Long-read single-cell genome, transcriptome and open chromatin profiling links genotype to phenotypes

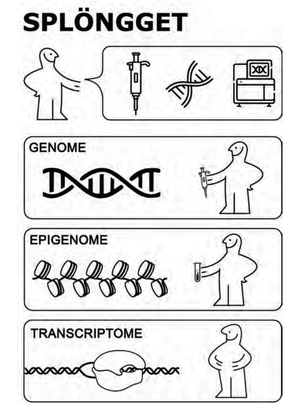
Alexandra Pančíková[1,2,3,4,*], Ruben Cools[1,2,3,*], Marios Eftychiou[1,2,3,5,*], Margo Aertgeerts[1,7], Joris Vande Velde[1,2,3], Heidi Segers[7,8], Jan Cools[1,6], Luuk Harbers[1,$], Jonas Demeulemeester[1,2,3,$,#]

1 VIB Center for Cancer Biology, VIB, Leuven, Belgium
2 Laboratory of Integrative Cancer Genomics, Department of Oncology, KU Leuven, Leuven, Belgium
3 VIB Center for AI and Computational Biology, VIB, Leuven, Belgium
4 Laboratory for Computational Biology, Department of Human Genetics, KU Leuven, Leuven, Belgium
5 Laboratory of Multi-omic Integrative Bioinformatics, Department of Human Genetics, KU Leuven, Leuven, Belgium
6 Laboratory of Molecular Biology of Leukaemia, Department of Human Genetics, KU Leuven, Leuven, Belgium
7 Paediatric Oncology, Department of Oncology, KU Leuven, Leuven, Belgium
8 Department of Paediatric Hematology and Oncology, UZ Leuven, Leuven, Belgium

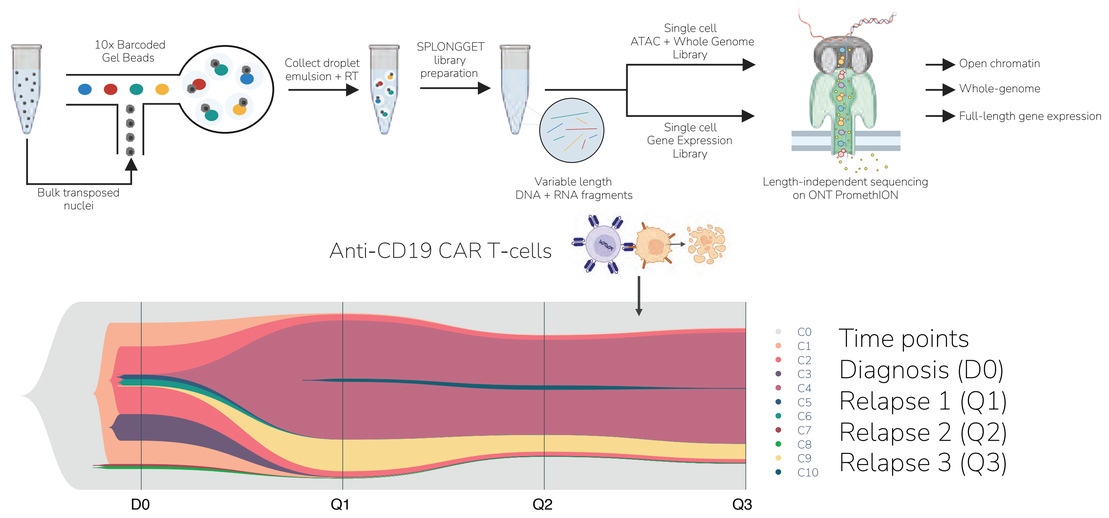
* Shared first authors
$ Jointly supervised
# Corresponding author

Background

Current single-cell multiomics methods typically provide limited genomic information, constraining genotype-phenotype studies. To address this gap, we developed **SPLONGGET** (**S**ingle-cell **P**rofiling of **LONG**-read **G**enome, **E**pigenome, and **T**ranscriptome), which integrates 10X Genomics barcoding with Oxford Nanopore sequencing to simultaneously profile **genome, chromatin accessibility, and full-length transcriptome** in thousands of **single cells**. By retaining all tagmentation fragments during library preparation, SPLONGGET delivers whole-genome coverage, supports target enrichment for effective single-cell genotyping, and remains backwards compatible with existing short-read workflows. SPLONGGET enables comprehensive calling of small variants, structural variants, and copy number alterations. Applying SPLONGGET to paediatric B-cell acute lymphoblastic leukaemia revealed clonal dynamics and the phenotypic effects of somatic variants. Notably, we evidence parallel evolution of immune escape variants with four distinct splice site mutations and loss of heterozygosity in the CAR T-cell therapy *CD19* target. In conclusion, SPLONGGET enables integrated high-throughput analysis of genetic variation and molecular phenotypes using off-the-shelf kits, offering a timely and powerful tool to study genetically heterogeneous samples, such as tumours but also ageing normal tissues



