

Enhancing SNV Detection in Clinical FFPE Data

Leveraging Mutational Signatures for Artifact Correction

Tara Friedrich

GitHub: [staracode/ffpe_code](https://github.com/staracode/ffpe_code)

Motivation

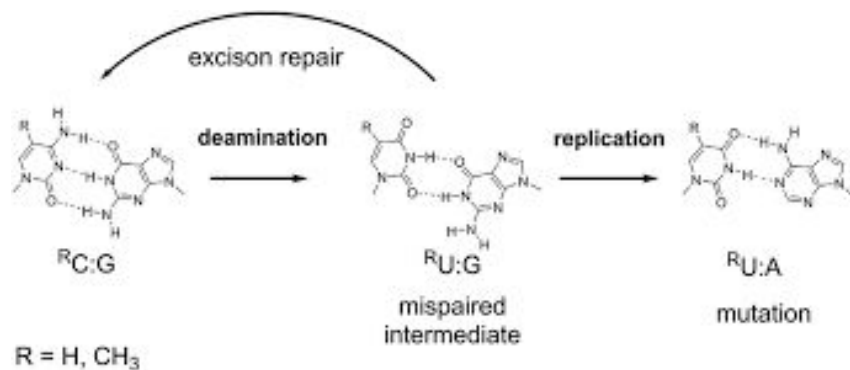
Why FFPE Data?

- Huge repositories of archived FFPE samples
- Easy preservation preferred by clinics
- Often intertwined well with histology and clinical metadata
- Opportunity to unlock insights from legacy data

FFPE Artifacts: The Core Challenge

What goes wrong with FFPE samples?

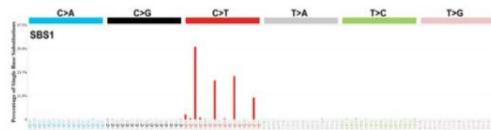
- DNA fragmentation
- Chemical modifications (e.g., cytosine deamination)
- Result: High false-positive SNV calls
- Especially problematic for C>T/G>A transitions



Can We Fix FFPE Artifacts?

Hypothesis: [Mutational signatures](#) (SBS Signature) can identify and correct FFPE noise

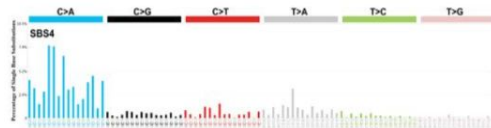
- Use statistical models to distinguish true mutations from artifacts
- Particularly useful in large-scale datasets



SBS1

Proposed Aetiology

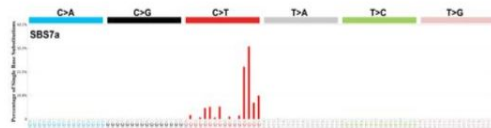
Spontaneous deamination of 5-methylcytosine
(clock-like signature)



SBS4

Proposed Aetiology

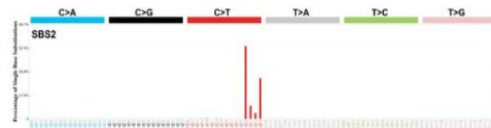
Tobacco smoking



SBS7a

Proposed Aetiology

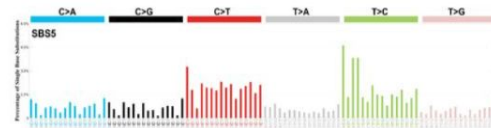
Ultraviolet light exposure



SBS2

Proposed Aetiology

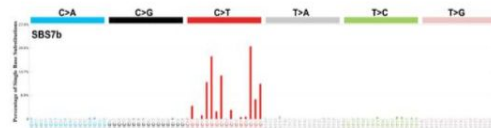
Activity of APOBEC family of cytidine deaminases



SBS5

Proposed Aetiology

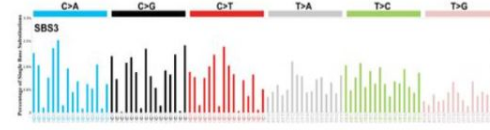
Unknown (clock-like signature)



SBS7b

Proposed Aetiology

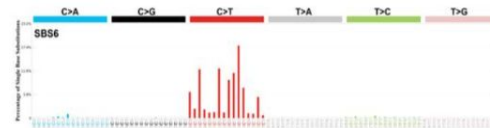
Ultraviolet light exposure



SBS3

Proposed Aetiology

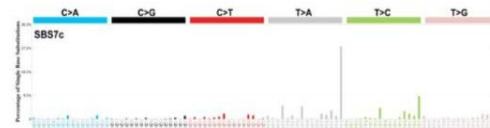
Defective homologous recombination DNA
damage repair



