Shifting from DNA-Based to RNA-Informed Somatic Mutation Detection: A Systematic Benchmark Across Different Tumor Types

Charlotte Stubbe (1, 2, 3), Jasper Anckaert (1), Philippe Decruyenaere (1, 2, 3), Fritz Offner (2, 3), Pieter Mestdagh (1, 3), Jo Vandesompele (1, 3)

- OncoRNALab, Department of Biomolecular Medicine, Center for Medical Genetics Ghent (CMGG), Ghent University
- Department of Internal Medicine and Pediatrics, Ghent University; Department of Hematology, Ghent University Hospital
- Cancer Research Institute Ghent (CRIG), Ghent, Belgium

1. Context

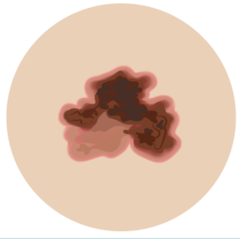
Benefits of tumor-only RNA variant calling

- cost-effective
- **actionable mutations**
- multi-omics approach
- no need for matched normal

Diffuse Large B-Cell Lymphoma (DLBCL)

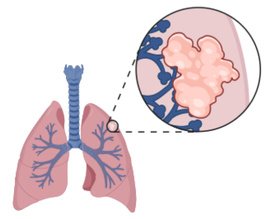
- most common subtype of **Non-Hodgkin Lymphoma**
- most prevalent lymphoid malignancy in adults
- **150.000** new diagnoses annually worldwide

Skin Cutaneous Melanoma (SKCM)



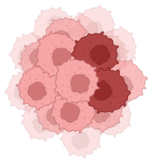
- most severe type of **skin cancer** but representing only 1 % of all types
- responsible for 80 % of skin cancer related deaths
- **325.000** new diagnoses annually worldwide

Lung Adenocarcinoma (LUAD)



- most common histological subtype of **lung cancer**
- 40 % of all lung cancer types
- **542.000 & 717.000** new diagnoses anually worldwide (male and female respectively)

What makes mutation calling in RNA "easier"?



Tumor purity

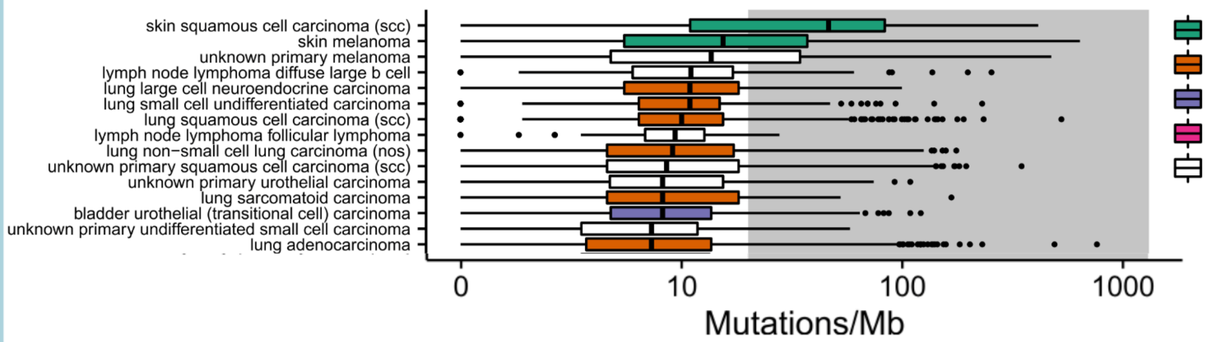
Tumor vs normal DNA

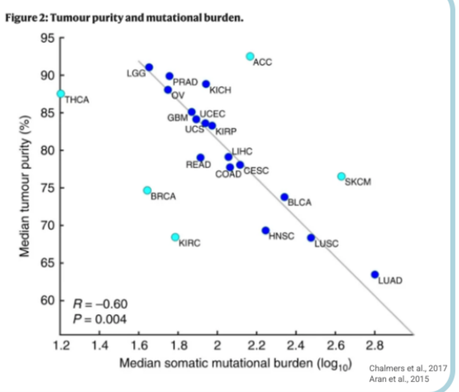
Variant Allele Frequency (VAF)

$$\frac{\text{Variant reads}}{\text{Total reads}} * 100$$

Tumor mutational burden (TMB)

