Genome Analysis 2021
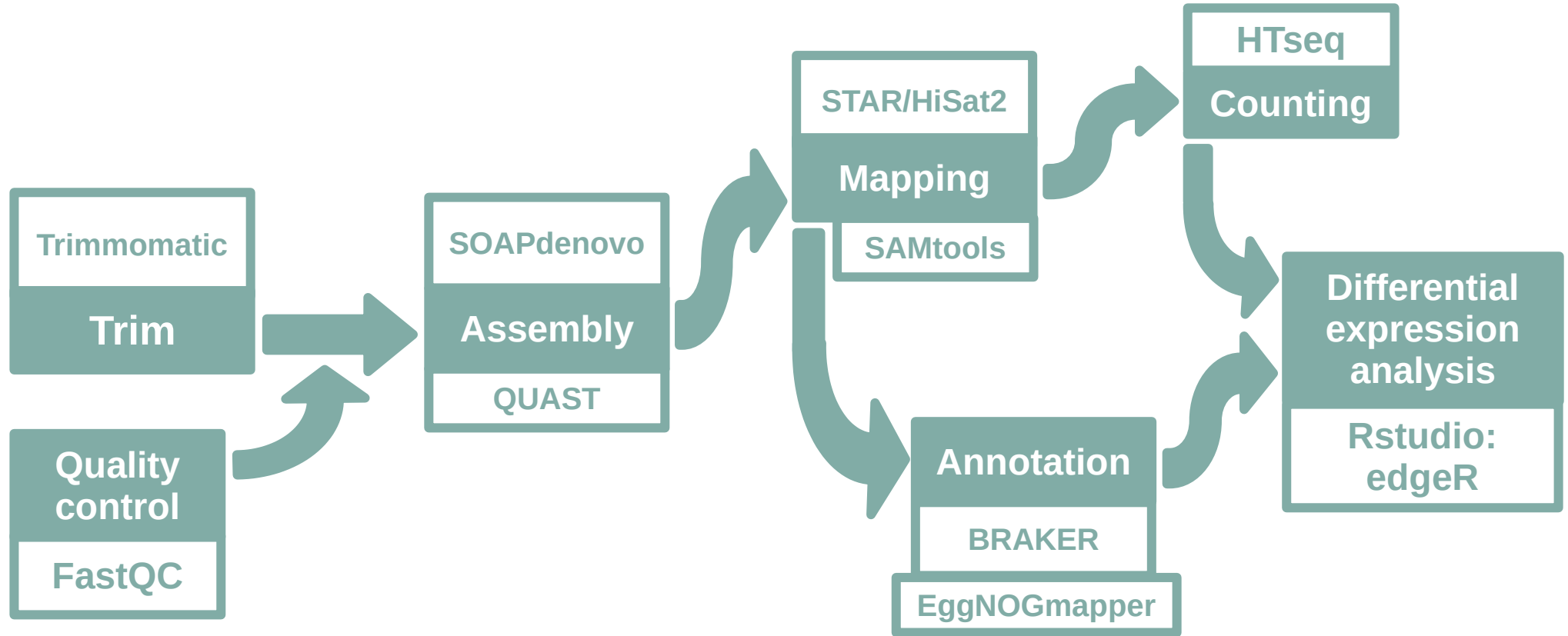
Paper II:

Transcriptomic and epigenomic characterization of the developing bat wing

Eckalbar et al.

