Intro and Best Practices: RNA-Seq

INTRODUCTION TO RNA-SEQ DATA ANALYSIS

BTEP SERIES 2017

RNA-Seq Applications

Differential Gene Expression

- Looks at genes that are at least at the detection limit of microarrays
- Most straightforward, requires less read depth (10-30 M reads)
- Can be more cost-effective than microarrays

Differential Transcript Expression (Isoform switching)

- Still confined to known transcripts / isoforms
- Complexity is in the assignment of exons to particular isoforms
- Many algorithms can differ in results

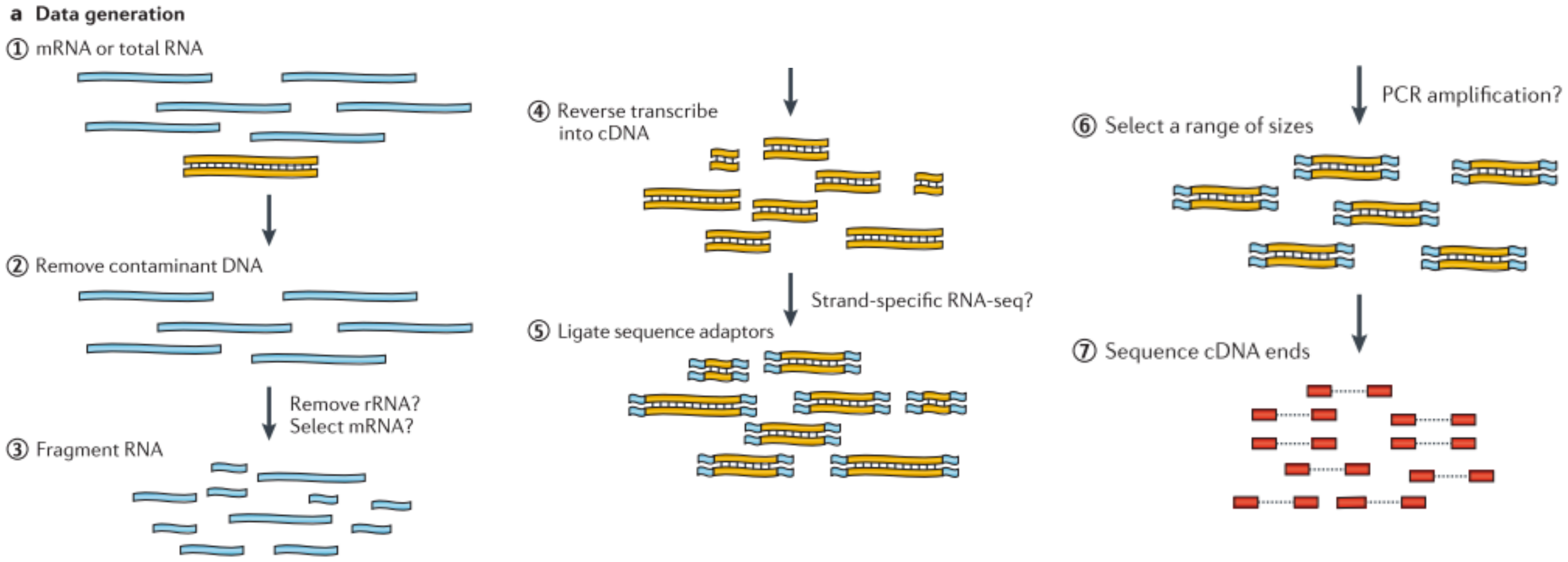
Transcript Discovery / Whole Transcriptome Profiling

- Interest is in looking for new isoforms or unannotated genes
- More complex in terms of bioinformatics analysis
- Can find false positives, depending on leniency of algorithm

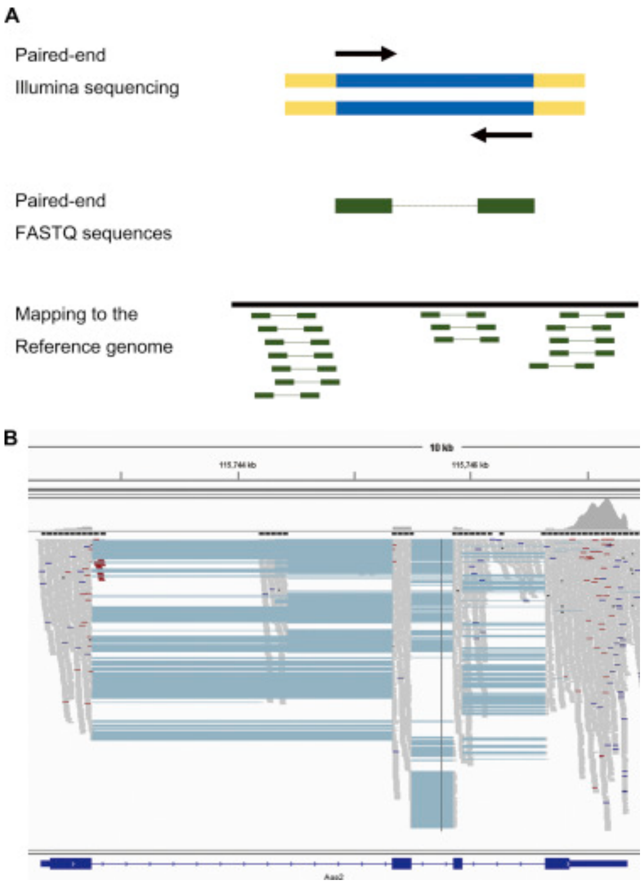
Others

- SNP/Somatic Variant/Gene Fusion Detection

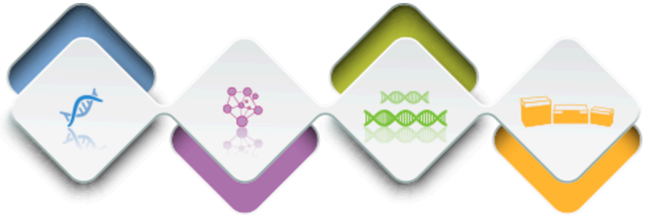
Method – Preparation

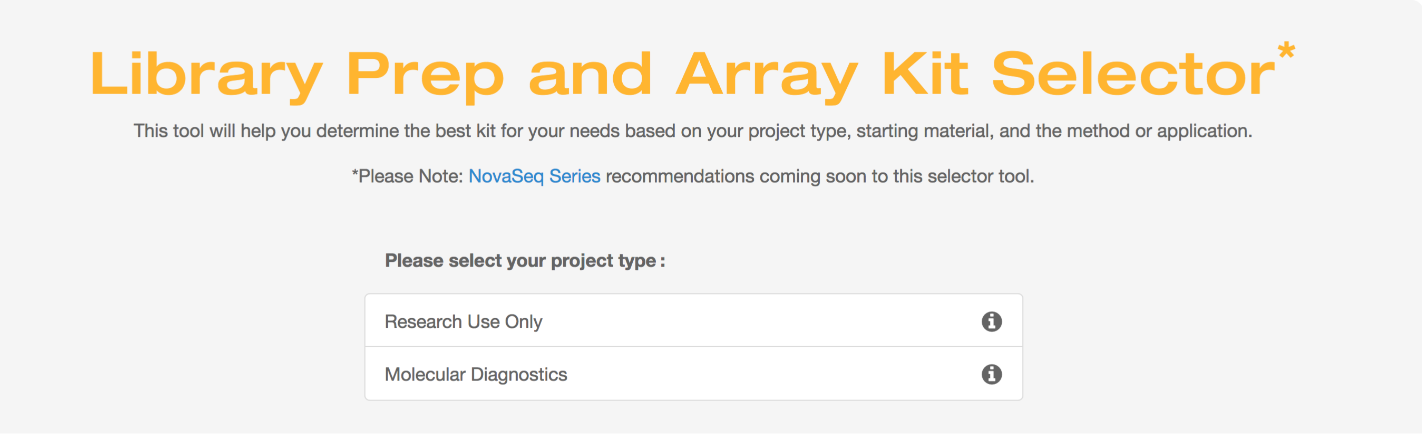


Paired-end Sequencing



Which method?





Library Prep and Array Kit Selector*

This tool will help you determine the best kit for your needs based on your project type, starting material, and the method or application.

*Please Note: NovaSeq Series recommendations coming soon to this selector tool.

Please select your project type :

- Research Use Only
- Molecular Diagnostics

https://www.illumina.com/library-prep-array-kit-selector.html

Which RNA type?

Library Kits available:

mRNA

Whole Transcriptome

Targeted

miRNA

Low Input

Ribosomal Profiling

Which library method?

	TruSeq RNA v2	TruSeq Stranded mRNA	TruSeq RNA Access
Input Amount	0.1 – 1ug High Quality Total RNA	NeoPrep: 25-100ng	10ng High Quality Total RNA
	10-400ng previously isolated mRNA	LT/HT: 0.1 – 1ug Total RNA	20ng Degraded Total RNA
FFPE Compatible	No	No	Yes
Capture Method	Oligo dT beads capture poly-A tail	Oligo dT beads capture poly-A tail	Capture probes targeting coding RNA sequence
Capture Content	Coding Transcriptome	Coding Transcriptome	Coding Transcriptome

Which library method?

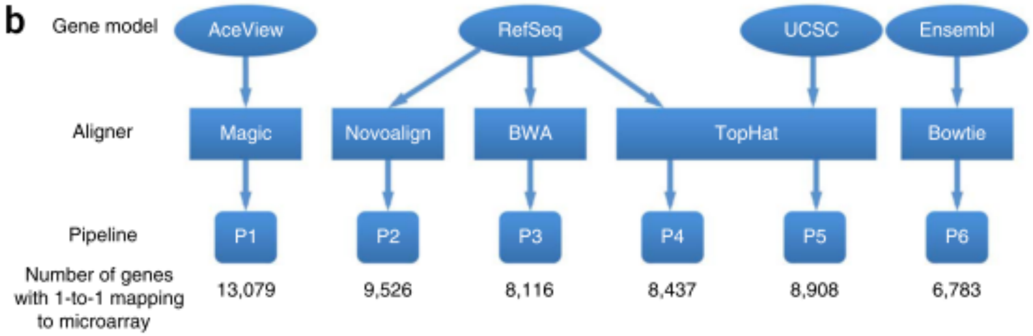
	TruSeq Stranded Total RNA Ribo-Zero H/M/R	Clontech SMART-Seq v4 Ultra Low Input RNA Kit + Nextera XT
Input Amount	0.1–1 µg of total RNA (mid to high-quality)	1–1,000 intact cells (or as little as 10 pg–10 ng of total RNA
FFPE Compatible	Yes	No
Capture Method	RT + Random Primers	cDNA Synthesis Using Template Switching Technology
Capture Content	Coding and Non-coding Transcriptome	Coding Transcriptome

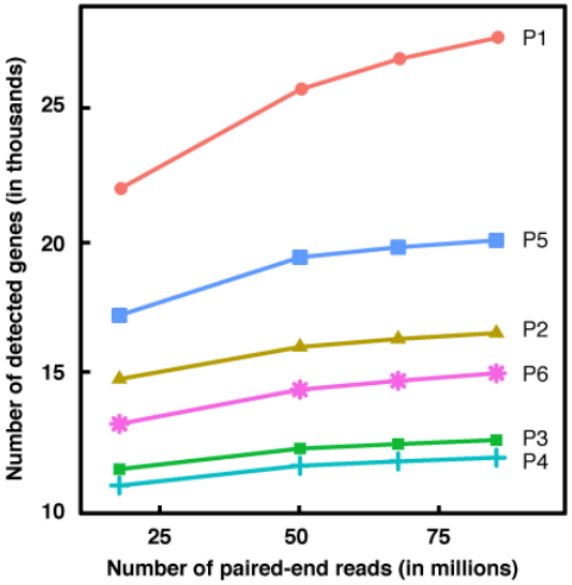
Cost:

	HiSeq 2500	**HiSeq 3000**
Raw reads per lane	400 M	600 M
Cost/sample*		
mRNA-Seq: 20 M PE	18 samples/lane: $126 + $100 = $226	27 samples/ lane: $58 + $100 = $158 (75 bp) $73 + $100 = $173 (150 bp)
Total RNA-Seq: 60 M PE	6 samples/lane: $378 + $126 = $500	9 samples/lane: $175 + $126 = $300 (75 bp) $244 + $126 = $370 (150 bp)

**Cost doesn't include 33% CCR subsidy*

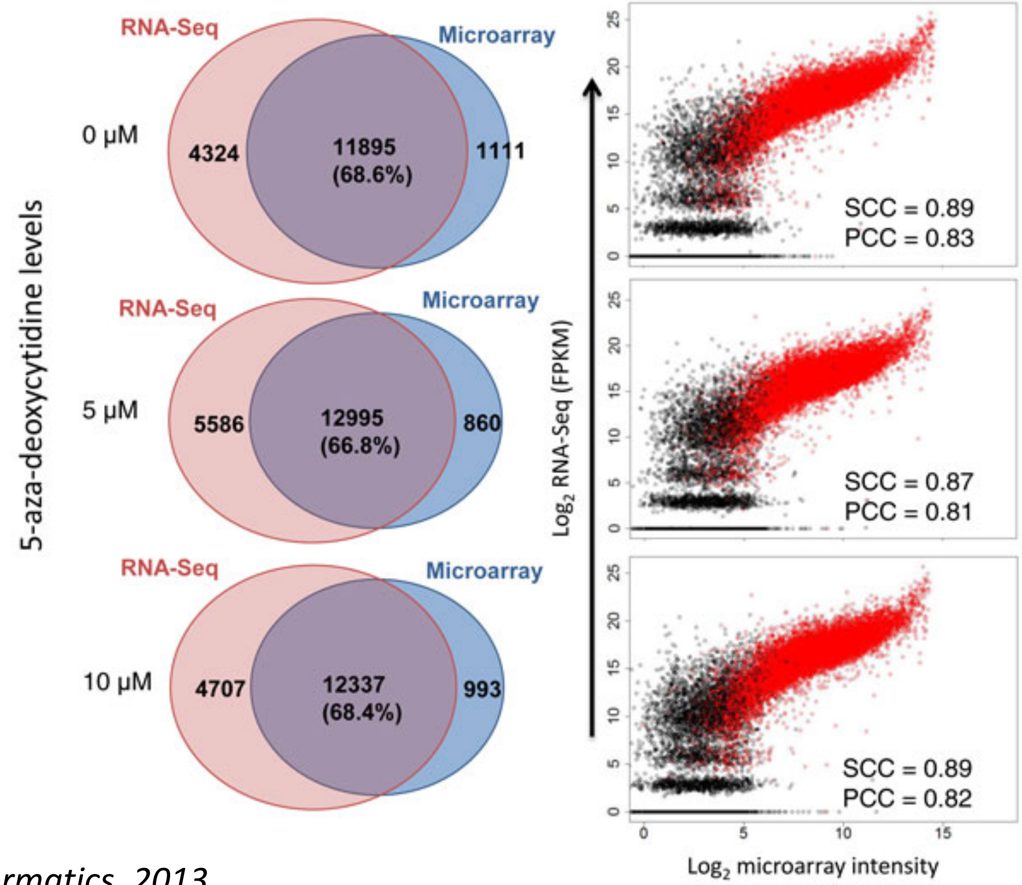
What sequencing depth is enough?





