RNA-Seq Module 3

Advanced RNA-Seq Analysis Topics and Trouble-Shooting

Kevin Silverstein PhD, John Garbe PhD and
Ying Zhang PhD,
Research Informatics Support System (RISS)
MSI

May 24, 2012

UNIVERSITY OF MINNESOTA
Driven to Discover℠

RNA-Seq Tutorials

- Tutorial 1: Introductory (Mar. 28 & Apr. 19)
  - RNA-Seq experiment design and analysis
  - Instruction on individual software will be provided in other tutorials
- Tutorial 2: Introductory (Apr. 3 & Apr 24)
  - Analysis RNA-Seq using TopHat and Cufflinks
- Tutorial 3: Intermediate (May 24)
  - Advanced RNA-Seq analysis topics and trouble-shooting
- Hands-on tutorials (Summer 2012)…

RNA-Seq Module 3

Advanced RNA-Seq Analysis Topics and Trouble-Shooting

Part I: Review and Considerations for Different Goals and Biological Systems (Kevin Silverstein, PhD)

Part II: Read Mapping Statistics and Visualization (John Garbe, PhD)

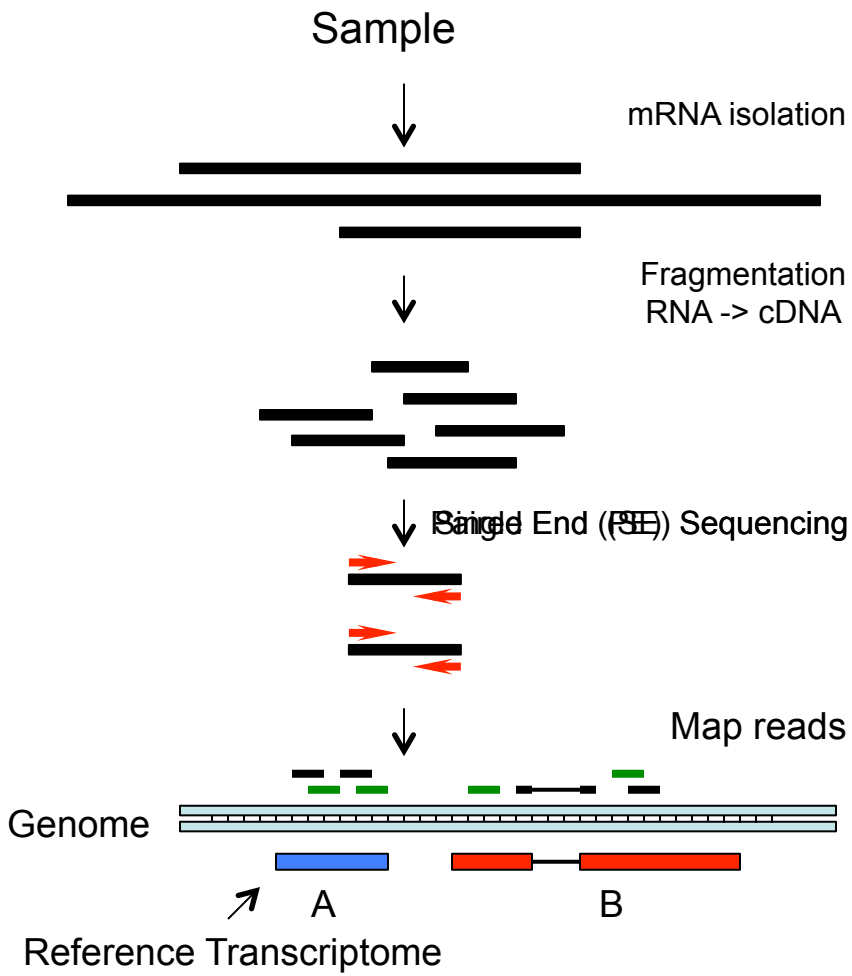
Part III: Post-Analysis Processing – Exploring the Data and Results (Ying Zhang, PhD)

Part I

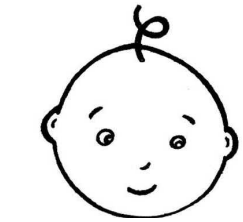
Review and Considerations for Different Goals and Biological Systems

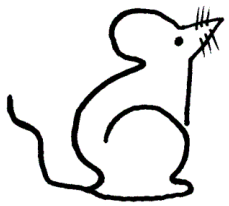
Kevin Silverstein, PhD

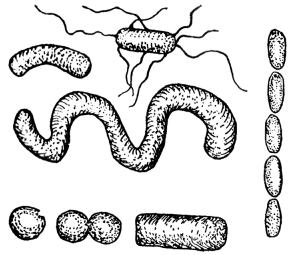
Typical RNA-seq experimental protocol and analysis



Steps in RNA-Seq data analysis depend on your goals and biological system







Step 1: Quality Control

Step 2: Data prepping

Step 3: Map Reads to Reference Genome/Transcriptome

Step 4: Assemble Transcriptome

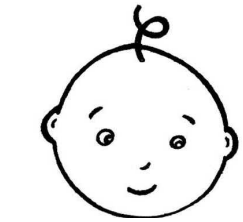
"microarray simulation"

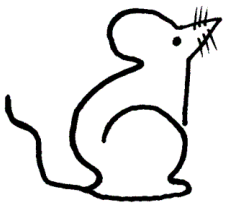
Discovery mode

Other applications:
De novo Assembly
Refine gene models

Identify Differentially Expressed Genes

Programs used in RNA-Seq data analysis depend on your goals and biological system





Step 1: Quality Control — FastQC

Step 2: Data prepping — Filter/Trimmer/Converter

Step 3: Map Reads to Reference Genome/Transcriptome — TopHat, GSNAP

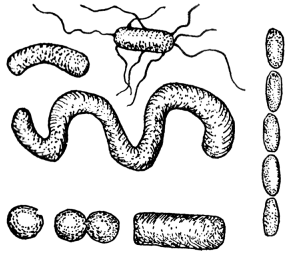
Step 4: Assemble Transcriptome — Cufflinks, Cuffmerge

IGV

Other applications: Refine gene models

Identify Differentially Expressed Genes — Cuffdiff

Programs used in RNA-Seq data analysis depend on your goals and biological system



Step 1: Quality Control — FastQC

Step 2: Data prepping — Filter/Trimmer/Converter

Step 3: Map Reads to Reference Genome/Transcriptome — **Bowtie, BWA**

Step 4: Assemble Transcriptome — Cufflinks, Cuffmerge

Glimmer
Artemis
Trinity, Velvet

Other applications:
De novo Assembly
Refine gene models

Identify Differentially Expressed Genes — Cuffdiff

Visualizing microbial data in Artemis

All mapped reads

Reverse reads

Forward reads

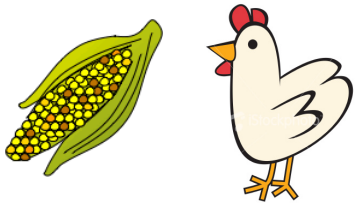
Strand-specific coverage

Forward genes

Reverse genes

Croucher NJ and Thomson NR. Curr Opin Microbiol. (2010) 13:619–624.

