University of Glasgow | College of Medical, Veterinary & Life Sciences

Transcriptomics

Graham Hamilton
Glasgow Polyomics

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Outline

Introduction to sequencing technologies

- High throughput sequencing
- Overview of the sequencing platforms

Experimental design

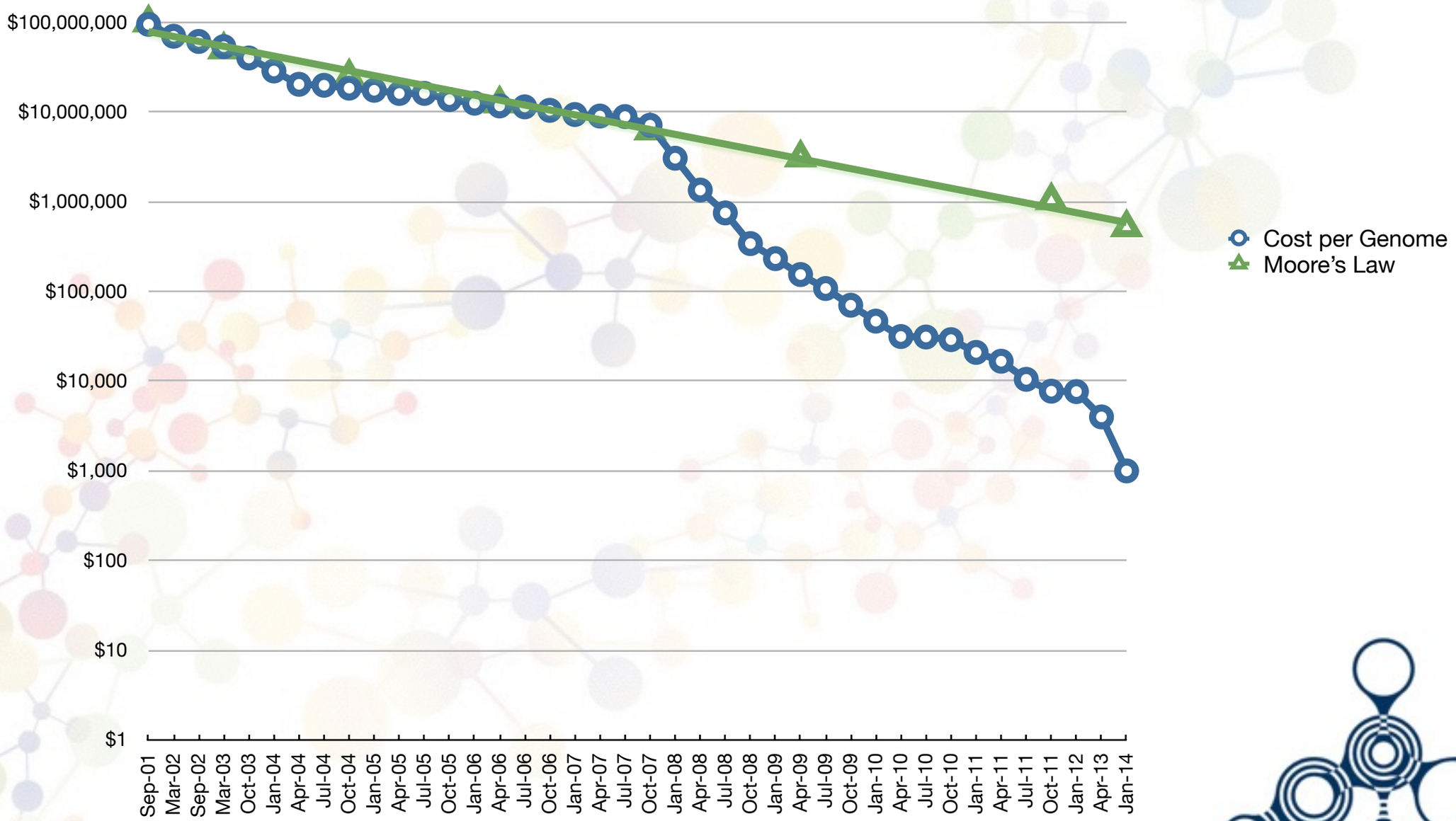
Analysis of transcriptomics data

Case studies

- Drosophila study
- No genome available

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

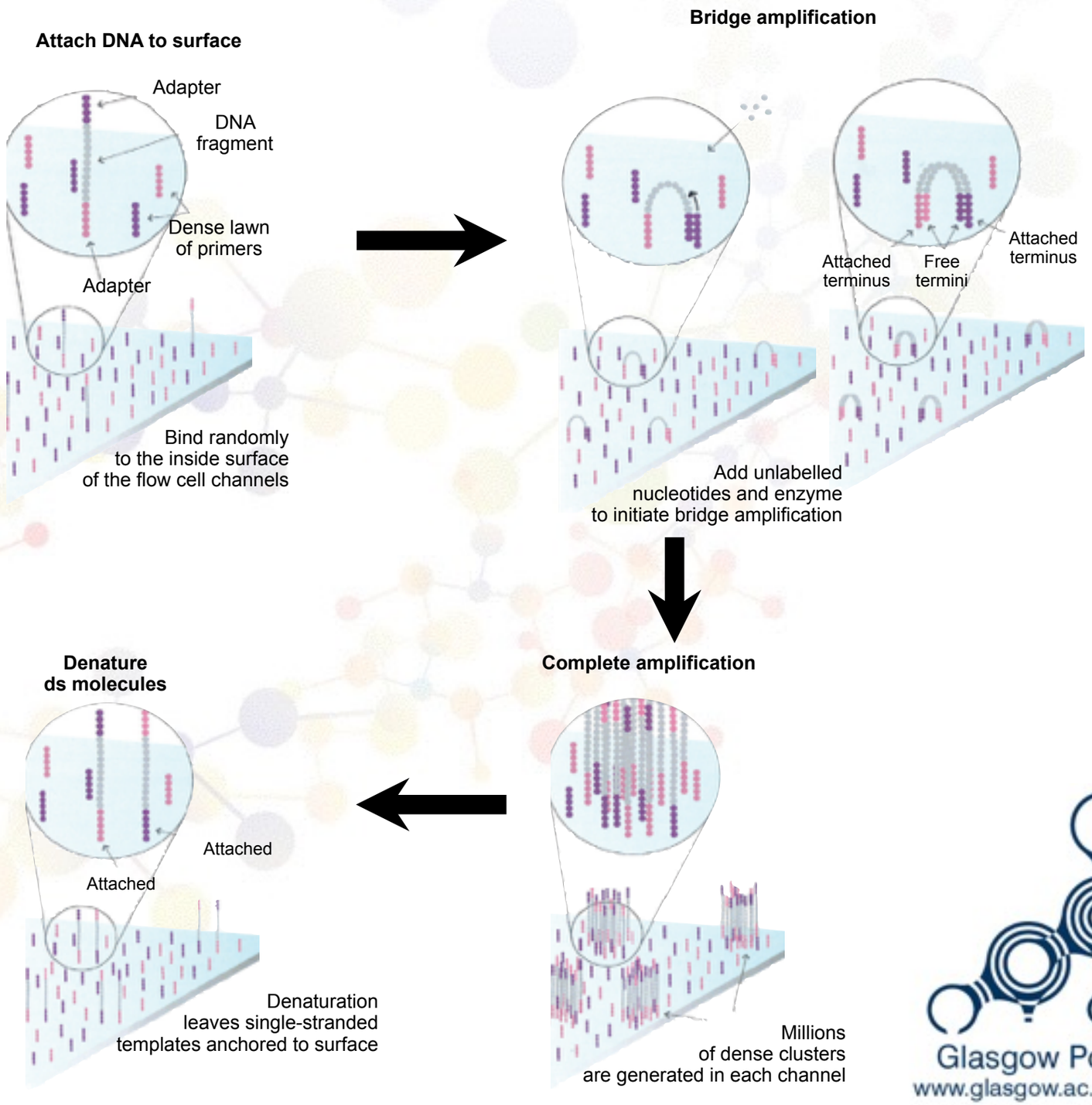
