

Lab 1b: GenVisR

What is GenVisR? ¹

- Bioconductor package ([link](#), [intro](#))
- Github repo ([link](#))
- Highly customizable, publication-quality graphics focused on cohort level genome analyses.
- The package is maintained by the Griffith lab.
- [license](#)

waterfall()

What does it do?

- Visualize the types of mutations within a cohort
- Visualize the **mutation burden** in a data set
- Visualize the proportion of samples with a mutated gene
- Visualize clinical data

Why do we care?

- Determine mutually exclusive or co-occurring genomic events
- Recognize patterns within clinical data
- See if the mutation burden conforms to expectations

Mutation Burden

Generically defined as the total number of mutations (changes) found in the DNA of cancer cells, but different studies use different formulas ²

The `Waterfall()` function uses the following formula to estimate mutation burden:

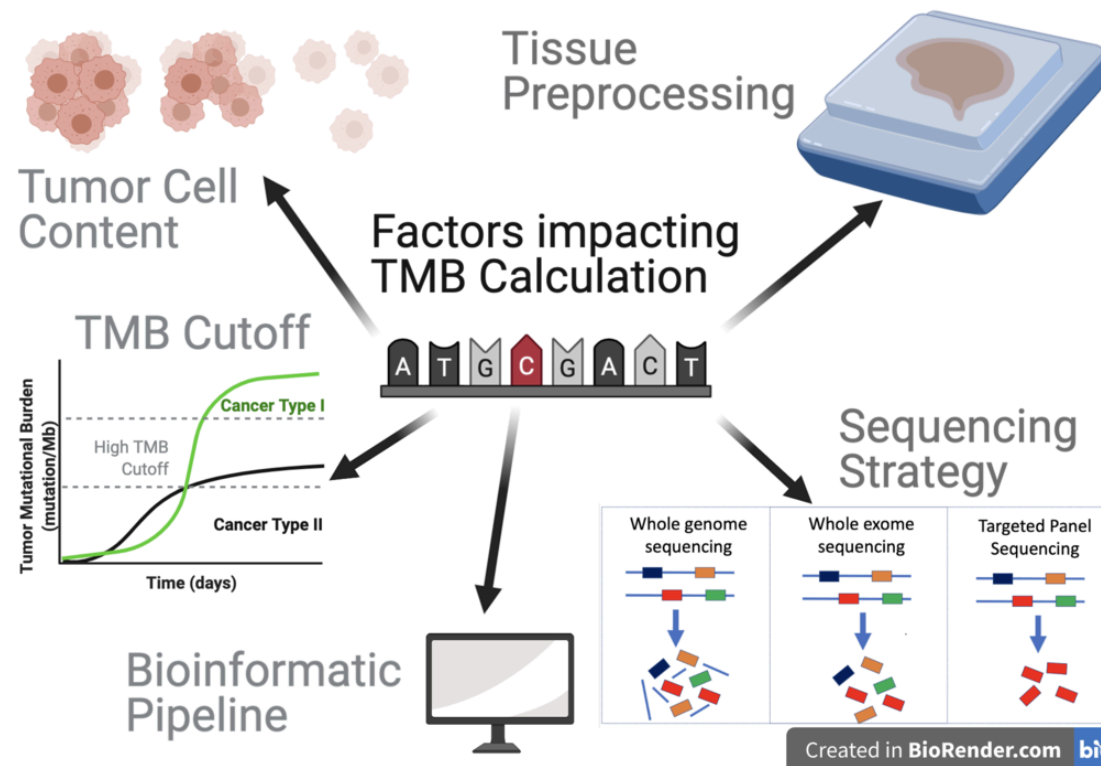
```
(# of mutations)/(coverage space) * 1,000,000
```

Why do we care?