Programs used in RNA-Seq data analysis depend on your goals and biological system



Step 1: Quality Control — FastQC

Step 2: Data prepping — Filter/Trimmer/Converter

Step 3: Map Reads to Reference Genome/Transcriptome — **TopHat, GSNAP**

Step 4: Assemble Transcriptome — Cufflinks, Cuffmerge

GeneMark, FGeneSH
Trinity, TransABySS
BLAT

Other applications:
De novo Assembly
Refine gene models

Identify Differentially Expressed Genes — Cuffdiff

Programs used in RNA-Seq data analysis depend on your goals and biological system

Step 1: Quality Control — FastQC

Step 2: Data prepping — Filter/Trimmer/Converter

Step 3:

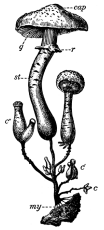
Step 4: Assemble Transcriptome — Cufflinks, Cuffmerge

Trinity, TransABySS

Other applications:
De novo Assembly
Refine gene models

Identify Differentially Expressed Genes — Cuffdiff

Programs used in RNA-Seq data analysis depend on your goals and biological system



Consider sequencing and assembling the genome first if not too big…

Step 1: Quality Control — FastQC

Step 2: Data prepping — Filter/Trimmer/Converter

Step 3: Map Reads to Reference Genome/Transcriptome — **TopHat, GSNAP**

Step 4: Assemble Transcriptome — Cufflinks, Cuffmerge

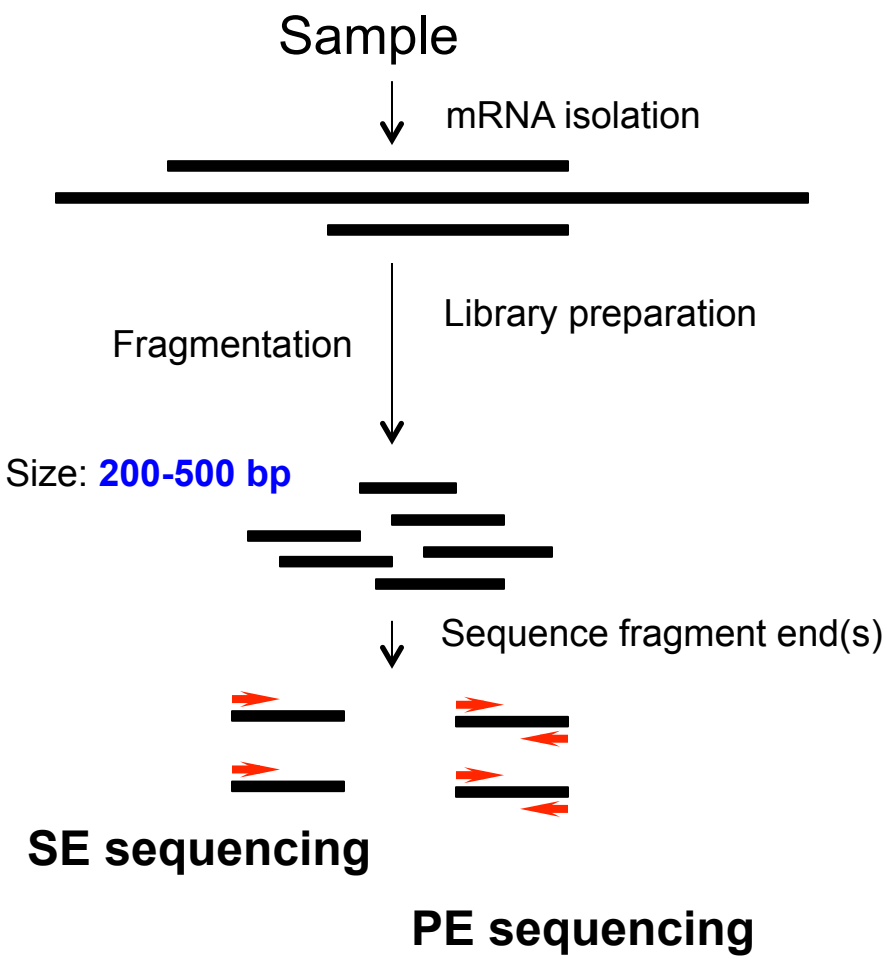
GeneMark, FGeneSH
Trinity, TransABySS
BLAT

Other applications:
De novo Assembly
Refine gene models

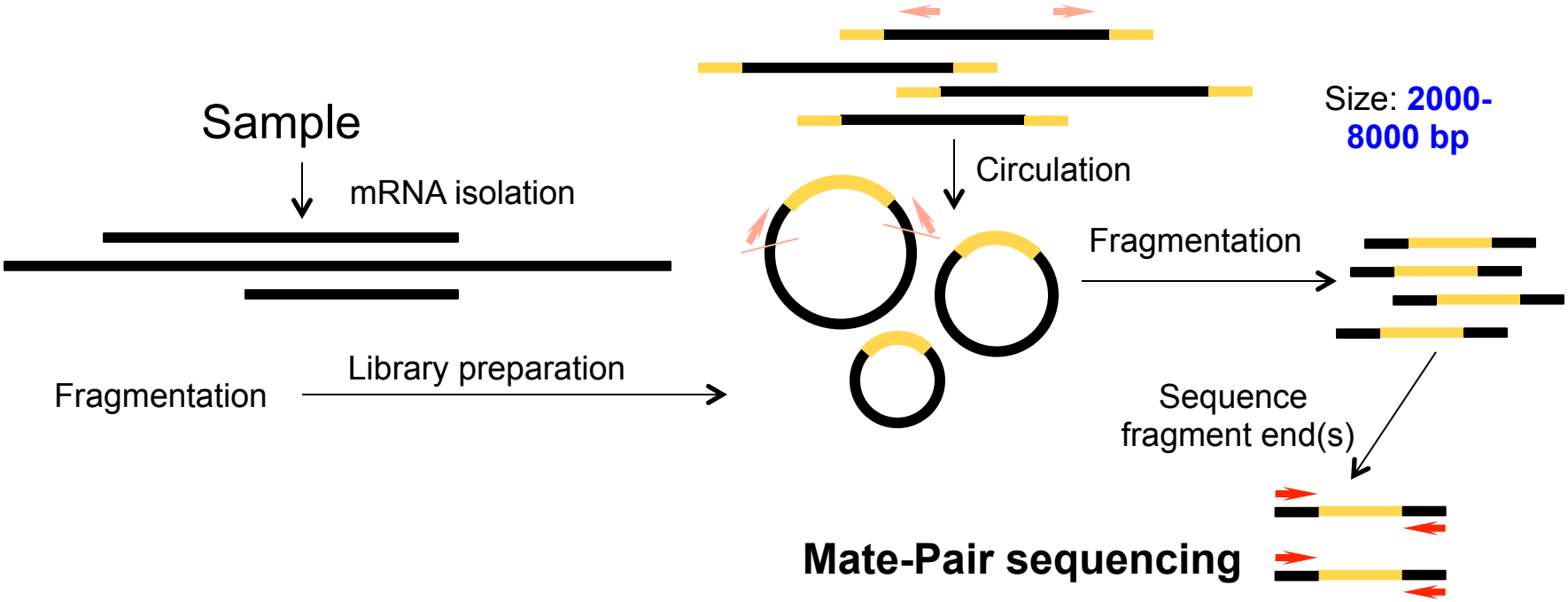
Identify Differentially Expressed Genes — Cuffdiff