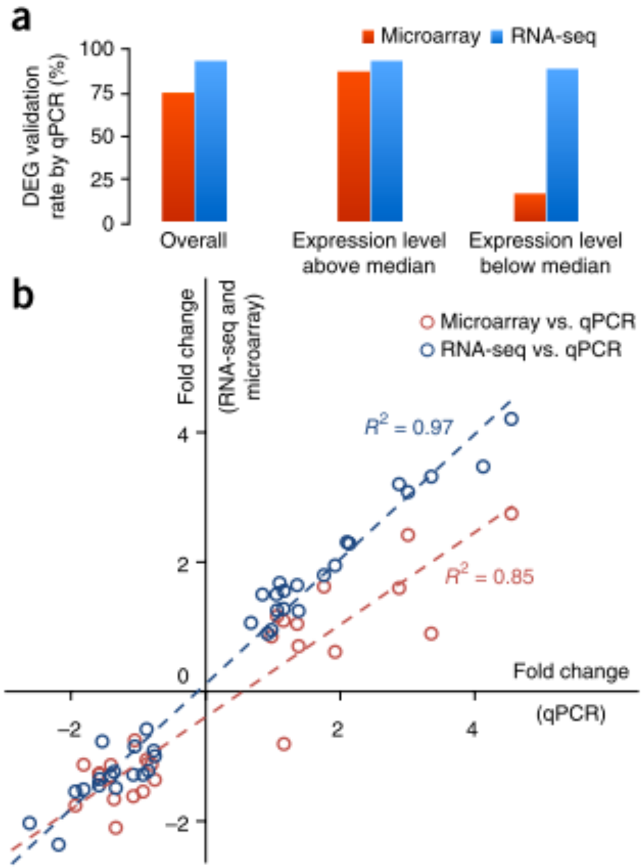
Wang et al, Nature Biotechnology, 2014

Comparison between Microarray and RNA-Seq



Xu et al, BMC Bioinformatics, 2013

Comparison between Affymetrix, RNA-Seq and qPCR



Wang et al, Nature Biotechnology, 2014

RNA-Seq or Microarray?

Current configuration for running samples on HiSeq 2500:

Whole Transcriptome profiling: Ribo-Zero

~25-50M PE reads

(6-12 samples/lane)

mRNA Profiling:

~10-20M PE reads

(18-36 samples/lane)

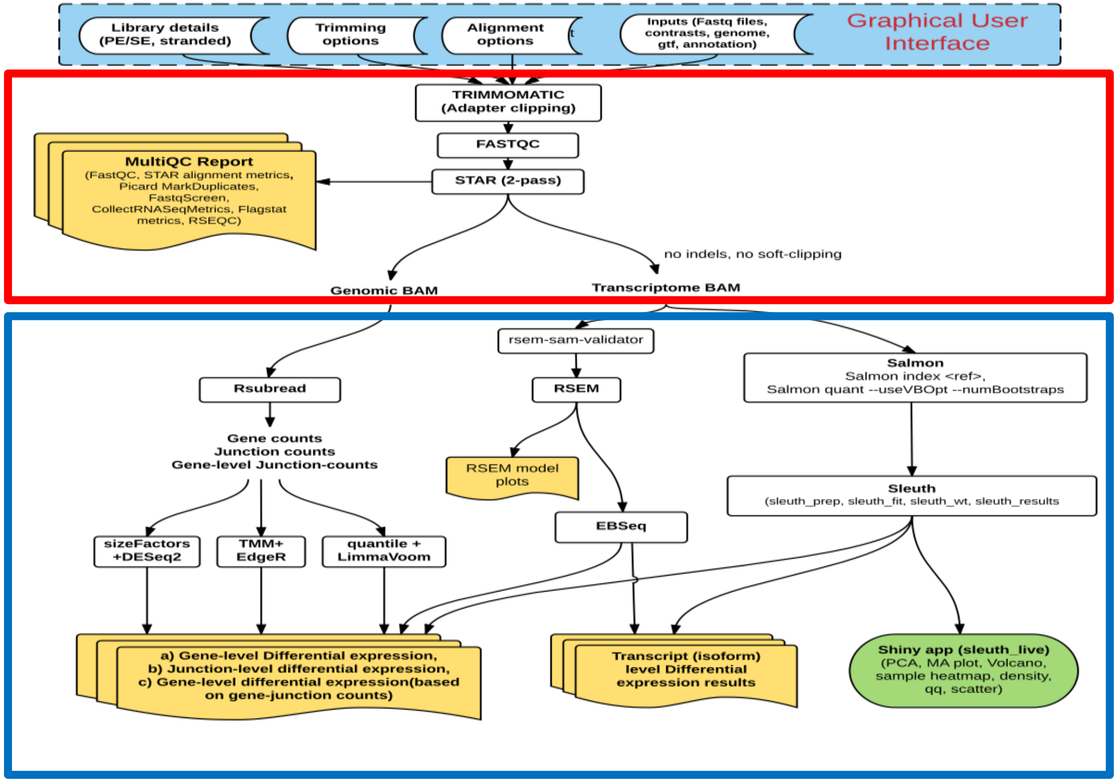
Microarray

- Pathways
- Genes or known transcripts
- Well-expressed

RNA-Seq

- Full transcriptome analysis
- Rare transcripts
- Splice variants
- Fusion transcripts

RNA-Seq Pipeline Workflow: CCBR Pipeliner



Library details (PE/SE, stranded)

Trimming options

Alignment options

Inputs (Fastq files, contrasts, genome, gtf, annotation)

Graphical User Interface

TRIMMOMATIC (Adapter clipping)

FASTQC

STAR (2-pass)

MultiQC Report (FastQC, STAR alignment metrics, Picard MarkDuplicates, FastqScreen, CollectRNASeqMetrics, Flagstat metrics, RSEQC)

no indels, no soft-clipping

Genomic BAM

Transcriptome BAM

rsem-sam-validator

Rsubread

RSEM

Salmon
Salmon index <ref>,
Salmon quant --useVBOpt --numBootstraps

Gene counts
Junction counts
Gene-level Junction-counts

RSEM model plots

Sleuth
(sleuth_prep, sleuth_fit, sleuth_wt, sleuth_results

EBSeq

sizeFactors +DESeq2

TMM+ EdgeR

quantile + LimmaVoom

a) Gene-level Differential expression,
b) Junction-level differential expression,
c) Gene-level differential expression(based on gene-junction counts)

Transcript (isoform) level Differential expression results

Shiny app (sleuth_live)
(PCA, MA plot, Volcano, sample heatmap, density, qq, scatter)

STEP1: INITIAL QC

STEP2: COUNTING & DEG

Data Types

Raw Reads:

- Fastq files: usually in .gz format

Aligned Reads:

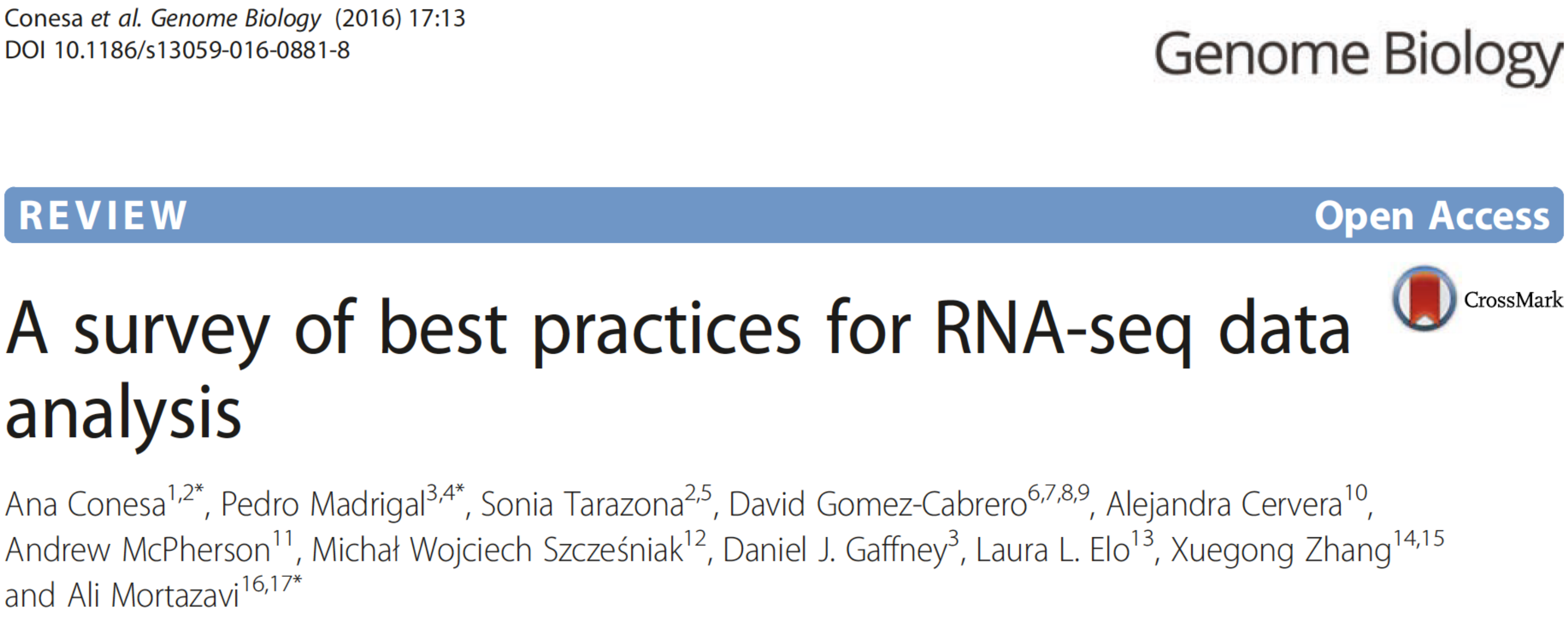
- SAM: Sequence Alignment/Map format
- BAM: binary version of SAM
- BAI: BAM index (for fast retrieval of BAM reads)

QC Report: MultiQC Report

- FastQC
- RSeQC
- Samtools
- Picard

Gene Counts and Differentially Expressed Genes (DEG) Reports

A good review:



Conesa *et al. Genome Biology* (2016) 17:13
DOI 10.1186/s13059-016-0881-8

Genome Biology

REVIEW Open Access

A survey of best practices for RNA-seq data analysis

Ana Conesa[1,2*], Pedro Madrigal[3,4*], Sonia Tarazona[2,5], David Gomez-Cabrero[6,7,8,9], Alejandra Cervera[10], Andrew McPherson[11], Michał Wojciech Szcześniak[12], Daniel J. Gaffney[3], Laura L. Elo[13], Xuegong Zhang[14,15] and Ali Mortazavi[16,17*]

Generic roadmap for expt design & analysis

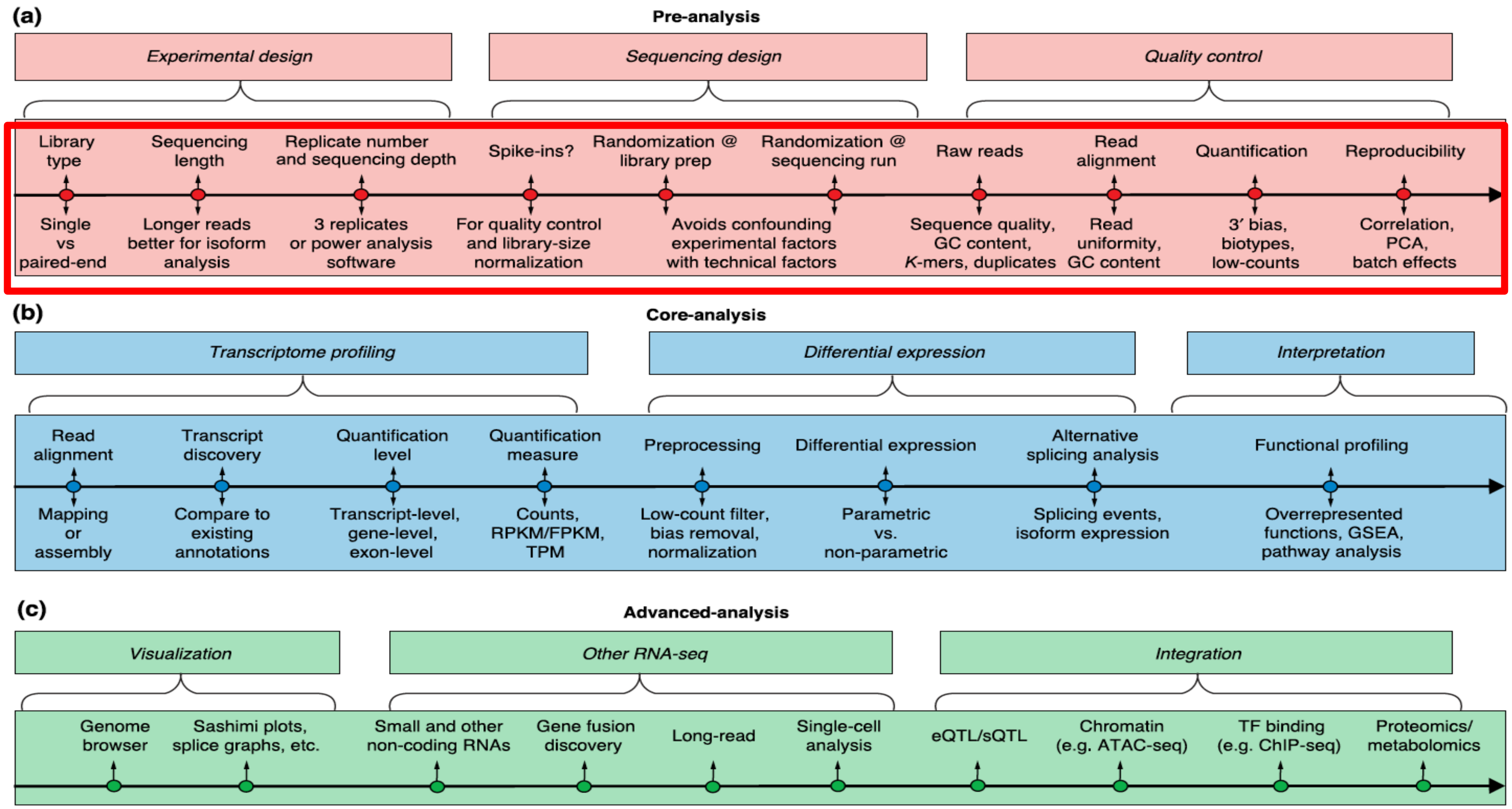

(a)
Pre-analysis
Experimental design
Sequencing design
Quality control
Library type
Single vs paired-end
Sequencing length
Longer reads better for isoform analysis
Replicate number and sequencing depth
3 replicates or power analysis software
Spike-ins?
For quality control and library-size normalization
Randomization @ library prep
Randomization @ sequencing run
Avoids confounding experimental factors with technical factors
Raw reads
Sequence quality, GC content, K-mers, duplicates
Read alignment
Read uniformity, GC content
Quantification
3′ bias, biotypes, low-counts
Reproducibility
Correlation, PCA, batch effects
(b)
Core-analysis
Transcriptome profiling
Differential expression
Interpretation
Read alignment
Mapping or assembly
Transcript discovery
Compare to existing annotations
Quantification level
Transcript-level, gene-level, exon-level
Quantification measure
Counts, RPKM/FPKM, TPM
Preprocessing
Low-count filter, bias removal, normalization
Differential expression
Parametric vs. non-parametric
Alternative splicing analysis
Splicing events, isoform expression
Functional profiling
Overrepresented functions, GSEA, pathway analysis
(c)
Advanced-analysis
Visualization
Other RNA-seq
Integration
Genome browser
Sashimi plots, splice graphs, etc.
Small and other non-coding RNAs
Gene fusion discovery
Long-read
Single-cell analysis
eQTL/sQTL
Chromatin (e.g. ATAC-seq)
TF binding (e.g. ChIP-seq)
Proteomics/ metabolomics