Cost of sequencing



Glasgow Polyomics
www.glasgow.ac.uk/polyomics

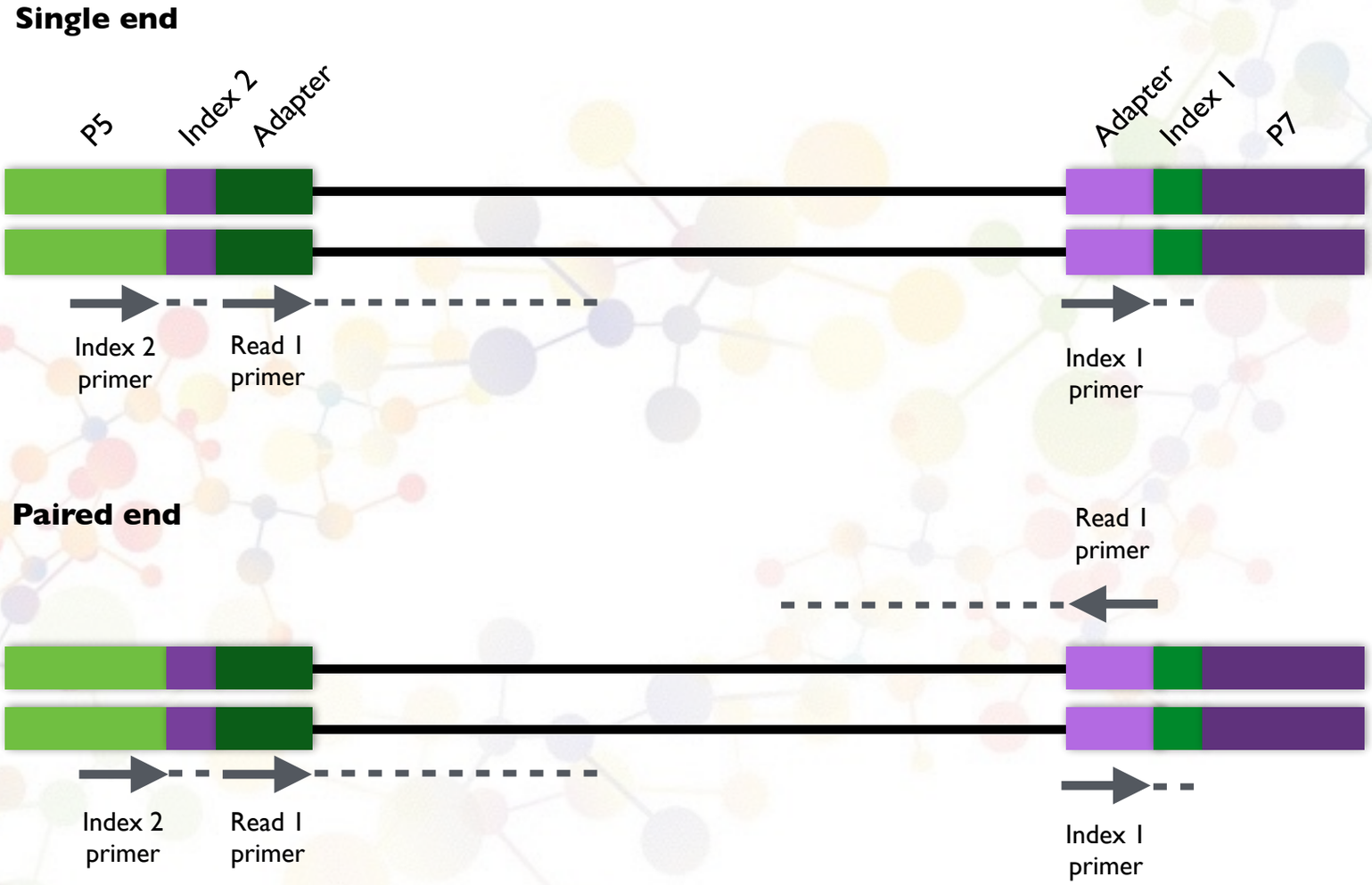
Illumina sequencing

- Cluster generation



Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Paired vs Single end sequencing



Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Illumina

- Base by base sequencing
- Reads all the same length
- Short reads
- High quality

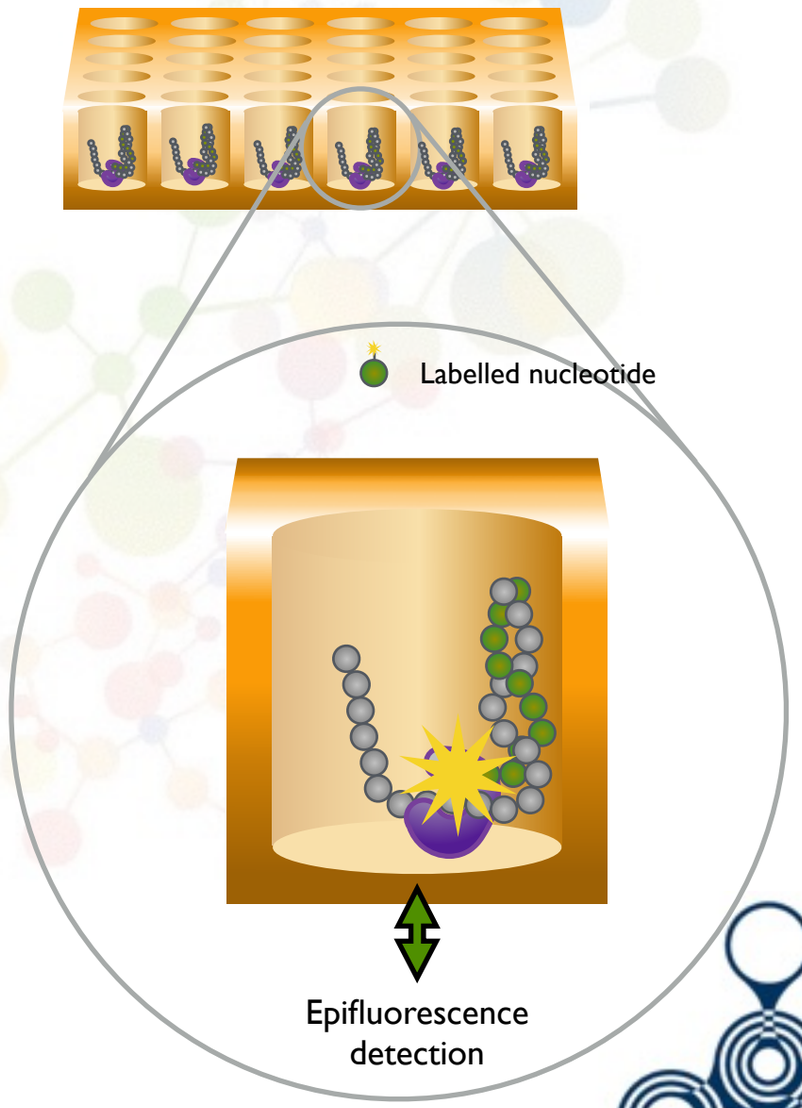
Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Pacific Biosciences

