LEIDEN UNIVERSITY MEDICAL CENTER

Large deletion detection

Jeroen F. J. Laros

Leiden Genome Technology Center

Department of Human Genetics

Center for Human and Clinical Genetics

Introduction

A simple pipeline for large deletions.

Goal: Find arbitrary large deletions.

Layout of the idea.

- Use an RNA aligner that can split reads.
- Make a plot of the areas spanned by split reads.
- Use a cut off to convert the plot to a bed track.

We chose GMAP for alignment.

- Able to split up both reads.
- Able to split a read into many pieces.

RNASeq

A problem with split reads.

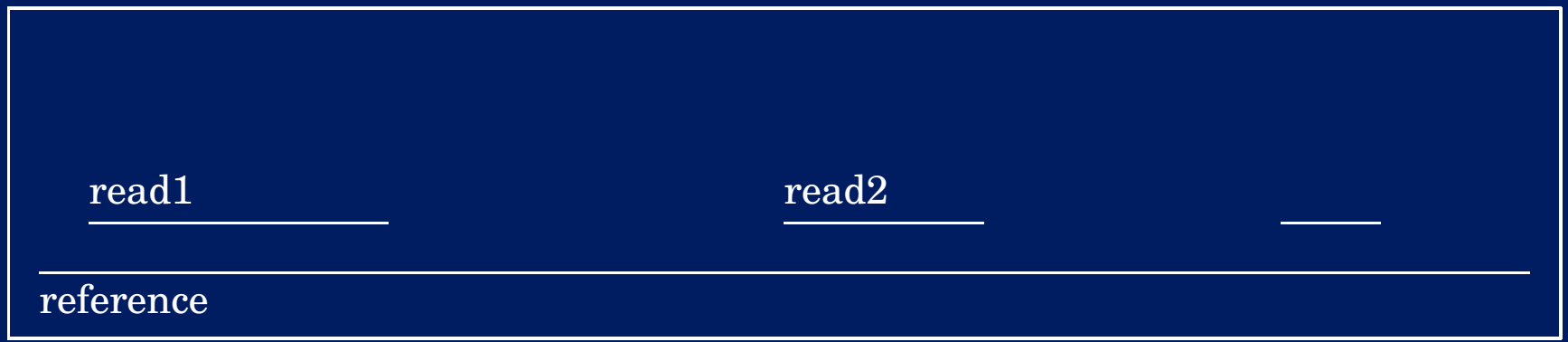

Figure 1: Mapped reads.

Figure 2: Coverage.

Experimental data

Normal RNASeq output.

