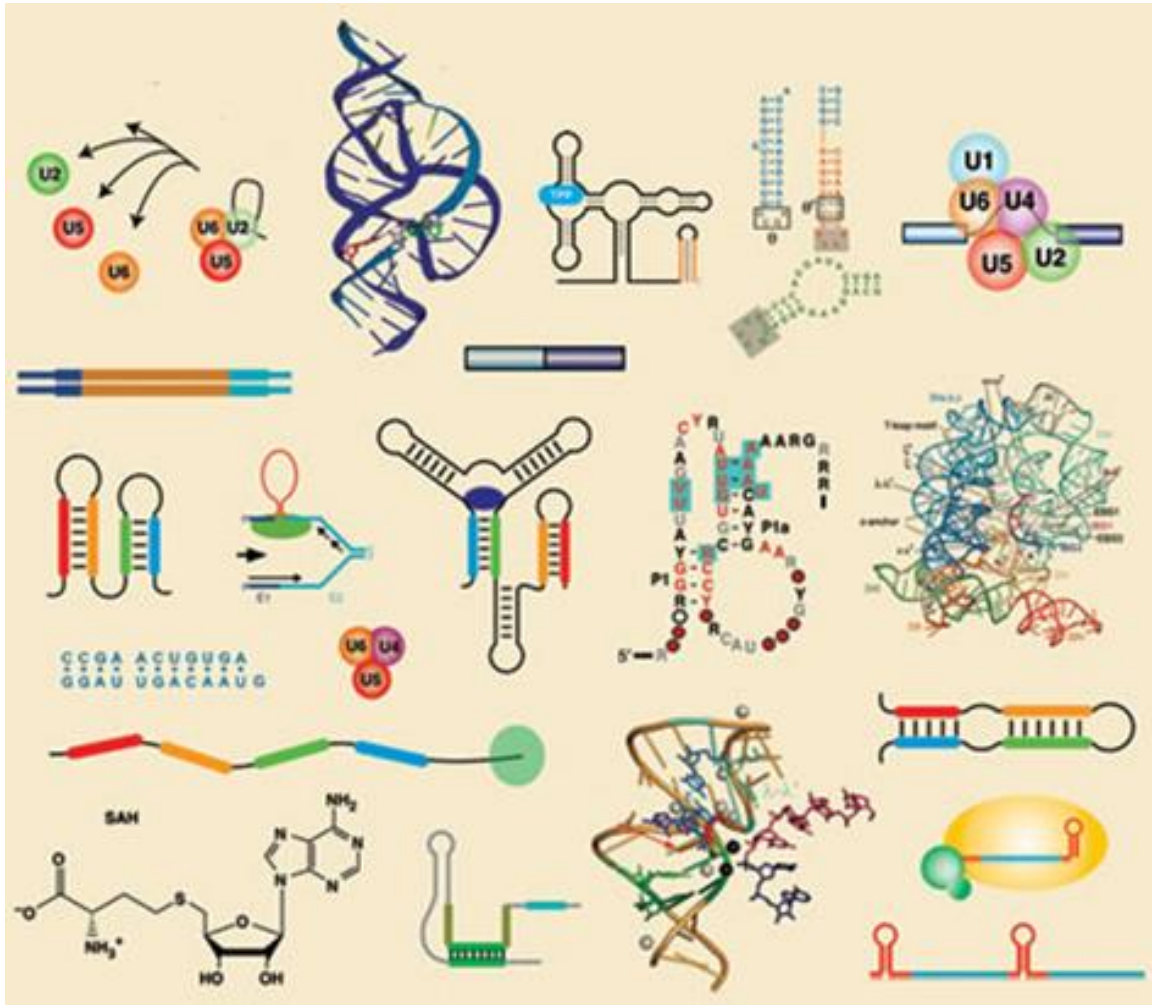


Gene Expression for Biomarker Discovery

BYRON REEVE & CAROLINE BELTRAN

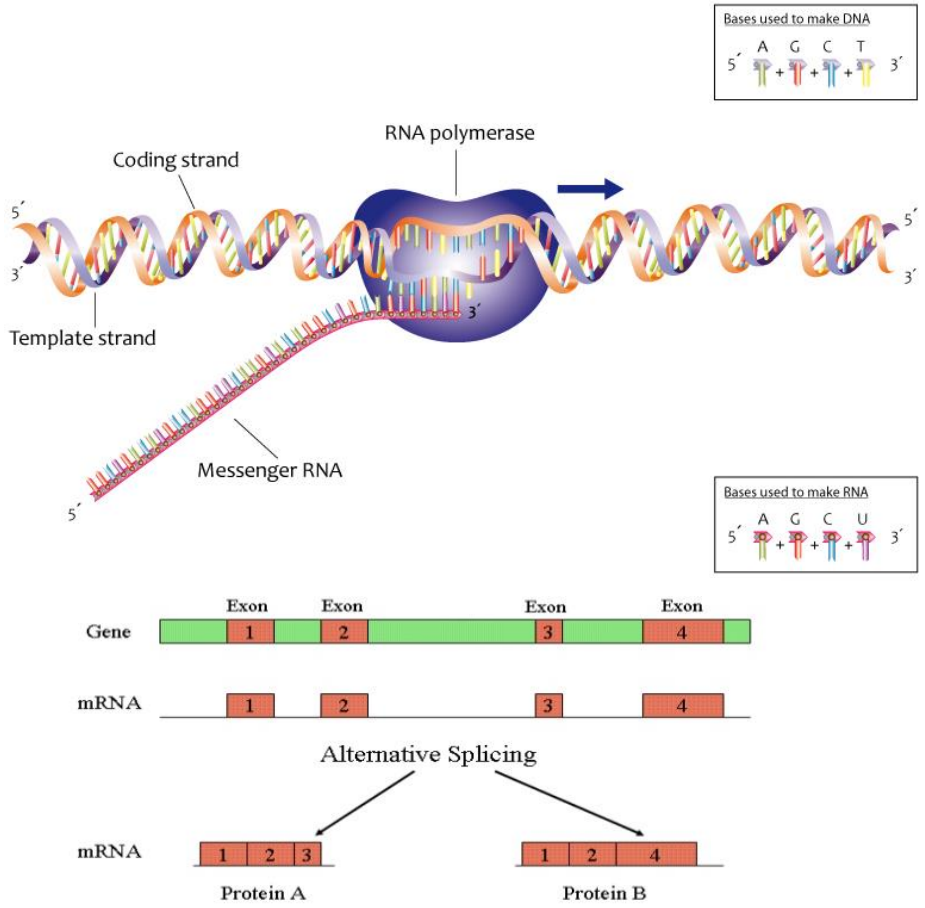
UWC BIOMARKERS MODULE

20 AUGUST 2019

Measuring gene expression

- Central dogma: DNA ⟶ RNA ⟶ Protein
- Formation of functional proteins starts with transcription of coding genes (DNA) into mRNA
- Nearly all cells contain the same set of genes but not all genes are expressed
  - Gene expression is dynamic
    - Different genes and quantities at different times
  - Same gene may be expressed differently in different conditions and cells (RNA editing, modification, alternative splicing…)
- Quantify level at which gene is expressed within cell, tissue or organism
  - Measure the **transcriptome**
  - Coding: mRNA