- Knowing the tumor mutational burden may help plan the best treatment.
- Tumors that have a high number of mutations appear to be more likely to respond to certain types of immunotherapy (e.g. [Keytruda](#)). Tumor mutational burden (TMB) is being used as a type of biomarker. ³

Mutation Burden

Factors that influence TMB calculation



- Greater tumor cell content and sequencing coverage enable more accurate TMB calculation.³

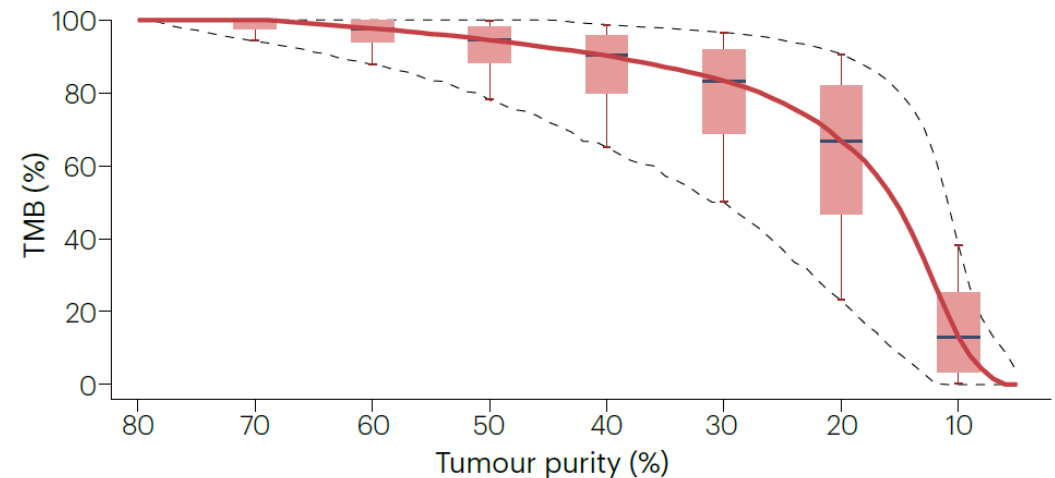
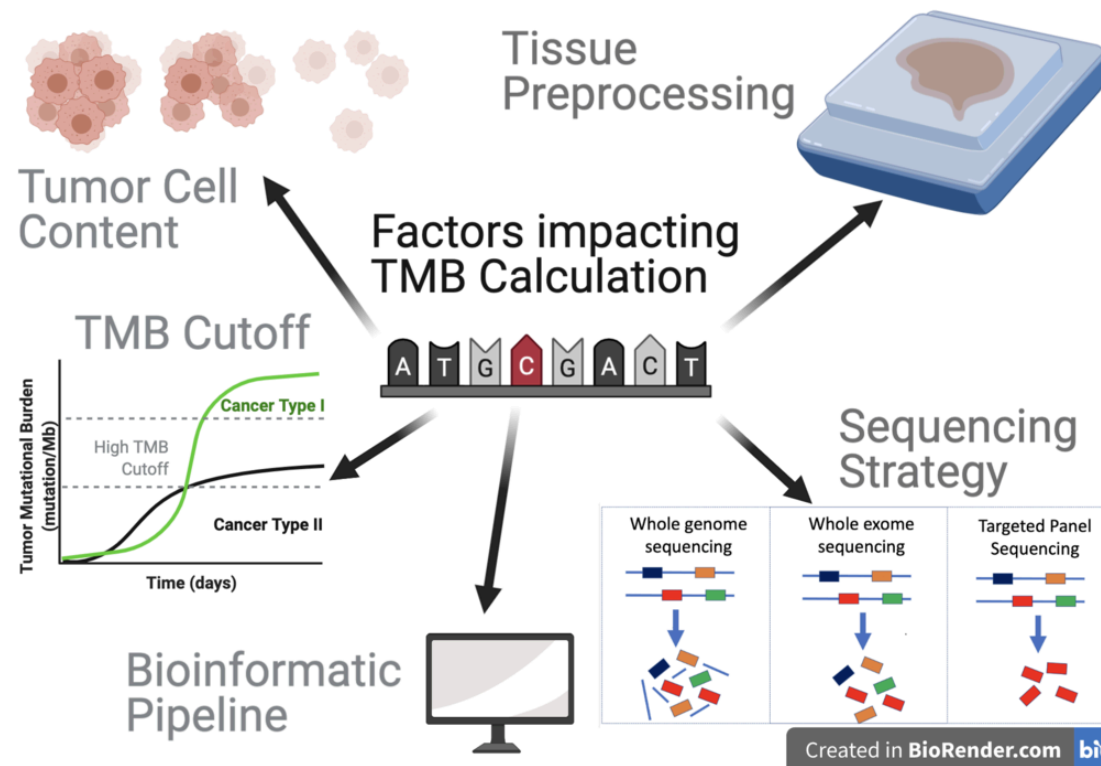


