

DAY 1 · PRACTICAL SESSION

Alignment & read-level quality control

Long-read transcriptome analysis — from raw reads to SQANTI-reads QC.

What we'll cover today



01

Experiment design & dataset

Why chromosome 8, and how the LRGASP Challenge 2 subset was built.

Background

02

Aligning long reads

minimap2, annotation-guided uLTRA, and learned splice models (minisplice).

Hands-on 1

03

SQANTI-reads

Read-level structural QC — categories & unique junction chains.

Hands-on 2

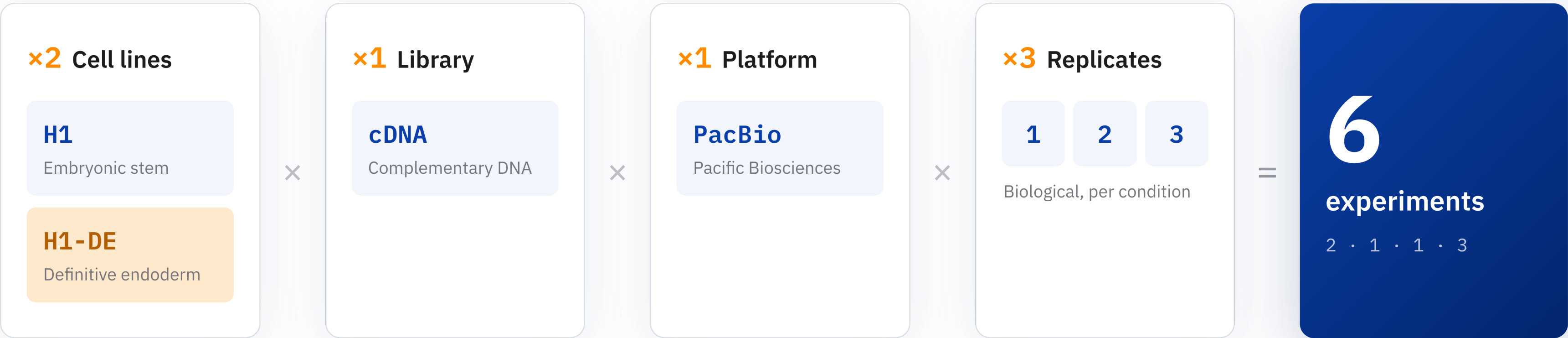
PART ONE

Experiment design & dataset

A subset of LRGASP Challenge 2, focused on endoderm differentiation markers on chromosome 8.

A controlled subset for benchmarking

The challenge measures transcript diversity and expression across platforms and cell states. Our slice is a tidy factorial design:



Why chromosome 8?

Chromosome 8 hosts two transcription factors that drive endoderm differentiation:

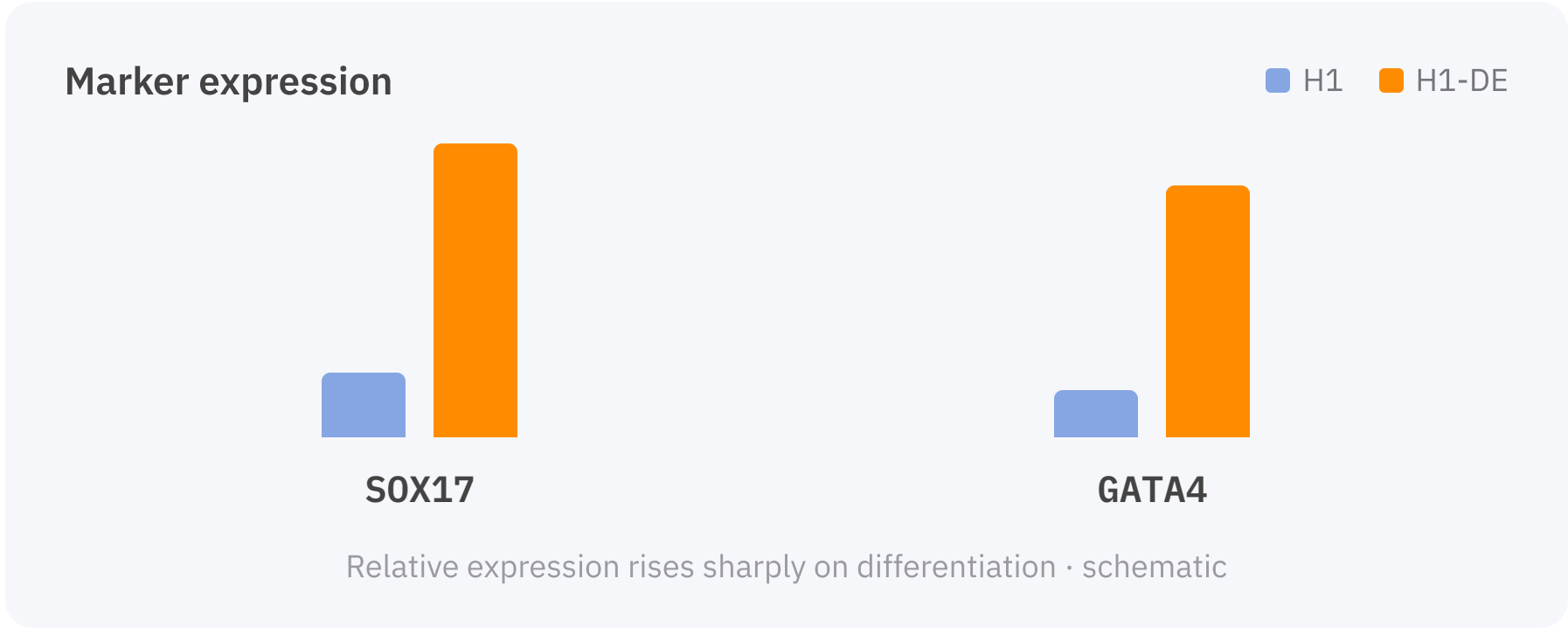
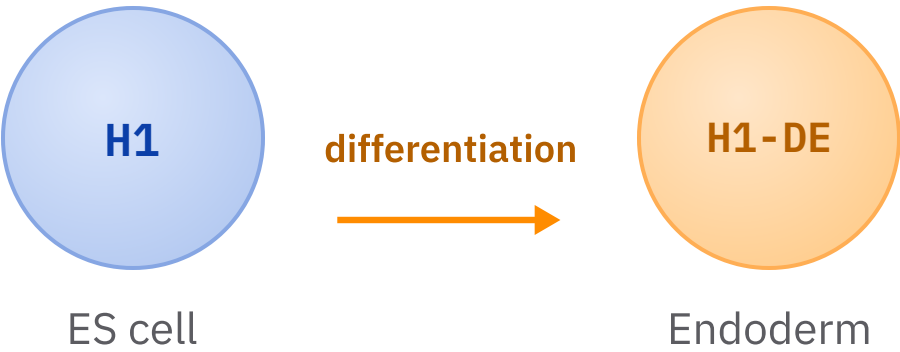
SOX17

8q11.23
Master regulator of definitive endoderm commitment

GATA4

8p23.1
Key driver of endodermal maturation

Both shift expression dramatically across the H1 → H1-DE transition — ideal markers for our analysis.





PART TWO • HANDS-ON 1

Aligning long reads

Three strategies: minimap2, annotation-guided uLTRA, and learned splice models.

Three ways to align long reads

minimap2

General-purpose

- ✓ Fast, splice-aware long-read aligner
- ✓ Built-in GT-AG splice consensus
 - › The base aligner everything builds on

uLTRA

Annotation-guided

- ✓ Two-pass collinear chaining of junctions
- ✓ Uses the reference GTF to place junctions
- ! **Relies on the reference annotation**