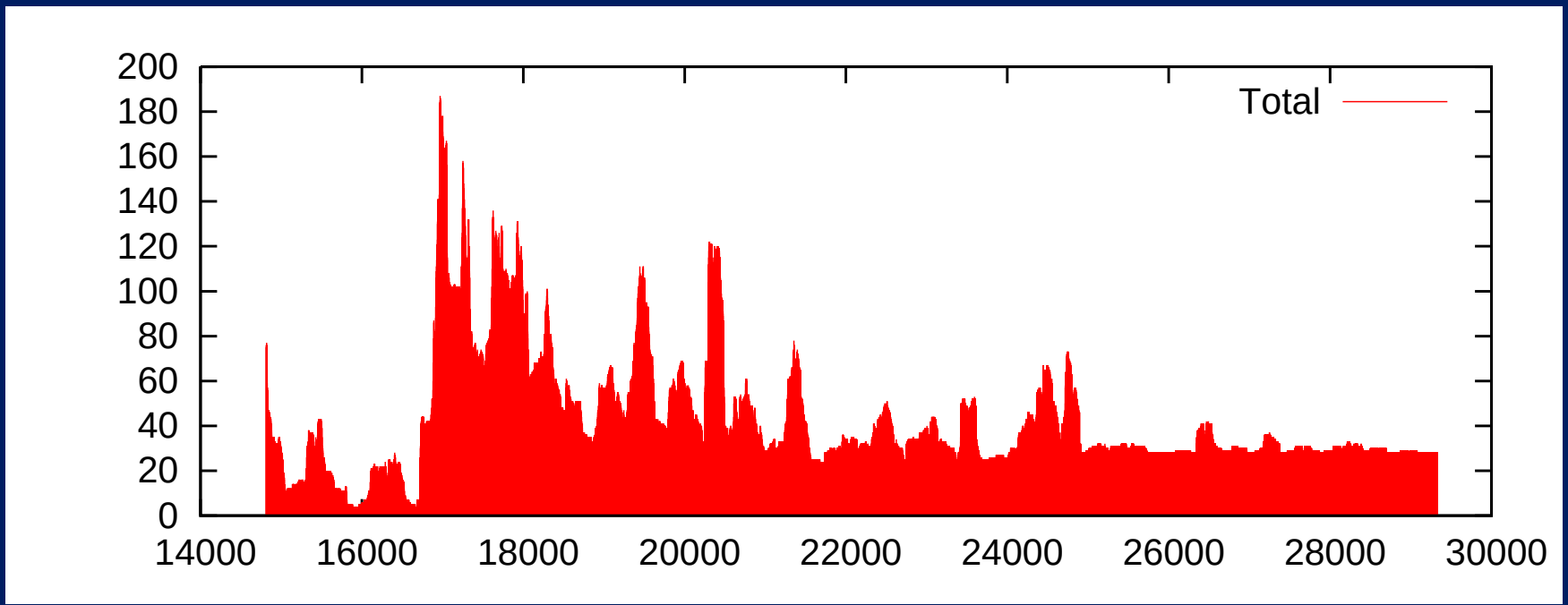

Figure 3: Total coverage.

Experimental data

Corrected RNASeq output.

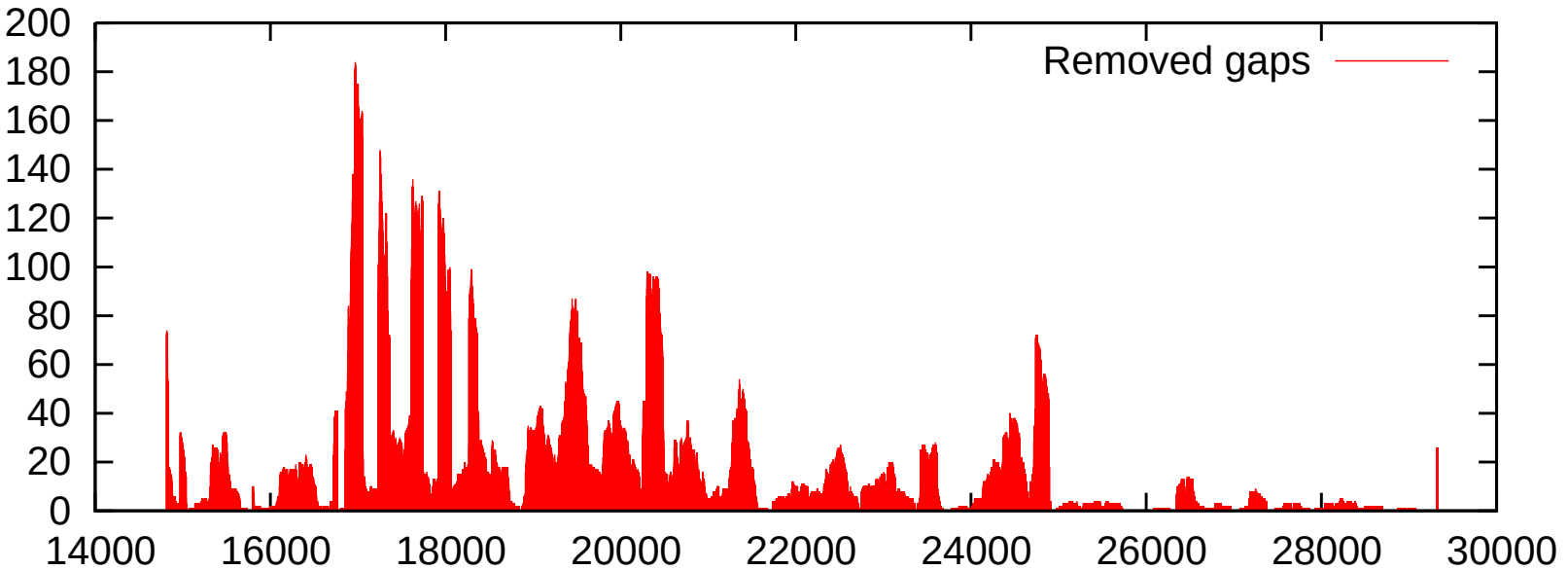

Figure 4: Removed gaps.

Experimental data

Visualisation of the gaps.