  - Non-coding: rRNA, tRNA, short and long non-coding RNAs

regulatory function



Gene expression XIST RNA and X Chromosome inactivation

Formation of functional proteins starts with transcription of coding genes into mRNA

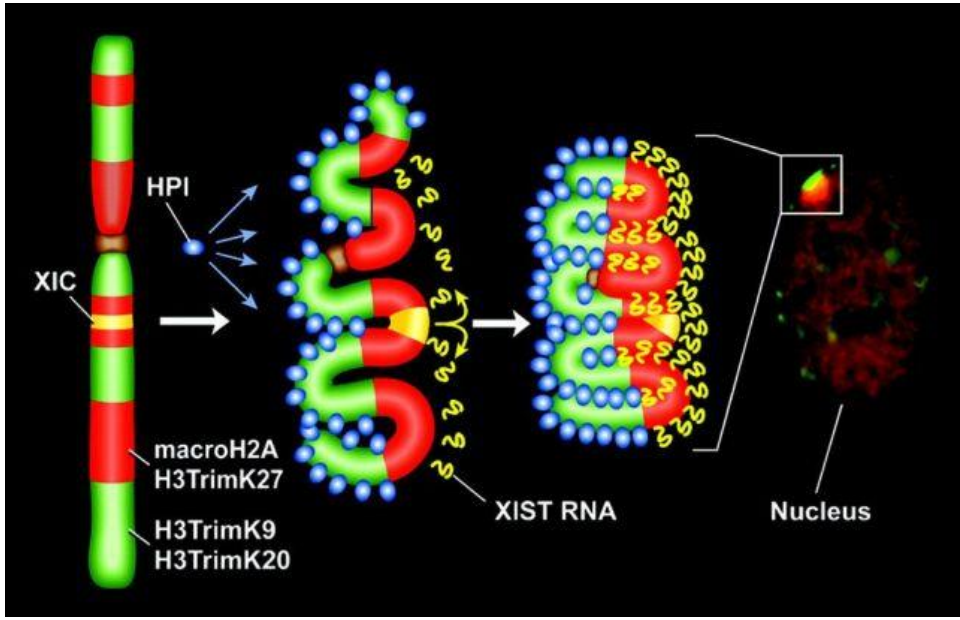
Nearly all cells contain the same set of genes but not all genes are expressed

- Gene expression is dynamic
- Same gene may be expressed differently in different conditions and cells

Quantify level at which gene is expressed within cell, tissue or organism

- Measure the **transcriptome**
- Coding: mRNA
- Non-coding: rRNA, tRNA, short and long non-coding RNAs

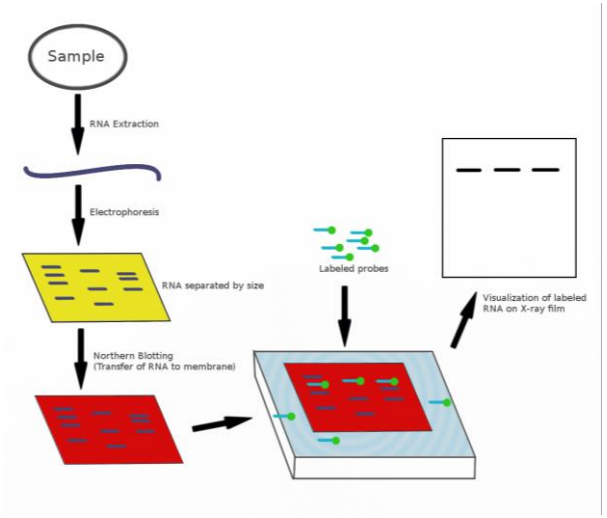
regulatory function

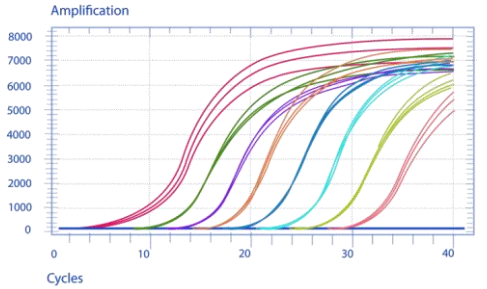


Available methods for transcriptomics

Subdivided into two categories:

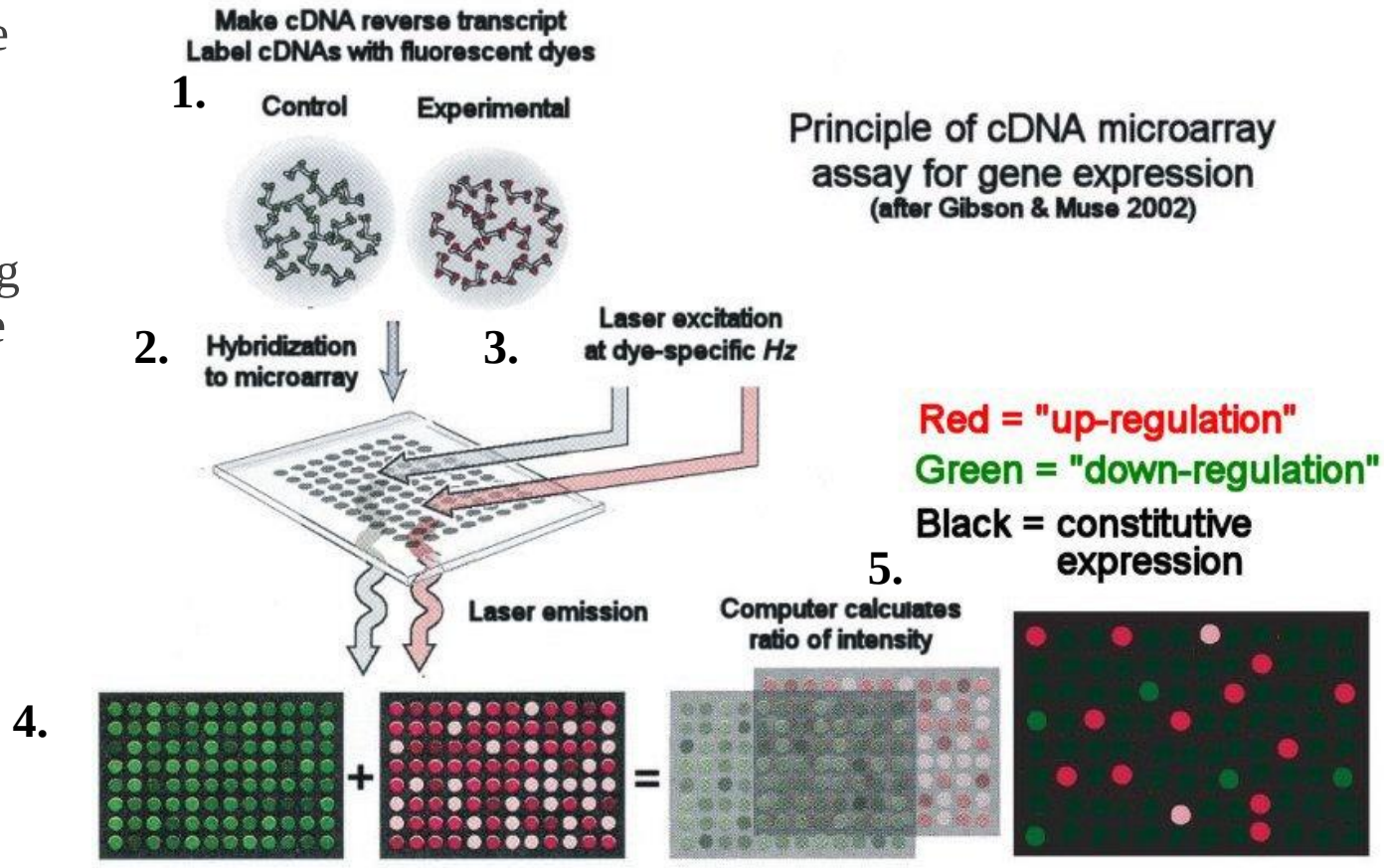
1. **Targeted**: Northern Blots (single genes), RT-qPCR (multiple genes)
2. **Untargeted**: global analysis of expression profiles
   - Gene microarrays:
     - Probes immobilized on surface
       - *a priori* sequence
       - Limited dynamic range
   - RNA-Sequencing: Whole Transcriptome Shotgun Sequencing [Next Generation Sequencing (NGS) of cDNA]





Gene Microarrays

- Analysis of gene expression whose sequence is **already known**
- Up to thousands of genes
- DNA fragments (probes) representing coding region of genes are immobilized onto surface
- cDNA labelled with dyes and hybridized
- Relative gene expression ratio of two fluorophores
- Cross hybridization
  - Need *a priori* knowledge
  - Difficult to detect low expression genes



www.mun.ca/biology/scarr/cDNA_microarray

RNA Sequencing

- Whole Transcriptome Shotgun Sequencing (WTSS)
  - Entire transcriptome [mRNA, tRNA, non-coding, splicing events and post-transcriptional modifications (PTMs)]
  - Mutations, single nucleotide polymorphisms (SNPs), insertions, deletions, copy number variations, chromosome rearrangements
- Small amount of starting material (~100 ng RNA)
  - Immense amount of data
- Large dynamic range and high sensitivity