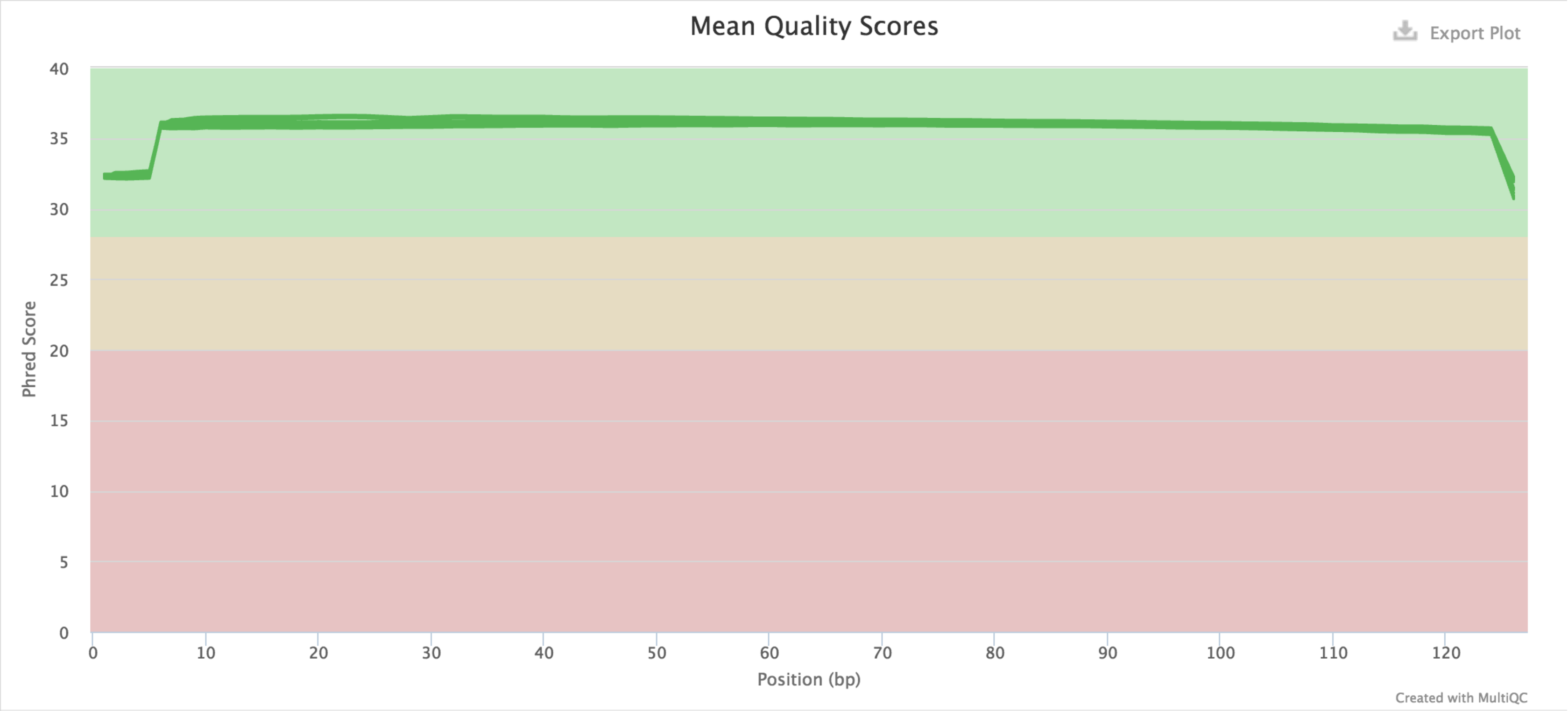
Fig. 1 A generic roadmap for RNA-seq computational analyses. The major analysis steps are listed above the lines for pre-analysis, core analysis and advanced analysis. The key analysis issues for each step that are listed below the lines are discussed in the text. **a** Preprocessing includes experimental design, sequencing design, and quality control steps. **b** Core analyses include transcriptome profiling, differential gene expression, and functional profiling. **c** Advanced analysis includes visualization, other RNA-seq technologies, and data integration. Abbreviations: *ChIP-seq* Chromatin immunoprecipitation sequencing, *eQTL* Expression quantitative loci, *FPKM* Fragments per kilobase of exon model per million mapped reads, *GSEA* Gene set enrichment analysis, *PCA* Principal component analysis, *RPKM* Reads per kilobase of exon model per million reads, *sQTL* Splicing quantitative trait loci, *TF* Transcription factor, *TPM* Transcripts per million

Pre-Alignment QC:

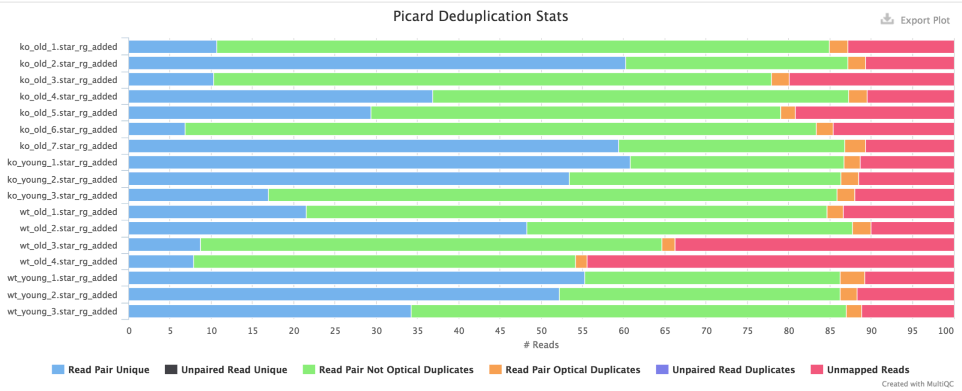
Quality control for the raw reads involves

1. analysis of sequence quality
2. GC content
3. presence of adaptors
4. overrepresented *k*-mers
5. duplicated reads in order to detect sequencing errors, PCR artifacts or contaminations

Pre-alignment QC: FastQC

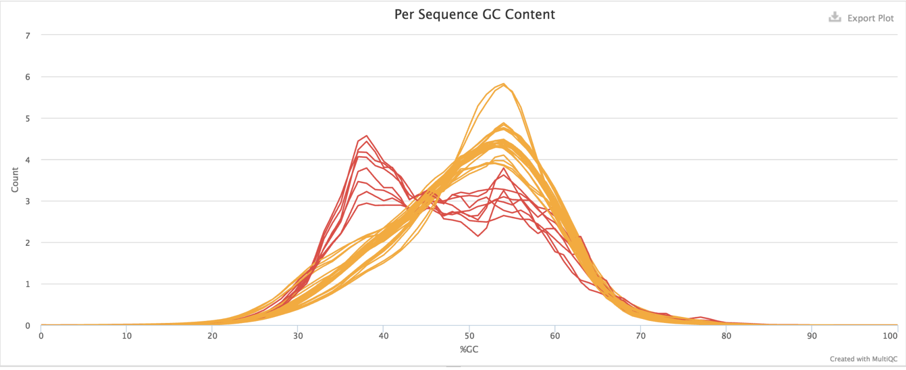


Duplication Rates

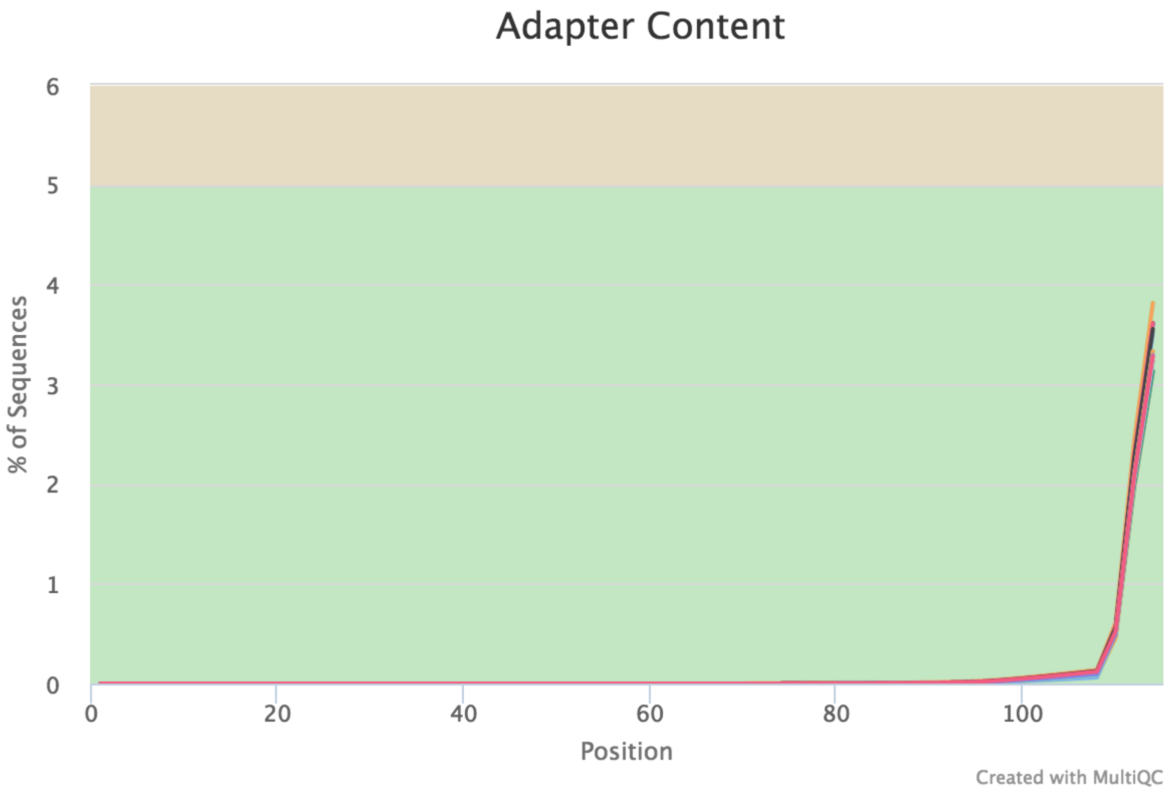


High duplication rates

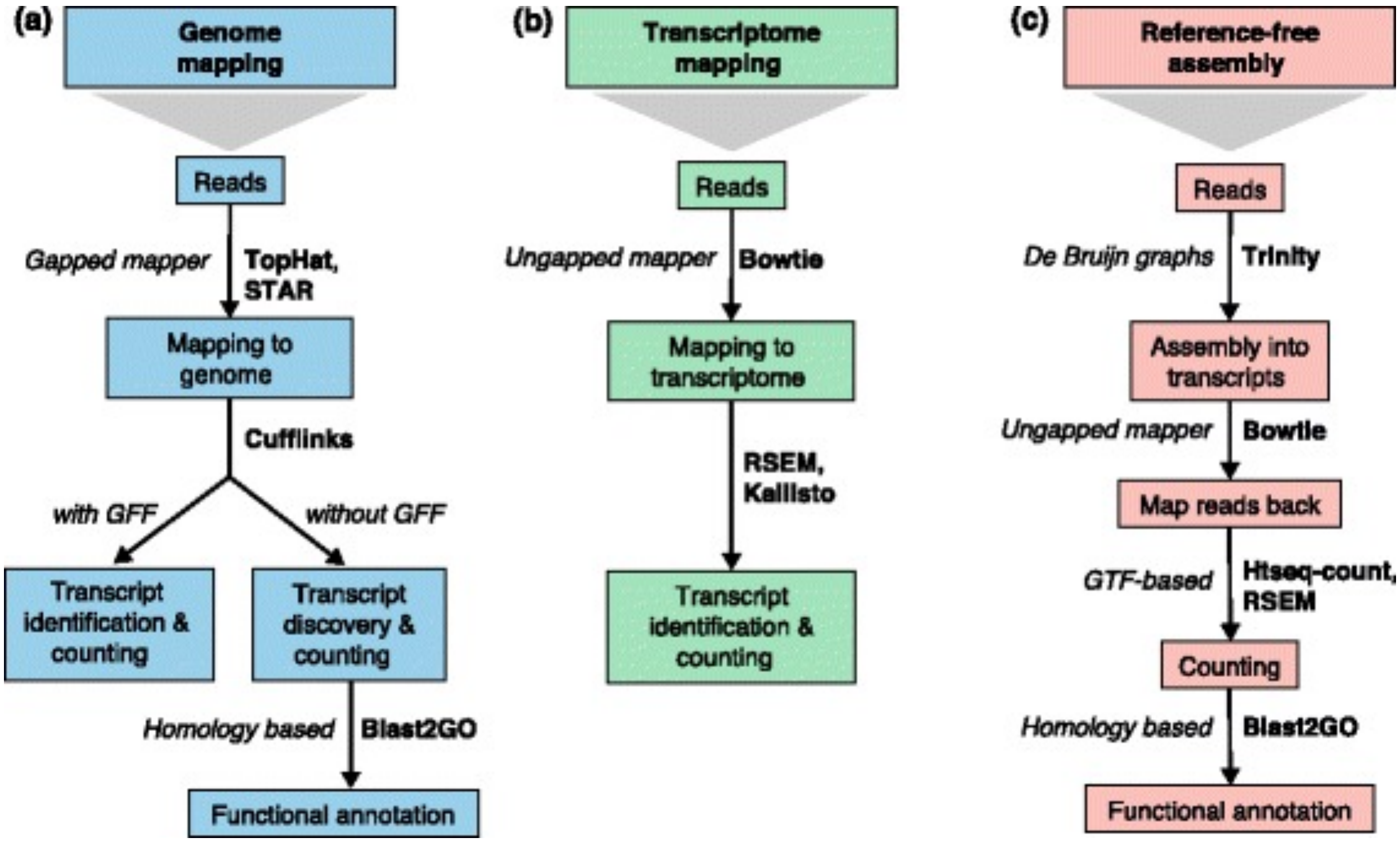
GC Bias



Adapter Content



Alignment methods



Post-alignment QC:

QC Metrics:

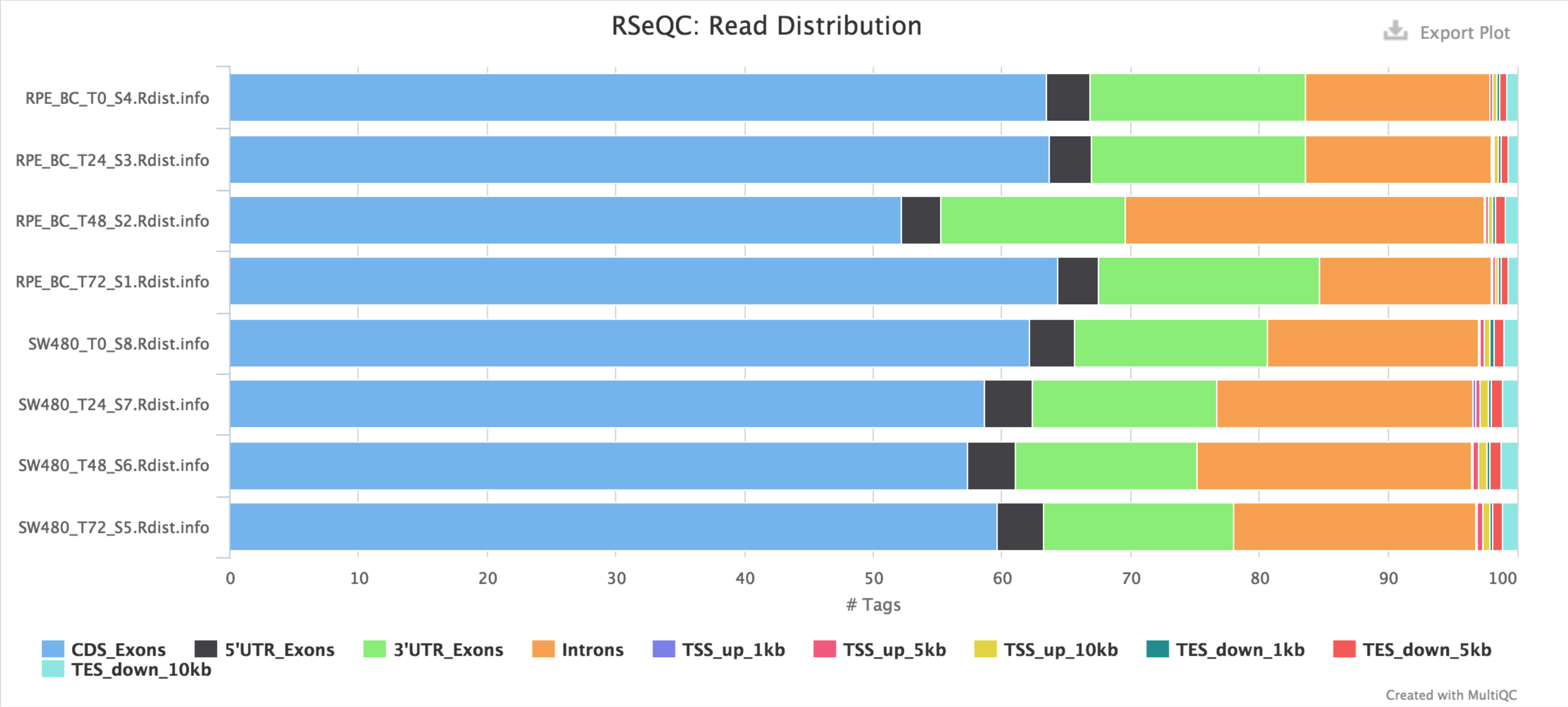
1. Alignment percentage: between 70 and 90 % of regular RNA-seq reads to map onto the human genome (depending on the read mapper used)