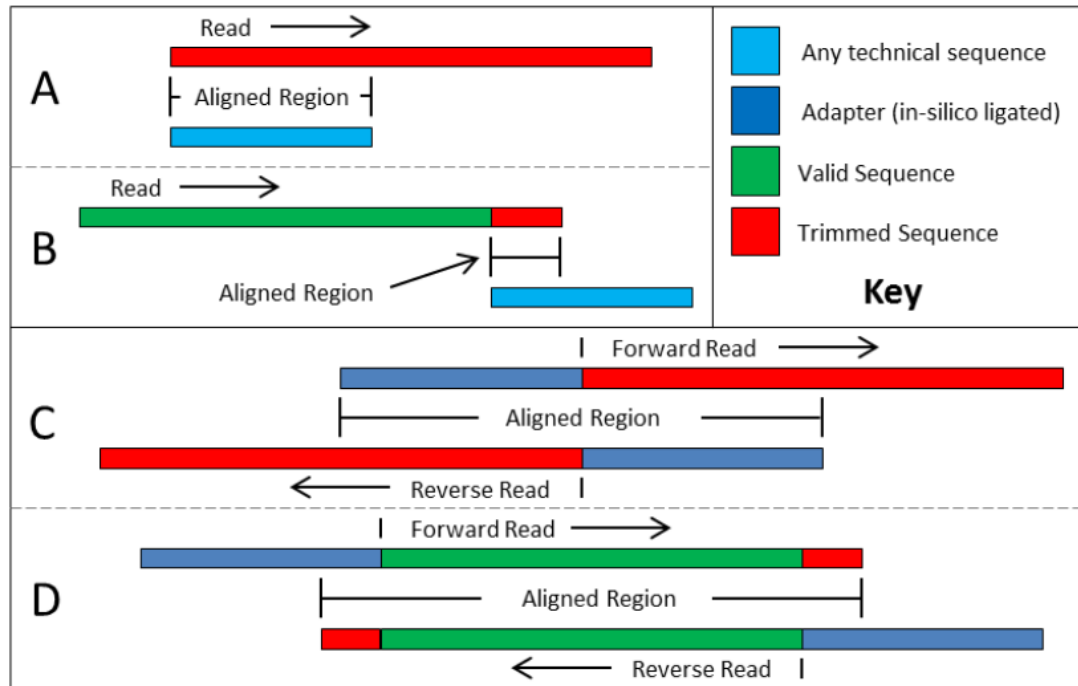Milena Regina Trabert

Workflow

Trimmomatic
Trim

Quality control
FastQC

SOAPdenovo
Assembly
QUAST

STAR/HiSat2
Mapping
SAMtools

HTseq
Counting

Annotation
BRAKER
EggNOGmapper

Differential expression analysis
Rstudio: edgeR

Trimmomatic & FastQC

ILLUMINA alignment parameters

Seed mismatches:………………. **2**

palindrome clip threshold:…….. **20**

simple clip threshold:…………… **7**

minimum adapter length:……….. **2**

keep both reads:…………….. **true**

LEADING & TRAILING:……….. **17**

SLIDINGWINDOW: length……… **5**

    min. avg. quality…………... **20**

MINLEN:…………………………. **20**

Assembly

SOAPdenovo

SOAPdenovo config file parameters

	Run 1	Run 2	Run 3
reverse_seq	0	1	1
asm_flag	3	3	1

reverse_seq: complementary reverse paired-end sequences (0=FR, 1=FF)
asm_flag: which parts of the reads are used (1=contig assembly, 3=contig and scaffold assembly)

Assembly

SOAPdenovo & QUAST

SOAPdenovo config file parameters

	Run 1	Run 2	Run 3
`reverse_seq`	0	1	1
`asm_flag`	3	3	1

Total length	141836091
Reference length	3990777
GC (%)	43.24
Reference GC (%)	36.99

`reverse_seq`: complementary reverse paired-end sequences (0=FR, 1=FF)

`asm_flag`: which parts of the reads are used (1=contig assembly, 3=contig and scaffold assembly)

Mapping

STAR (+ indexer)

Samtools functionality integrated

Unmapped: 69.41%

HiSat2 (+ indexer)

Samtools used separate (pipe)

Unmapped: 61,88%

Mapping

STAR (+ indexer)

Samtools functionality integrated

Unmapped: 69.41%

HiSat2 (+ indexer)

Samtools used separate (pipe)

Unmapped: 61,88%

Gene content in scaffold 1?

Mapping

STAR (+ indexer)

Samtools functionality integrated

Unmapped: 69.41%

HiSat2 (+ indexer)

Samtools used separate (pipe)

Unmapped: 61,88%

Gene content in scaffold 1?

Scaffold 1
576

Scaffold 2
2210

Scaffold 4
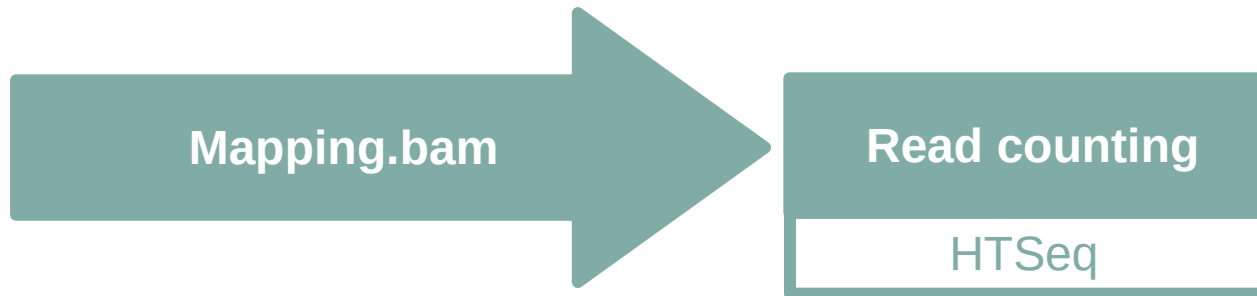4933

Mapping

STAR (+ indexer)