RNA Seq Workflow

Review | Open Access

A survey of best practices for RNA-seq data analysis

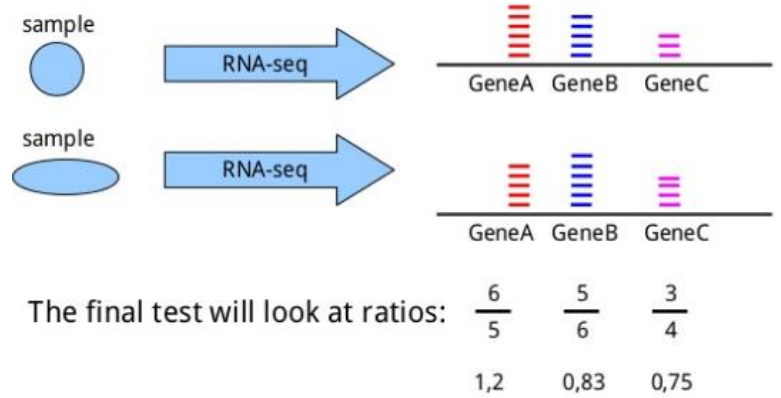
Ana Conesa, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szcześniak, Daniel J. Gaffney, Laura L. Elo, Xuegong Zhang and Ali Mortazavi

Genome Biology 2016 **17**:13

https://doi.org/10.1186/s13059-016-0881-8 | © Conesa et al. 2016

Published: 26 January 2016

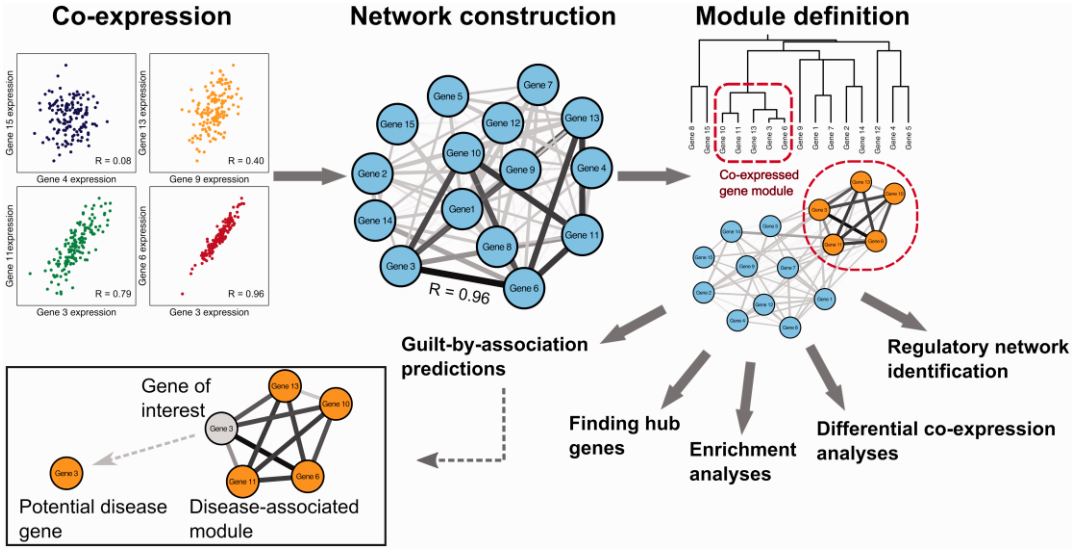
1. Library Preparation
   - RNA Isolation: Dnase treat, RNA quality (RIN), quantify
   - RNA Selection/Depletion (90% = rRNA)
     - "As is"
     - Poly A selection: mature processed mRNAs
     - Ribosomal depletion
   - cDNA synthesis (RNA reverse transcribed to cDNA)
   - Sequence purification (remove low quality sequences)
2. Sequencing: NGS platform
3. Analysis
   - Transcriptome assembly (*de novo* or sequence guided)
   - Differential gene expression: count number of reads that mapped to each locus in transcriptome assembly
   - Co-expression networks/functional analysis