2. Uniformity of read coverage on exons and the mapped strand

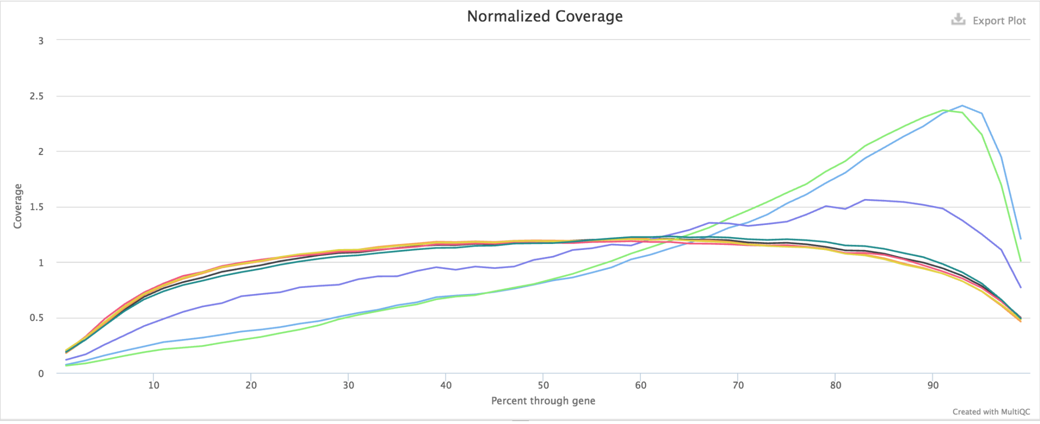
3. Reproducibility among replicates and for possible batch effects (PCA)

4. Contamination: rRNA and microbial RNAs should not be present

Post-alignment QC: RSeQC

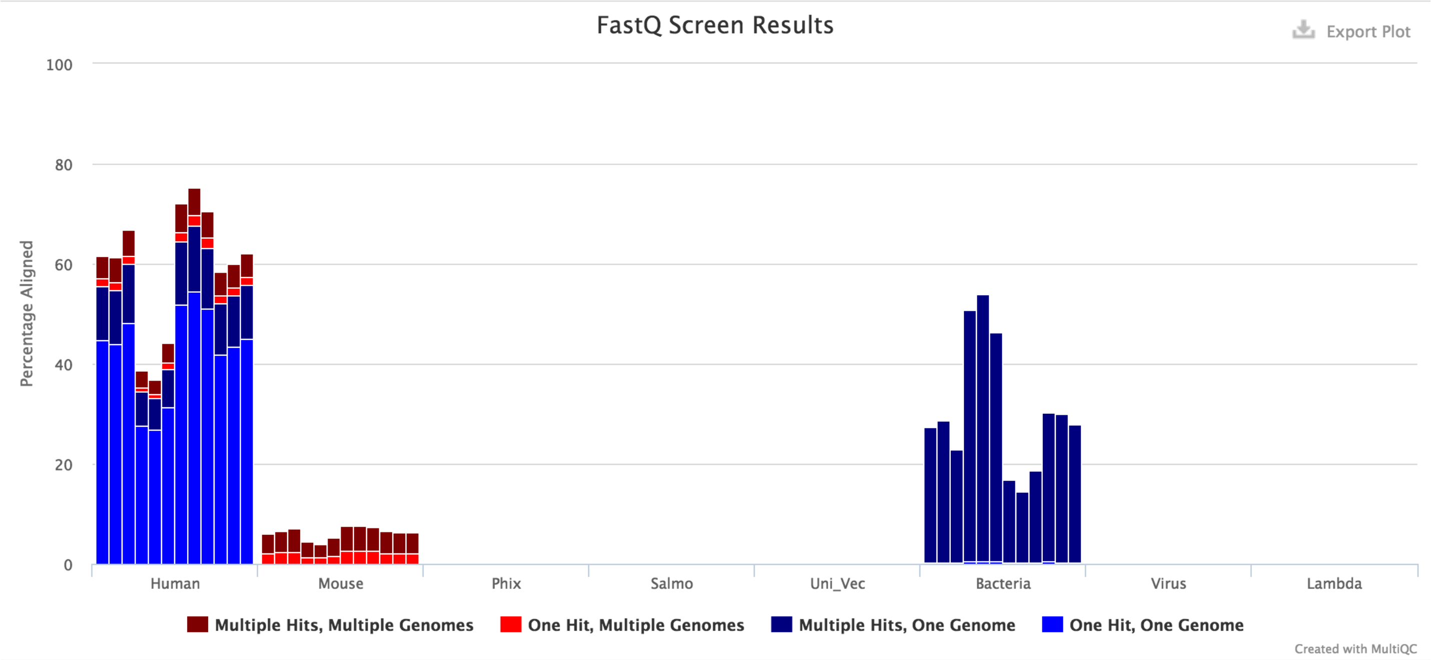


QC: Poor RNA Quality
(RIN > 7, for FFPE or degraded, use total RNA-Seq)

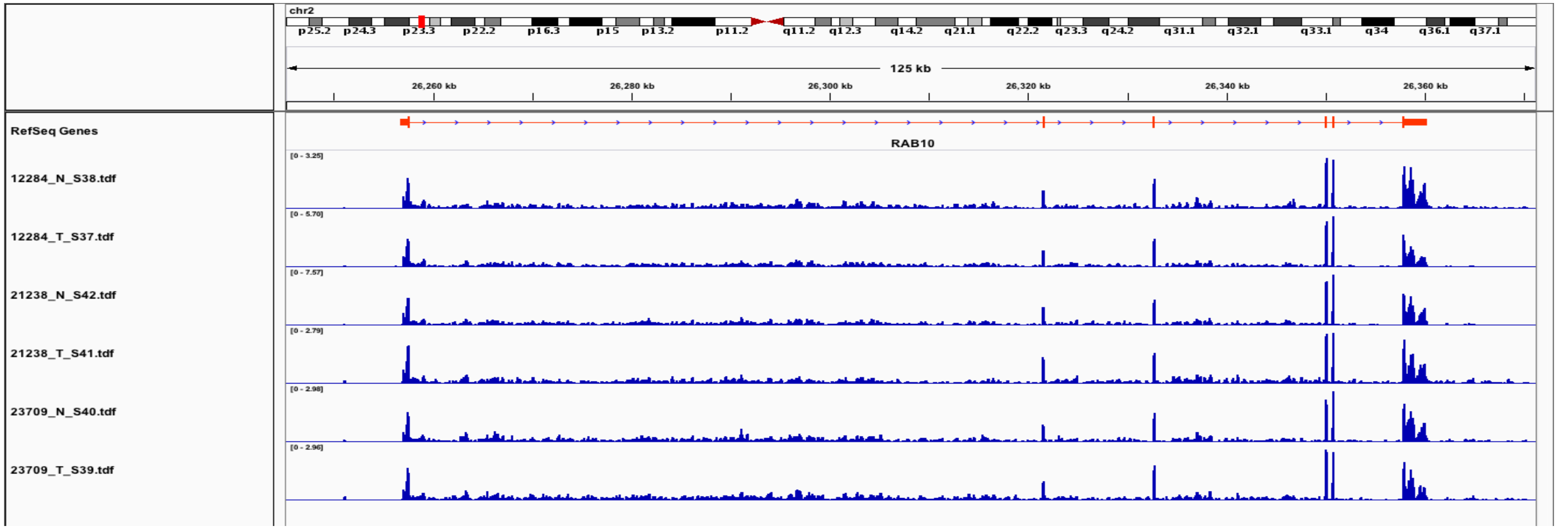


Degraded RNA showing 3’ bias in coverage

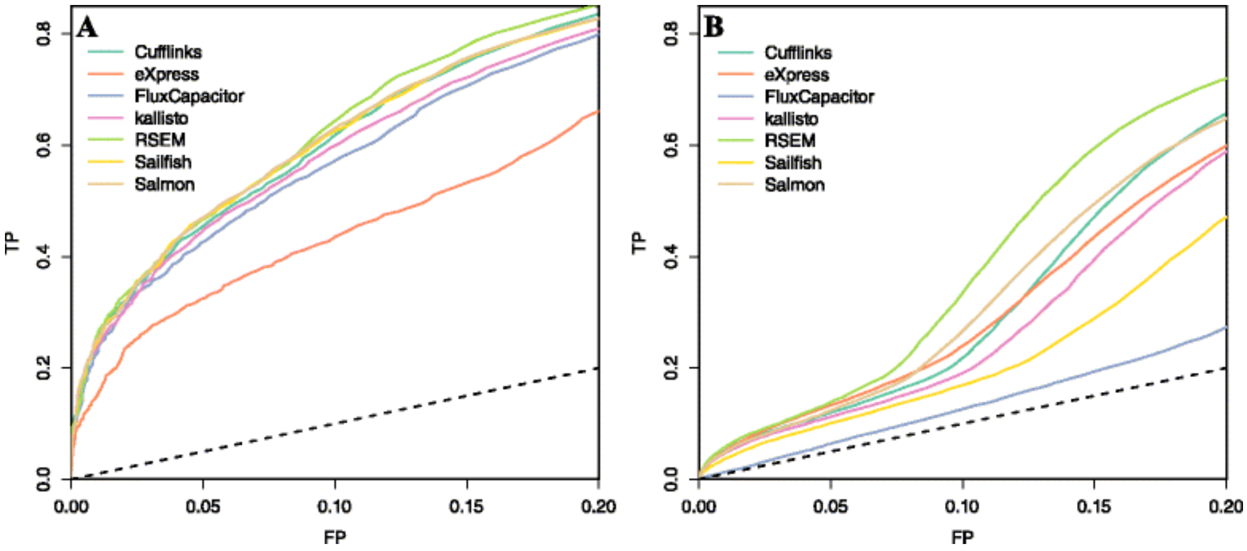
QC: Contamination



Intronic Reads



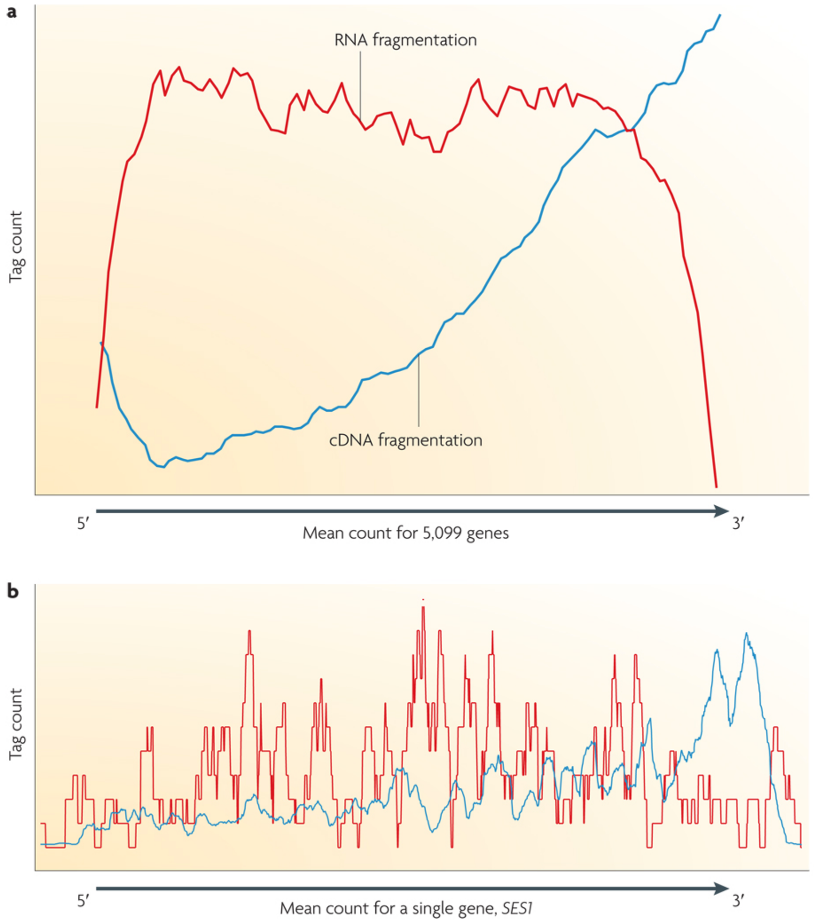
Which gene-counting method?



“Using two independent datasets, we assessed seven competing pipelines. Performance was generally poor, with two methods clearly underperforming and RSEM slightly outperforming the rest.”

Teng, et al. A benchmark for RNA-seq quantification pipelines
Genome Biol. 2016; 17: 74.

Gene coverage for short reads



Gene Expression Data

Not Normalized:

- Raw Counts: number of reads that align to a particular feature

Normalized:

- CPM (or log CPM): Counts per Million Reads
  - For relative gene expression

Within-sample Normalization:

- RPKM: Reads per Kilobase exon per Million Reads
  - For single-end reads
- FPKM: Fragments per Kilobase exon per Million Reads
  - For paired-end reads
- TPM: Transcripts per base normalized by all transcripts per base per Million
  - estimated fraction of transcripts made up by a given isoform or gene

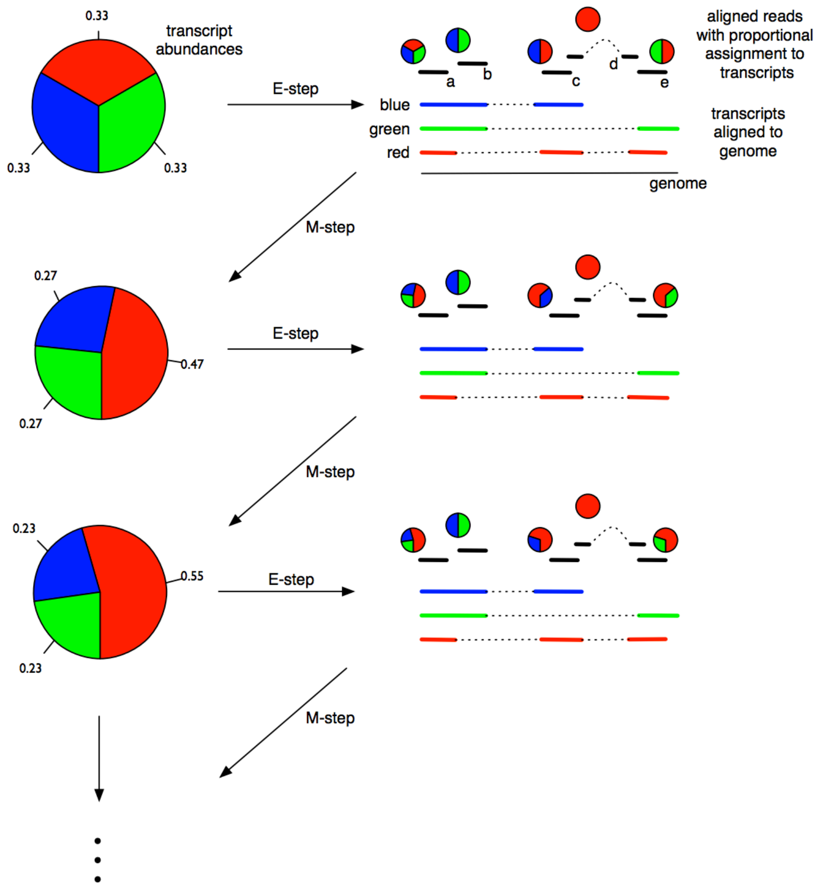
Normalization methods

1. Total count (TC): Gene counts are divided by the total number of mapped reads (or library size) associated with their lane and multiplied by the mean total count across all the samples of the dataset.
2. Upper Quartile (UQ): Very similar in principle to TC, the total counts are replaced by the upper quartile of counts different from 0 in the computation of the normalization factors.
3. Median (Med): Also similar to TC, the total counts are replaced by the median counts different from 0 in the computation of the normalization factors.
4. DESeq: This normalization method is included in the DESeq Bioconductor package, using a "reference sample" by taking, for each gene, the geometric mean of the counts in all samples.
5. Trimmed Mean of *M*-values (TMM): Trimmed mean of M values (TMM) between each pair of samples. This normalization method is implemented in the edgeR Bioconductor package.
6. Quantile (Q): First proposed in the context of microarray data, this normalization method consists in matching distributions of gene counts across lanes.
7. Reads Per Kilobase per Million mapped reads (RPKM): This approach was initially introduced to facilitate comparisons between genes within a sample and combines between- and within-sample normalization.