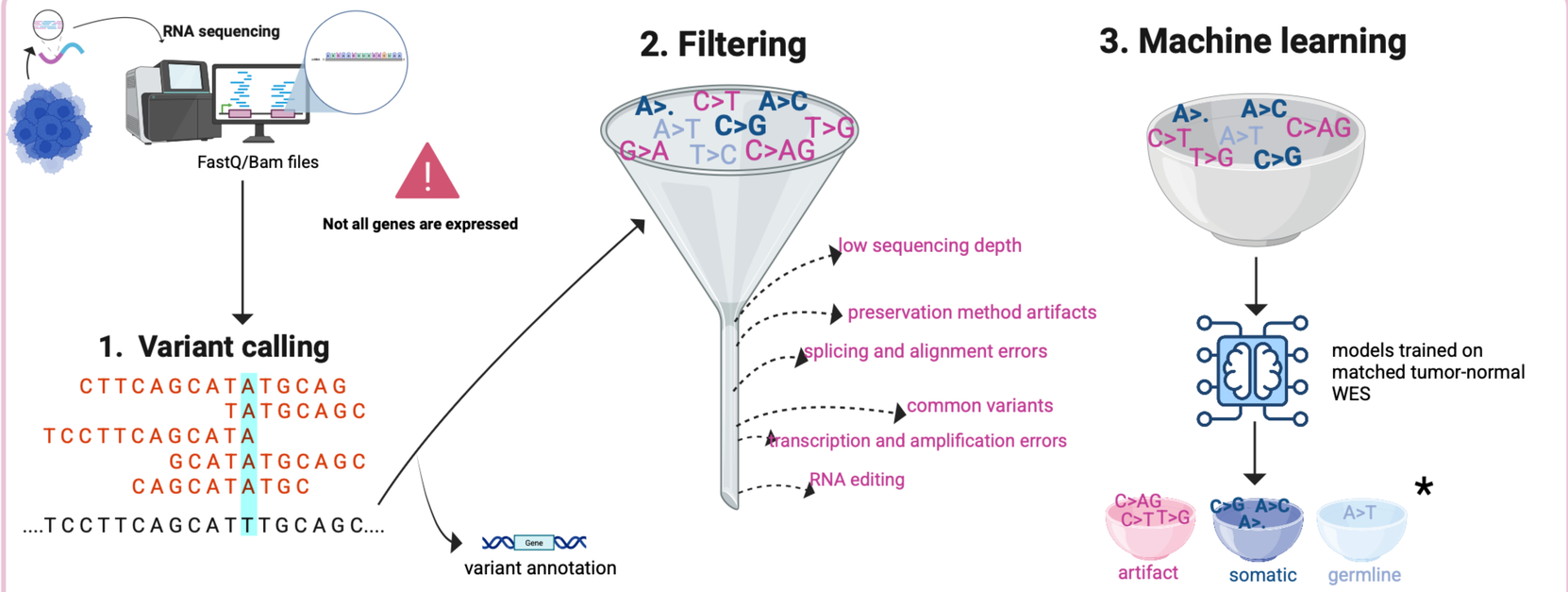
Total number of non-silent somatic mutations within a cancer's cell genome

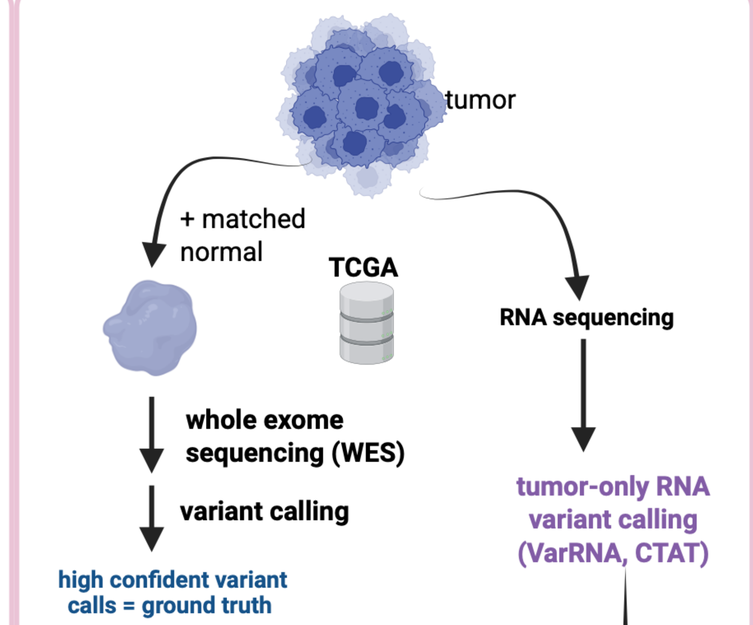




RNA variant calling in tumor-only samples enables the detection of expressed, functionally active alterations, facilitating precision medicine approaches