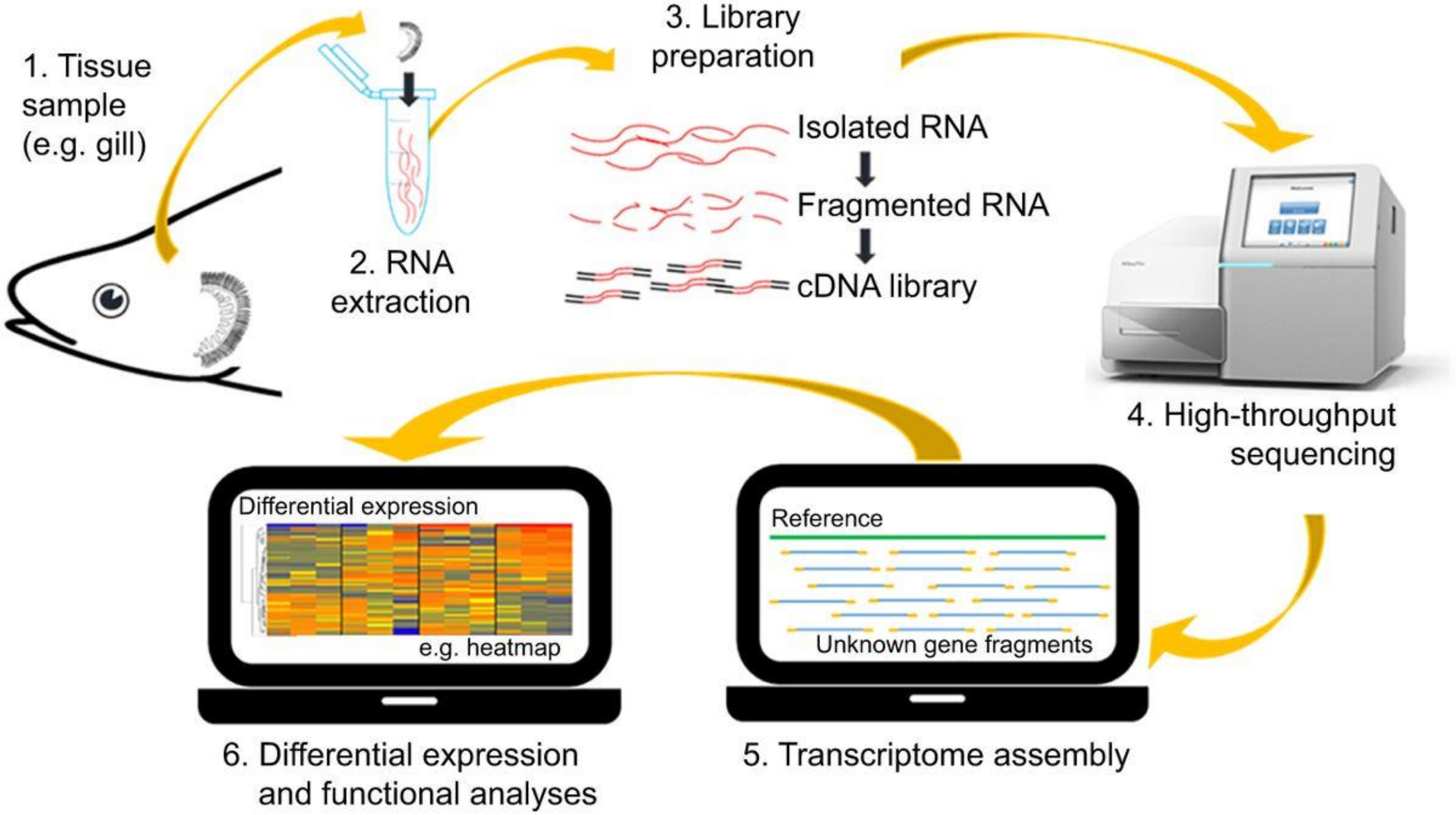
https://doi.org/10.1093/bib/bbw139

RNA Seq Workflow

1. Library Preparation
   - RNA Isolation: Dnase treat, RNA quality (RIN), quantify
   - RNA Selection/Depletion (90% = rRNA)
     - "As is"
     - Poly A selection: mature processed mRNAs
   - cDNA synthesis (RNA reverse transcribed to cDNA)
2. Sequencing: NGS platform
3. Analysis
   - Transcriptome assembly (*de novo* or sequence guided)
   - Gene expression: count number of reads that mapped to each locus in transcriptome assembly
   - Co-expression networks