Methods for Quantification and Differential Gene Expression

1. Raw counts:
   - Gene level: subread, HTSeq
2. Normalized counts and DEG:
   - Gene level: EdgeR, DESeq2, Limma-voom, RSEM
   - Transcript level: RSEM, Salmon, Kallisto, Sailfish

Expectation Maximization



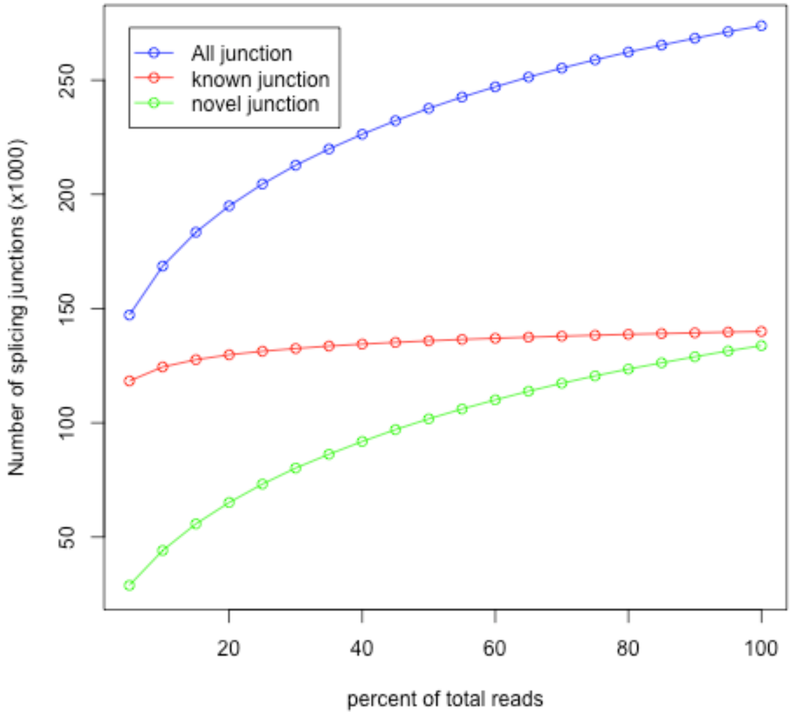
Splice Variant Quantification

Either with a reference or de novo, the complete reconstruction of transcriptomes using short-read Illumina technology remains a challenging problem, and in many cases de novo assembly results in tens or hundreds of contigs accounting for fragmented transcripts.

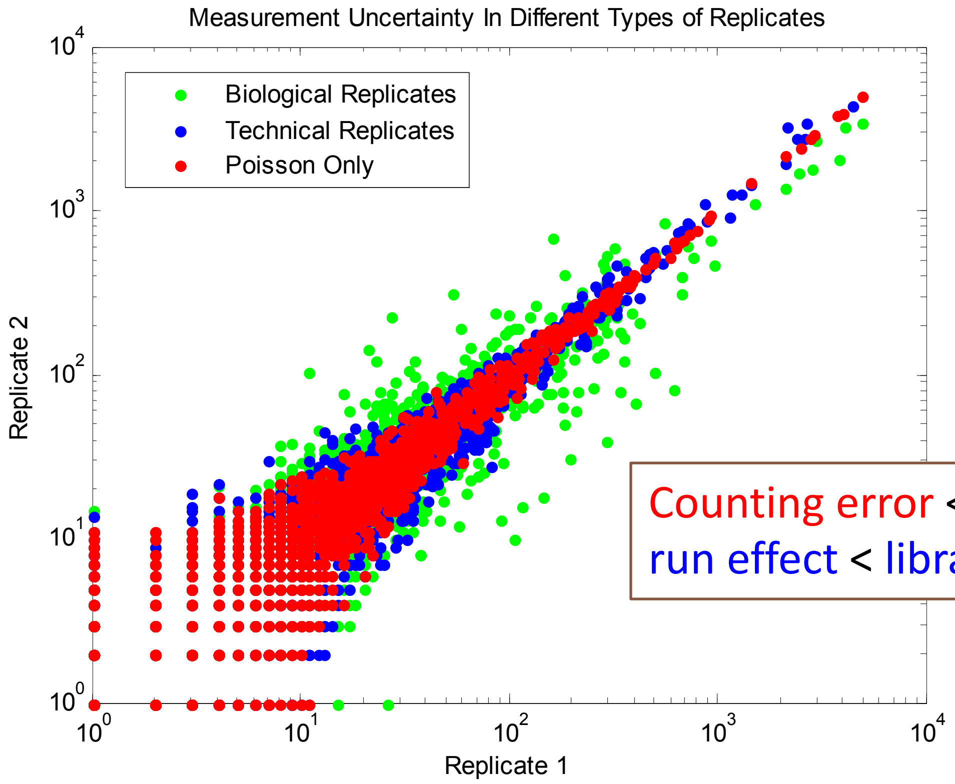
Emerging long-read technologies, such as SMRT from Pacific Biosciences, provide reads that are long enough to sequence complete transcripts for most genes

Conesa et al., Genome Biol. 2016

Junction Counts



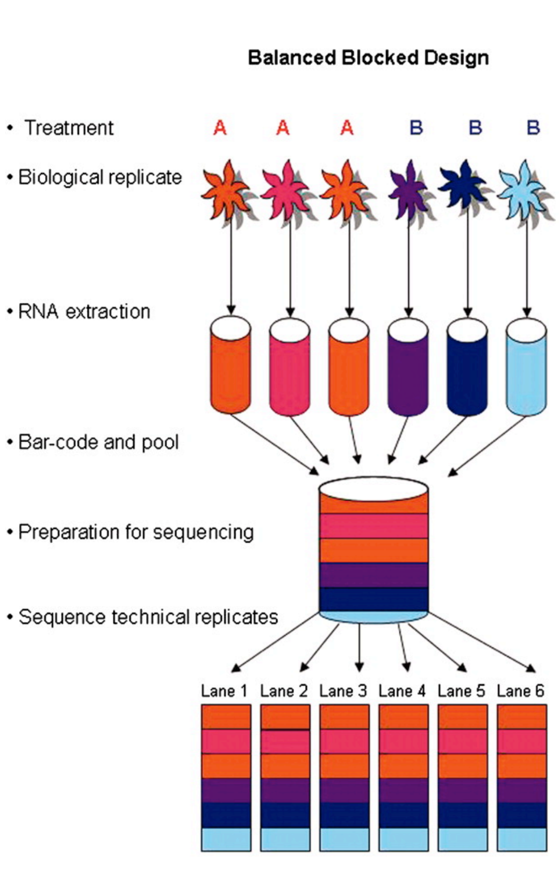
Types of variance

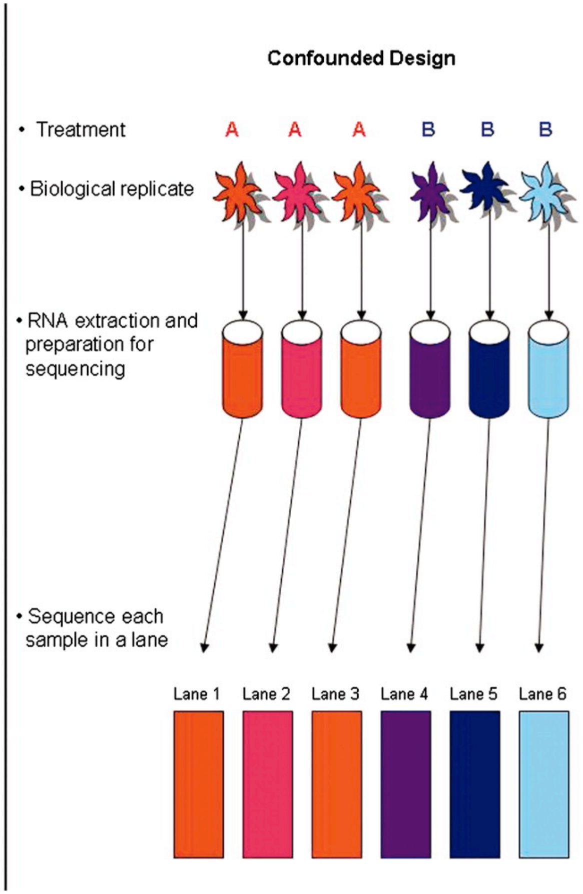


Counting error < Technical effects (lane effect < run effect < library prep effect) << biological effect

Busby et al, Bioinformatics 2013
Marioni et al, Genome Res 2008

Experimental Design: avoiding lane effects

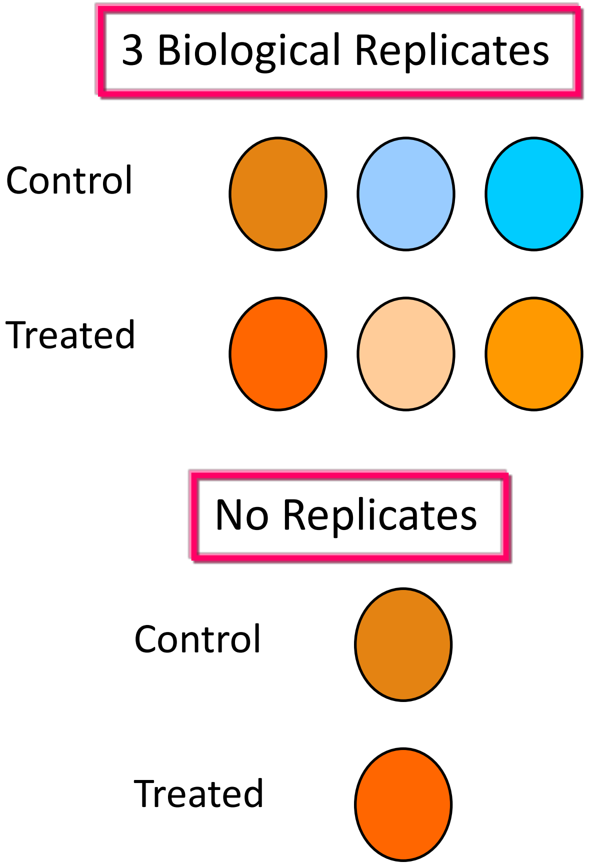




- does not permit partitioning of batch and lane effects from the estimate of within-group biological variability

Auer and Doerge, Genetics 2010

What happens when I run a single sample per treatment group?



3 Biological Replicates

Control

Treated

No Replicates

Control

Treated

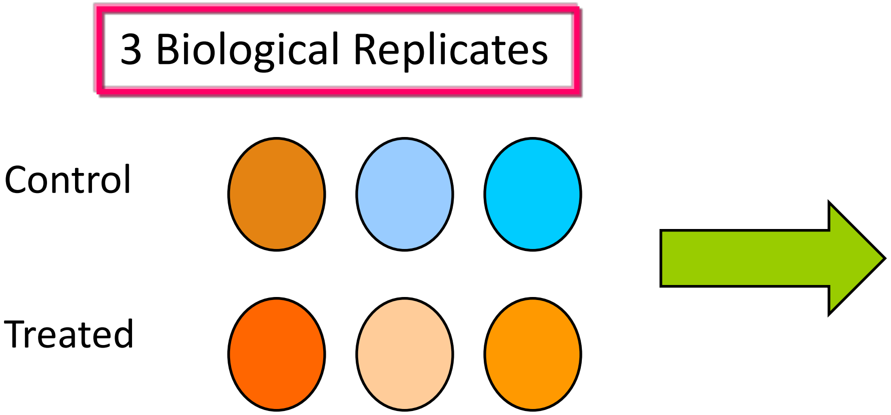
Sorted by p-value

- lowest p values signify genes that are stable (low within group variance
- can set false positive/false negative rate cutoffs
- can prioritize genes for validation
- *more expensive up front but can cut down cost (time and resources) in the long run*

Sorted by fold change

- could be a highly variable gene with no biological relevance at all
- no idea of false positive/false negative rate
- might need to validate larger number of genes on replicate samples (more effort downstream)
- *inexpensive, but likely to be more costly (time and resources) in the long run*

Consequences of running biological vs. technical replicates

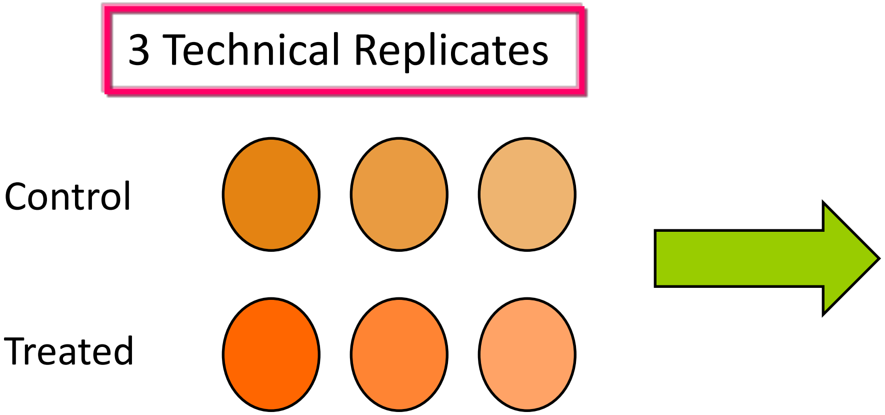


3 Biological Replicates

Control

Treated

- Captures variation among individuals, animals, culture conditions
- Larger variance within each group
- Larger p values (fewer significant genes)
- Decreased false positive rates
- Higher validation/reproducibility rate