Real time single molecule sequencing

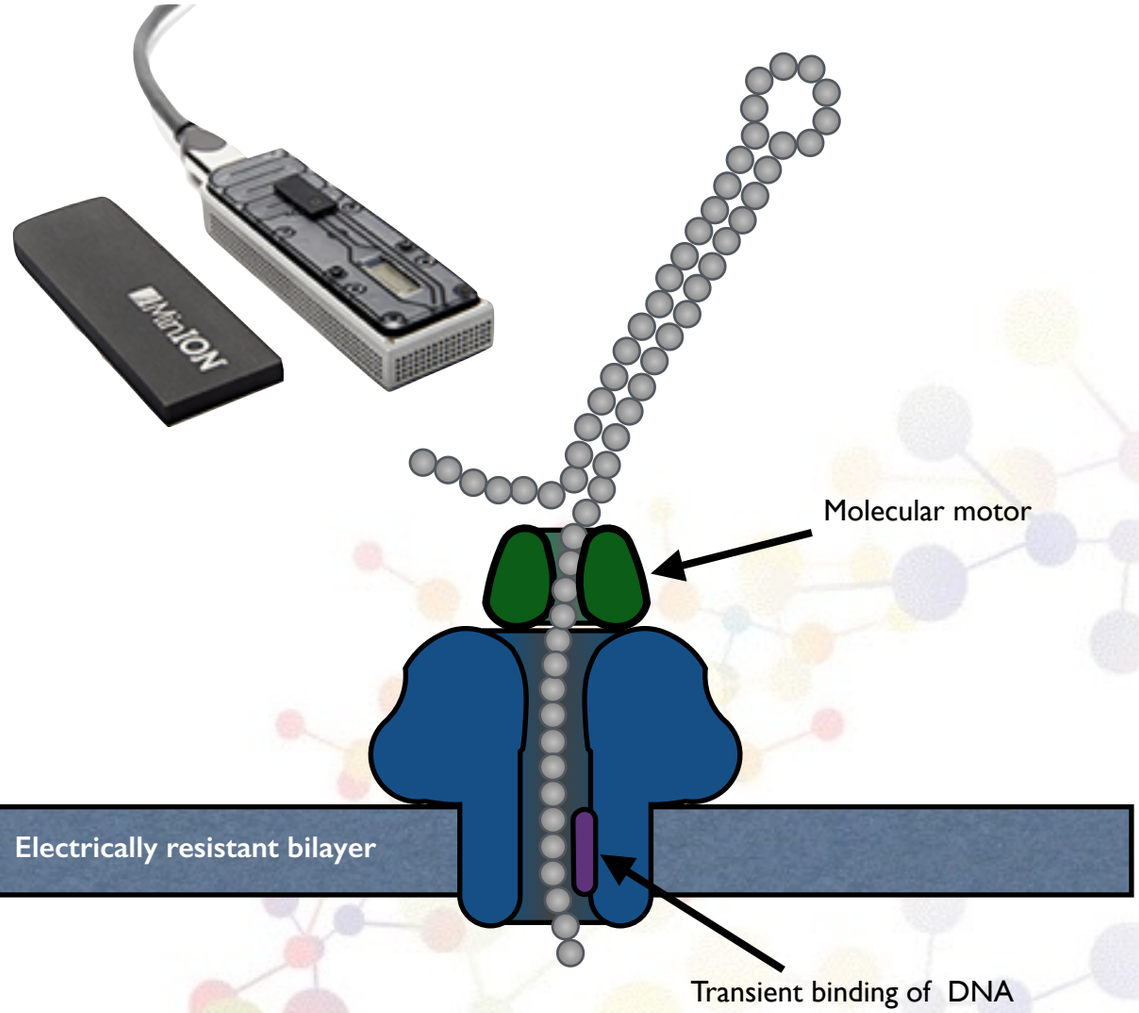
Long reads up to 20kb

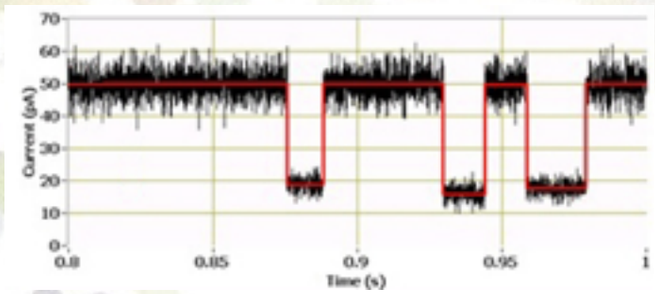
Direct sequencing of modified bases e.g. methylated Cytosine



Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Oxford Nanopore





Real time single molecule sequencing

Long read lengths
10's of kilo bases

Can detect modified DNA bases

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Many applications

Total RNA

micro RNA

Immuno precipitated RNA

Template preparation & Sequencing

Data processing, base calling & storage

Mapping to reference genome

***De novo* assembly**

Transcript differential expression

Novel transcript discovery

Differential isoform usage

micro RNA analysis

RIP-Seq analysis

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Experimental design

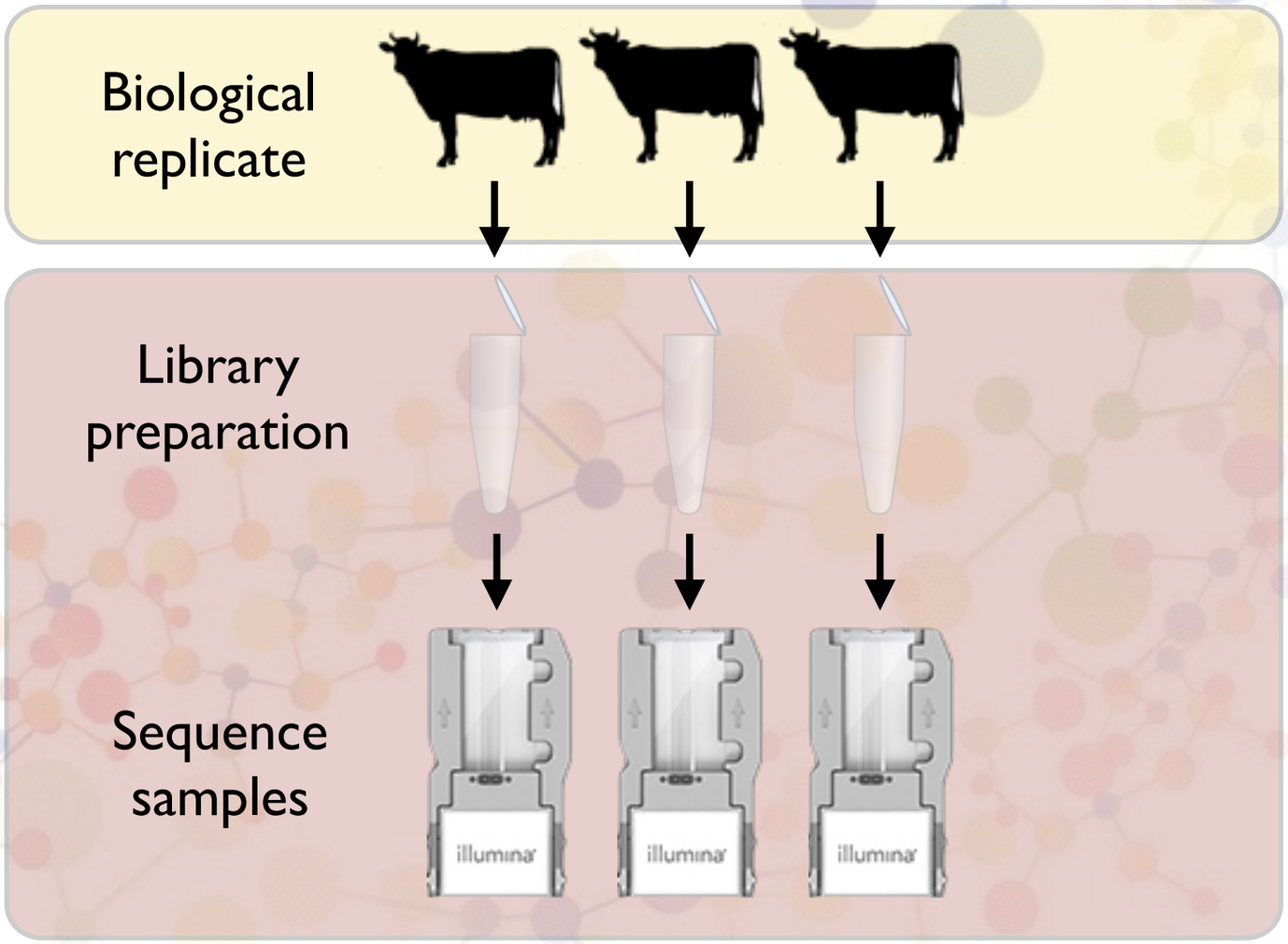
- Key to an expression study
- Ensure that the questions can be answered within experimental constraints
  - Cost
  - Availability of RNA