Library construction and sequencing design decisions

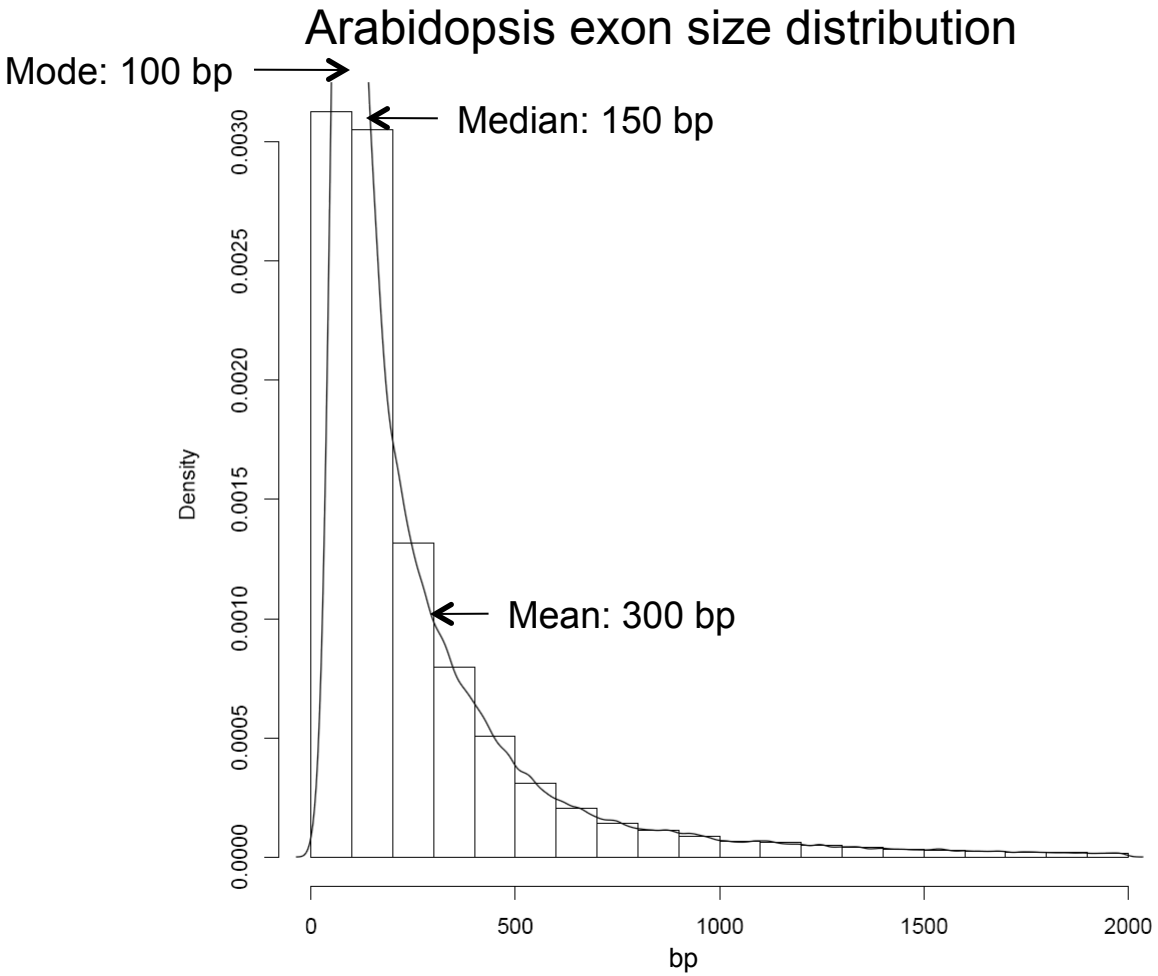
Library type (SE/PE) and insert size



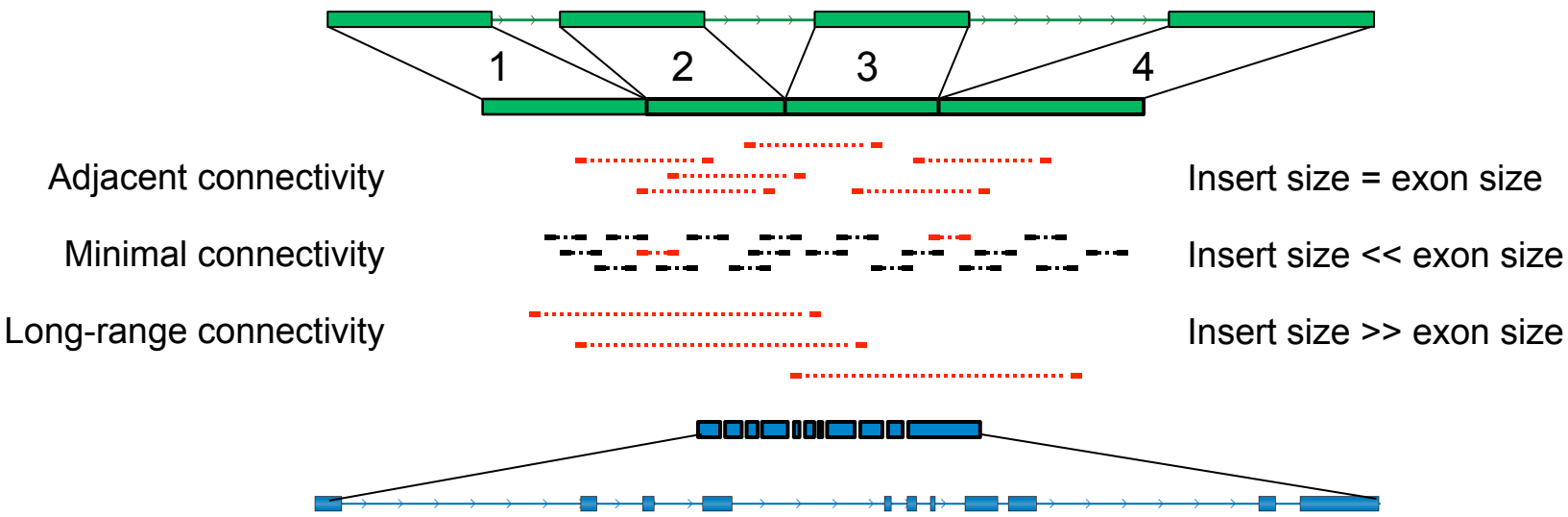
Library type (Mate-pair) and insert size



Optimal library size depends on goals and organism: ***exon size***



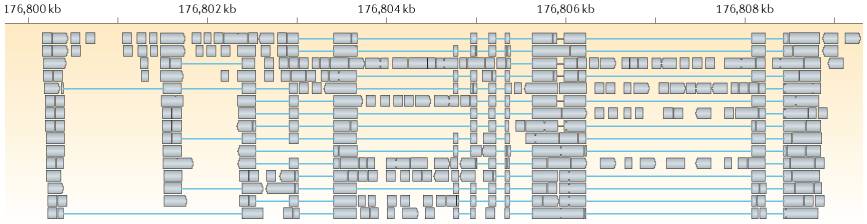
Optimal library size depends on goals and organism: ***exon size***



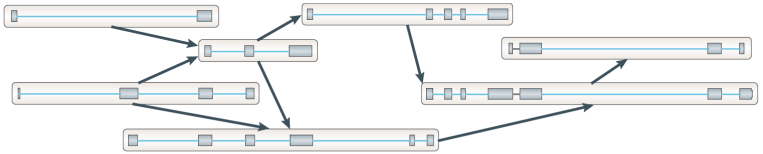
One size doesn't fit all: organisms can differ in exon size distribution

How does connectivity play into the analysis?

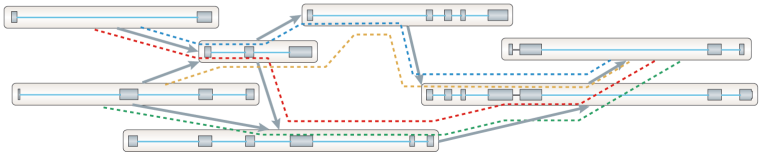
1. splice-align reads to the genome



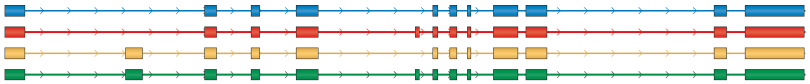
2. Build a graph representing alternative splicing events