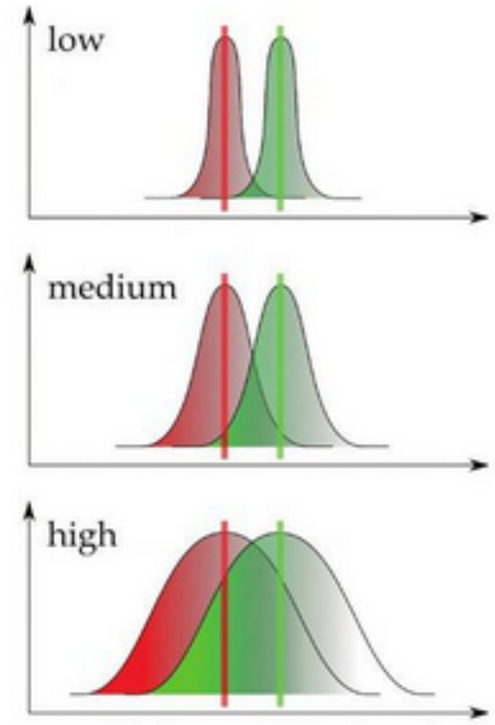
3 Technical Replicates

Control

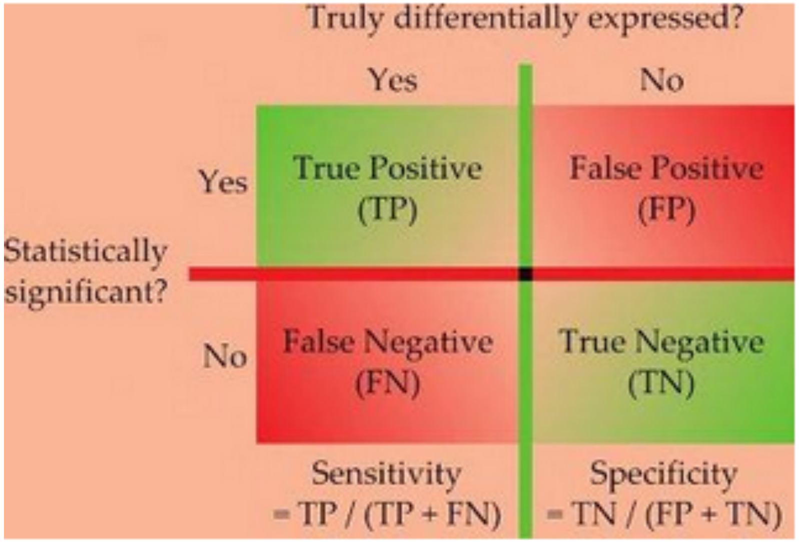
Treated

- Captures variation secondary to array or sample processing conditions
- Small variance within each group
- Smaller p values (more significant genes)
- Increased false positive rates (not capturing true biological variation)
- Lower validation/reproducibility rate

Statistical Tests



Statistical tests provide p values, which are a measure of whether they are significant or not



		Truly differentially expressed?	
		Yes	No
Statistically significant?	Yes	True Positive (TP)	False Positive (FP)
	No	False Negative (FN)	True Negative (TN)
		Sensitivity = TP / (TP + FN)	Specificity = TN / (FP + TN)

2 types of error:

Type 1 error: Calling a gene change statistically significant when it is not (α), false positive

Type 2 error: Calling a gene not significantly changed when it is (β), false negative

Samples vs Read depth

If on a tight budget, deciding between number of replicates vs sequencing depth, always higher replicates with lower sequencing depth leads to higher statistical power

- 3M reads x 10 replicates = 30M reads yields 52% power
- 10M reads x 3 replicates = 30M reads yields 33% power

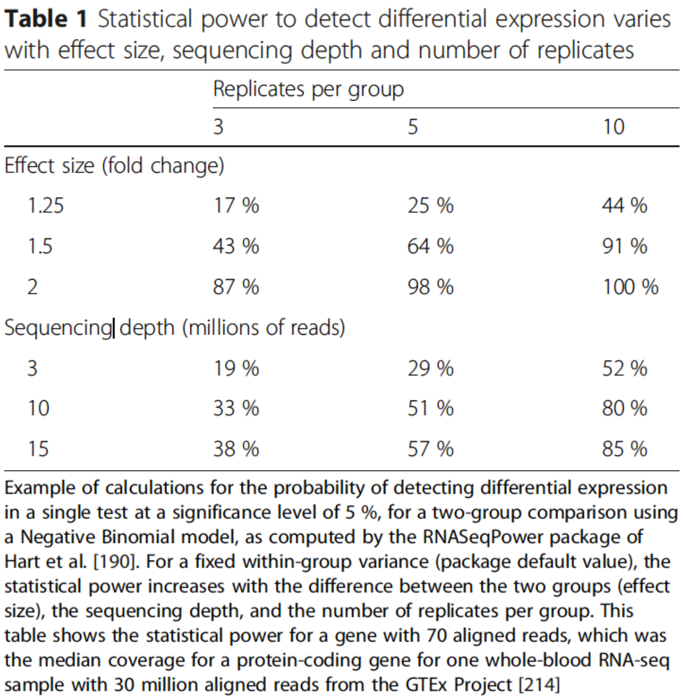


Table 1 Statistical power to detect differential expression varies with effect size, sequencing depth and number of replicates

	Replicates per group		
	3	5	10
Effect size (fold change)			
1.25	17 %	25 %	44 %
1.5	43 %	64 %	91 %
2	87 %	98 %	100 %
Sequencing depth (millions of reads)			
3	19 %	29 %	52 %
10	33 %	51 %	80 %
15	38 %	57 %	85 %

Example of calculations for the probability of detecting differential expression in a single test at a significance level of 5 %, for a two-group comparison using a Negative Binomial model, as computed by the RNASeqPower package of Hart et al. [190]. For a fixed within-group variance (package default value), the statistical power increases with the difference between the two groups (effect size), the sequencing depth, and the number of replicates per group. This table shows the statistical power for a gene with 70 aligned reads, which was the median coverage for a protein-coding gene for one whole-blood RNA-seq sample with 30 million aligned reads from the GTEx Project [214]

Best Practices

1. Factor in at least 3 replicates (absolute minimum), but 4 if possible (optimum minimum). Biological replicates are recommended rather than technical replicates.

2. Always process your RNA extractions at the same time. Extractions done at different times lead to unwanted batch effects.

3. There are 2 major considerations for RNA-Seq libraries:

If you are interested in coding mRNA, you can select to use the mRNA library prep. The recommended sequencing depth is between 10-20M paired-end (PE) reads. Your RNA has to be high quality (RIN > 8).

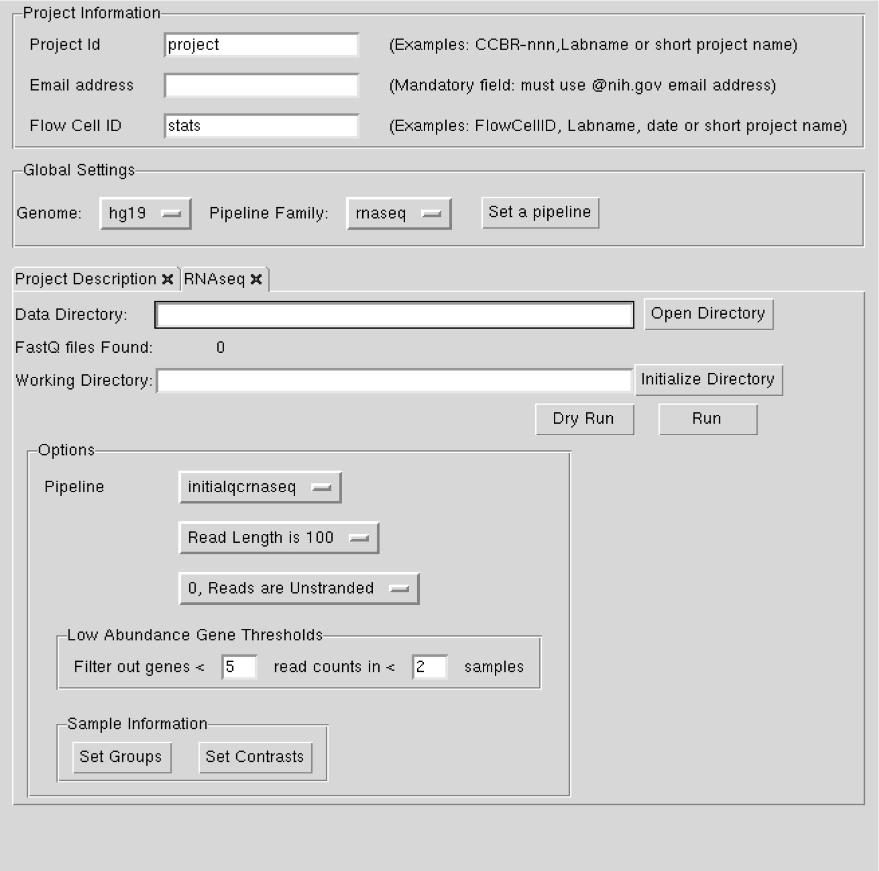
If you are interested in long noncoding RNA as well, you can select the total RNA method, with sequencing depth ~25-60M PE reads. This is also an option if your RNA is degraded.

4. Ideally to avoid lane batch effects, all samples would need to be multiplexed together and run on the same lane. This may require an initial MiSeq run for library balancing. Additional lanes can be run if more sequencing depth is needed.

5. If you are unable to process all your RNA samples together and need to process them in batches, make sure that replicates for each condition are in each batch so that the batch effects can be measured and removed bioinformatically.

https://bioinformatics.cancer.gov/content/rna-seq

CCBR Pipeliner
(QC Report, DEG Analysis)



Project Information

Project Id: project (Examples: CCBR-nnn,Labname or short project name)

Email address: (Mandatory field: must use @nih.gov email address)

Flow Cell ID: stats (Examples: FlowCellID, Labname, date or short project name)

Global Settings

Genome: hg19

Pipeline Family: rnaseq

Data Directory:

FastQ files Found: 0

Working Directory:

Options

Pipeline: initialqcrnaseq

Read Length is 100

0, Reads are Unstranded

Low Abundance Gene Thresholds

Filter out genes < 5 read counts in < 2 samples

Sample Information

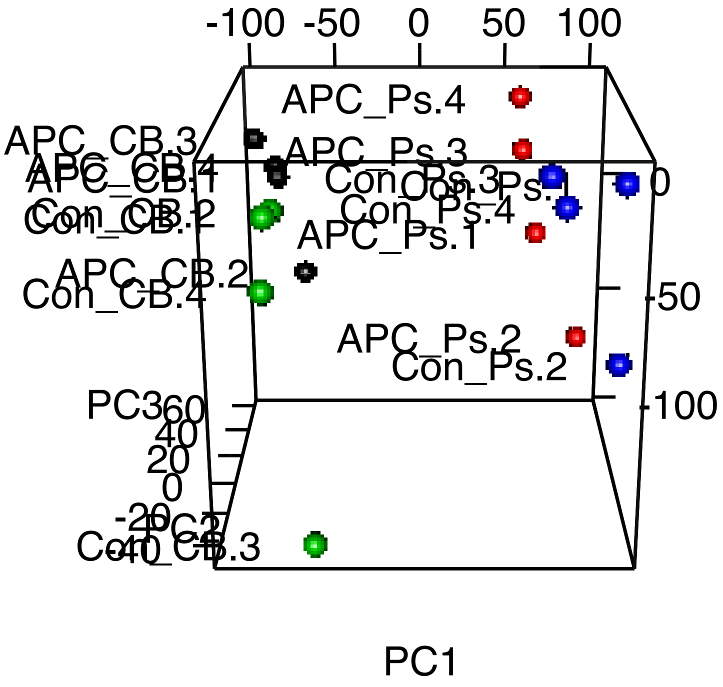
Validation Methods

- **Quantitative RT-PCR**
  - well-accepted gold standard
  - housekeeping gene - use microarray data instead of GAPDH, Beta-actin

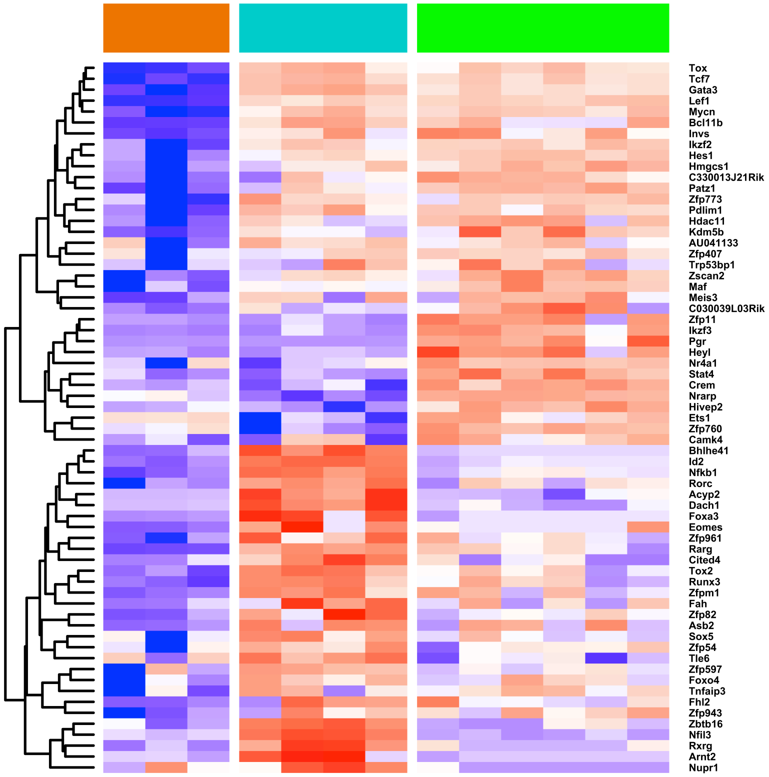
- **NanoString**
  - Multiplex assay, for several genes simultaneously
  - design based on microarray probes – increase validation
  - especially well-suited for large number of samples
  - use a number of housekeeping genes rather than a single gene

- **FISH**
  - Fluorescence in situ hybridization
  - single cell level
  - Localization especially for heterogeneous samples