Image Source

Mutation Burden

Factors that influence TMB calculation

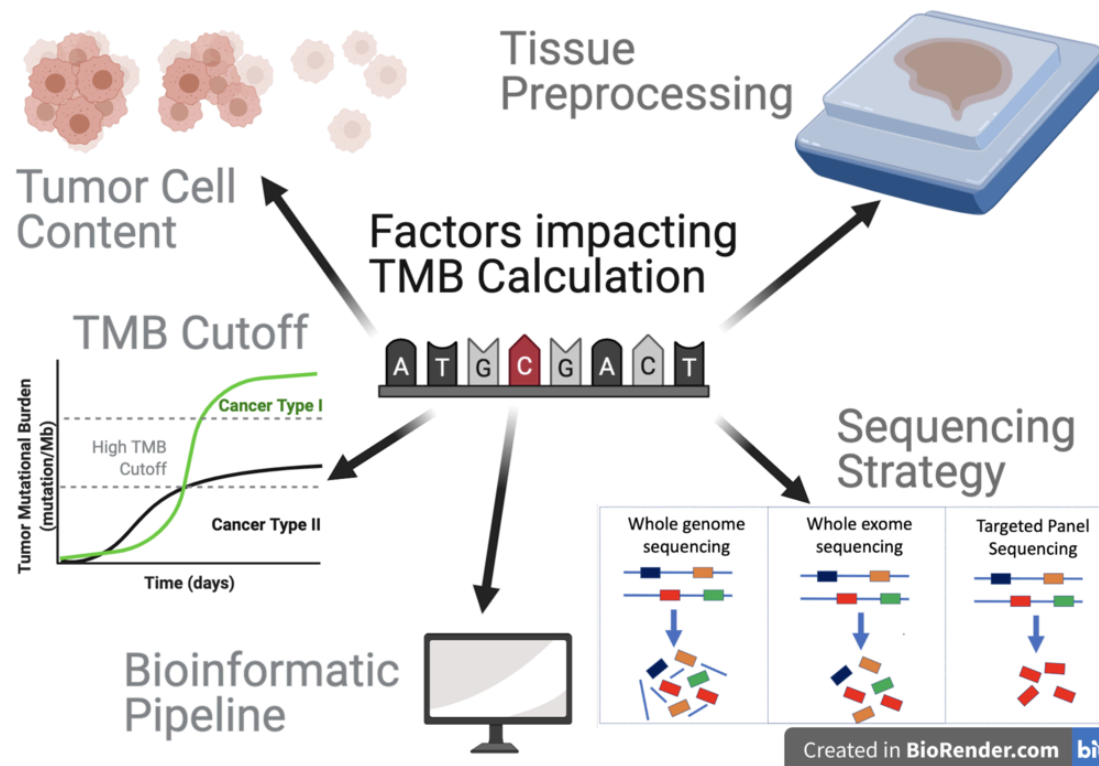


- Tumor tissues are fixated in formalin to preserve tissue and cellular morphology in the formalin-fixed paraffin-embedded (FFPE) protocols have sequence artifacts that can affect TMB calculations.³