3. Traverse the graph to assemble variants

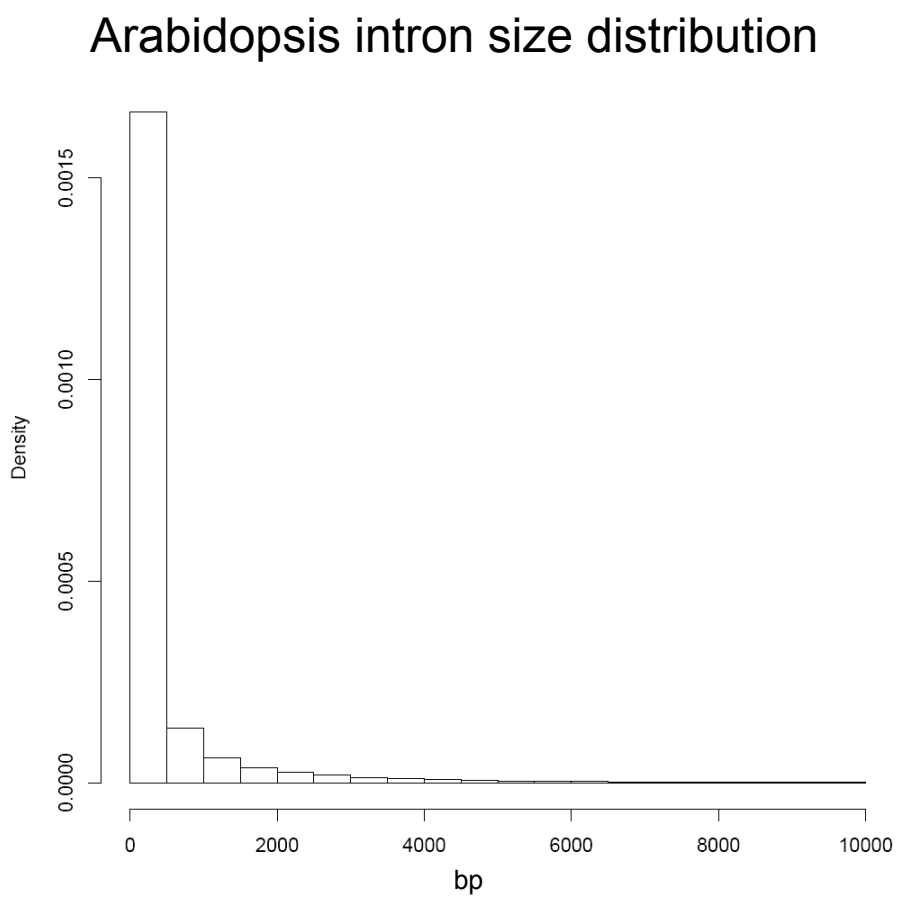


4. Assemble isoforms



Martin JA and Wang Z. Nat Rev Genet. (2011) 12:671–682.

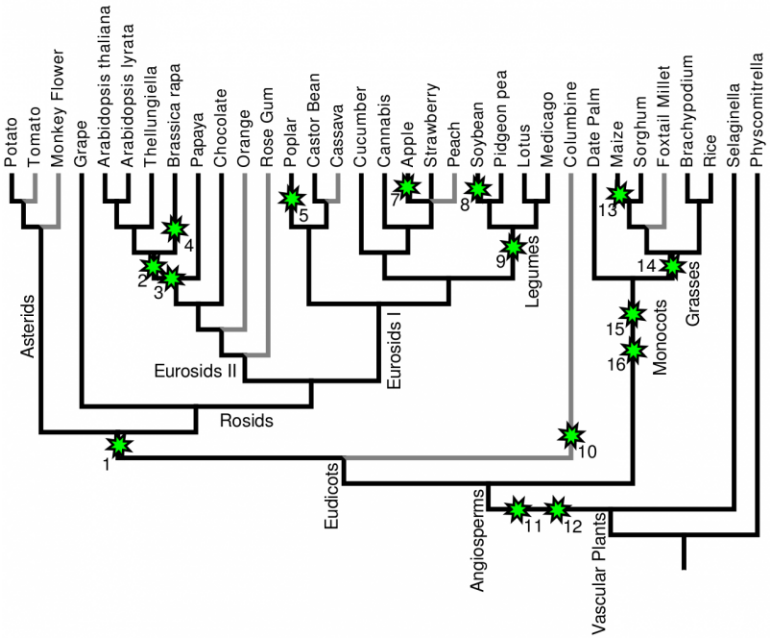
Some algorithms (e.g., tophat) exhaustively look for candidate splices in a specified distance pegged to the expected intron size distribution (default 70-500,000)



Why not just leave the defaults? (e.g., 70-500,000 bp)

- ~3500 Arabidopsis introns < 70 bp
- Huge increase in computation time
- Will accumulate spurious long-range splice junctions

Many plant genomes have undergone ancient Whole Genome Duplications (WGDs)

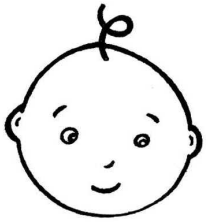


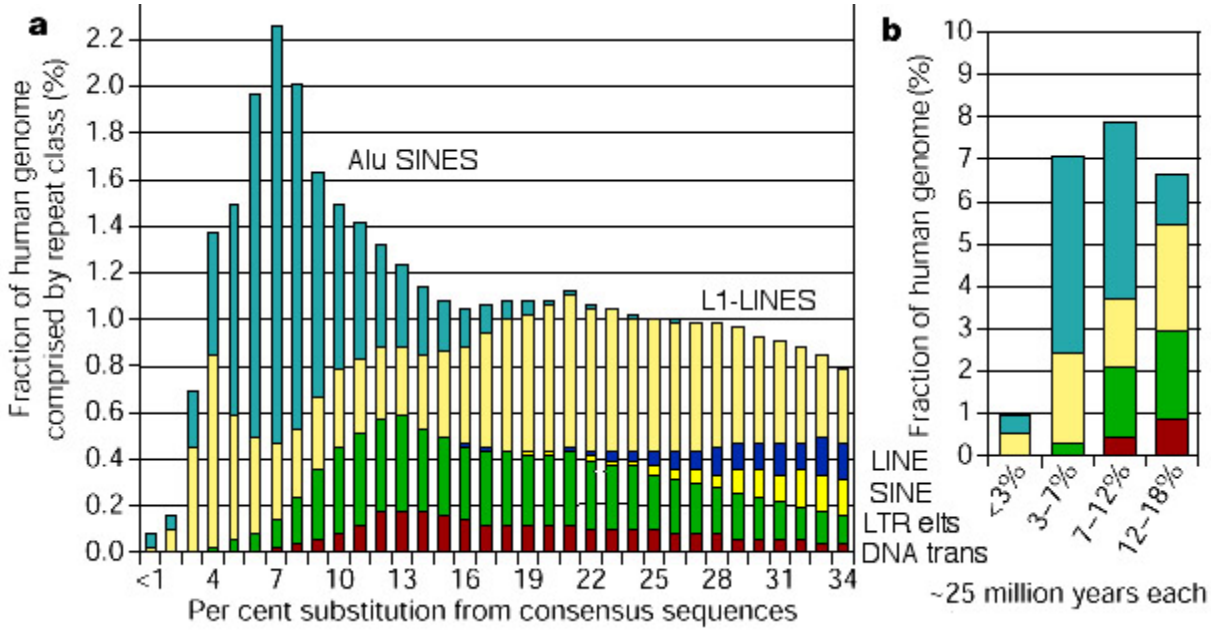
WGD event

http://genomevoloution.org

- Difficulty mapping uniquely to related gene family members
- Abundance levels (e.g., FPKMs) can become skewed for members of large gene families
- Both PE strategies and longer reads help to distinguish paralogs

Some genomes are rife with repetitive elements

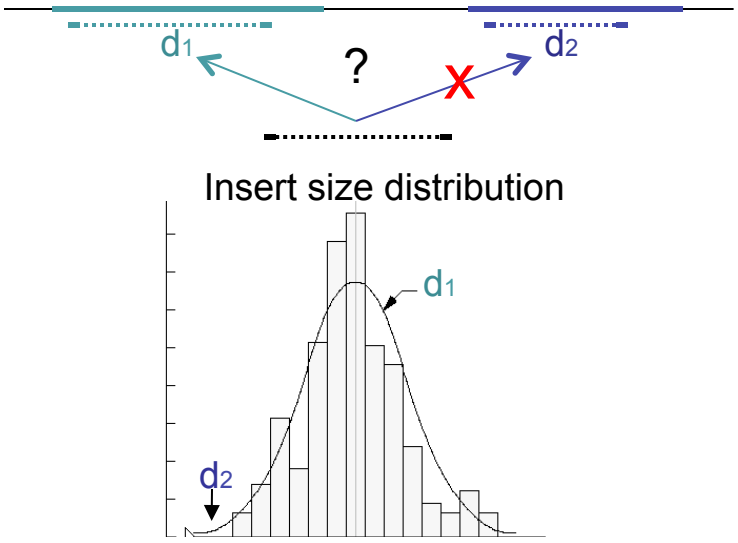




http://genomevoloution.org

- 50%, 65% of the human and maize genome are repeat elements, respectively (repbase, Kronmiller et al., Plant Phys 2008;)
- PE, mate-pair strategies and multiple insert sizes help to uniquely map repeats
- Long reads can help for small-scale or simple repeats

Why is PE crucial for repetitive genomes and those with paralogous gene families?