SBS6

Proposed Aetiology

Defective DNA mismatch repair



SBS7c

Proposed Aetiology

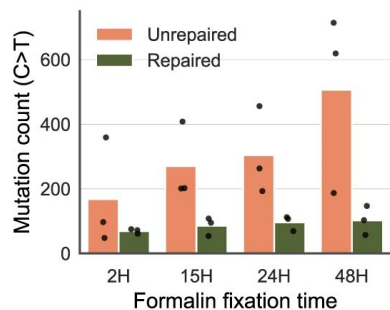
Ultraviolet light exposure

Foundational Work – Paper 1

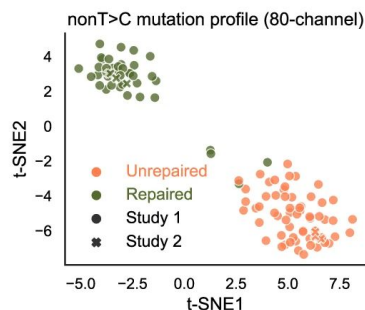
[Wong et al., Nature Communications \(2022\)](#)

- Used signature deconvolution to separate artifact vs biological signal
- Distinguished between **repaired** (UDG treatment) and **unrepaired** FFPE
- Enabled better variant calling accuracy

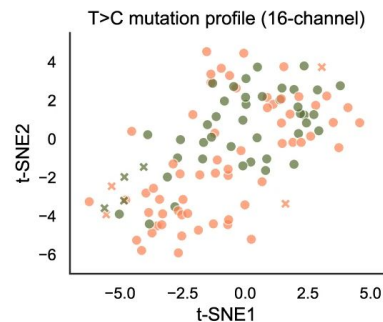
a



b

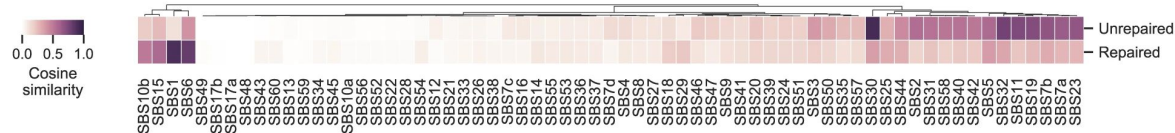


c

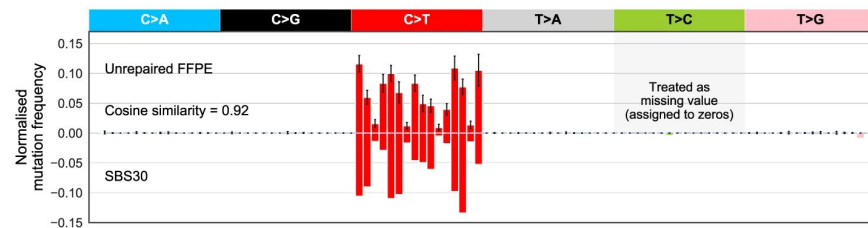


FFPE signature matches cosmic signatures

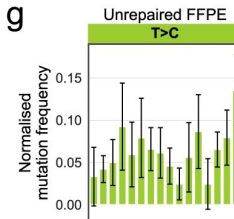
d



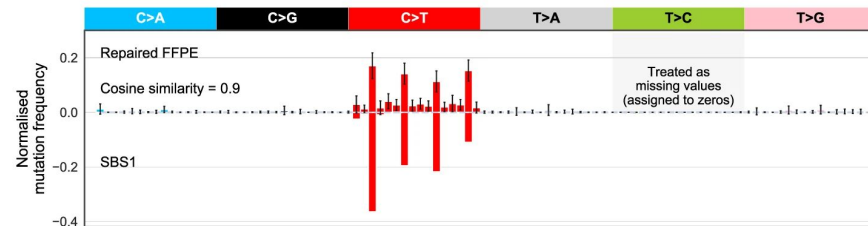
e



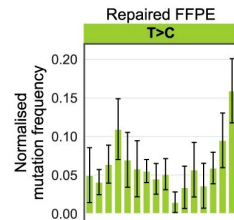
g



f



h



Follow-up Work – Paper 2

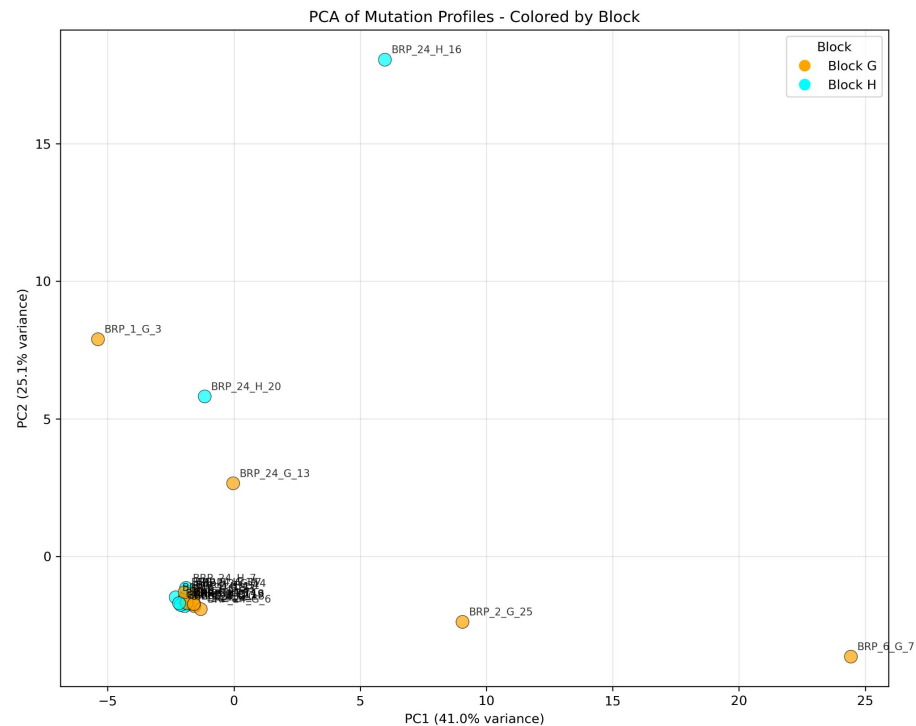
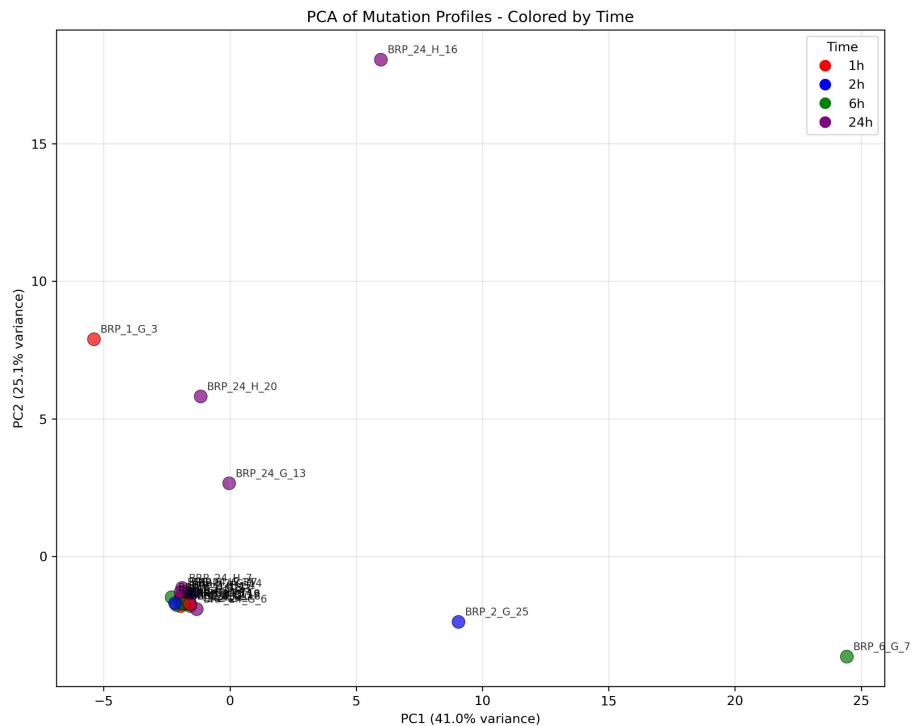
Genome Biology (2022)