Image Source

Mutation Burden

Factors that influence TMB calculation

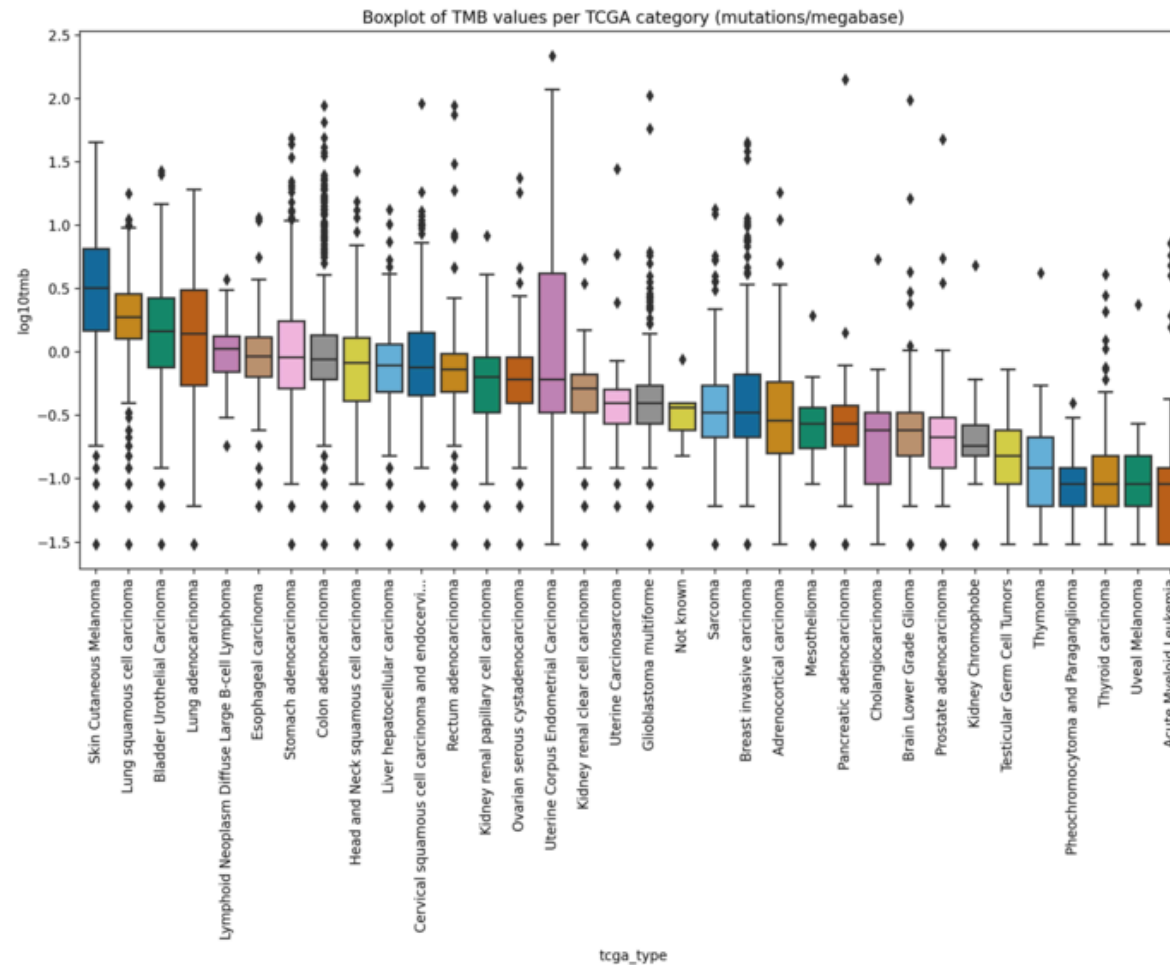


- The greater number of genes (e.g. WES, WGS) sequenced the greater the confidence of TMB estimation.³
- TMB calculations generally filter out synonymous variants and germline variants.³ **Did your variant calling pipeline do this?**
 - The choice of variant callers and other software in the downstream analyses may also affect TMB calculation.³