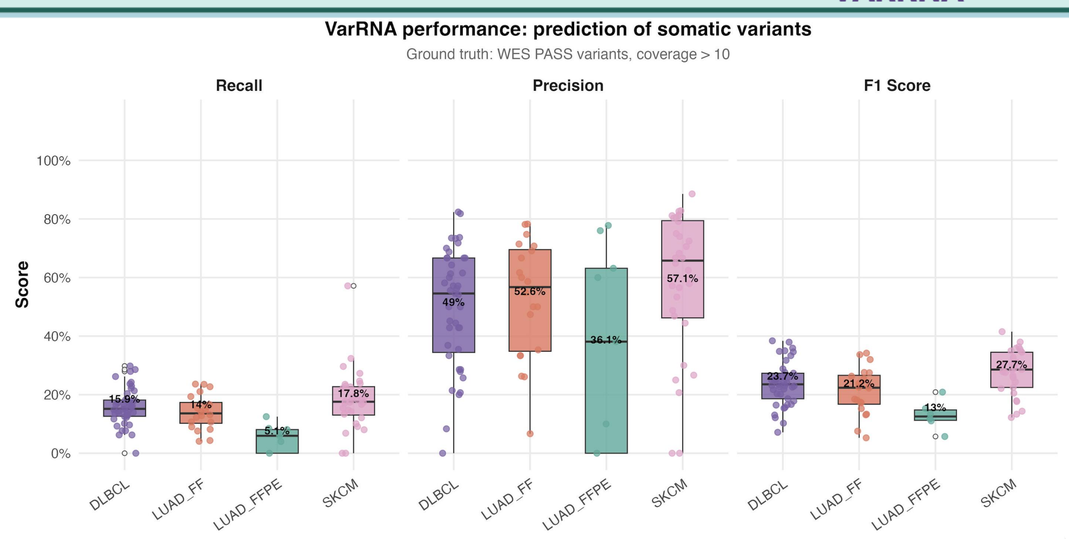
2. Methodology

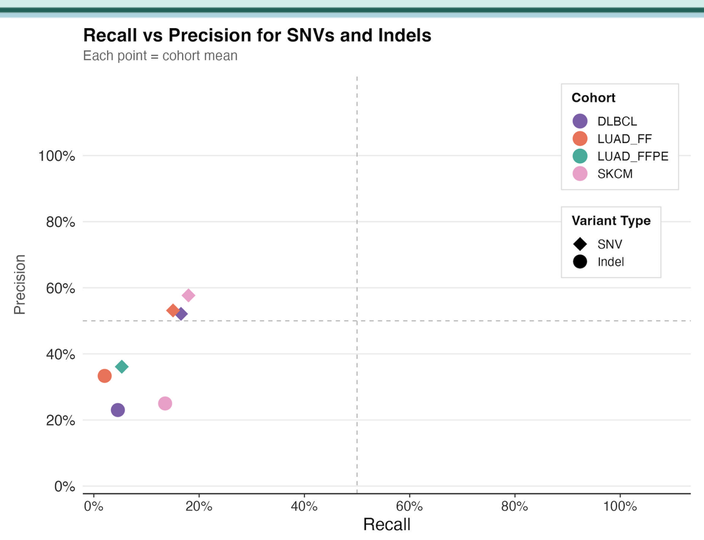


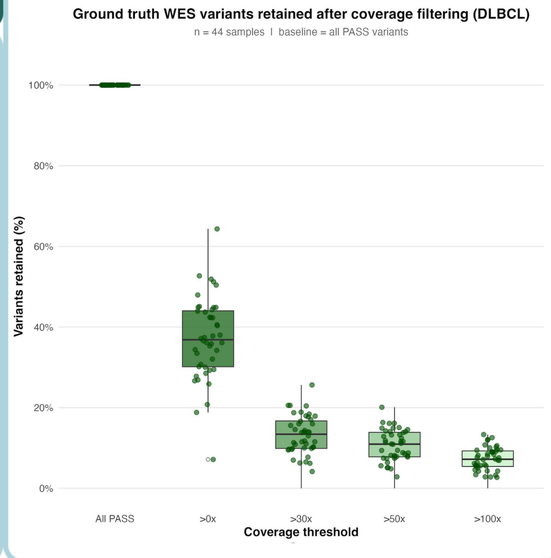


3. Results

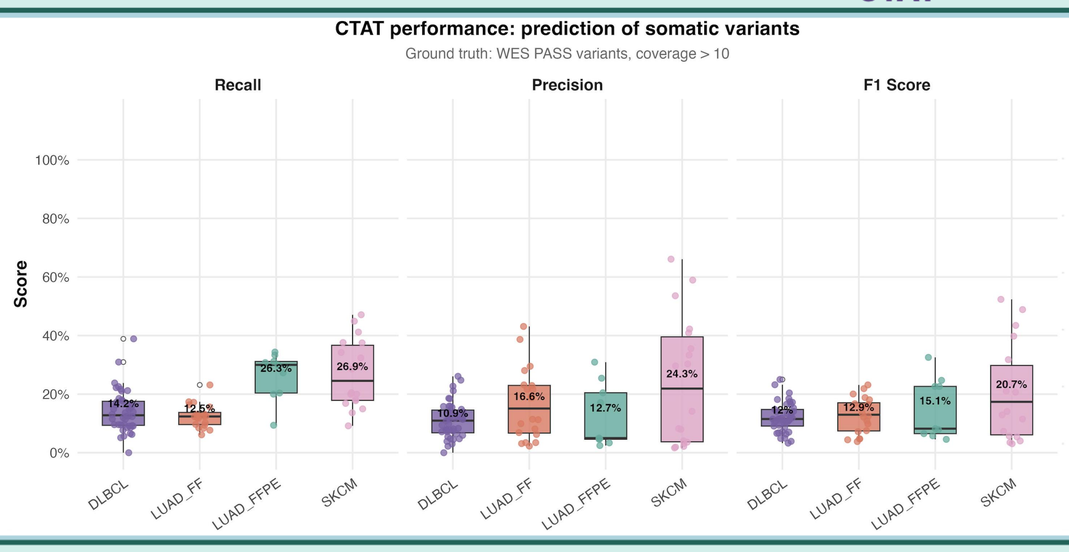
VARRNA

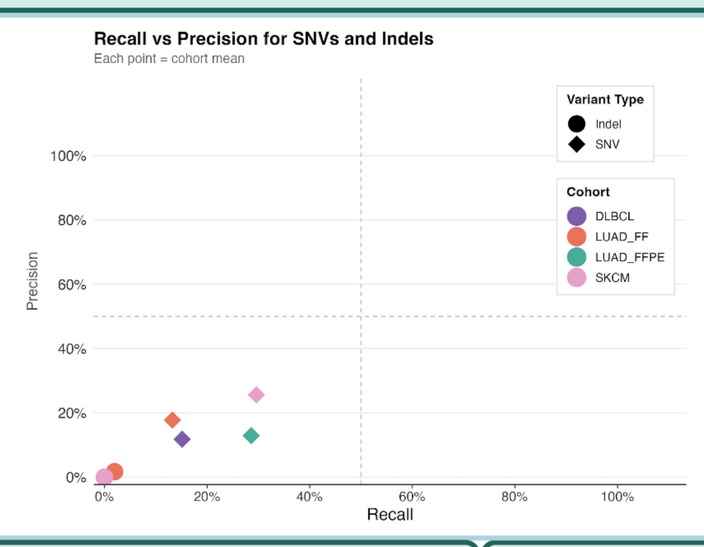


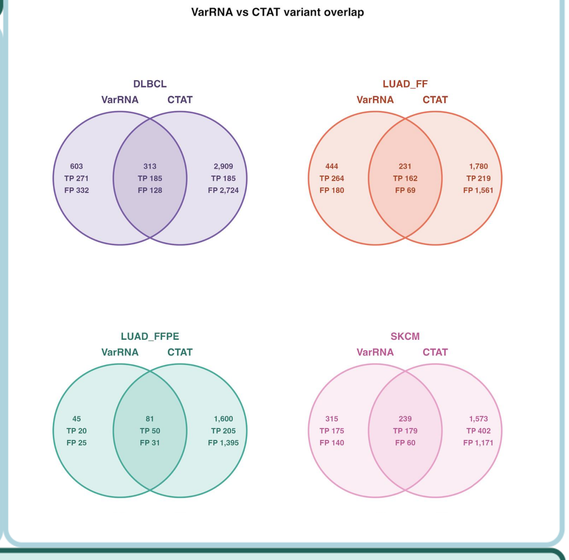




CTAT







4. Conclusions & future directions

- **VarRNA outperforms CTAT across cohorts**, with diminished performance in FFPE samples driven by reduced recall
- Both methods achieve **higher accuracy for SNVs than for indels**
- **FFPE samples show overall reduced performance**, likely driven by artefacts such as cytosine deamination
- **Limited concordance between variant calling tools highlights substantial inter-tool variability in detected mutations**

- **Optimal VarRNA performance requires cohort-specific model training** to account for heterogeneity in cancer types, preservation methods, library preparation, and sequencing technologies
- **A third, independent variant calling pipeline is required to enable robust benchmarking**

5. References

1. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Medicine. 2017;9(1):34.
2. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nature Communications. 2015;6(1):8971.
3. Bollas A, Gaither J, Schieffer KM, White P, Mardis ER. Variant calling from RNA-Seq data reveals allele- specific differential expression of pathogenic cancer variants. Communications Medicine. 2025;5(1):202.
4. Fangal VD. CTAT mutations: a machine learning based rna-seq variant calling pipeline incorporating variant annotation, prioritization, and visualization: Harvard University; 2020.

Illustrations created with BioRender.com