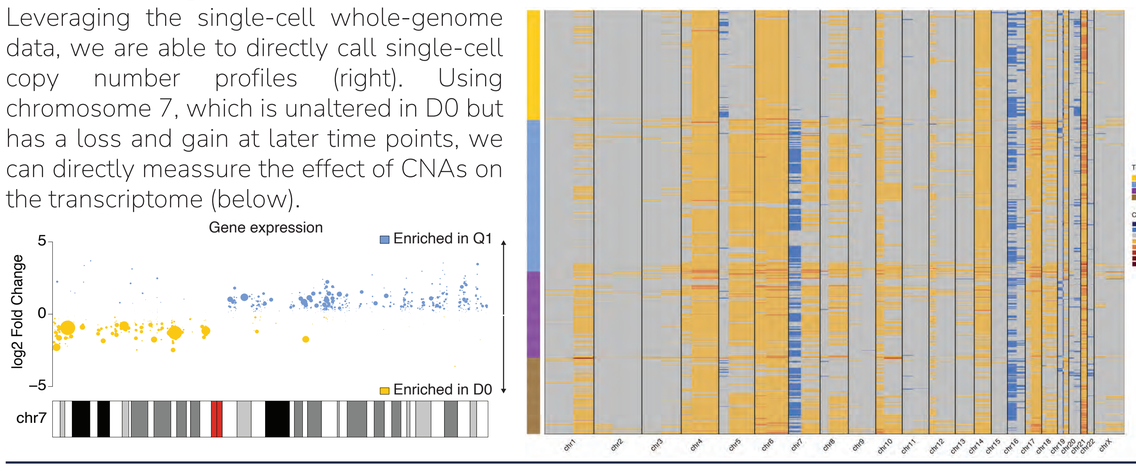
SPLONGGET schematic



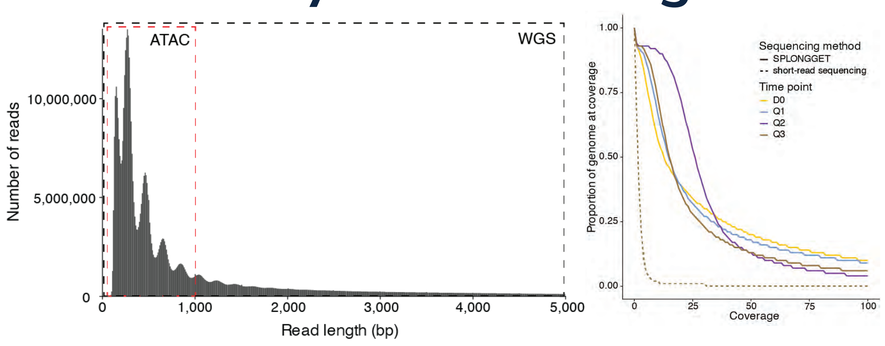
SPLONGGET leverages the widely available 10X Genomics Single Cell Multiome kit to tag genomic fragments and transcripts with consistent barcode pairs. We then read out genomic fragments of all fragment sizes, as well as full length cDNA using Nanopore sequencing (top). We applied SPLONGGET to bone marrow samples from a patient affected by paediatric B-cell Acute Lymphoblastic Leukaemia (B-ALL) across four different timepoints: at diagnosis and at three relapses (bottom).

Single-cell copy number alterations affect gene expression

Leveraging the single-cell whole-genome data, we are able to directly call single-cell copy number profiles (right). Using chromosome 7, which is unaltered in D0 but has a loss and gain at later time points, we can directly meassure the effect of CNAs on the transcriptome (below).

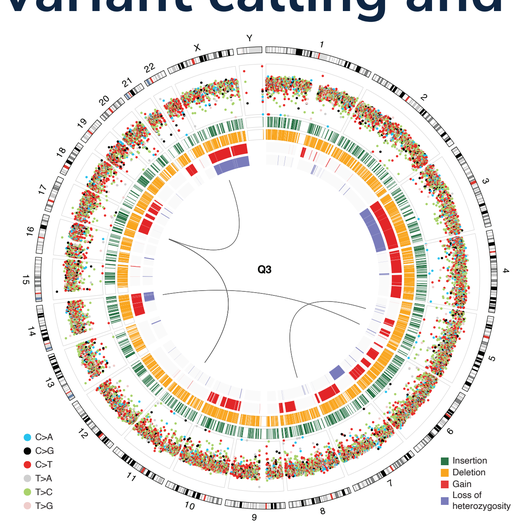


SPLONGGET captures chromatin accessibility and whole-genome

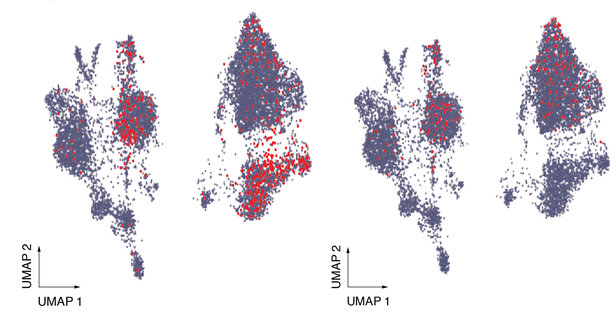


SPLONGGET captures chromatin accessibility and whole-genome information from the same library (left). Resulting in an increase of 12-15 fold in whole-genome coverage compared to short read Multiome data from the same library (right).