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Variation in RNA-Seq experiments

- Biological variation
- Technical vaiation

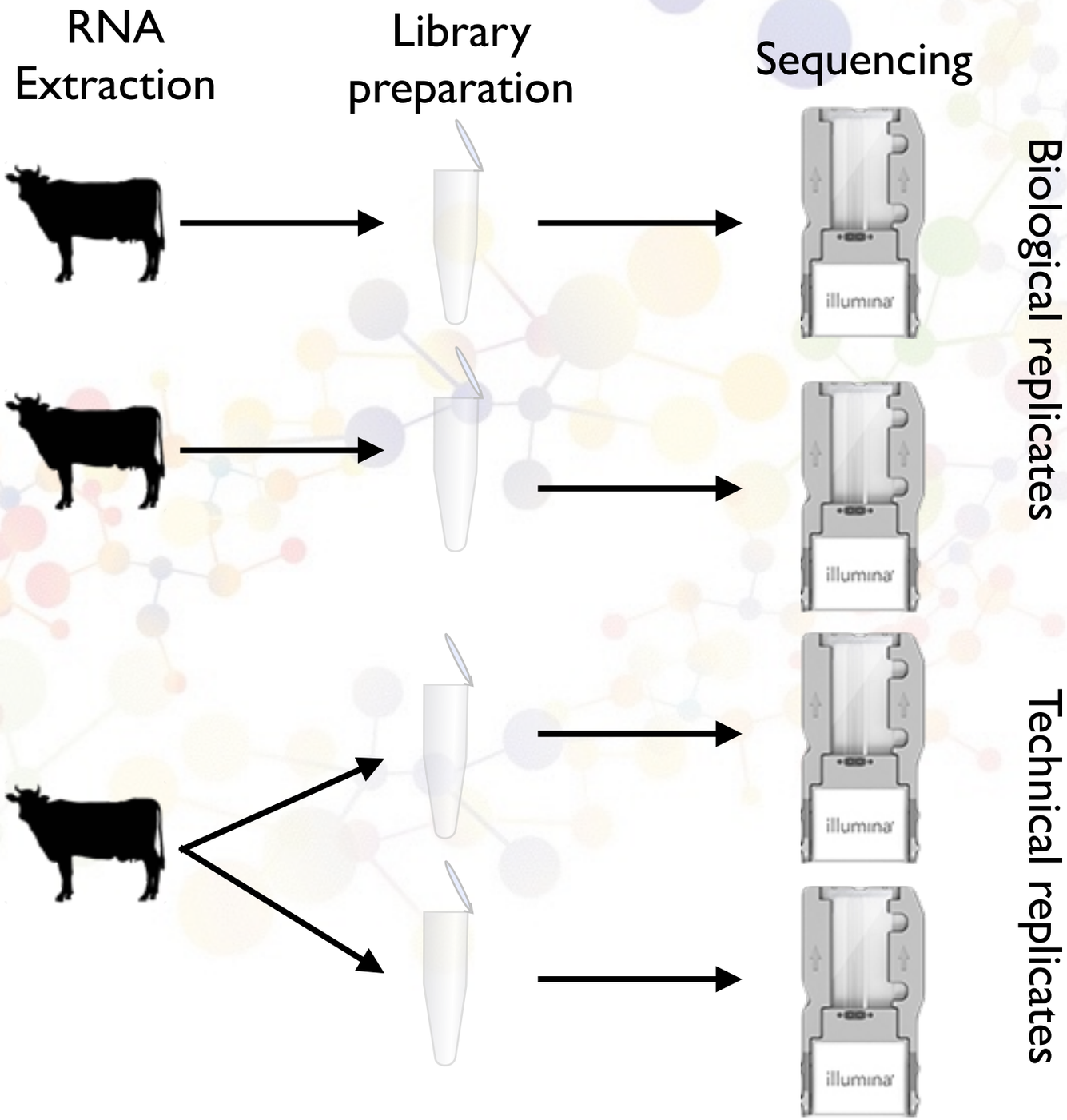
Levels of variation



Flowcell effect < Library prep effect << Biological effect

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Replicates



Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Replicates

- Biological replication is essential
  - How many replicates?
- Technical variation
  - Barcode and multiplex

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Sequencing depth

- Difficult to predict
- Estimates can be based on the length of the transcriptome and the expected dynamic range of transcript abundances
- Greater sequence depth increases sensitivity to detect smaller changes and lower expressed transcripts

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

What do I need to provide?

Poly A selection

- 1 µg total RNA
- Prepare using tri reagent or Qiagen Total RNA kit
- Recommend paired end sequencing

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Ribosomal reduction

- 10ng to 400ng ribosomal reduced RNA
- Ribosomal reduction kits from Qiagen, Illumina, Thermo, NEB etc
- Needs optimisation
- Prepare using tri reagent or Qiagen Total RNA kit
- Recommend paired end sequencing
- More reads