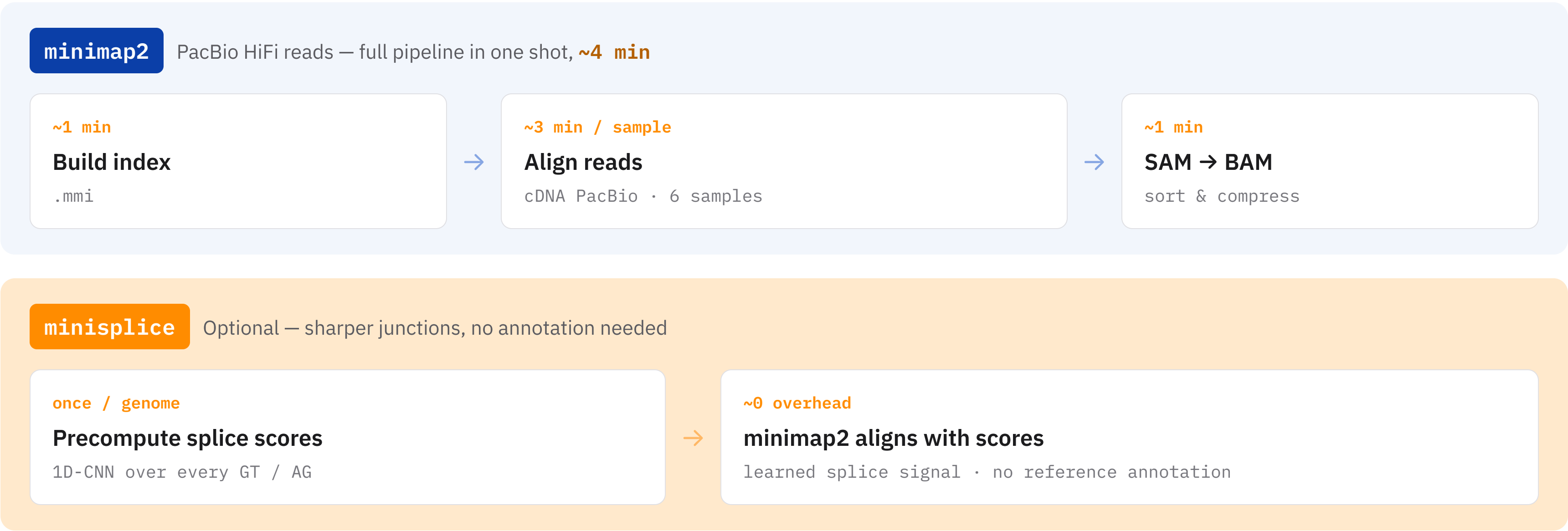
minisplice

Deep learning · augments minimap2

- ✓ 1D-CNN learns splice signals from sequence
- ✓ Feeds splice scores into minimap2
- ✓ **No reference annotation needed**

The key difference. uLTRA leans on the reference annotation, so it is biased toward what is already annotated. minisplice learns splice signals directly from the genome sequence — sharpening noisy long-read junctions without that bias.