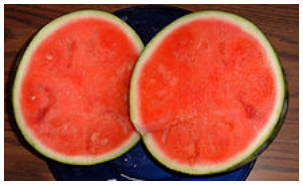


2 x 50 bp is better than 1 X 100 bp for most applications and systems.

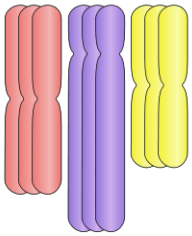
Sequencing depth needed depends on transcriptome size and the project goals

- **Sequencing Depth** is the average read coverage of target sequences
  - Sequencing depth = total number of reads X read length / estimated target sequence length
  - Example, for a 5MB transcriptome, if 1Million 50 bp reads are produced, the depth is 1 M X 50 bp / 5M ~ 10 X
- Average coverage may be misleading, since expression levels can vary more than 5 orders of magnitude!
- Differential expression requires less depth than assembly, gene model refinement and structural variant discovery.

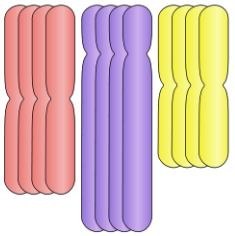
Polyploidy is particularly problematic



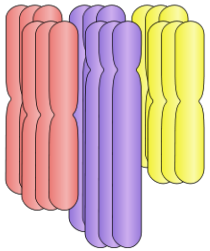
Triploid (3N)



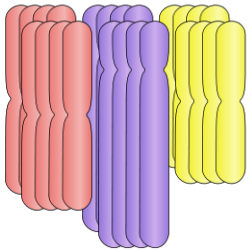
Tetraploid (4N)



Hexaploid (6N)

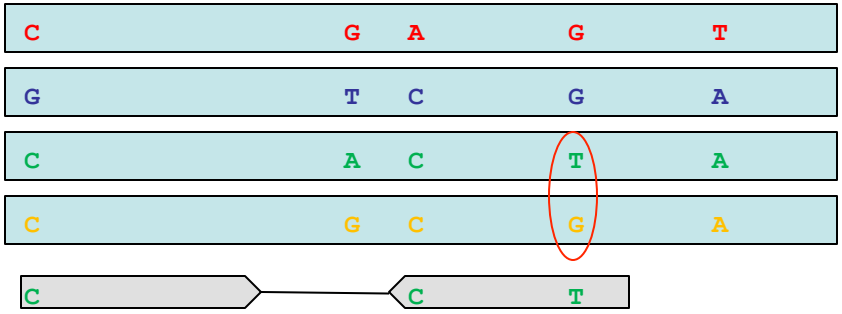


Octaploid (8N)



- Difficult to distinguish alleles from paralogs
- Genome assembly often intractable
- Need care in design of transcriptome experiment

Certain applications and biological systems will require special design considerations for maximal resolution



- **Polyploid genomes may require long reads, multiple insert sizes and custom software to distinguish among highly similar alleles at each locus.**
- **Ditto for those who wish to interrogate allele-specific differential expression (e.g., maternal or paternal impriting).**

Genome size characteristics (iGenomes)

Species	Number of genes	Transcriptome size (Mbp)	Mode|Avg **exon** size	**Intron** size range (1%|99%)	% genome repetitive	% genes in families*
Homo sapiens	29230	70.1	100|300	77|107000	47	20
Mus musculus	24080	61.4	100|300	78|100000	44	NA
*Gallus gallus***	4906	11.1	100|230	73|120000	10	NA
Drosophila melanogaster	18436	30.1	150|450	30|25000	32	7
Caenorhabditis elegans	23933	28.0	110|220	43|8000	4	24
Arabidopsis thaliana	27278	51.1	70|300	46|4900	9	35
Saccharomyces cerevisiae	6692	8.9	75|1200	20|2600	1	36
*Escherichia coli****	4290	0.6	NA	NA	3	52

* % genes with at least one paralog in the COG database (unicellular) or included in the COG lineage specific expansion (LSE) list. (These percentages are likely systematic underestimates)
** Poor annotation is suspected for iGenomes UCSC-based Gallus gallus (galGal3)
*** http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/E/Esch.coli.html; ecocyc; Gur-Arie, Genome Res 2000;.

Summary of Library Construction and Sequencing Decisions

	1	2	3	4
Project Goals:	*De novo* Assembly of transcriptome	Refine gene model	Differential Gene Expression	Identification of structural variants
Library Type:	PE, Mated PE	PE, SE	PE	PE, Mated PE
Sequencing Depth:	Extensive (> 50 X)	Extensive	Moderate (10 X ~ 30 X)	Extensive

- SE may be OK for (3) DGE if you have a good annotation and a simple genome.
- Strand-specific library creation may be necessary for organisms with a large percentage of genes that overlap on opposite strands (e.g. bacteria, yeast), or if you're interested in antisense regulation.

ENCODE RNA-Seq guidelines

Sample Replicates and Pooling Decisions

	1	2	3	4
Project Goals	*De novo* Assembly of transcriptome	Refine gene model	Differential Gene Expression	Identification of structural variants
Pooling OK?	No	Yes	No	Yes, for discovery
Biological Replicates?	Yes	Yes, if not pooling	Yes	Yes, if not pooling

- Pooling may be advisable if RNA is limited or if not interested in biological variability.

As a general rule, the following biological replicates are advisable for DGE:

- 3+ for cell lines and pooled samples
- 5+ for inbred lines (e.g., BL6 mice, NILs, RILs)
- 20+ for human samples

istockphoto.com

Part II

Read Mapping Statistics and Visualization

John Garbe, PhD

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Mapping Statistics

How well did my sequence library align to my reference?

Mapping Statistics

- Mapping Output
  - SAM (text) / BAM (binary) alignment files
  - Summary statistics (per read library)
    - % reads with unique alignment
    - % reads with multiple alignments
    - % reads with no alignment
    - % reads properly paired (for paired-end libraries)
    - Mean and standard deviation of insert size

SAM specification: http://samtools.sourceforge.net/SAM1.pdf

Mapping Statistics

- SAM Tools
- Picard
- Tophatstats

Mapping Statistics – SAMtools

- Galaxy
  - NGS: SAM Tools -> flagstat
- MSI Command line
  - Module load samtools
  - samtools flagstat accepted_hits.bam

Mapping Statistics – SAMtools

- SAMtools output

```
% samtools flagstat accepted_hits.bam
31443374 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 duplicates
31443374 + 0 mapped (100.00%:-nan%)
31443374 + 0 paired in sequencing
15771038 + 0 read1
15672336 + 0 read2
15312224 + 0 properly paired (48.70%:-nan%)
29452830 + 0 with itself and mate mapped
1990544 + 0 singletons (6.33%:-nan%)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Mapping Statistics – Picard

- Galaxy
  - NGS: Picard (beta) -> SAM/BAM Alignment Summary Metrics
- Command line:
  - module load picard-tools
  - java *-Xmx2g* -jar CollectAlignmentSummaryMetrics.jar INPUT=accepted_hits.bam OUTPUT=stats.txt

Mapping Statistics – Picard

- Picard output