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

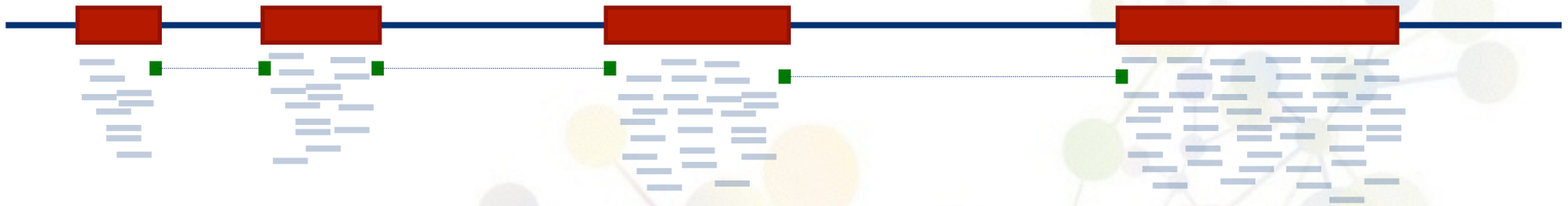
Alignment

Alignment-Tophat

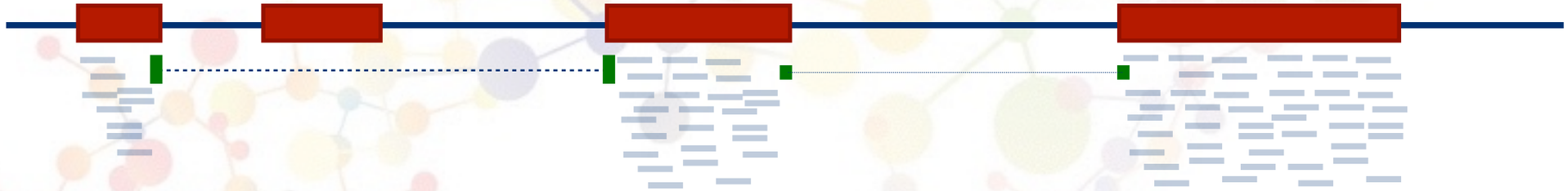
Genomic
reference
sequence

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

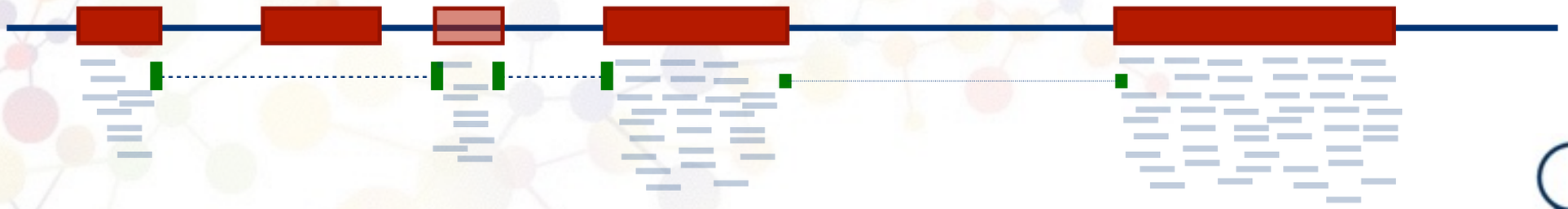
Alignment-Tophat



Exon skipped



Novel exon

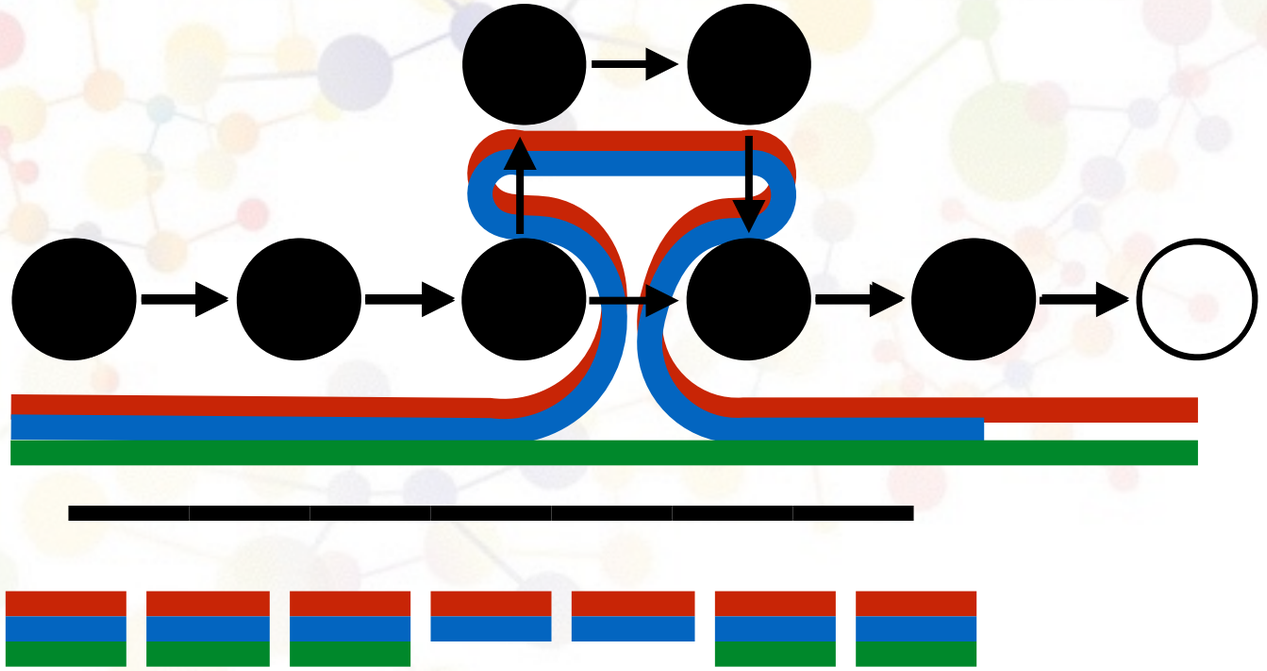


Glasgow Polyomics
www.glasgow.ac.uk/polyomics

HISAT

- Similar to Tophat
- Two types of index
  - An index of the whole genome
  - Multiple indexes of overlapping ~64kb chunks of the genome

Kallisto



Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Differential expression

- Several sources of variability
- Batch effects
- Over dispersion
- Uncertainty in count origin

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Software tools

Cufflinks - Differential analysis at transcript level

DESeq2 - Pools information across genes. Shrinkage estimation for dispersion and fold changes

Voom - Transforms transcript count data to logCPM and maps each observation to the mean variance

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Novel insecticide target discovery by RNA Sequencing