From index to aligned reads





PART THREE • HANDS-ON 2

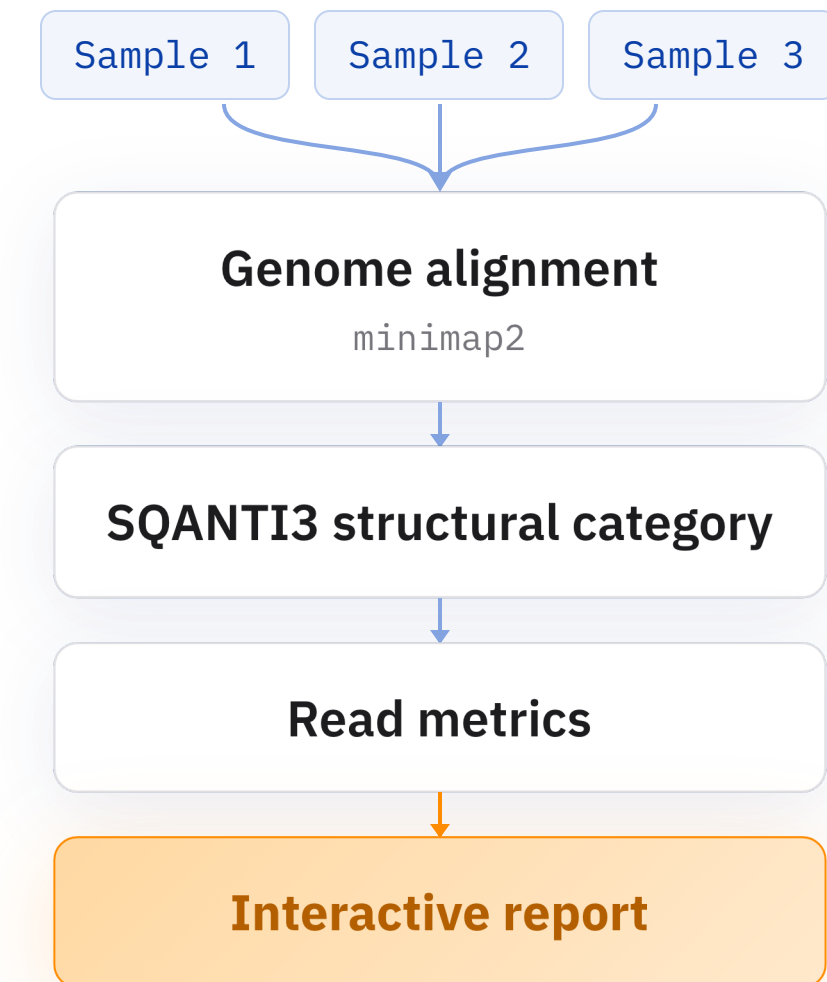
SQANTI-reads

Read-level structural QC — categories, junction usage, and interactive reports.

What is SQANTI-reads?

A read-centric extension of SQANTI3.

- Brings structural classification down to the **single-read** level
- Jointly evaluates reads from **many samples**
- Summarises categories, splicing patterns & junction usage
- Interactive plots flag outliers & under-annotated genes



