BIOPLATFORMS
AUSTRALIA

Alignment and splice junction identification

RNA-Seq Hands-on Practical

Gayle Philip, VLSCI

Curtin University, Perth.
12th -14th July 2016

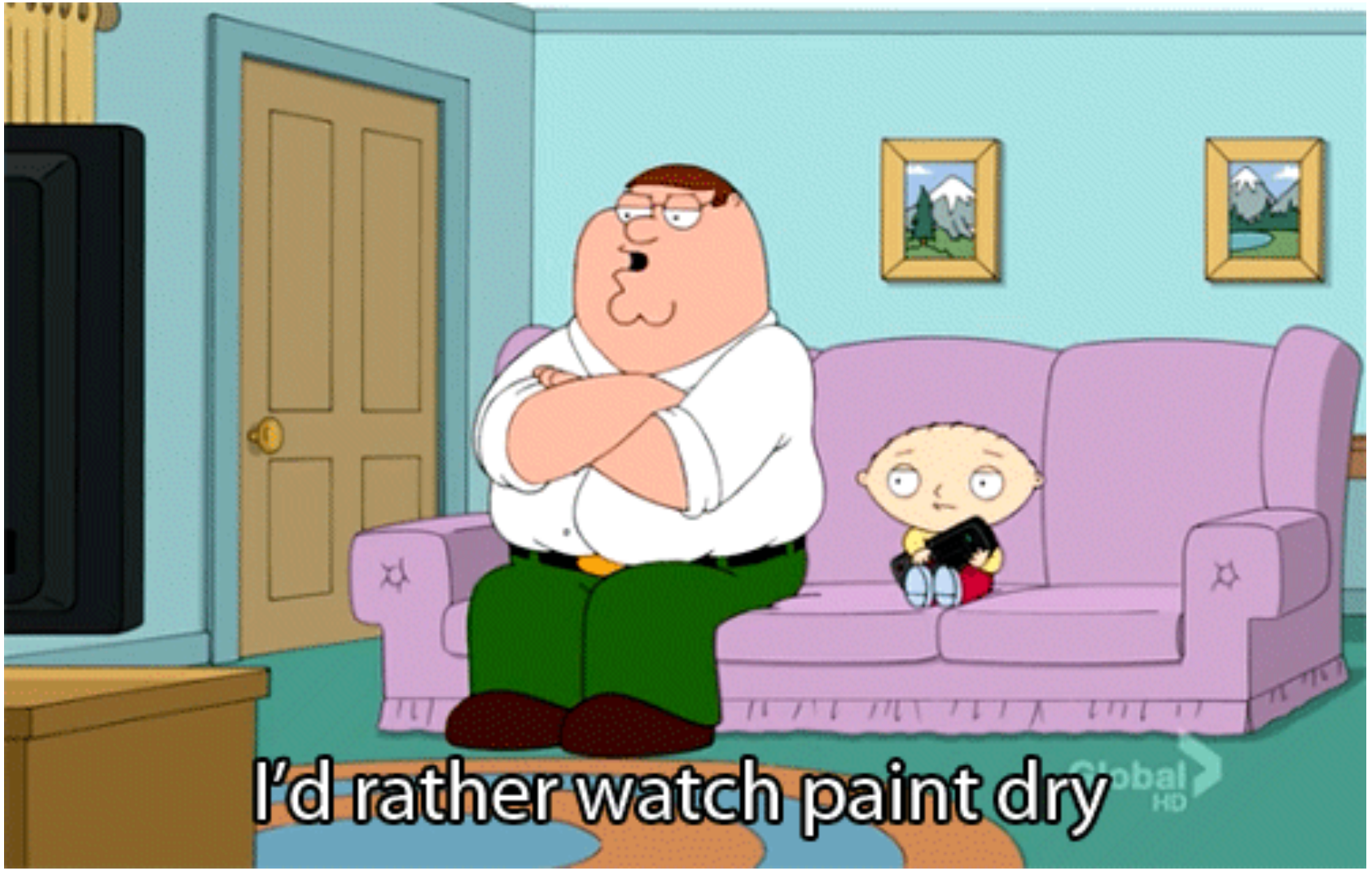
EMBL-EBI

Introduction

- The goal of this hands-on session is to perform some basic tasks in the analysis of RNA-seq data. We will start by aligning RNA-seq data to the zebrafish genome using Tophat2.
  - Prepare the environment
  - Tophat alignment
  - Viewing the alignment using IGV

Preparing the alignment

- Sample
  - Zebrafish (*Danio rerio*)
  - 2 conditions : 2cells and 6h
- Type of sequencing
  - 76bp pair-end from polyA selected RNA
  - Illumina
- File name and formats
  - Fastq file
  - 2cells 1.fastq and 2cells 2.fastq (RNA-seq data of a 2-cell zebrafish embryo)
  - 6h 1.fastq and 6h 2.fastq (RNA-seq data of zebrafish embryos 6h post fertilization)


I'd rather watch paint dry
Global
HD

DO NOT Run!

The command to create an index is as follows. You DO NOT need to run this command yourself - we have done this for you.

```
1 bowtie2-build genome/Danio_rerio.Zv9.66.dna.fa genome/ZV9
```

• Some commands take a **LONG** time to run. For example, indexing a genome can take a few hours to finish. **The genome index has been pre-computed for you.**

• Sometimes, the **same** command has to be **repeated** multiple times on different input files. E.g. Data from case and control conditions have to be aligned independently in the same manner. Pre-computed files have been provided for some of the samples.

Alignment

- Reference genome
  - Danio_rerio.Zv9.66
  - **Index has already been created**

- Alignment program and parameters
  - **The 2cells data is pre-aligned**
  - Alignment needs to be done for 6h dataset

Alignment Visualisation in IGV

- The Integrative Genomics Viewer (IGV) is able to provide a visualisation of read alignments given a reference sequence and a BAM file.
  - Visualise the alignment
  - Look at the splice junctions

- Interpreting the alignment in IGV
  - http://www.broadinstitute.org/igv/AlignmentData