- Used a well-characterized **diploid male lymphoblast cell line**, processed into FFPE cell blocks
- Tested **four different formalin fixation times** (1 hr, 2 hrs, 6 hrs, 24 hrs) and compared to flash frozen.
- Generated **96 FFPE sections**, distributed across four labs for targeted sequencing using four commercial oncopanels (AZ650, BRP, ILM, TFS)
- Inner FFPE sections (excludes top and bottom 200 μm) retain sequencing integrity—surface sections do not

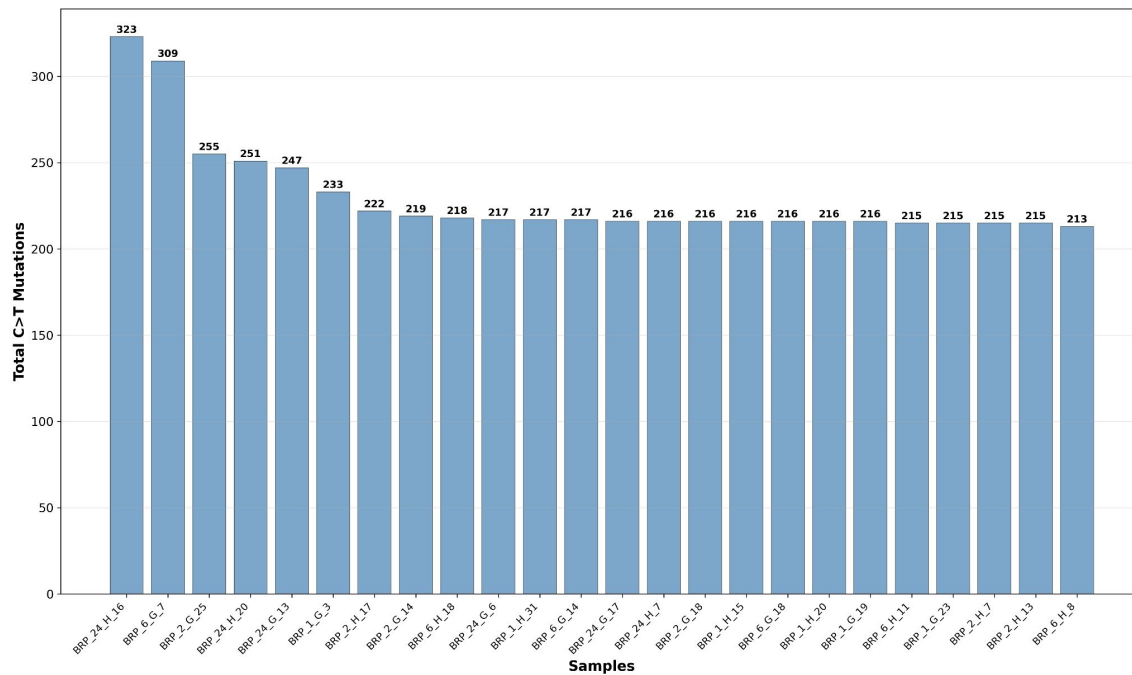
My Contribution

- Built a reproducible pipeline: [GitHub Repo](#)
- Parsed and visualized FFPE mutation counts
- Compared repaired vs unrepaired signatures
- Applied model from first paper to data from second paper