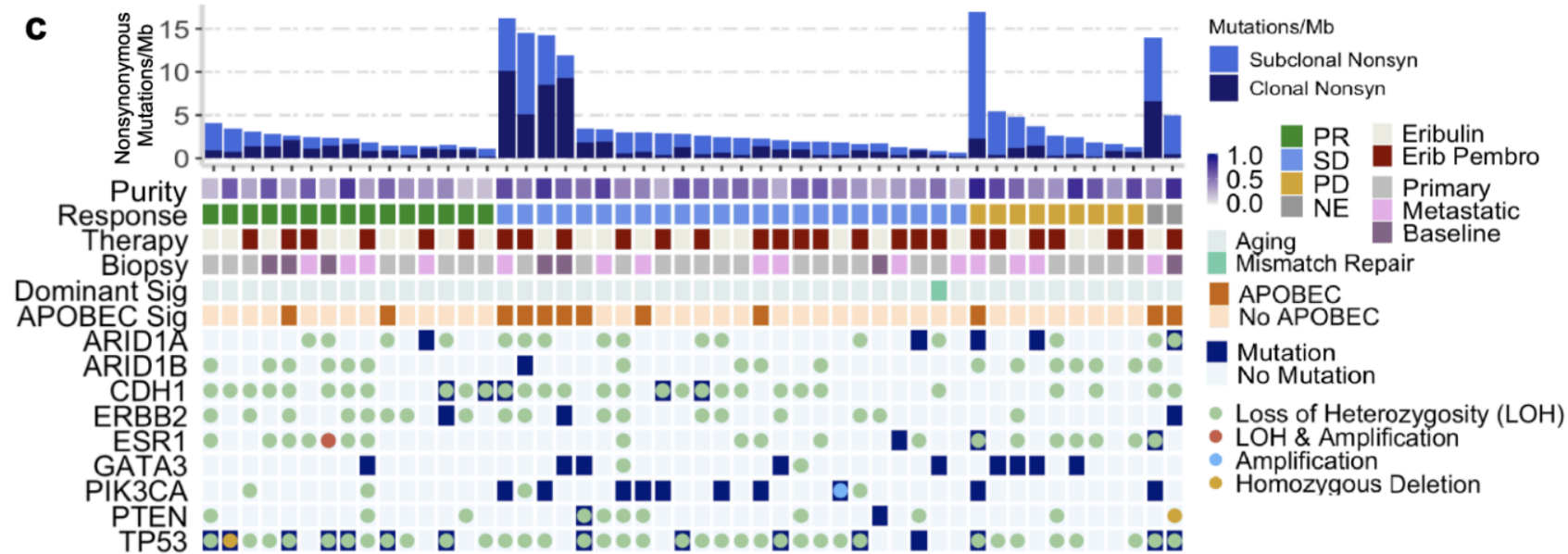
Image Source

TMB Cut-off

TMB varies by cancer type, so there is no single best TMB cut-off threshold



How to read a waterfall plot

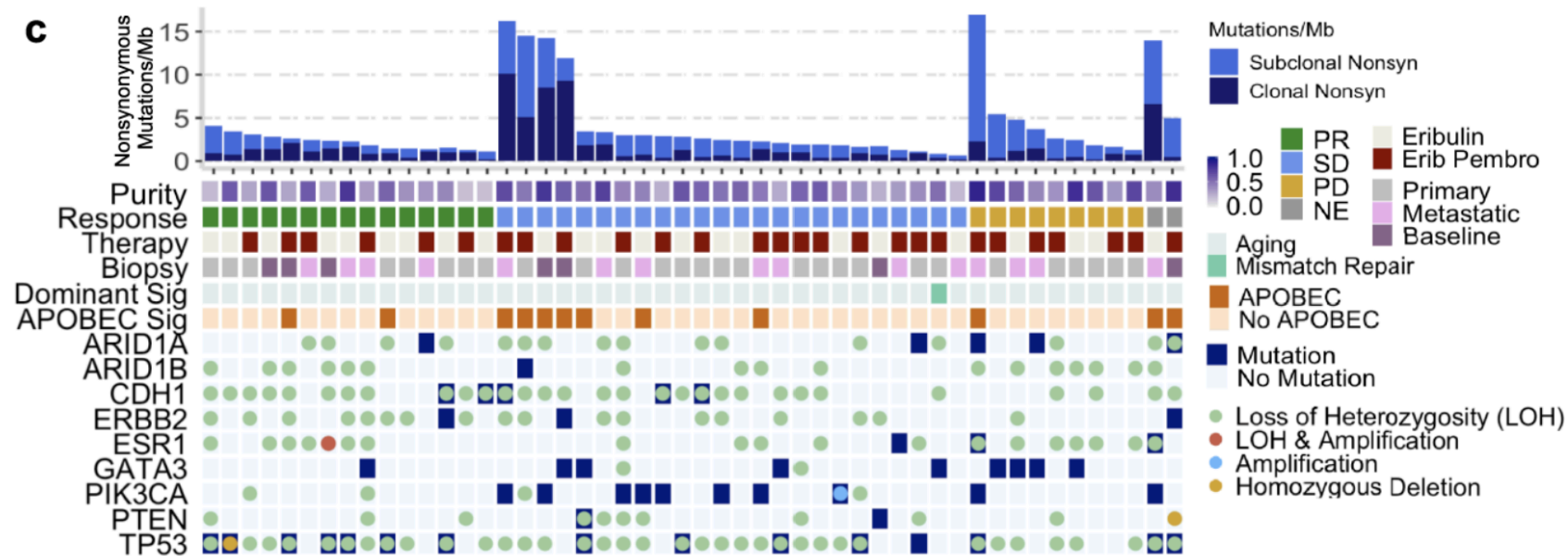


NE=not evaluable;
PD=progressive disease;
PR=partial response;
SD=stable disease

Source: [Keenan et al. Nat Commun \(2021\)](#)