Background

Insects dominant life form

Disease vectors

- malaria, yellow fever, sleeping sickness

15-25% loss to GDP worldwide

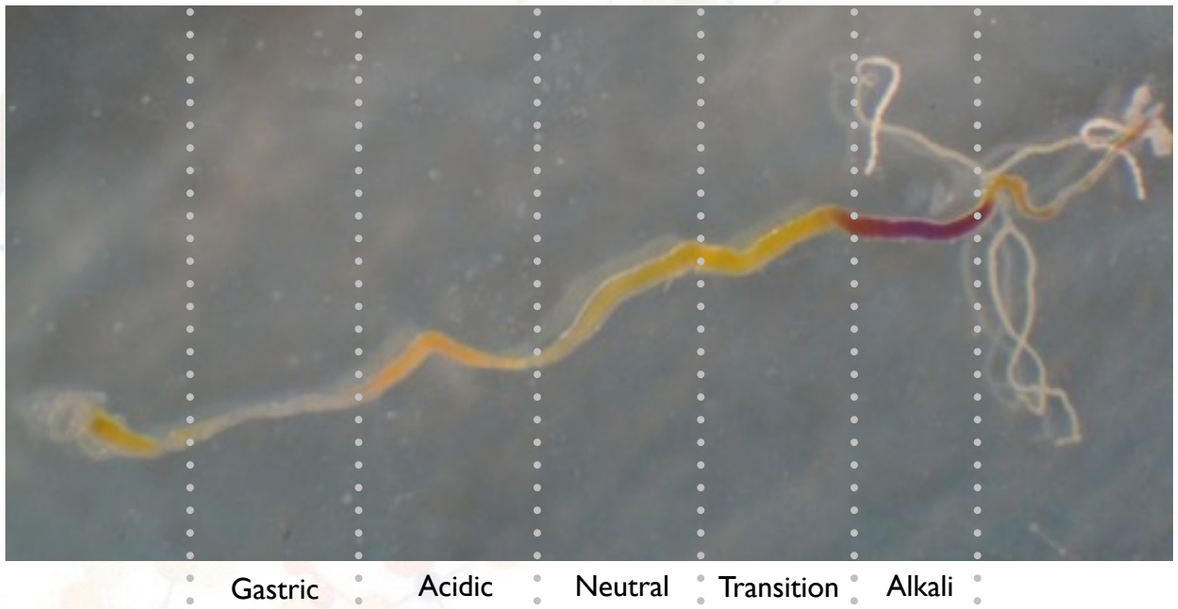
Few new pesticides

Must not harm pollinators

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

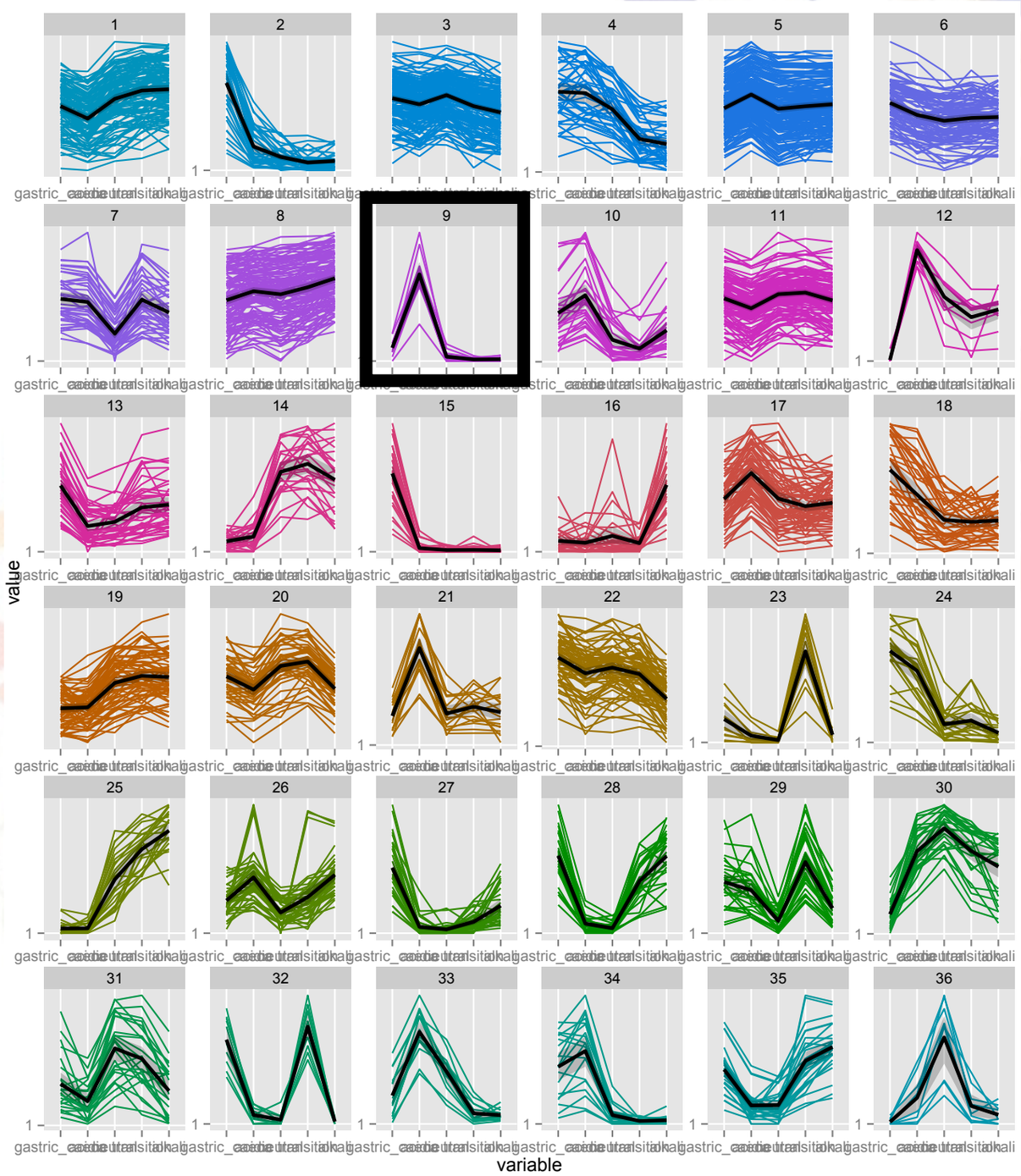
Transcriptomics

- RNA isolated from 5 distinct regions of the digestive tract
- Samples were sequenced
- Compare sample to discover differently expressed transcripts.



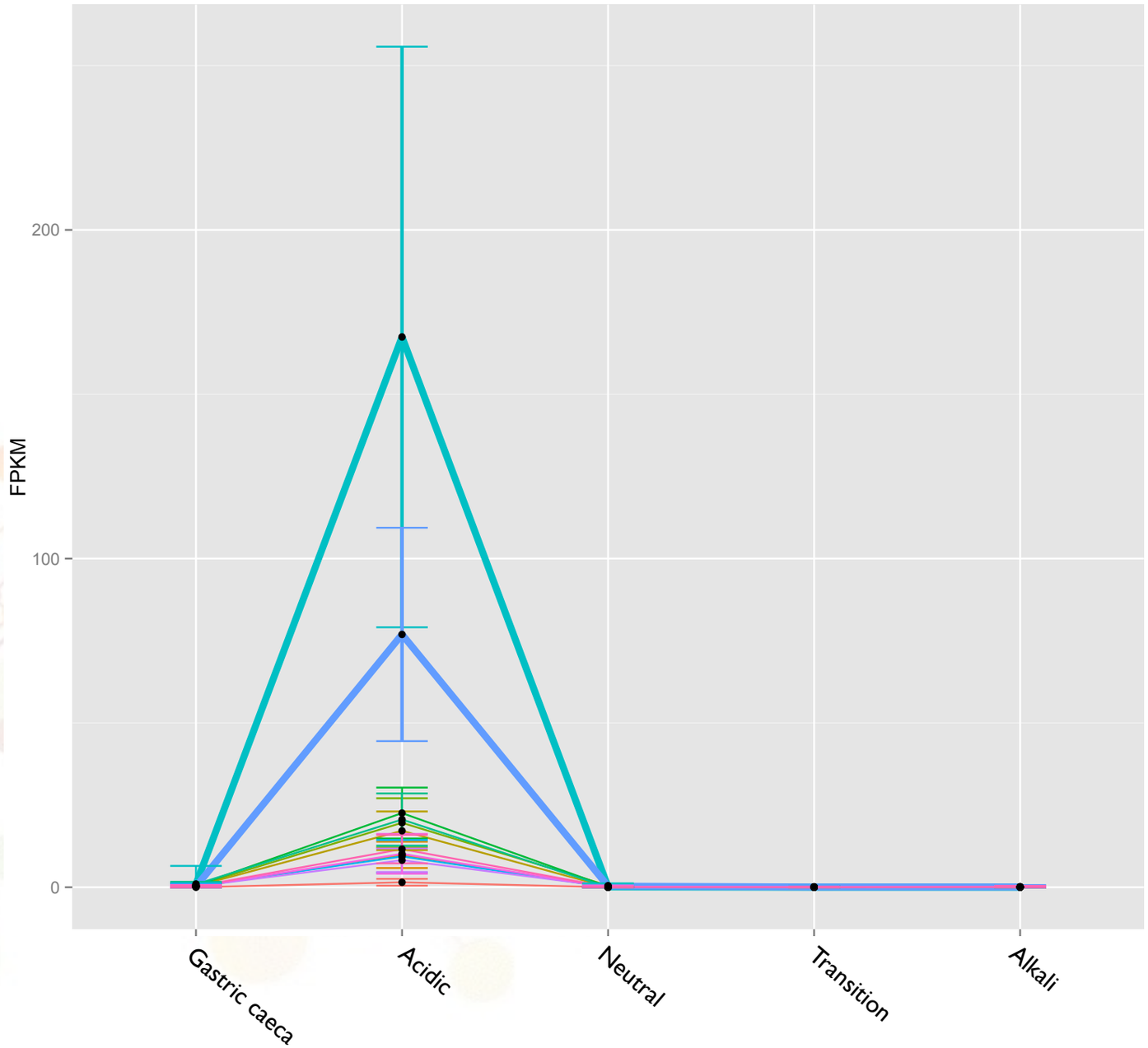
Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Transcriptomics