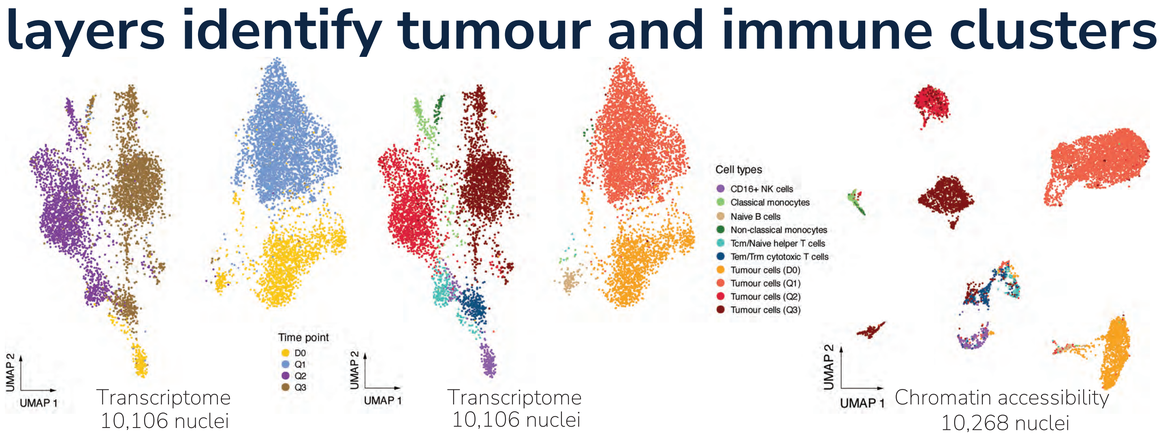
SPLONGGET enables comprehensive variant calling and genotyping



Long-read whole-genome data allows us to capture the full range of genomic variation. Such as, SNVs, indels and SVs (left). Furthermore, by using a targeted approach, we can enrich the original SPLONGGET library for variants of interest, allowing us to genotype a large number of cells (below).



Chromatin accessibility and transcriptome layers identify tumour and immune clusters

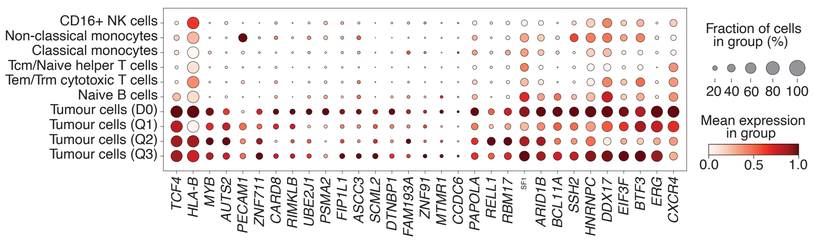


SPLONGGET results in high quality transcriptome (left two UMAPs) and chromatin accessiblity data (right), allowing accurate identification of cell types present in a sample. Furthermore, downstream processing is fully compatible with conventional 'short-read' tools such as Scanpy and Seurat.

SPLONGGET uncovers mechanism of anti-CD19 CAR T-cell resistance

Using long-reads whole-genome data, we were able to identify reads belonging to the Chimeric Antigen Receptor (CAR) construct and reassamble this contig. Remapping reads to the reassambled contig allowed us to identify cells with CAR integration (left). Next, we investigated potential ways of CAR T-cell therapy escape. We identifed loss of heterozygozity of the *CD19* locus, along with 4 previously unreported splice site mutations, leading to intron retention and complete loss of CD19 protein expression (right).

Combining modalities uncovers cancer-specific regulatory programs



SCENIC+ allows us to uncover enhancer driven gene regulatory networks (eGRNs) from the paired transcriptomic and chromatin accessibility data. We were able to identify eGRNs and transcription factors enriched in cancer cells, such as *TCF4*.

Conclusions

- We developed SPLONGGET, a single-cell long-read multiome method, providing simulatenous genome, open chromatin, and transcriptome data from thousands of cells.
- We show that we are able to profile the transcriptome and chromatin accessibility landscape, alongside identifying genomic variation.
- Leveraging SPLONGGET, we identify the mechanism of CAR T-cell therapy escape in a patient with B-ALL.
- SPLONGGET is backwards compatible with short-read sequencing.
- Using the targeted aproach allows for highly accurate genotyping of single cells.

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

VIB-KU LEUVEN CENTER FOR CANCER BIOLOGY

VIB KU LEUVEN