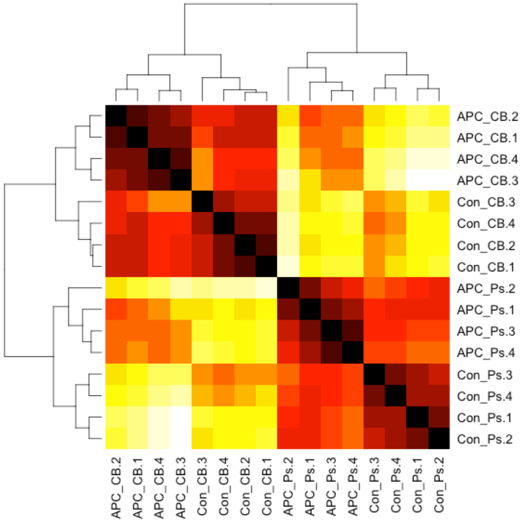
Visualization: PCA

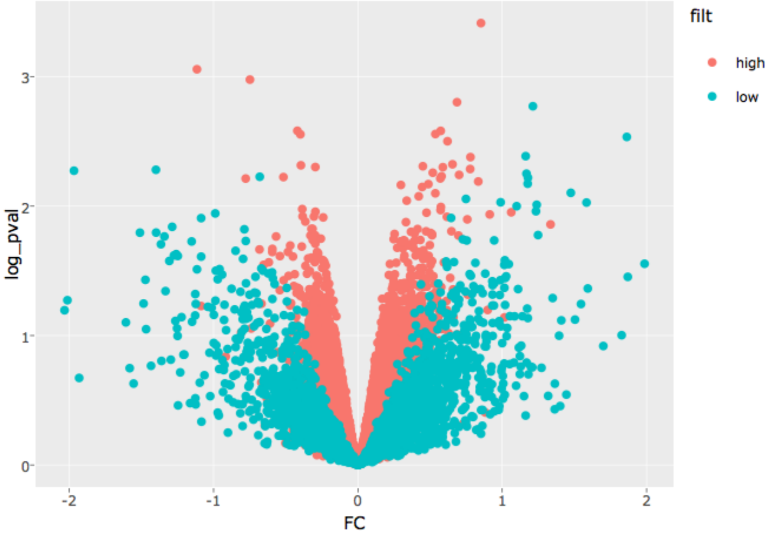


Visualization: Hierarchical Clustering

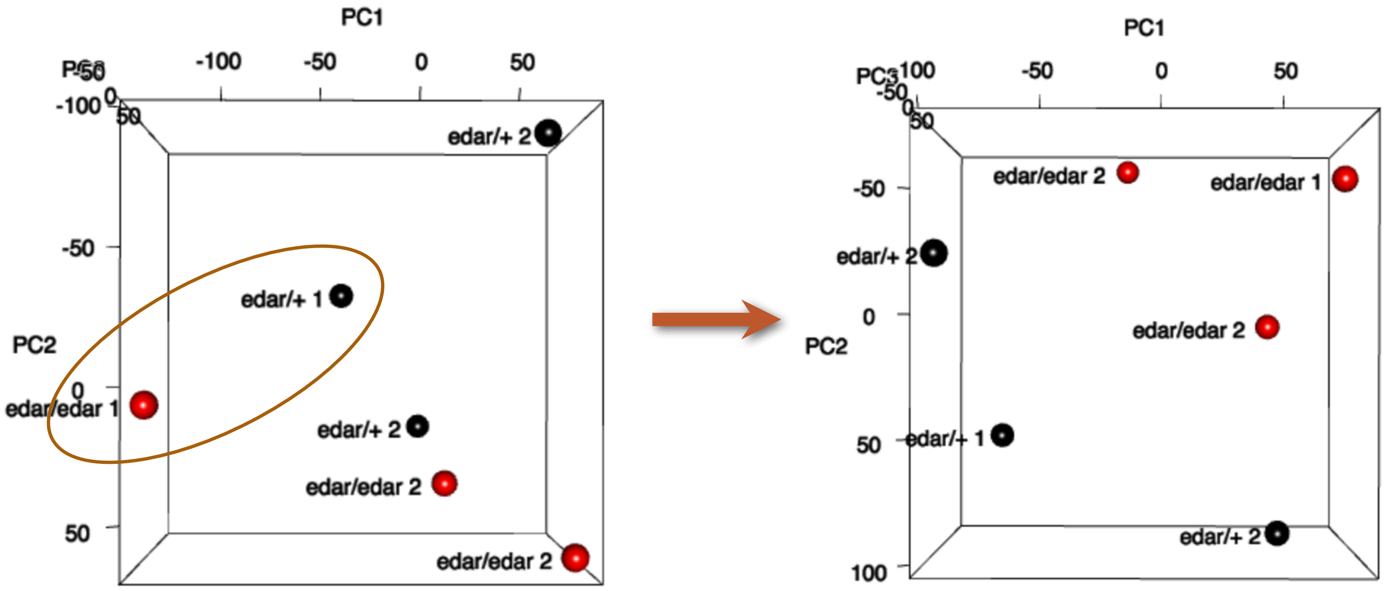


Visualization: Others



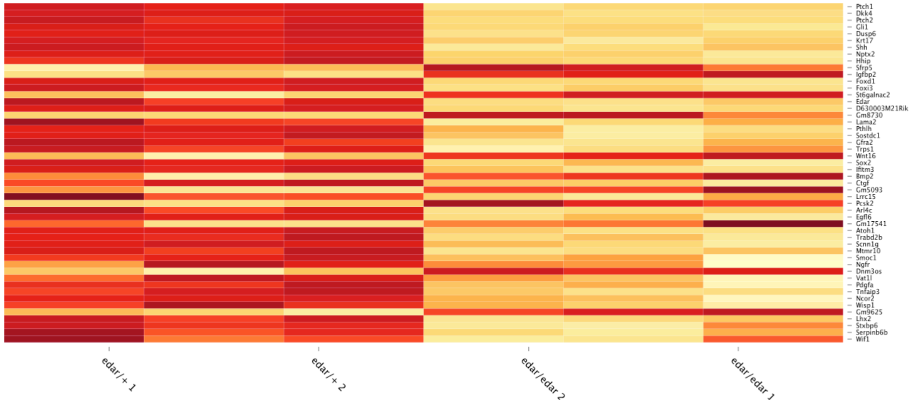


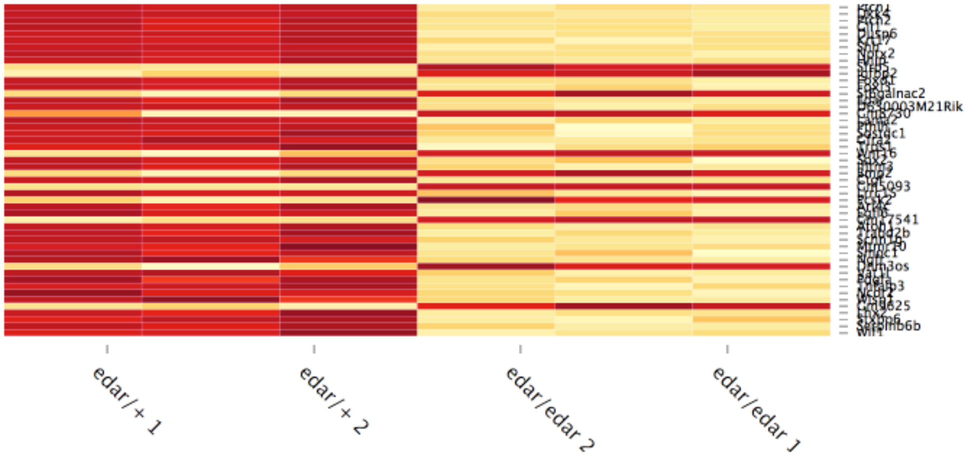
QC: Batch Effects



Litter effect: used batch removal

Visualization: Effect of batch removal

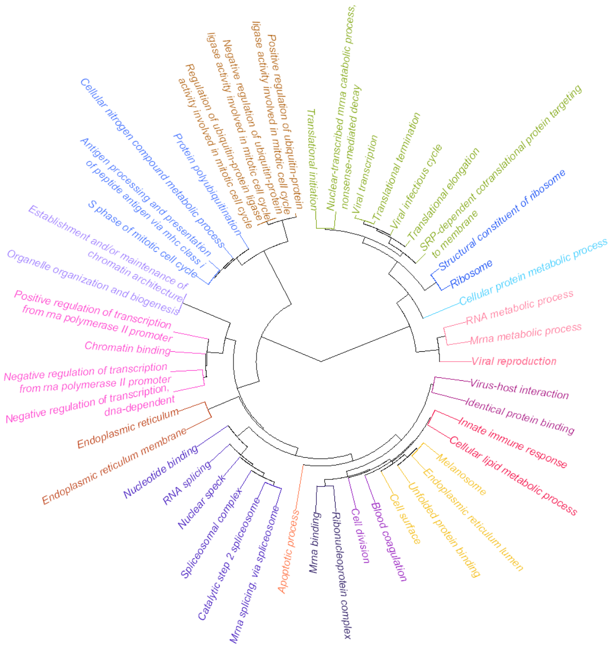




Gene Ontology Enrichment Analysis

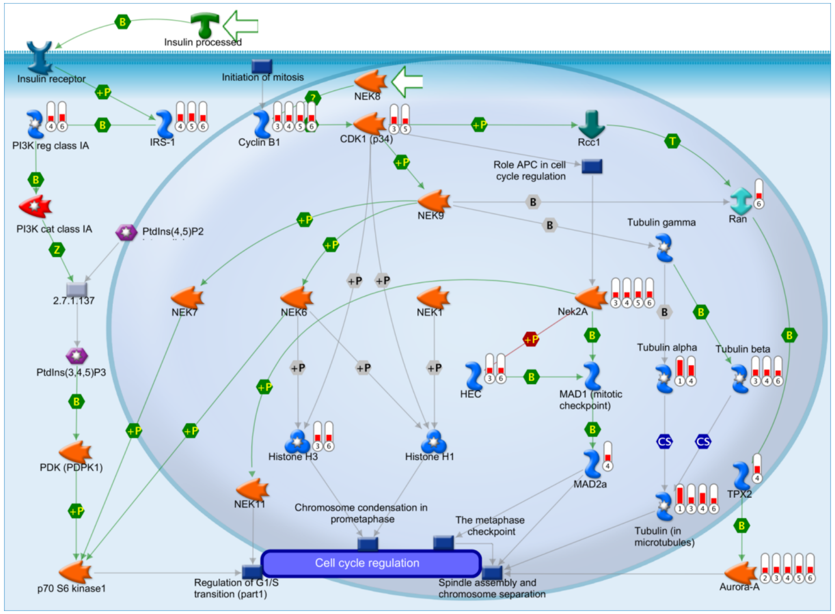
Are the differentially expressed genes in my microarray experiment concentrated in pathways or gene ontology categories which are biologically meaningful?

- Use hypergeometric distribution or similar test to look for interesting patterns


Translational initiation
Nuclear-transcribed mrna catabolic process, nonsense-mediated decay
Viral transcription
Translational termination
Viral infectious cycle
Translational elongation
SRP-dependent cotranslational protein targeting to membrane
Structural constituent of ribosome
Ribosome
Cellular protein metabolic process
RNA metabolic process
Mrna metabolic process
Viral reproduction
Virus-host interaction
Identical protein binding
Innate immune response
Cellular lipid metabolic process
Melanosome
Endoplasmic reticulum lumen
Unfolded protein binding
Cell surface
Blood coagulation
Cell division
Ribonucleoprotein complex
Mrna binding
Apoptotic process
Mrna splicing, via spliceosome
Catalytic step 2 spliceosome
Spliceosomal complex
Nuclear speck
RNA splicing
Nucleotide binding
Endoplasmic reticulum membrane
Endoplasmic reticulum
Negative regulation of transcription, dna-dependent
Negative regulation of transcription from rna polymerase II promoter
Chromatin binding
Positive regulation of transcription from rna polymerase II promoter
Organelle organization and biogenesis
Establishment and/or maintenance of chromatin architecture
S phase of mitotic cell cycle
Antigen processing and presentation of peptide antigen via mhc class i
Cellular nitrogen compound metabolic process
Protein polyubiquitination
Regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle
Negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle
Positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle

Pathway Analysis

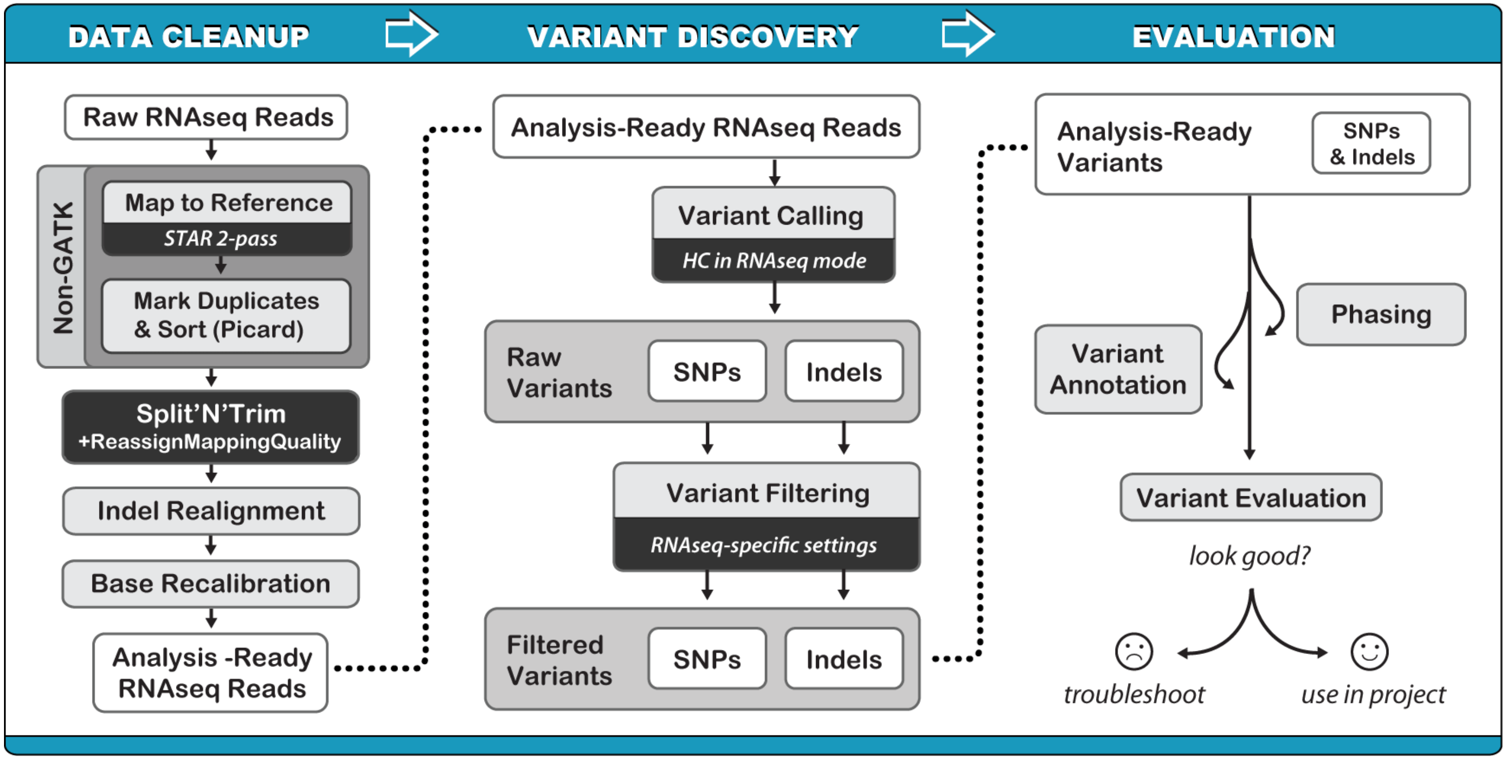
- Free software such as GSEA (Gene Set Enrichment Analysis) and DAVID use public pathway or gene ontology repositories (e.g. Kegg, GO, Reactome, GEO datasets, etc.)
- Many commercial platforms (Ingenuity Pathway Analysis, GeneGo Metacore, Pathway Studio) use curated information which are more comprehensive than public pathway databases



Commercial Bioinformatics Tools available @ CCR

Software	Vendor	Application
Partek Genomics Suite	Partek	Statistical analysis, Cluster Analysis, Pathways
Nexus Expression	BioDiscovery	Statistical analysis, Cluster Analysis
iPathwayGuide	iPathwayGuide	Pathway Analysis
IPA®	Ingenuity Systems	Pathway Analysis (Web-based)
METACORE	Thomson Reuters	Pathway Analysis (Web-based)
PATHWAY STUDIO™	Elsevier	Pathway Analysis (Web-based)
genomatix	Genomatix	Promoter Analysis