- Some samples have high mutational loads dominated by clonal mutations (suggesting early events in tumor evolution) or subclonal mutations (indicating later events).

How to read a waterfall plot

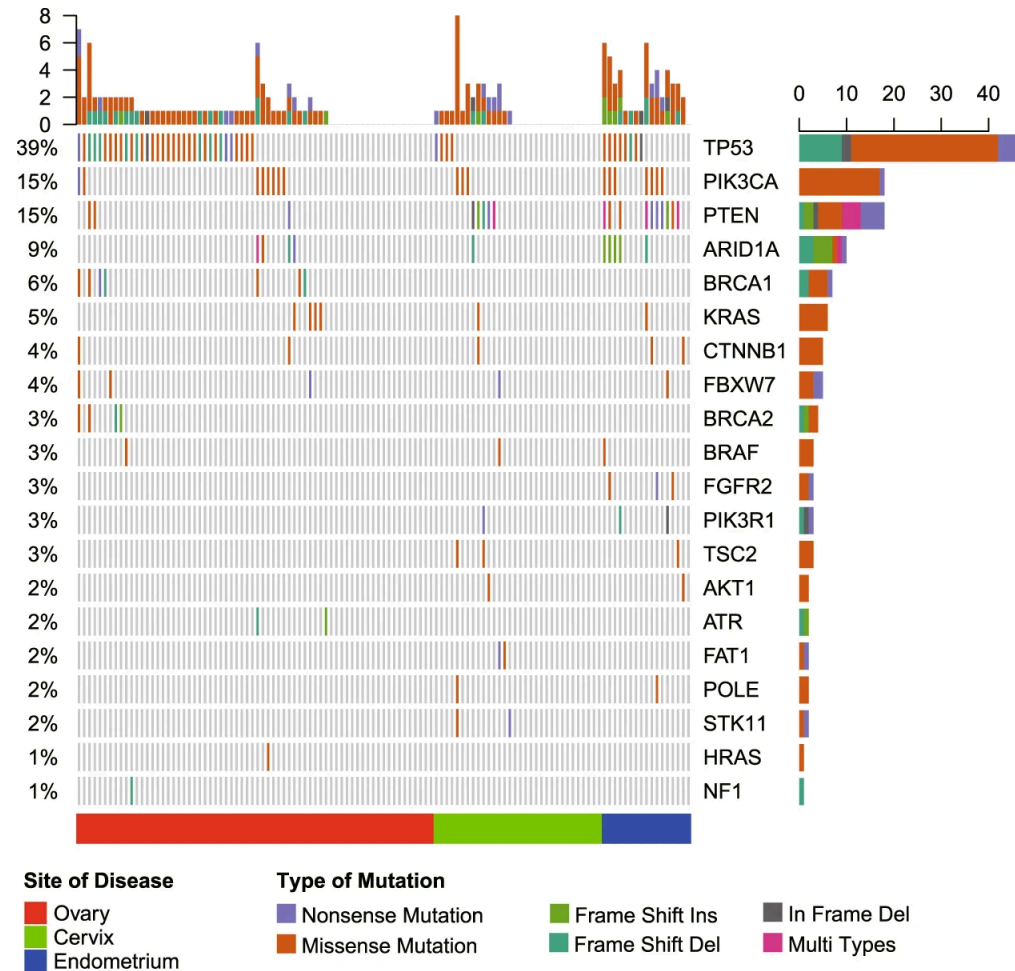


NE=not evaluable;
PD=progressive disease;
PR=partial response;
SD=stable disease

Source: [Keenan et al. Nat Commun \(2021\)](#)