Samtools functionality integrated
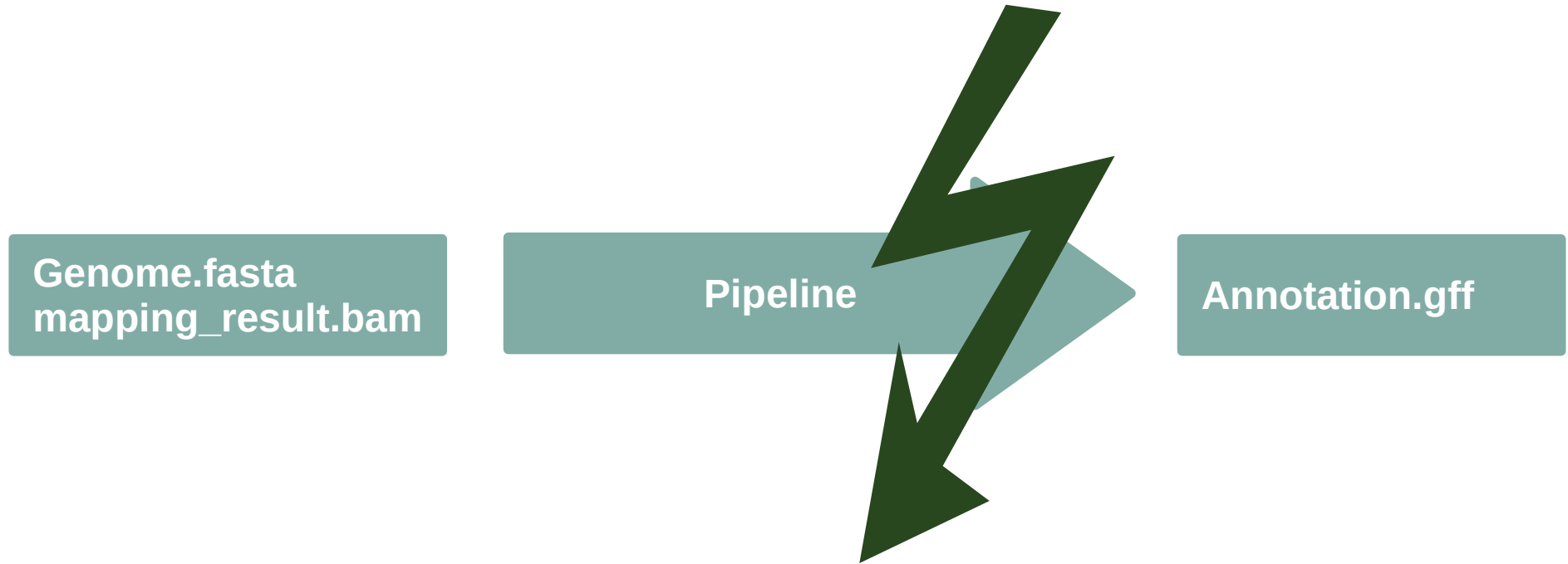
Unmapped: 69.41%

HiSat2 (+ indexer)

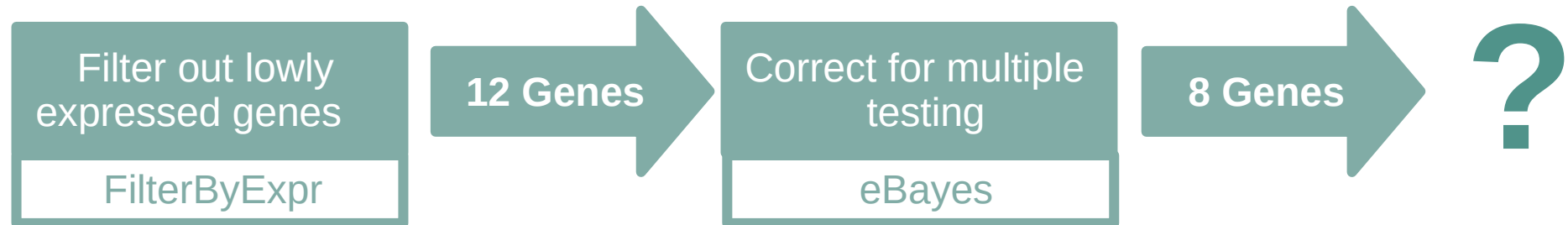Samtools used separate (pipe)