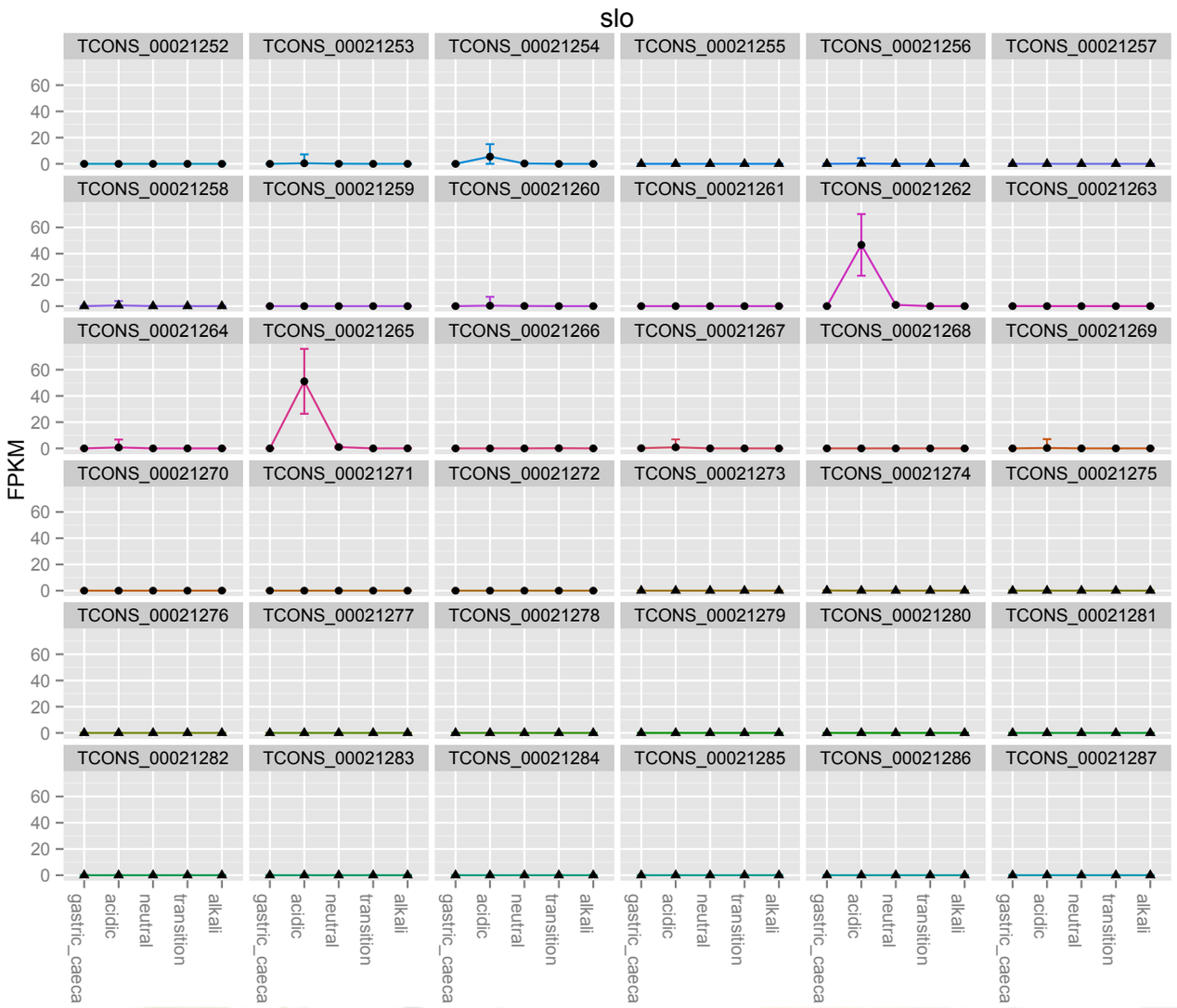
Glasgow Polyomics
www.glasgow.ac.uk/polyomics