https://doi.org/10.1093/bib/bbw139



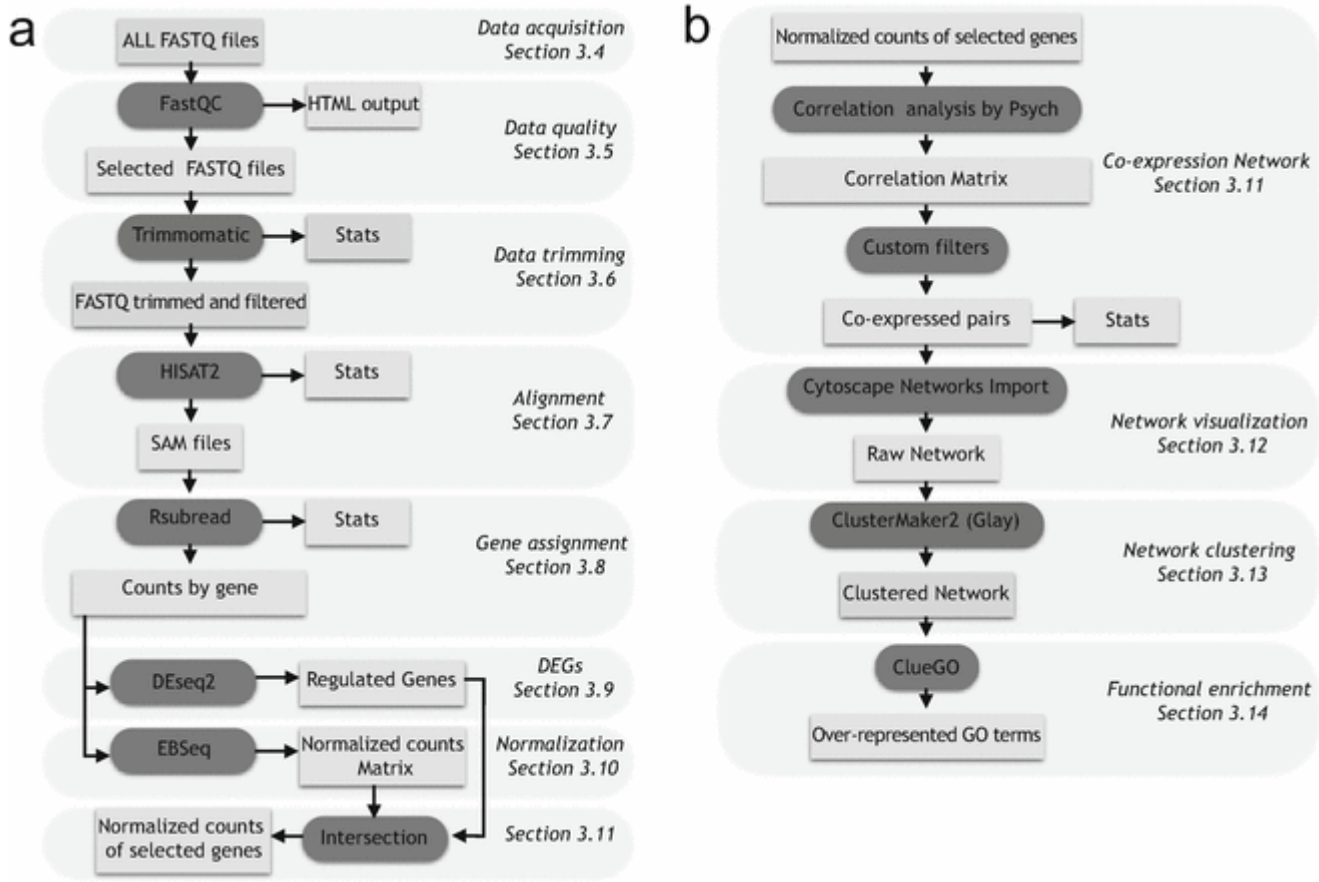
doi: 10.1242/jeb.148833

RNA-Seq Blog
Transcriptome Sequencing Research & Industry News

RNA-seqlopedia

Bioinformatics

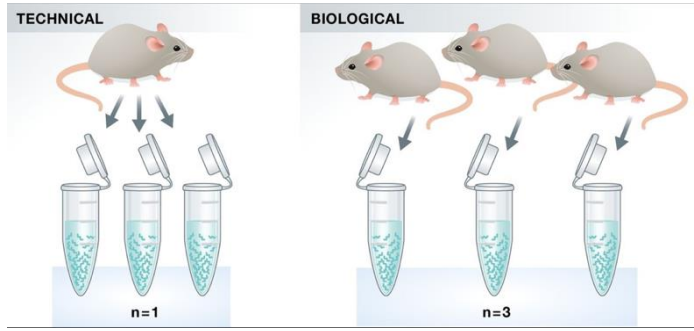
- Bulk of your time
- Highly complex
- Quality control of raw reads: pre-processing, filtering, trimming
- Quality control of mapping
- Quality control of counts
- Statistical analysis for differential expression
- Multivariate statistical analysis/visualization to assess transcriptome-wide differences among samples
- Biological insight
- Tools are constantly evolving!

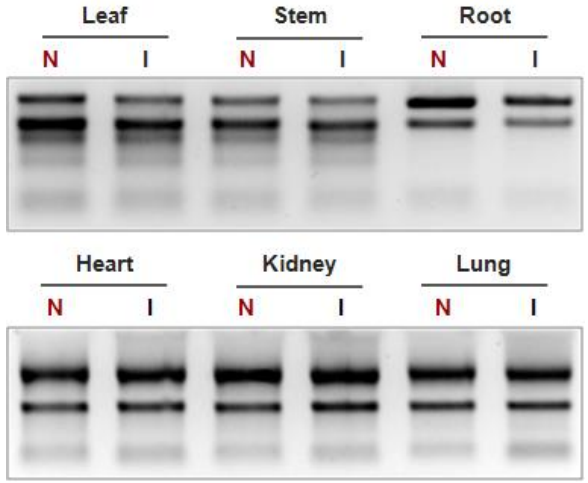


Contreras-López O, Moyano TC, Soto DC, Gutiérrez RA. (2018) **Step-by-Step Construction of Gene Co-expression Networks from High-Throughput Arabidopsis RNA Sequencing Data**. *Methods Mol Biol* 1761:275-301.