```
CATEGORY                TOTAL_READS
FIRST_OF_PAIR           14739626
SECOND_OF_PAIR          14653925
PAIR                    29393551
```

Mapping Statistics – tophatstats

- Galaxy
  - MSI -> tophatstats
- Command line
  - module load tophatstats

Mapping Statistics – tophatstats

- Tophatstats output (paired-end reads)

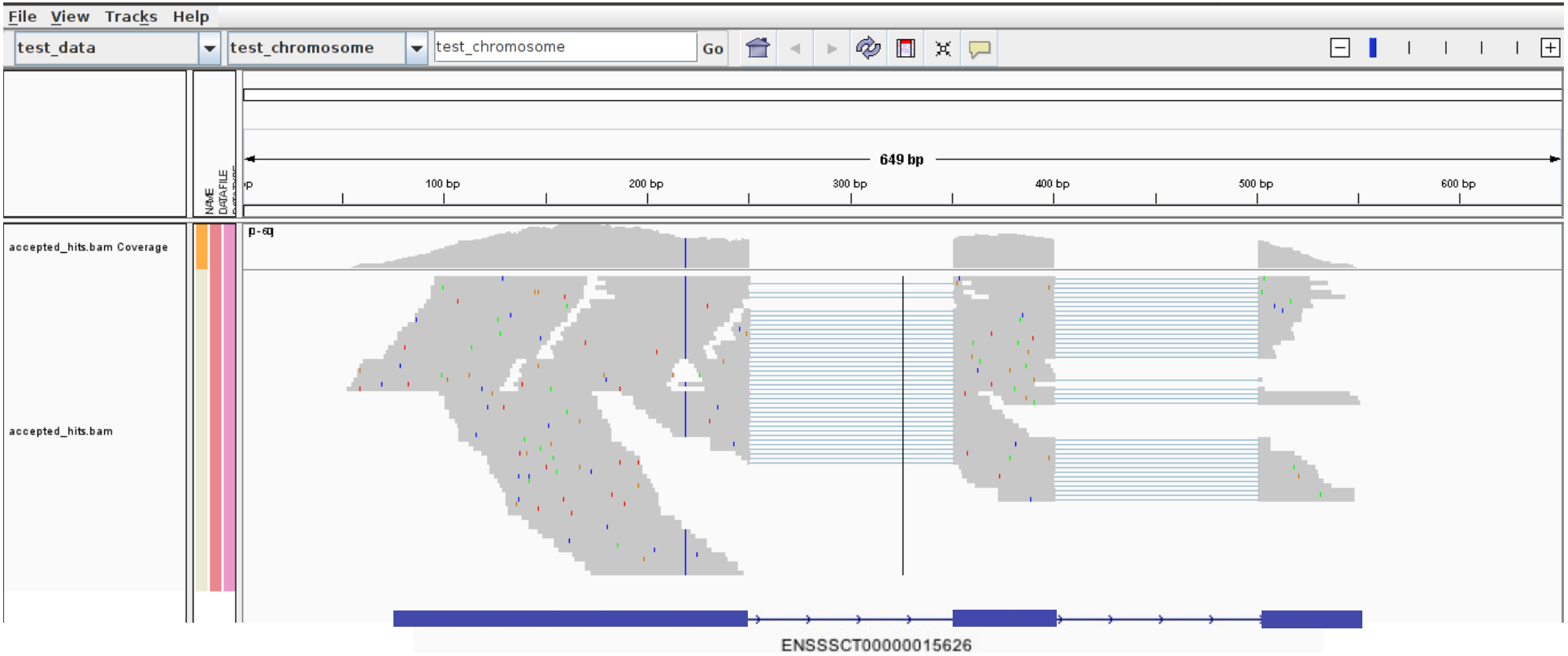
```
% tophatstats.pl accepted_hits.bam L1_R1_sample1.fastq
Input files: accepted_hits.bam    L1_R1_sample1.fastq
250000 total read pairs in fastq file
120004 (48.00%) read pairs mapped with correct insert size
       (116869 with unique alignments)
50536 (20.21%) read pairs mapped with wrong insert size
       (49351 with unique alignments)
24368 (9.75%) read pairs with only one read in the pair mapped
       (23544 with unique alignments)
55092 (22.04%) read pairs with no mapping
60.13bp average inner distance between read pairs
```

Mapping Visualization

- Integrative Genomics Viewer (IGV)
  - Fast genome browser
  - Supports array-based and next-generation sequence data, and genomic annotations
  - Free Java program

http://www.broadinstitute.org/igv/home

Mapping Visualization

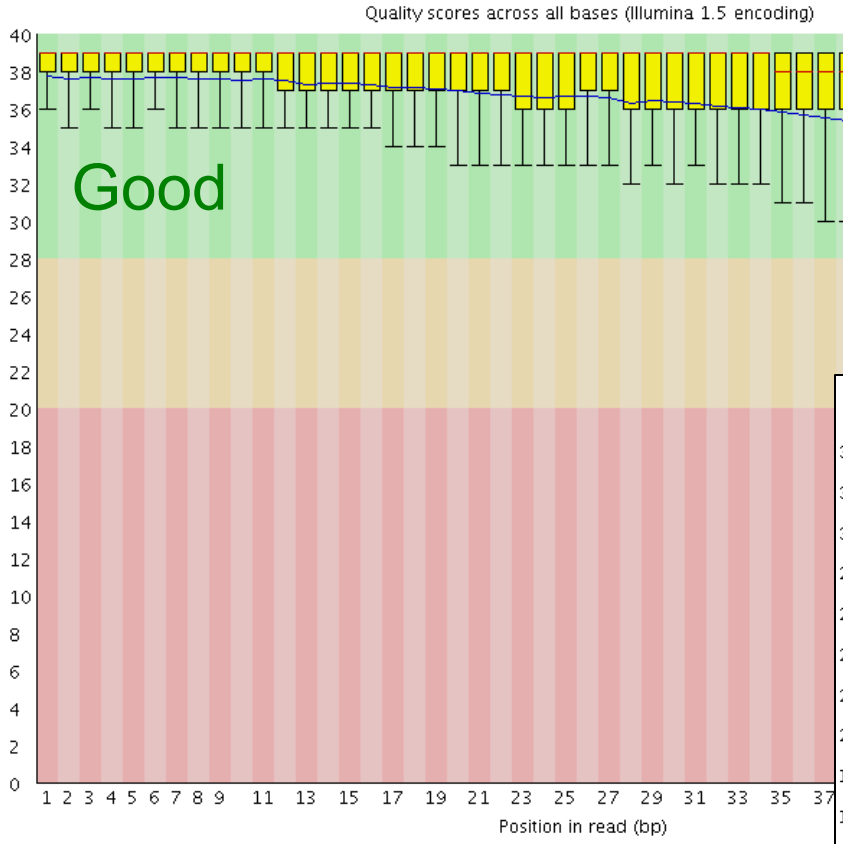


Bam file viewed with IGV

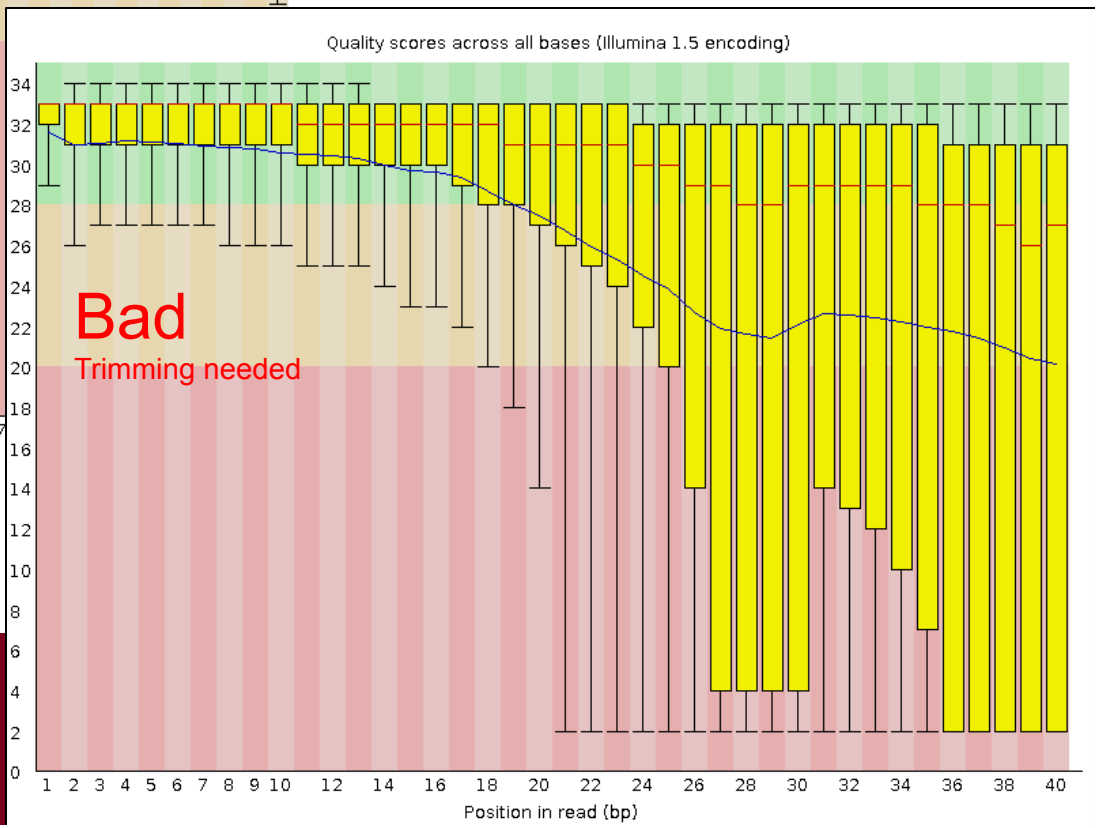
Causes of poor mapping

- Improper alignment parameters
- Poor quality sequence library
- Contaminated sequence library
- Poor quality reference
- Repetitive genome
- Divergence between sequenced population and reference
- Mislabeled samples
- Corrupted files
- Short read length
- Poor choice of mapping software
- Bug in mapping software
- …

Poor Quality Library



Poor quality read library decreases mapping performance