Variant Calling



Gene Fusion

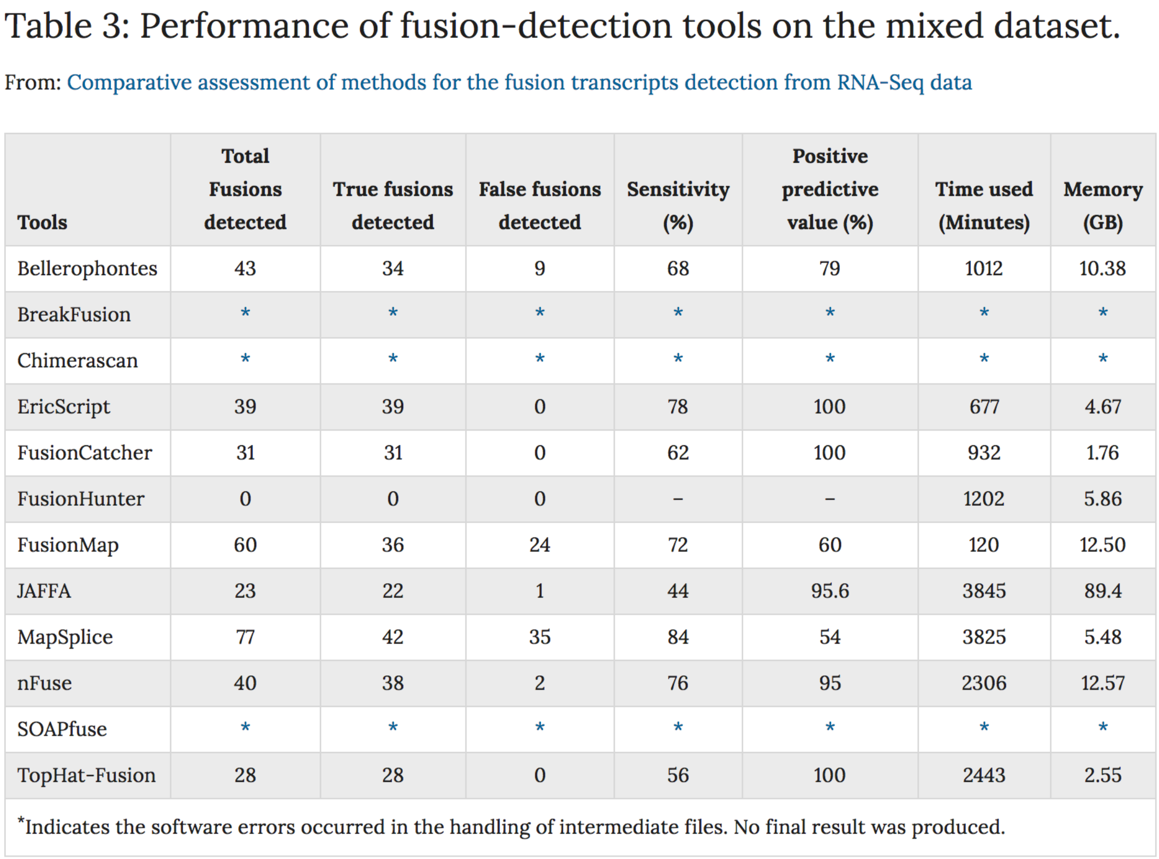


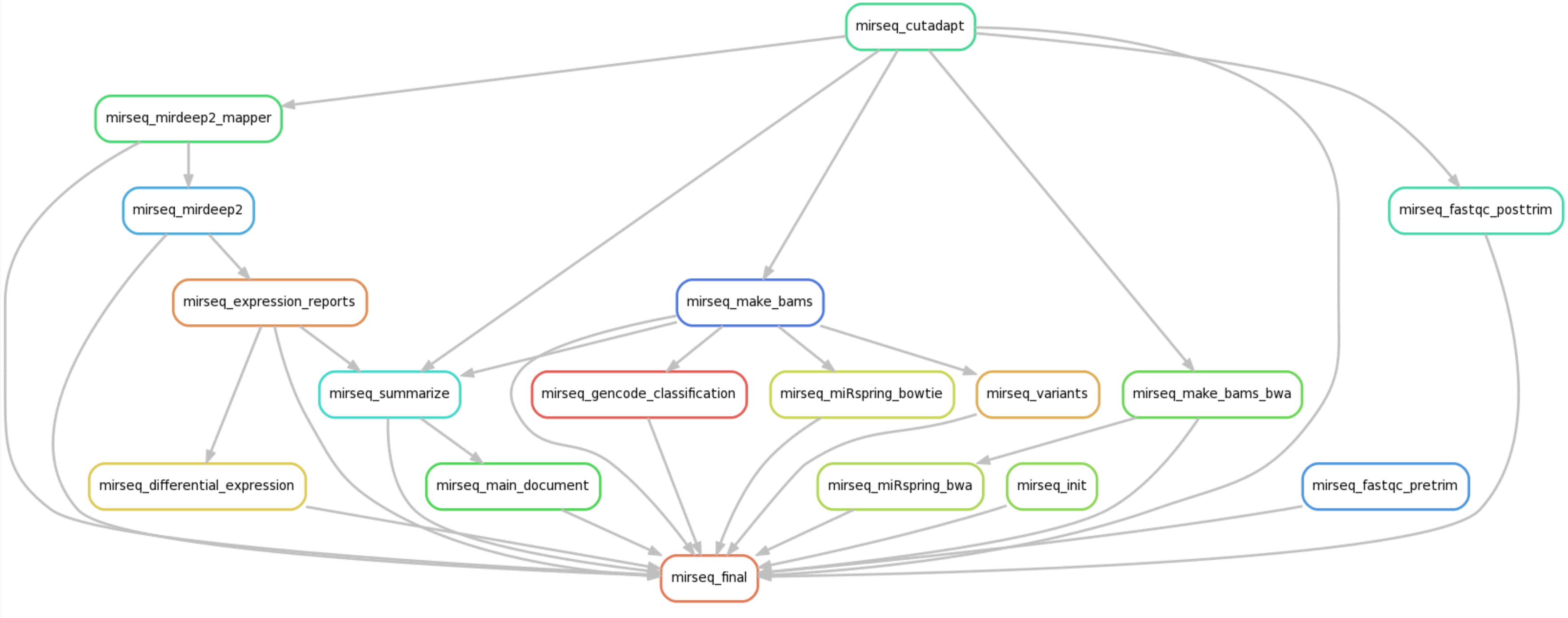
Table 3: Performance of fusion-detection tools on the mixed dataset.

From: Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data

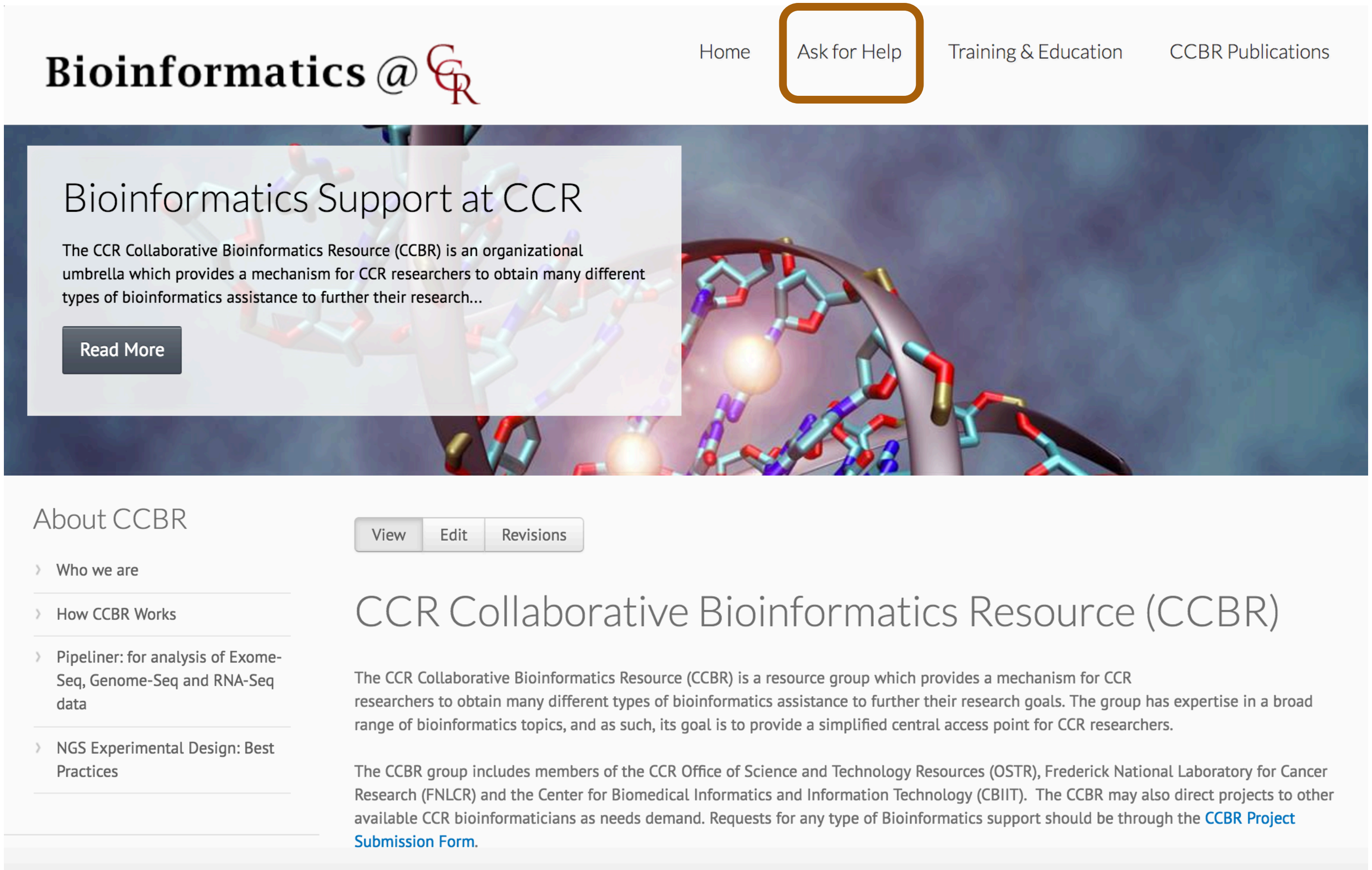
Tools	Total Fusions detected	True fusions detected	False fusions detected	Sensitivity (%)	Positive predictive value (%)	Time used (Minutes)	Memory (GB)
Bellerophontes	43	34	9	68	79	1012	10.38
BreakFusion	*	*	*	*	*	*	*
Chimerascan	*	*	*	*	*	*	*
EricScript	39	39	0	78	100	677	4.67
FusionCatcher	31	31	0	62	100	932	1.76
FusionHunter	0	0	0	–	–	1202	5.86
FusionMap	60	36	24	72	60	120	12.50
JAFFA	23	22	1	44	95.6	3845	89.4
MapSplice	77	42	35	84	54	3825	5.48
nFuse	40	38	2	76	95	2306	12.57
SOAPfuse	*	*	*	*	*	*	*
TopHat-Fusion	28	28	0	56	100	2443	2.55

*Indicates the software errors occurred in the handling of intermediate files. No final result was produced.

miRNA-Seq


mirseq_cutadapt
mirseq_mirdeep2_mapper
mirseq_mirdeep2
mirseq_fastqc_posttrim
mirseq_expression_reports
mirseq_make_bams
mirseq_summarize
mirseq_gencode_classification
mirseq_miRspring_bowtie
mirseq_variants
mirseq_make_bams_bwa
mirseq_differential_expression
mirseq_main_document
mirseq_miRspring_bwa
mirseq_init
mirseq_fastqc_pretrim
mirseq_final

bioinformatics.cancer.gov



Bioinformatics @ CCR

Bioinformatics Support at CCR

The CCR Collaborative Bioinformatics Resource (CCBR) is an organizational umbrella which provides a mechanism for CCR researchers to obtain many different types of bioinformatics assistance to further their research...

About CCBR

- Who we are
- How CCBR Works
- Pipeliner: for analysis of Exome-Seq, Genome-Seq and RNA-Seq data
- NGS Experimental Design: Best Practices

CCR Collaborative Bioinformatics Resource (CCBR)

The CCR Collaborative Bioinformatics Resource (CCBR) is a resource group which provides a mechanism for CCR researchers to obtain many different types of bioinformatics assistance to further their research goals. The group has expertise in a broad range of bioinformatics topics, and as such, its goal is to provide a simplified central access point for CCR researchers.

The CCBR group includes members of the CCR Office of Science and Technology Resources (OSTR), Frederick National Laboratory for Cancer Research (FNLCR) and the Center for Biomedical Informatics and Information Technology (CBIIT). The CCBR may also direct projects to other available CCR bioinformaticians as needs demand. Requests for any type of Bioinformatics support should be through the CCBR Project Submission Form.

CCBR support includes:

Consulting on experimental design, help with analysis and interpretation of biological data produced by large-scale genomics technologies including Next-generation sequencing (RNA-Seq, Exome-Seq, ChIP-Seq, Whole genome Sequencing), and microarrays

Support for the development of methods for new technologies provided by the Office of Science and Technology Resources (OSTR)

Provide training classes to CCR scientists focusing on software used in the analysis of their own data

CCBR Members

Office of Science and Technology Resources (OSTR)

Maggie Cam (Head)

Center for Biomedical Informatics and Information Technology (CBIIT)

Chunhua Yan

Ying Hu

Richard Finney

Frederick National Laboratory of Cancer Research (Leidos)

Parthav Jailwala (Manager)

Fathi Elloumi

Justin Lack

Bong-Hyun Kim

George Nelson

Alexei Lobanov

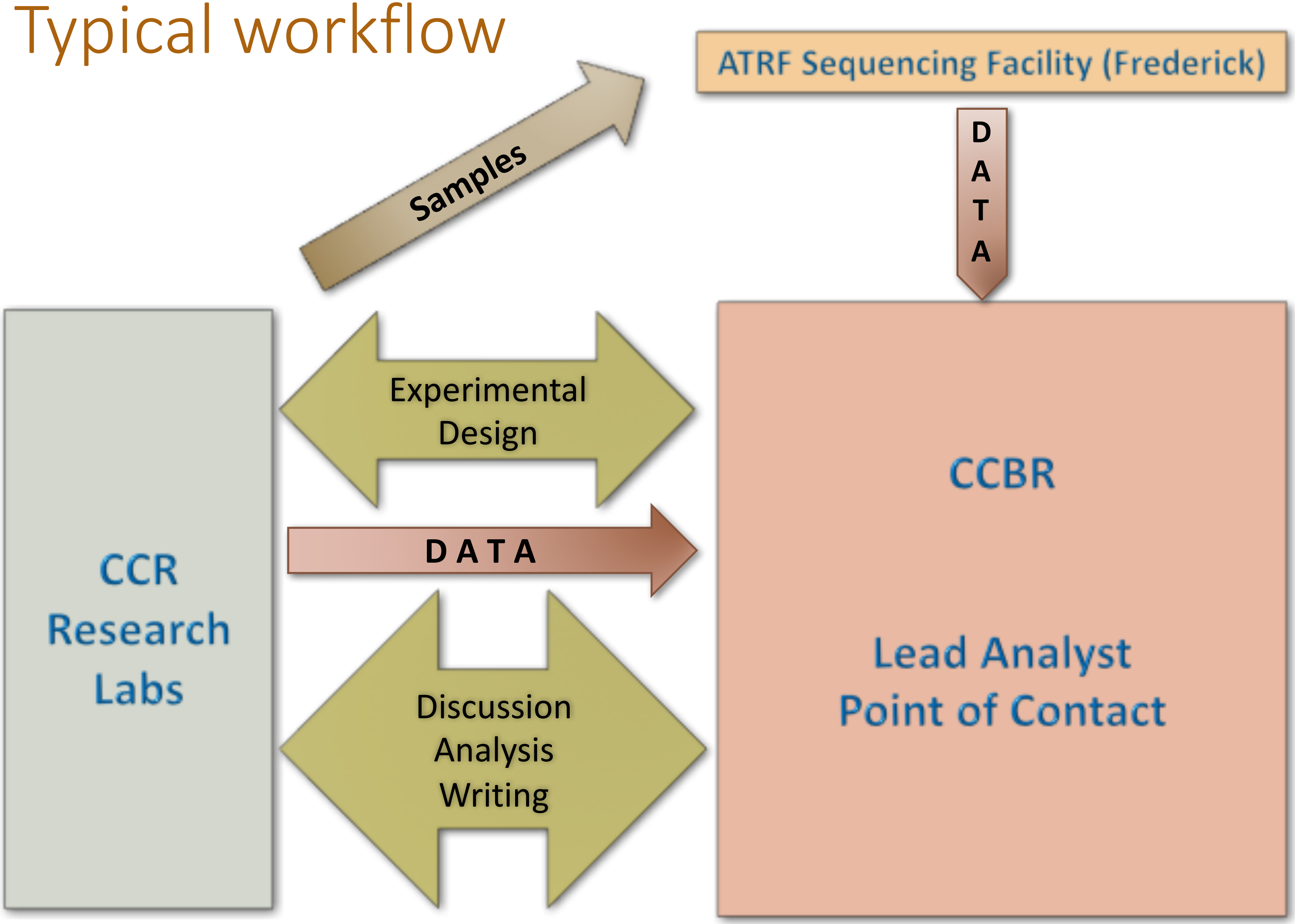
Jack Chen

Ashley Walton

Vishal Koparde

Soon to be part of CDSL (CCR Cancer Data Science Lab)

Typical workflow


ATRF Sequencing Facility (Frederick)
Samples
DATA
Experimental Design
DATA
Discussion Analysis Writing
CCR Research Labs
CCBR
Lead Analyst Point of Contact

Take Home Message:

While you are planning your RNA-Seq experiment (not after), please come talk to us.

CCBR@mail.nih.gov