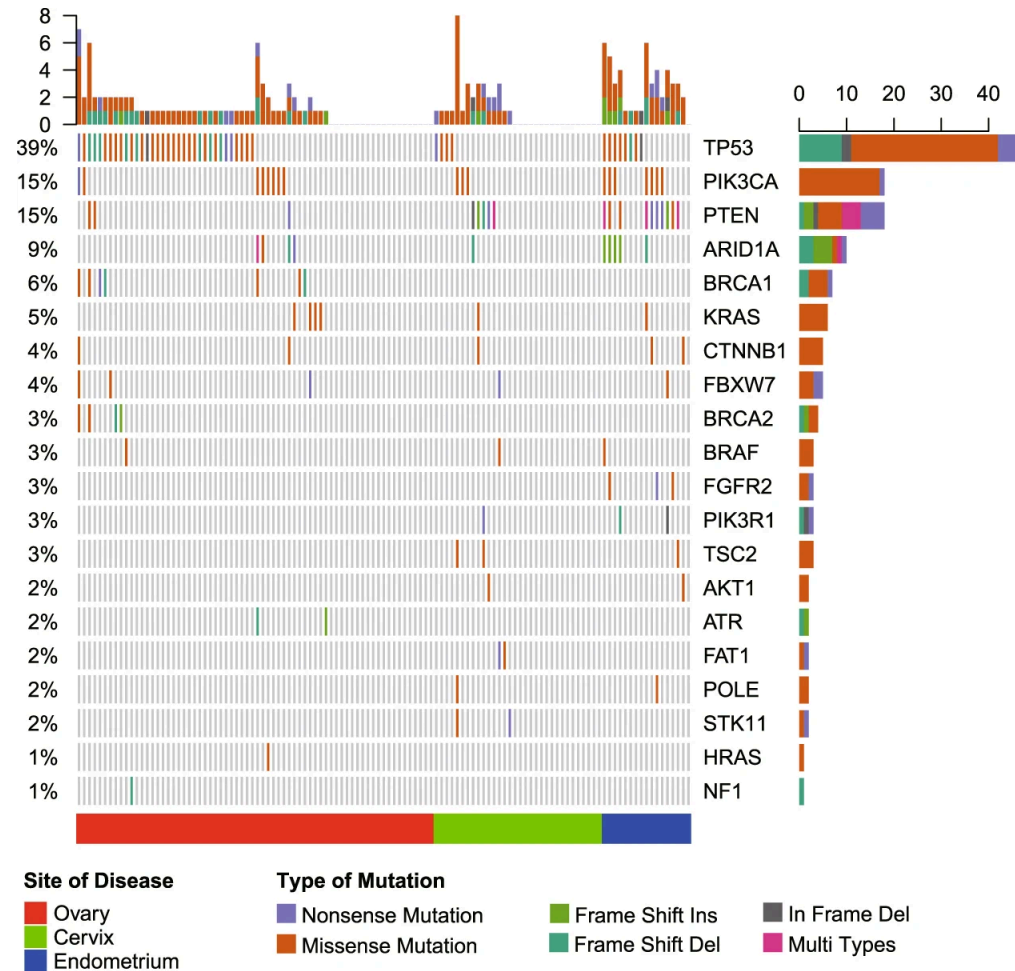
- Genes like PIK3CA and GATA3 show frequent mutation (high % samples with mutation)

How to read a waterfall plot



- The mutations in TP53 and PIK3CA are mostly missense mutations.
- BRCA1 and BRCA2 mutations occurred in affected ovary tissues but not in cervix and endometrium tissues.

How to read a waterfall plot



- TP53 is the most frequently mutated gene. It has a high number of missense mutations (orange) and some nonsense mutations (purple), which is consistent with its role as a tumor suppressor gene. Most of the cancers demonstrate a biallelic loss-of-function mutation in TP53.⁴

Ti/Tv plot

What does it do?

