspring

De novo mutations

- On average, humans acquire **~74 *de novo* single nucleotide variants** (SNVs) per genome per generation.
- The de novo mutational load seems **correlated with paternal** (as opposed to maternal) **age**.
- The rate of de novo mutations seems higher in individuals with genetic diseases, particularly **sporadic disorders** such as **intellectual disability and autism**.
- De novo mutations tend to be **more deleterious** than inherited variation because they **haven't undergone** the same level of **evolutionary selection**.
- Because true de novo mutations occur randomly (and newly) in individuals, there's **no database like dbSNP** to guide discovery. We must instead **rely on** deeper sequence **coverage**, better **algorithms**, and ultimately, orthogonal **validation**.
- The best evidence to **implicate a gene** requires looking across a significant number of samples, to find genes that:
  - Harbor **mutations in multiple (unrelated) cases** with a similar phenotype, and
  - **Lack** similarly damaging mutations in populations of **unaffected individuals**

Veltman JA, & Brunner HG (2012). De novo mutations in human genetic disease. Nature reviews. Genetics, 13 (8), 565-75 PMID: 22805709

THANK YOU.