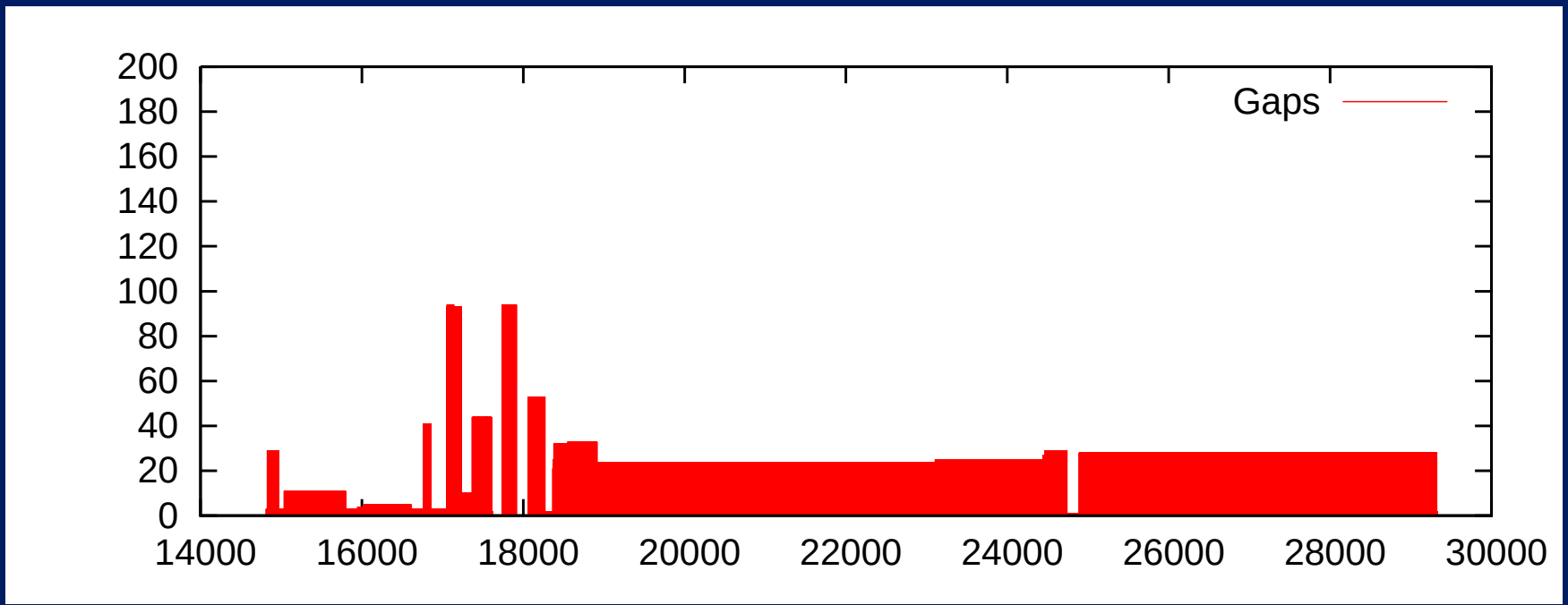

Figure 5: Only the gaps.

Experimental data

List of deletions.

chr1	14829	14969
chr1	15038	15795
chr1	16765	16857
chr1	17055	17605
chr1	17742	17914
chr1	18061	18267
chr1	18366	24737
chr1	24891	29320
chr1	126240	165883
chr1	165942	167963
chr1	168165	169048
chr1	237955	238417
chr1	322228	324287
chr1	324345	324438
chr1	415975	417159
chr1	418969	584389

Listing 1: Final output.

Pipeline.

Use this idea for GoNL data.

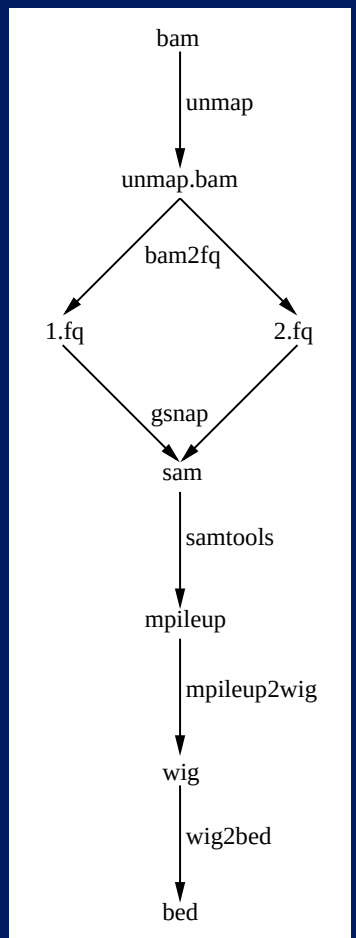

Figure 6: Concept pipeline.

Tools:

- unmap
  - Unmapped.
  - Single-end mapped.
  - Soft-clipped.
- bam2fq
- samtools
- mpileup2wig
  - Extract gaps.
- wig2bed
  - Convert everything above a certain threshold.

Questions?

Acknowledgements:

Martijn Vermaat
Michiel van Galen
Johan den Dunnen