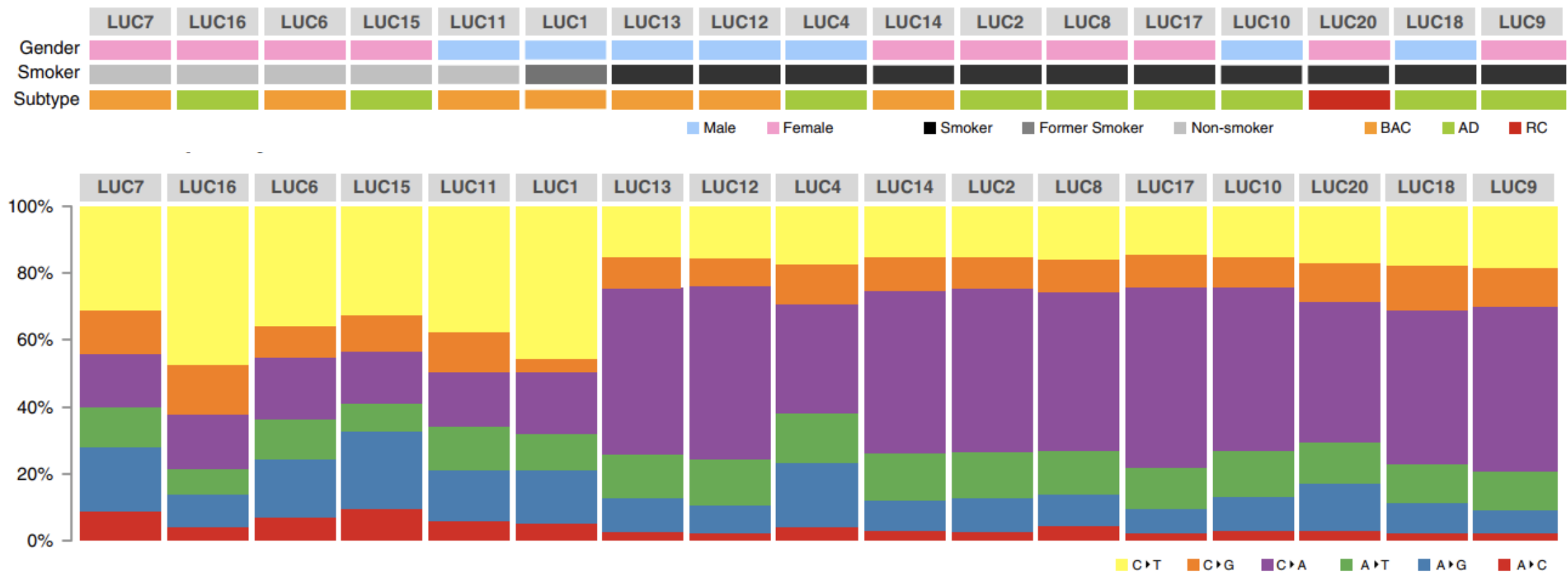
- Visualize Transition and Transversion proportions
- Visualize Transition and Transversion frequencies
- Compare these proportions with expectations

Why do we care?

- Visualize mutation profile patterns
- Ex. Smoking tends to increase G>T or C>A transversions due to oxidative damage
- Practical application verify smoking status of patients with lung carcinomas

Ti/Tv plot

Clinical Data



Source: Tan *et al.* Cell Reports Medicine (2023)

Lab 1b Instructions

- Lab 1b runs on R kernel in a Jupyter notebook, but you may create an `.R` / `.rmd` file to run your code using the RStudio IDE on CoCalc (see instructions [here](#))
- Do make sure your final code is in `lab_1b.ipynb` . This is version that I'll collect.
- You may work in groups, but the final submission is individual.
- If you complete the lab early, you may work on **Group Assignment #1**.

Lab 1b Correction: Mutation Burden in Waterfall plot

"If a cancer has a high tumor mutation burden level of **10 mut/Mb** or greater, it's more likely to respond to immunotherapy medication. This means that a specific drug type—immune checkpoint inhibitors—may activate the immune system to help the body recognize cancer cells. With more mutations, the higher the chance that one of the mutations will be able to identify and target cancer cells."

"Cancers with low tumor mutation burden levels of **10 mut/Mb** or lower may be less successful in activating the immune system to identify and target cancer cells."

Ignore the numeric TMB cut-off threshold