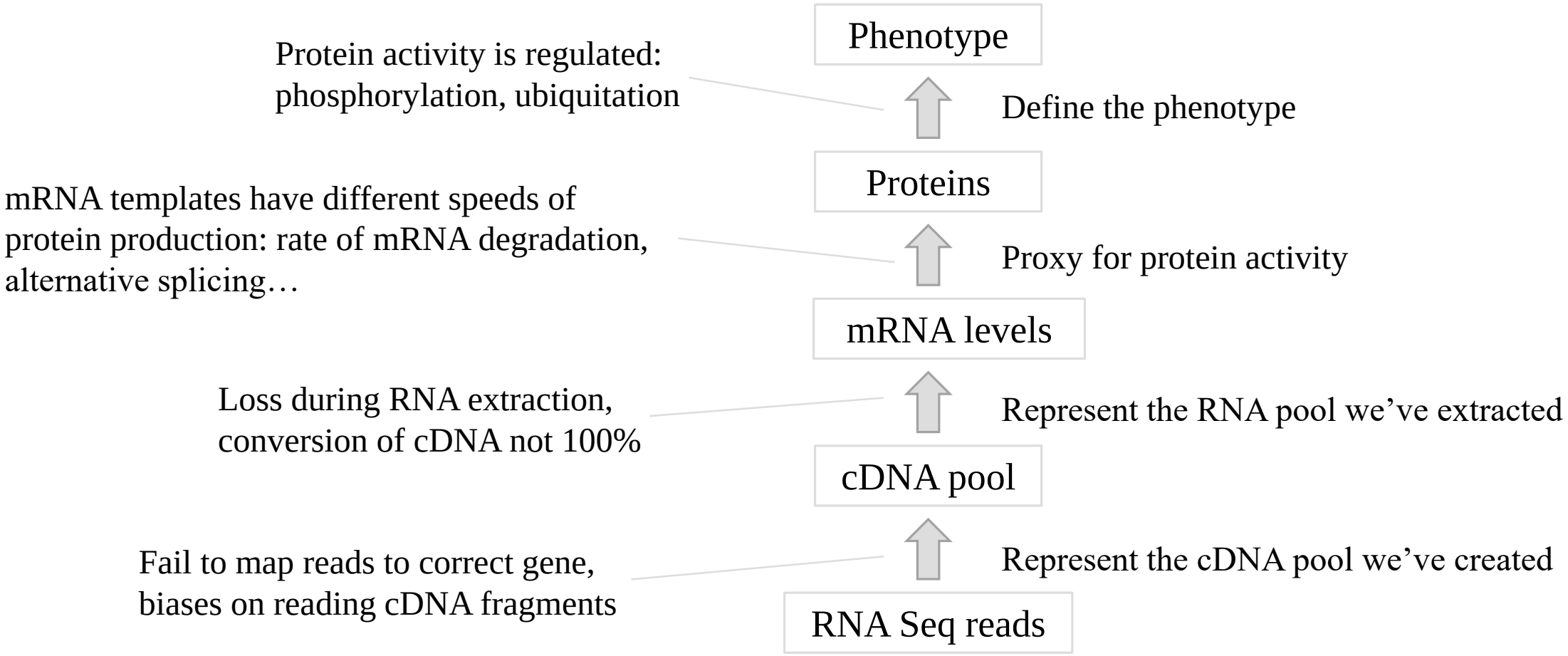
Sampling Design

- What is your research question?
  - Tissue to target?
  - Homogeneous sampling
  - Replicates (biological and technical)
  - RNA yield?
- Stabilize RNA
  - Snap freezing: immediate storage at -80°C
  - RNA Later/Trizol/RNA Later
- RNA is not a stable molecule:
  - Success is dependent on pure and intact RNA
  - Act fast
  - RNase-free tubes and reagents



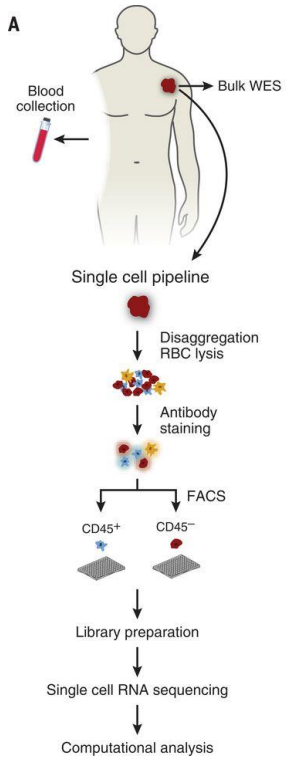


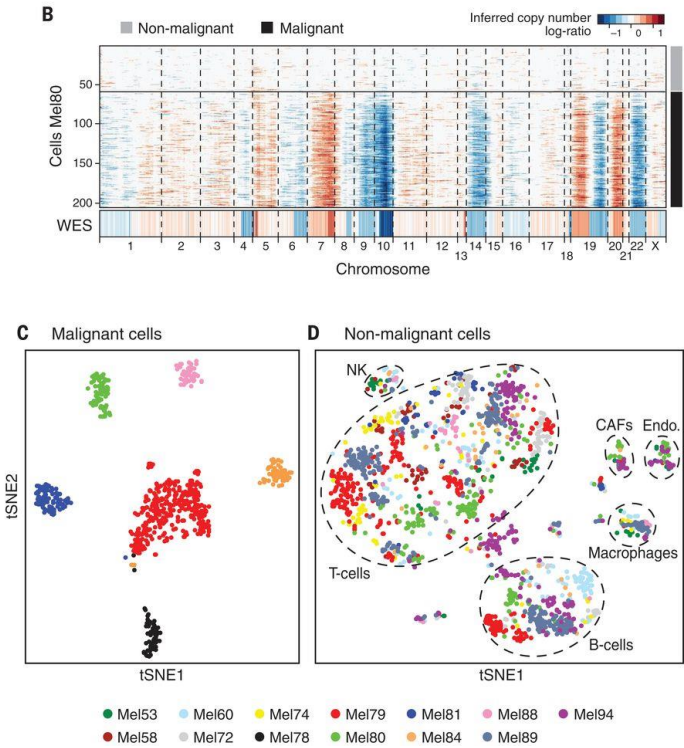
Assumptions



Single Cell RNA Seq

- Bulk sequencing = average expression profile for all the constituent cells
  - Fail to identify if a change in the expression profile is due to a change in regulation or composition in which one cell type arises to dominate the population
- RNA profiling of single cells
- Isolate single cells from tissue/cell suspension (e.g. Fluidigm)
  - Microfluidics/FACS
  - Laser capture microdissection
- Hundreds of single cells in one experiment
- Spatial and cellular context of cells with their transcriptomes