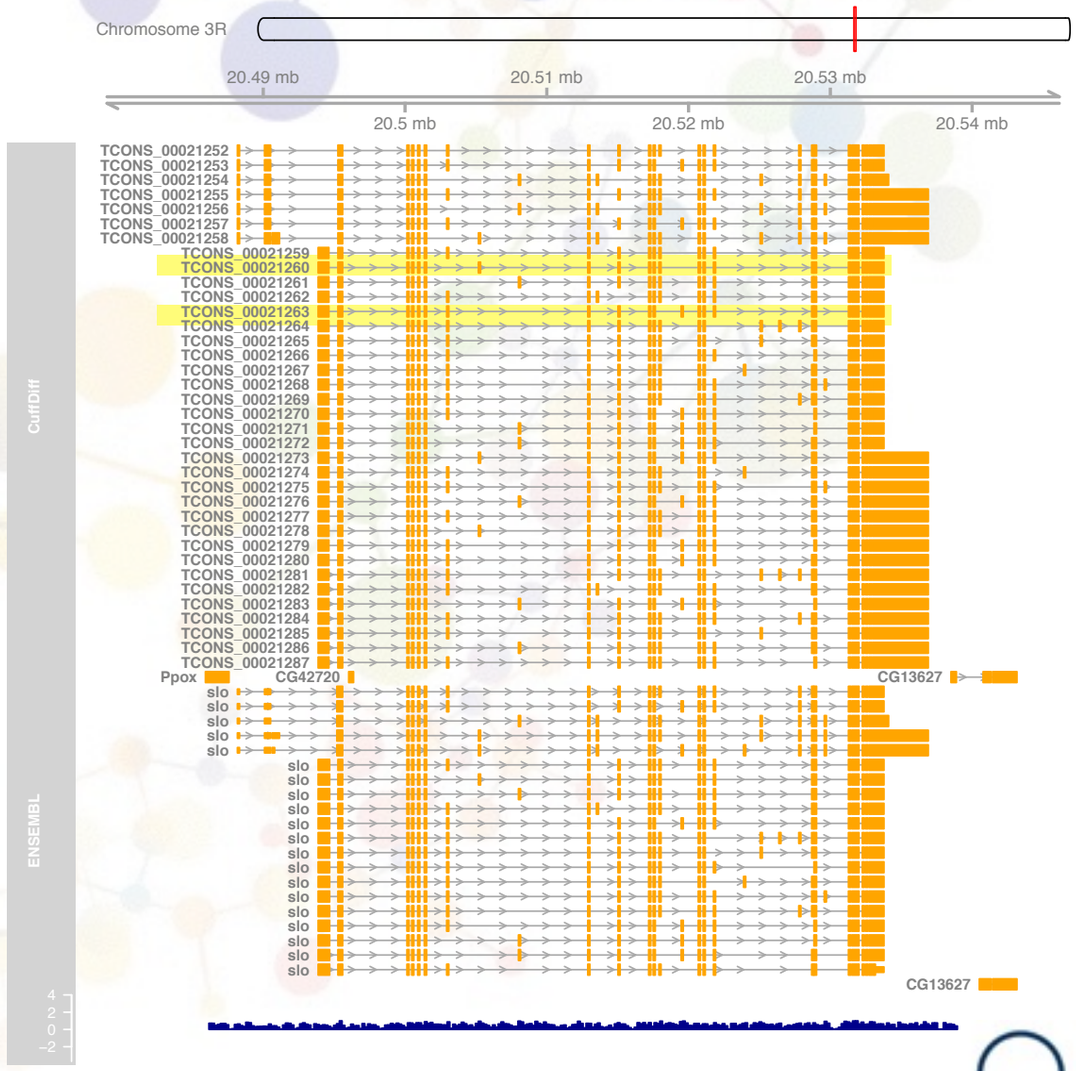
Transcriptomics



Glasgow Polyomics
www.glasgow.ac.uk/polyomics

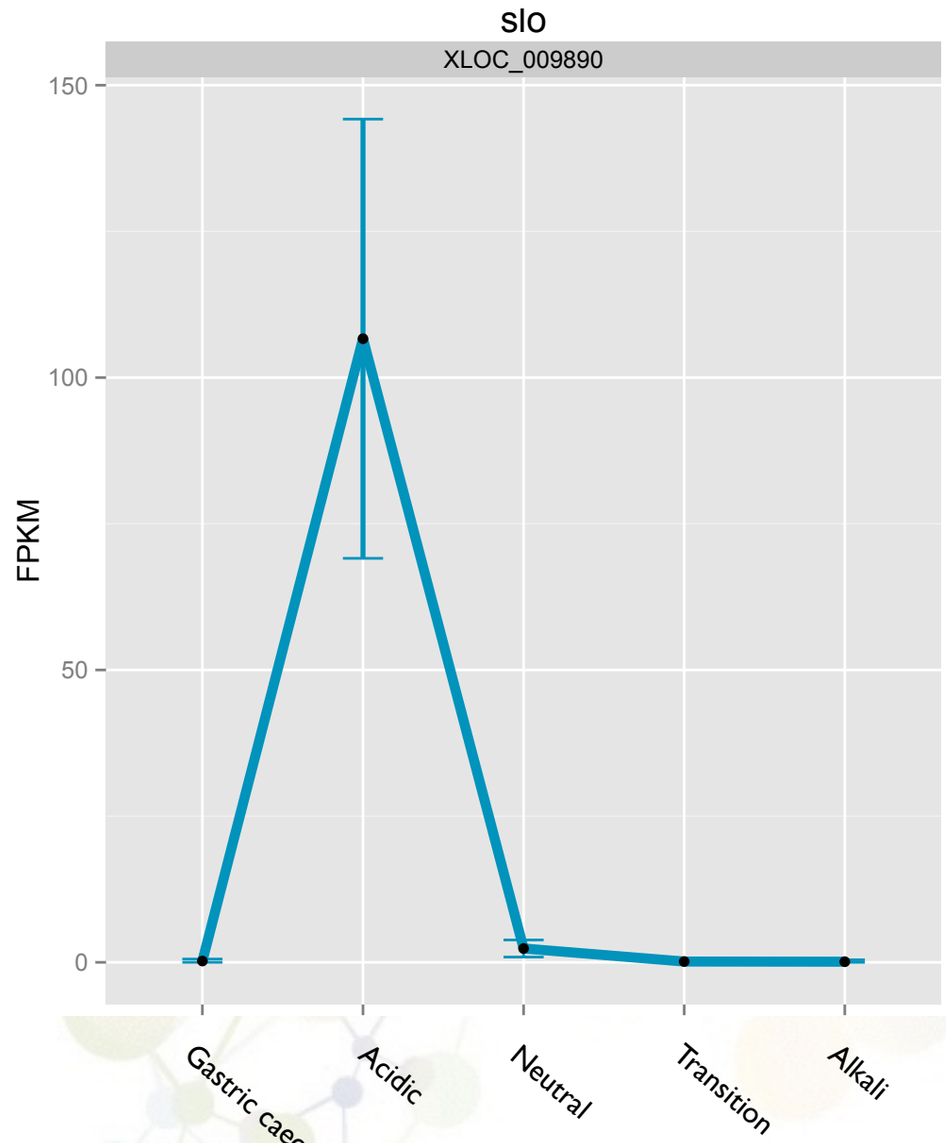
RNA-Seq

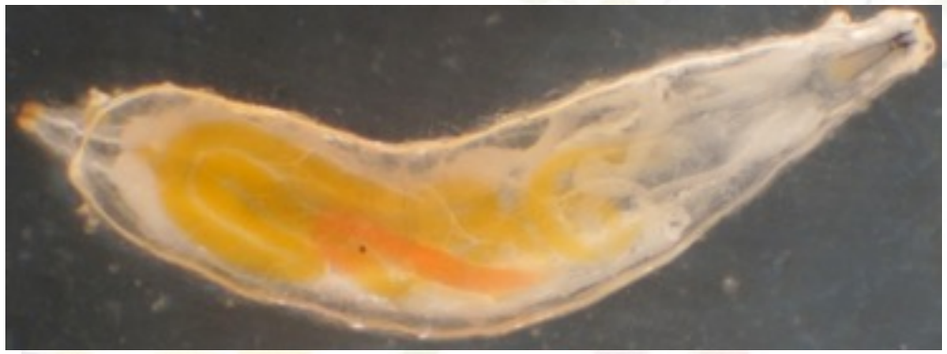




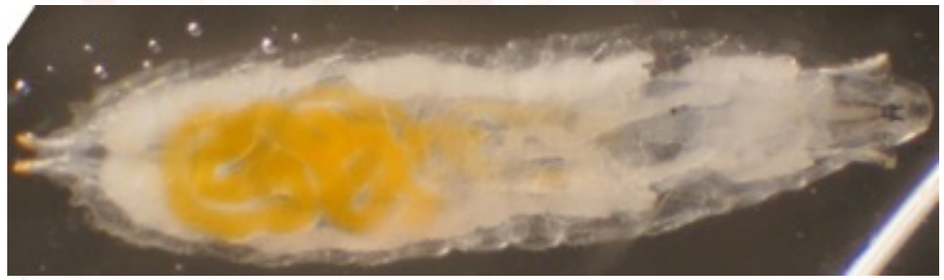
Glasgow Polyomics
www.glasgow.ac.uk/polyomics

What does this transcript do?





Control



Slowpoke RNAi

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

RNA-Seq without a reference

Trinity-
Transcriptome assembly

Sequence RNA

Align to a related reference genome

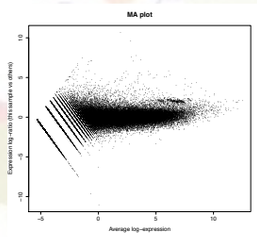
de novo assembly

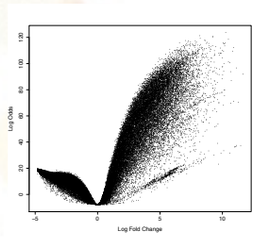
Assembled transcriptome

Annotation

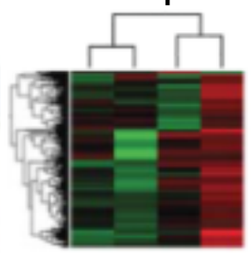
Abundance estimation

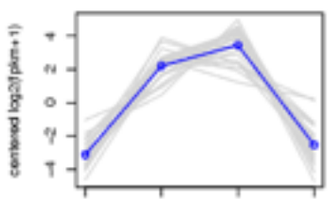
Identify differentially expressed transcripts





Expression patterns, transcript clusters





Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Glasgow Polyomics
www.glasgow.ac.uk/polyomics

Glasgow Polyomics
University of Glasgow
Contact: Allison Jackson
Allison.Jackson@glasgow.ac.uk

Glasgow Polyomics
www.glasgow.ac.uk/polyomics