What goes in, what comes out

INPUTS what you provide

CSV

Design file
sampleID · file_acc

GTF

Reference annotation
GTF / GFF3

SEQ

Reads
one of two modes

Fast

SQANTI3-QC dirs

Simple

fastq · GTF + FASTA



SQANTI-READS

- Classify reads
- Summarise metrics
- Visualise QC



OUTPUTS generated for you

txt

reads_classification.txt
+ jxn_string, jxnHash

csv

design.csv
+ classification & junction files

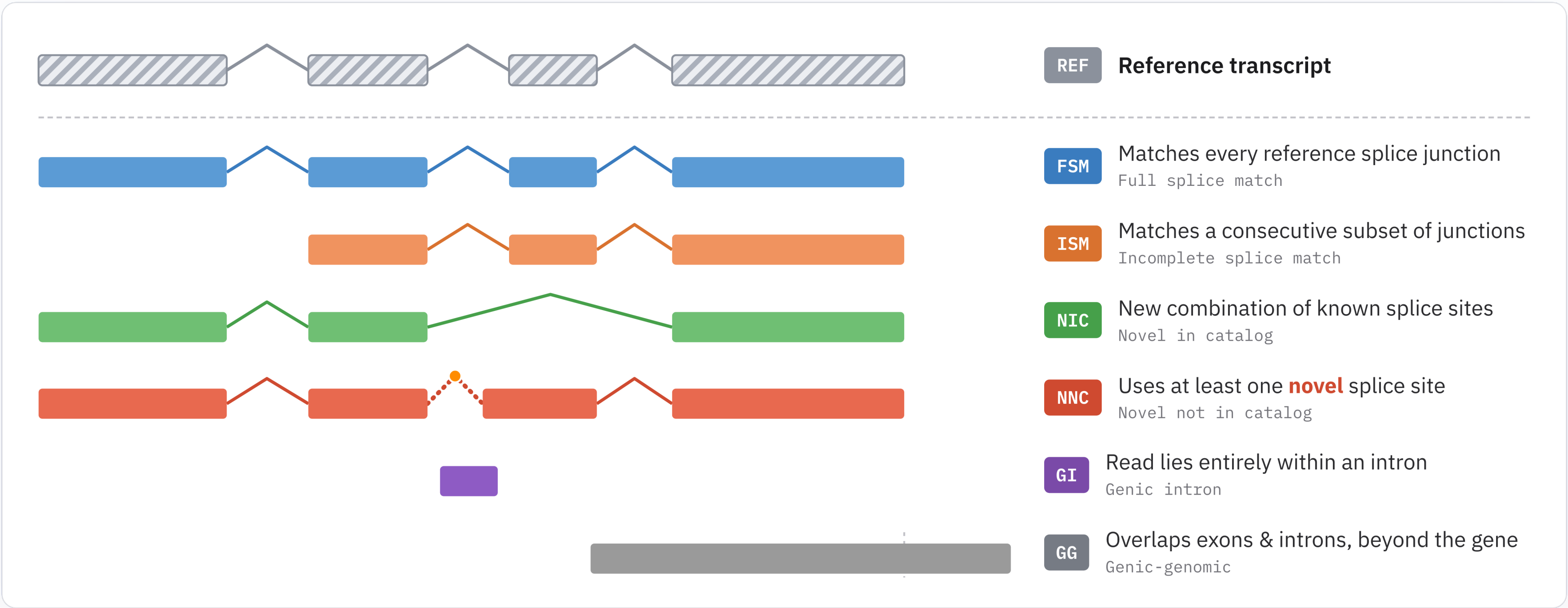
csv

Summary tables
gene · ujc counts · length · CV

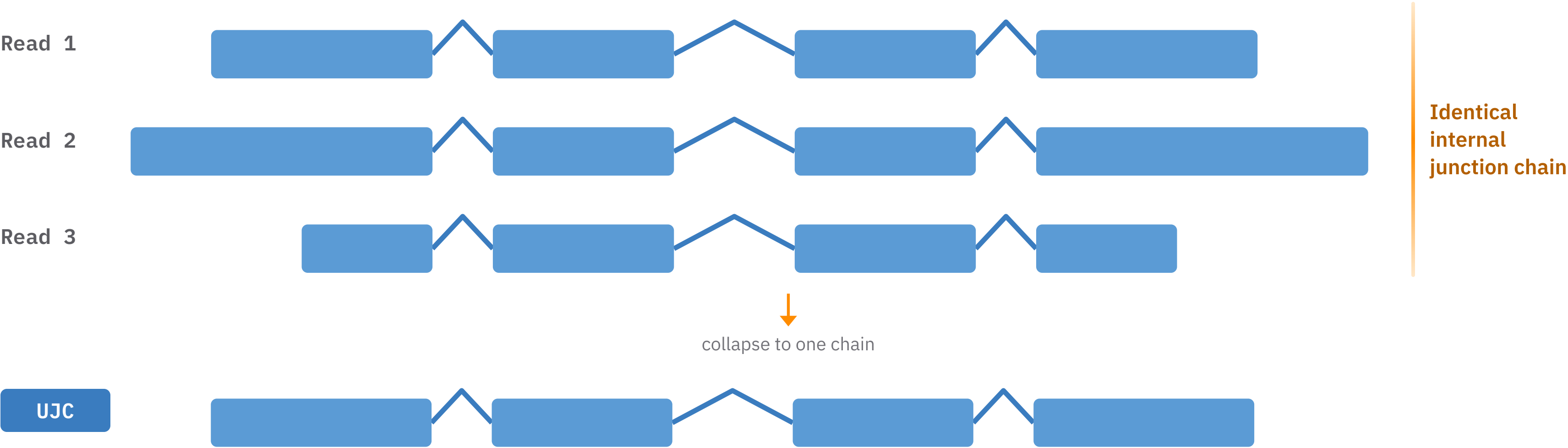
PDF

QC & annotation plots
PDF + optional HTML report

SQANTI3 structural categories



Unique junction chains



Reads with variable start/end (TSS/TTS) but the same ordered splice junctions collapse into a single **Unique Junction Chain**.

Your hands-on task

What you'll do

- 1 Run the minimap2 alignment workflow
- 2 Compare alignment approaches & quality
- 3 Process alignments through SQANTI-reads
- 4 Explore categories & junction patterns
- 5 Generate QC reports & visualisations

What you'll take away

- Trade-offs between aligners & splice models
- Why read-level quality control matters
- Reading SQANTI3 structural categories
- Spotting junction-usage patterns across samples
- Recognising technical artifacts in lrrNA-seq

Day 1 in summary

DESIGN

Experiment design

LRGASP subset on chr 8 endoderm markers

DATASET

Read preparation

Systematic chr 8 extraction & validation

ALIGN

Read alignment

minimap2, uLTRA & learned splice models

QC

Quality assessment

SQANTI-reads for read-level QC

STRUCTURE

Structural categories

FSM · ISM · NIC · NNC and junction chains

SKILLS

End-to-end workflow

Raw reads to quality-controlled alignments



END OF DAY 1

Thank you

Questions about the practical session?

WEB longtrec.eu

COURSE conesalab.github.io/VALT_Summer_Course

EMAIL longtrec@csic.es



Funded by
the European Union