Unmapped: 61,88%

Mapping.bam → **Read counting**

HTSeq

Structural Annotation

BRAKER

Genome.fasta
mapping_result.bam

Pipeline

Annotation.gff

Functional Annotation

EggNOGmapper

Extract protein fasta from annotation.gff

Cufflinks & Emboss

Protein.fasta

Functional Annotation

EggNOGmapper

Differential expression analysis

Rstudio, edgeR

Filter out lowly expressed genes

FilterByExpr

12 Genes →

Correct for multiple testing

eBayes

8 Genes →

?

Differential expression analysis

Multidimentional scaling
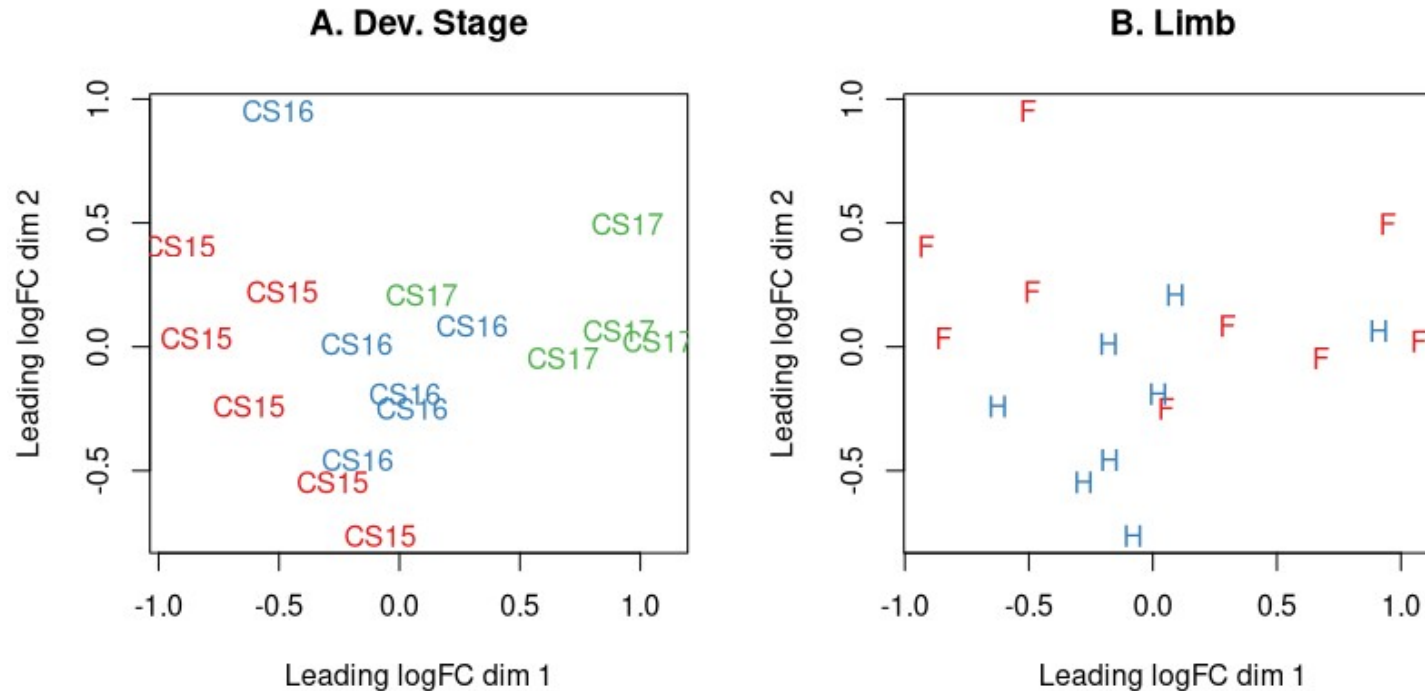
Differential expression analysis

PCA, clustering, Heatmap

PC1: 89%, PC2: 6%