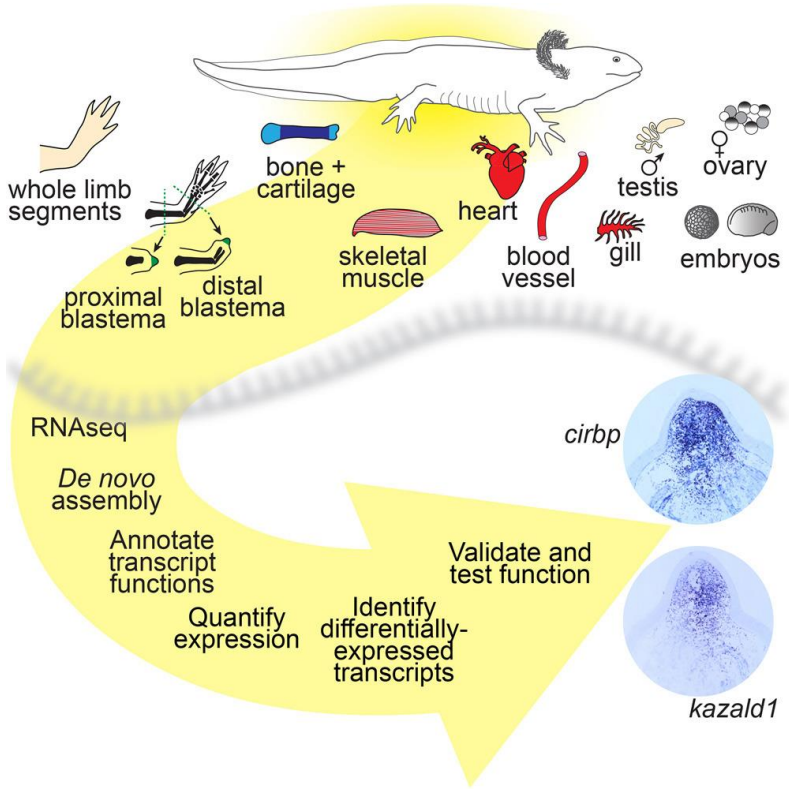




DOI: 10.1126/science.aad0501

Validation

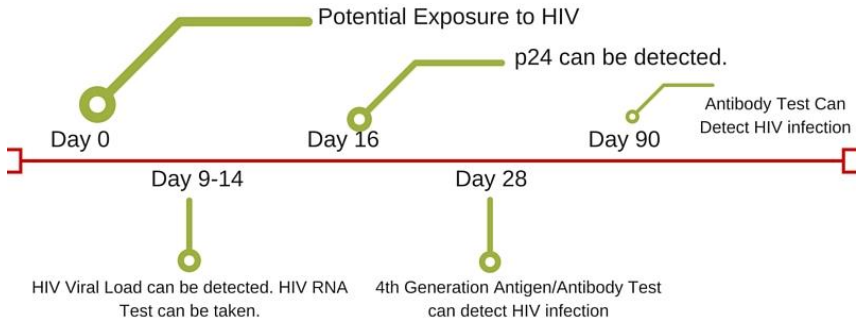
- Whole-transcriptome experiments must be validated by **independent technique**
  - RT-qPCR
  - *In situ* hybridization
  - Functional validation: gene knock-down/rescue experiment
  - Proteomics
- Validation is important especially if you do not have many biological replicates
- Target choice is important:
  - Sequence depth
  - Pathway relevance
  - New biological replicates



DOI: https://doi.org/10.1016/j.celrep.2016.12.063

Application in the clinic

Pathogen detection



AlloMap

- Non-invasive gene expression-based blood test for clinical care of heart transplant recipients
- Quantified score for the risk of rejection based on the measurement of expression of 20 genes

"Personalised medicine"

"multi-gene Fingerprints"

REVIEWS

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Translating RNA sequencing into clinical diagnostics: opportunities and challenges

Table 1 | **Selected examples of current RNA-based clinical tests**

RNA biomolecule	Method	Examples	Use
Viral RNA	qRT-PCR	• Influenza virus[68] • Dengue virus[69] • HIV[70] • Ebola virus[71]	Viral detection and typing
mRNA	qRT-PCR	• AlloMap (CareDx; heart transplant)[15,16] • Cancer Type ID (BioTheranostics)[143]	Diagnosis
	Microarray	Afirma Thyroid Nodule Assessment (Veracyte)[116]	Diagnosis
	qRT-PCR	• OncotypeDx (Genome Health; breast, prostate and colon cancer)[144–147] • Breast Cancer Index (BioTheranostics)[148] • Prolaris (Myriad; prostate cancer)[136]	Prognosis
	Digital barcoded mRNA analysis	Prosigna Breast Cancer Prognostic Gene Signature (Nanostring)[149]	Prognosis
	Microarray	• MammaPrint (Agendia; breast cancer)[134] • ColoPrint (Agendia; colon cancer)[150] • Decipher (Genome Dx; prostate cancer)[151]	Prognosis
miRNA	Microarray	Cancer Origin (Rosetta Genomics)[152]	Diagnosis
Fusion transcript	qRT-PCR	AML (*RUNX1–RUNX1T1*)[18]	Diagnosis
	qRT-PCR	*BCR–ABL1* (REF. 21)	Monitoring molecular response during therapy
	qRT-PCR (exosomal RNA)	ExoDx Lung (ALK) (Exosome Dx)[161]	Fusion detection
	RNA-seq	FoundationOne Heme[2,3]	Fusion detection

AML, acute myeloid leukaemia; *BCR*, breakpoint cluster region; miRNA, microRNA; qRT-PCR, quantitative reverse transcription PCR; RNA-seq, RNA sequencing; *RUNX1*, runt-related transcription factor 1; *RUNX1T1*, runt-related transcription factor1 translocated to 1 (cyclin D related).

Summary

- Transcriptomics offers insights into how gene expression dictates and controls cell physiology
- RNA Seq offers a versatile, comprehensive solution with high reproducibility and resolution without “prior knowledge”
- Bioinformatics of transcriptomic data is challenging and complex
- Experimental design should be carefully considered and appropriate number of replicates calculated
- Transcriptomics are being used more routinely in disease management