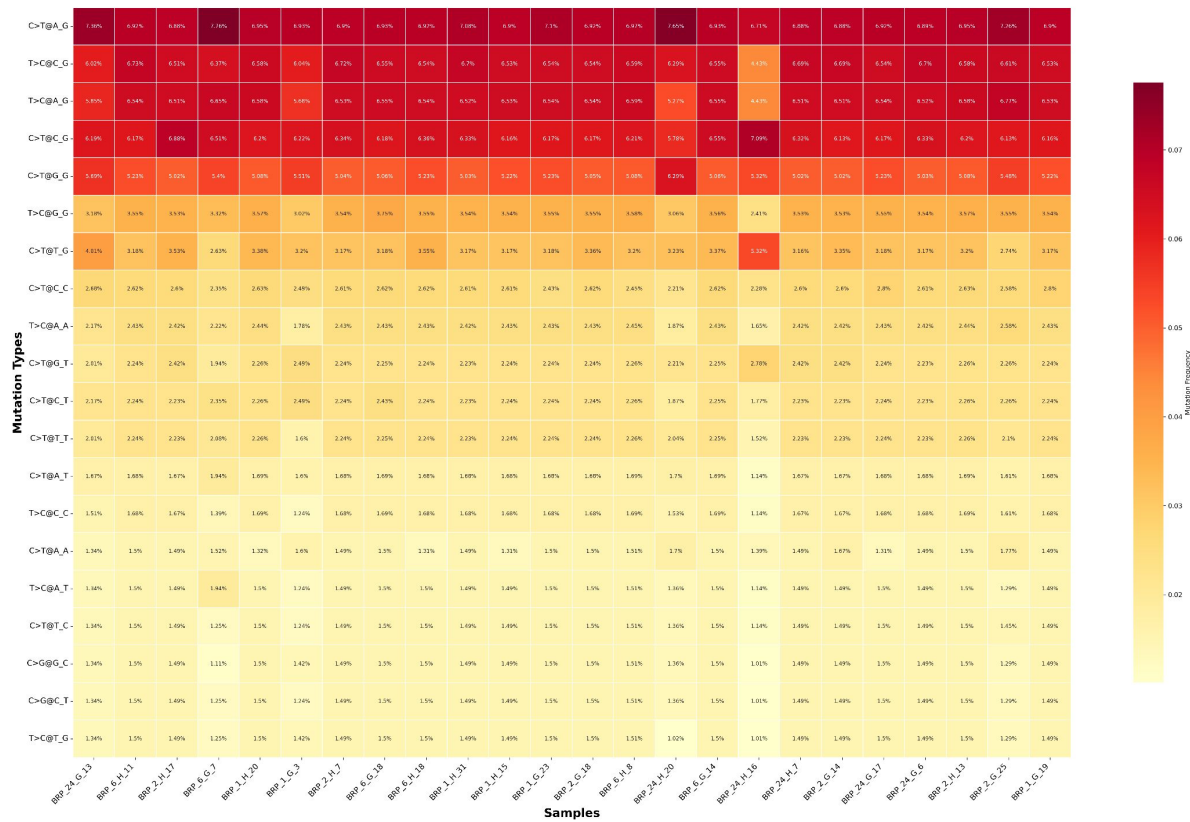
Sample QC



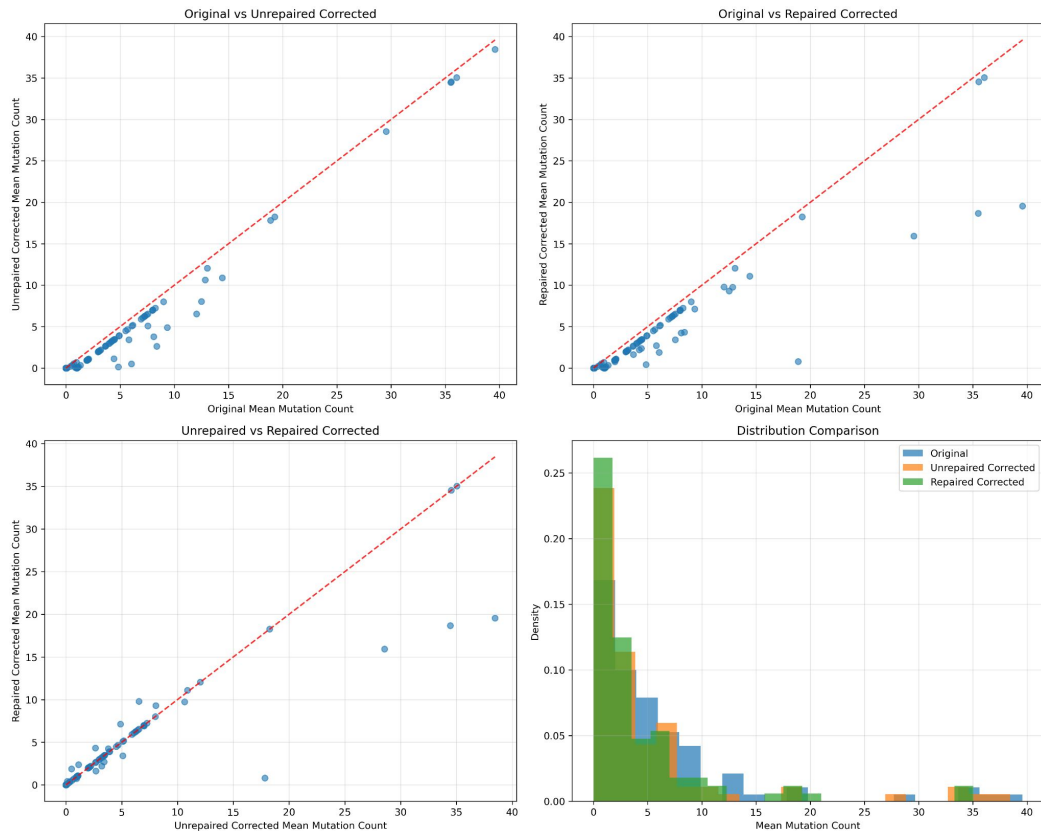
C >T transitions by sample



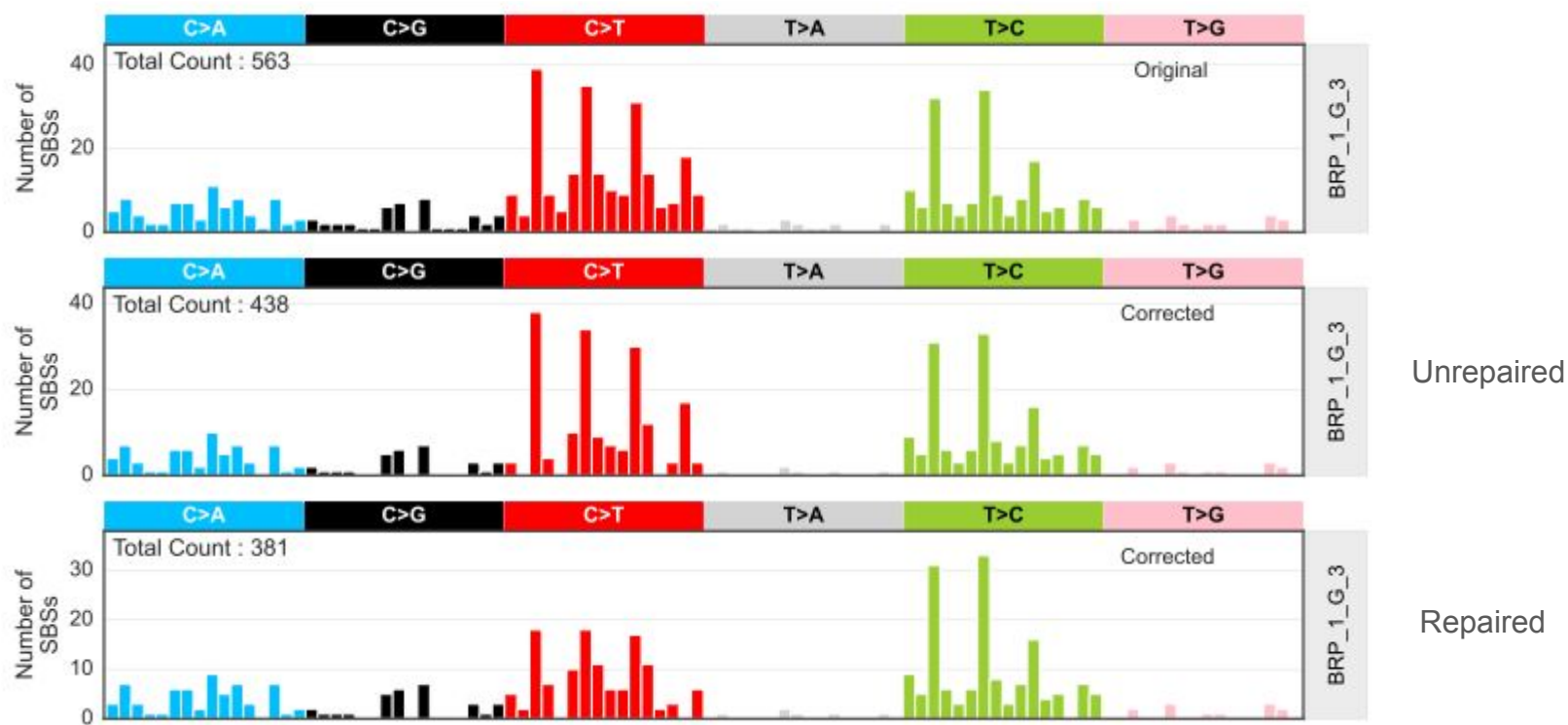
Top 20 Mutation Context Frequencies



Effect of FFPE signature correction



Effect of FFPE signature correction



Limitations & Future Work

- FFPEsig less effective on **low-SNR** samples
 - FFPEsig was shown to be less effective in cancers with less signal-to-noise
- Missing information about samples (eg. block position) and UDG treatment status
- Missing Flash Frozen truth set
- Potential: build a classifier to flag likely FFPE-induced artifact or flag samples with excess damage

Thank You

Questions?

Tara Friedrich

[GitHub: staracode/ffpe_code](#)