Bug in software

Tophat 2.0.0	Tophat 2.0.1	
35%	48%	mapped, properly paired
33%	20%	mapped, wrong insert size
10%	9%	singleton
22%	22%	no mapping

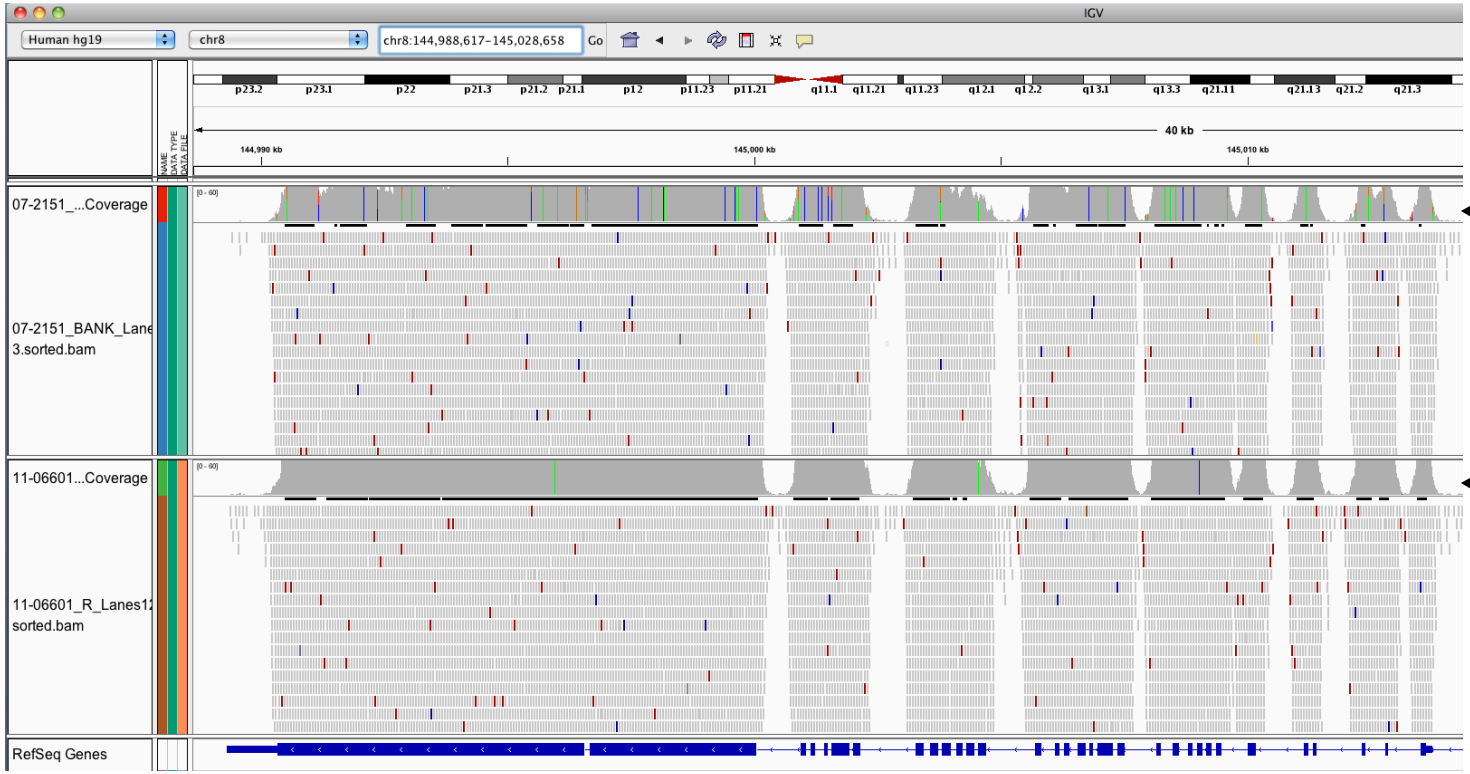
New “bugfix” release of Tophat improves mapping performance

Poor Quality Reference

Sus scrofa 9.2	Sus scrofa 10.2	
46%	48%	mapped, properly paired
17%	20%	mapped, wrong insert size
9%	9%	singleton
26%	22%	no mapping

Mapping performance improves due to improvement in Pig genome build

Divergence between sequenced population and reference



Large and small sequence divergence between two human samples and the human reference genome

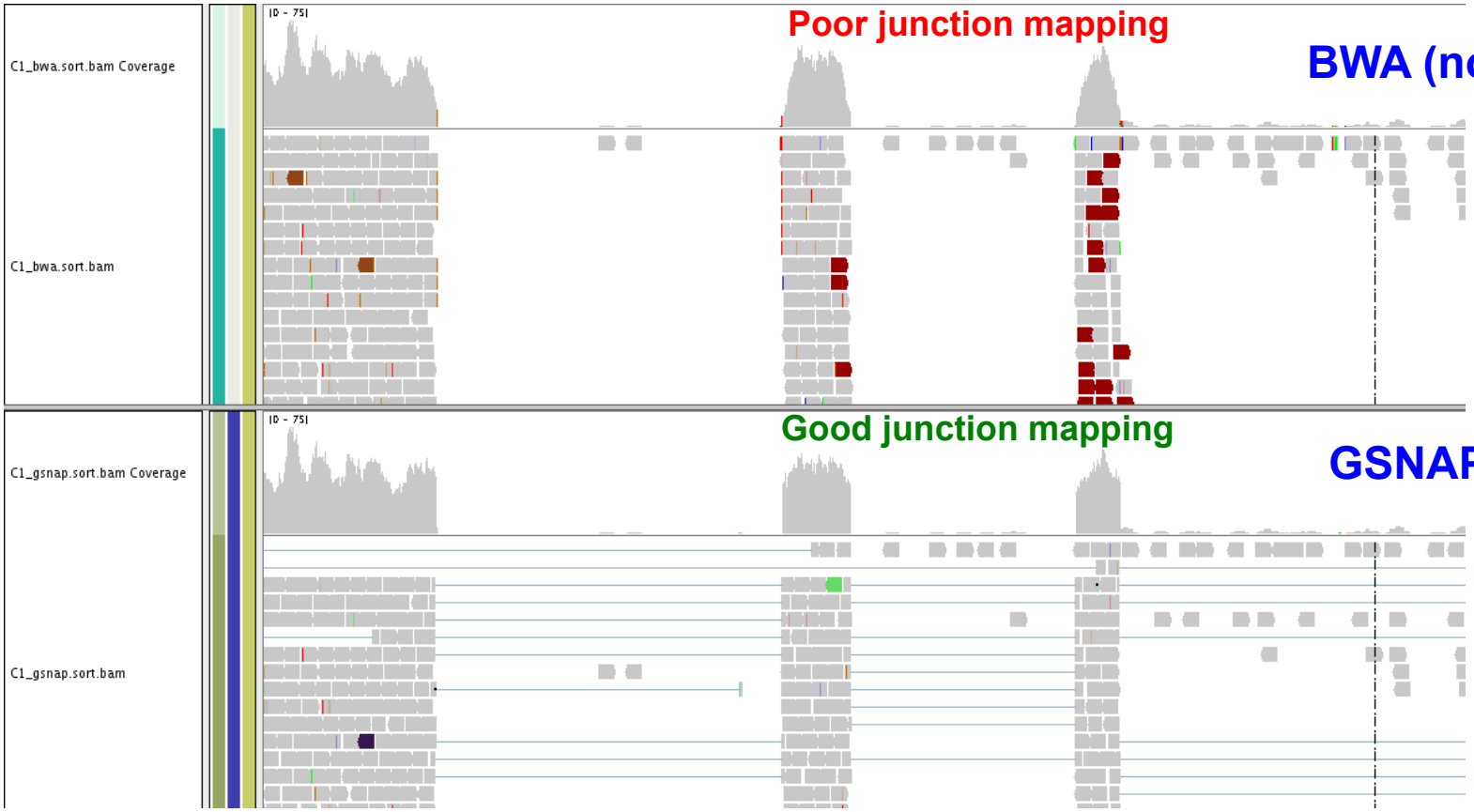
Contaminated sequence library

Overrepresented sequences

Sequence	Count	Percentage	Possible Source
GTATTACAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCG	820428	2.8366639370528275	Illumina Paired End PCR Primer 2 (100% over 43bp)
GTATACAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGT	749728	2.5922157461699773	Illumina Paired End PCR Primer 2 (100% over 44bp)
CGGTTCAGCAGGAATGCCGAGATCGGAAGAGCGGTTCAGCAGGAATGCCG	648852	2.243432780066747	Illumina Paired End Adapter 2 (100% over 31bp)
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGATCGGAAGAGCGGTTCAG	176765	0.6111723403310748	Illumina Paired End PCR Primer 2 (97% over 36bp)
ACGTCGTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCG	143840	0.4973327832615156	Illumina Paired End PCR Primer 2 (100% over 43bp)
GTATTCAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGT	124281	0.42970672717272257	Illumina Paired End PCR Primer 2 (100% over 44bp)
GTATCAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTA	99207	0.34301232917842867	Illumina Paired End PCR Primer 2 (100% over 45bp)
GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGT	96289	0.33292322279941655	Illumina Paired End PCR Primer 2 (100% over 50bp)
CGGAAGAGCGGTTCAGCAGGAATGCCGAGATCGGAAGAGCGGTTCAGCAG	93842	0.3244626185124245	Illumina Paired End PCR Primer 2 (96% over 33bp)
CGTTACGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCG	75370	0.26059491013918545	Illumina Paired End PCR Primer 2 (100% over 43bp)
CGTACGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGT	63691	0.22021428183196043	Illumina Paired End PCR Primer 2 (100% over 44bp)
ACGTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTAT	56765	0.19626734873359242	Illumina Paired End PCR Primer 2 (100% over 46bp)
TACTGTAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCG	42991	0.14864317078139472	Illumina Paired End PCR Primer 2 (100% over 43bp)

FastQC output showing ~10% adapter contamination

Poor choice of mapping software



Improper alignment parameters

Correct inner distance (60)	Incorrect inner distance (220)	
48%	43%	mapped, properly paired
20%	25%	mapped, wrong insert size
9%	10%	singleton
22%	22%	no mapping

Incorrect “inner mate pair distance” parameter decreases mapping performance

Corrupted files

Correct fastq file	Corrupted fastq file	
48%	22%	mapped, properly paired
20%	46%	mapped, wrong insert size
9%	10%	singleton
22%	22%	no mapping

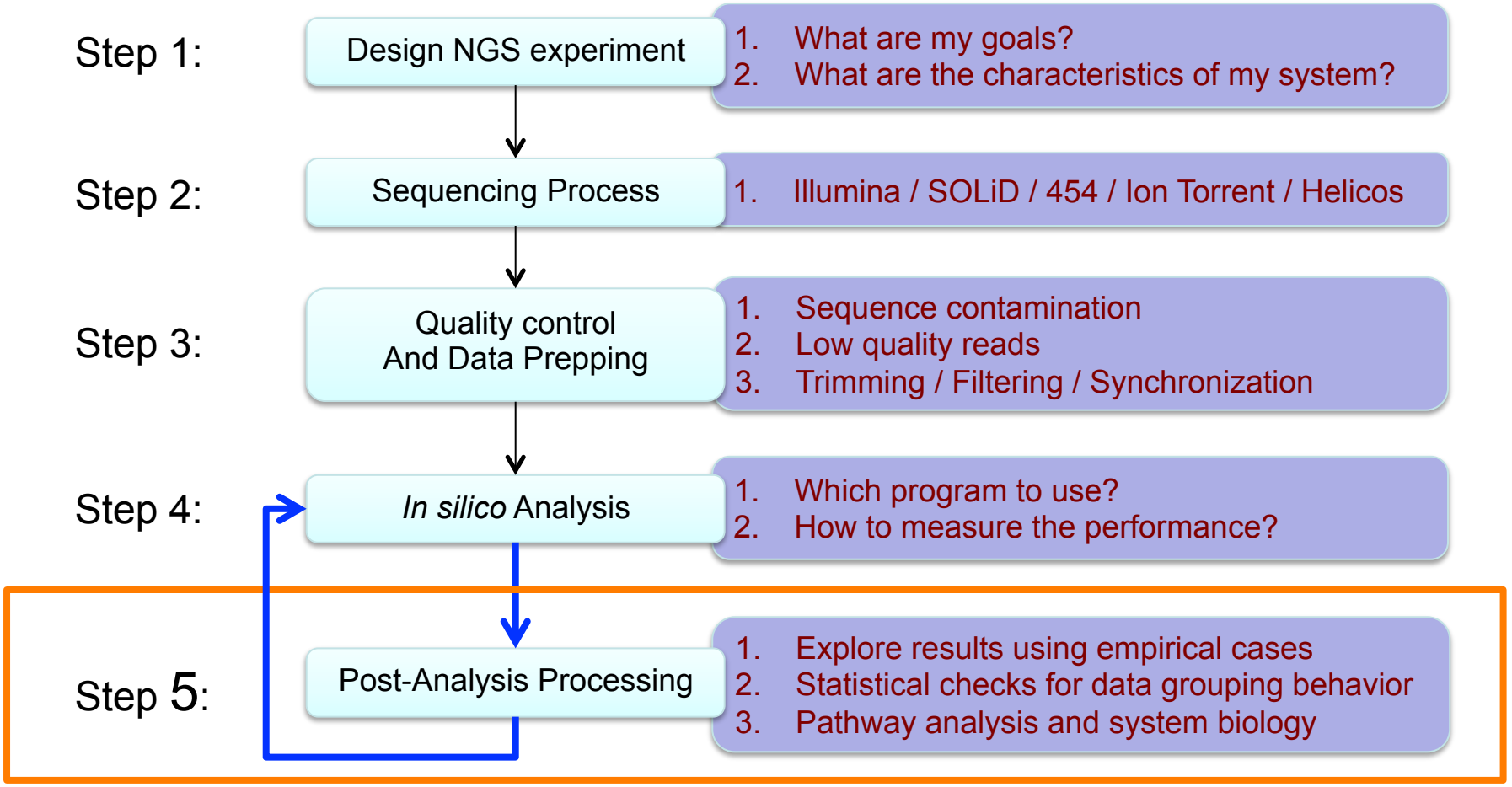
Unsynchronized paired-end fastq file decreases percentage of properly-paired reads

Part III

Post-Analysis Processing - Exploring the Data and Results

Ying Zhang, PhD

Workflow of a typical NGS project



Widely-used Tools in Data Exploring

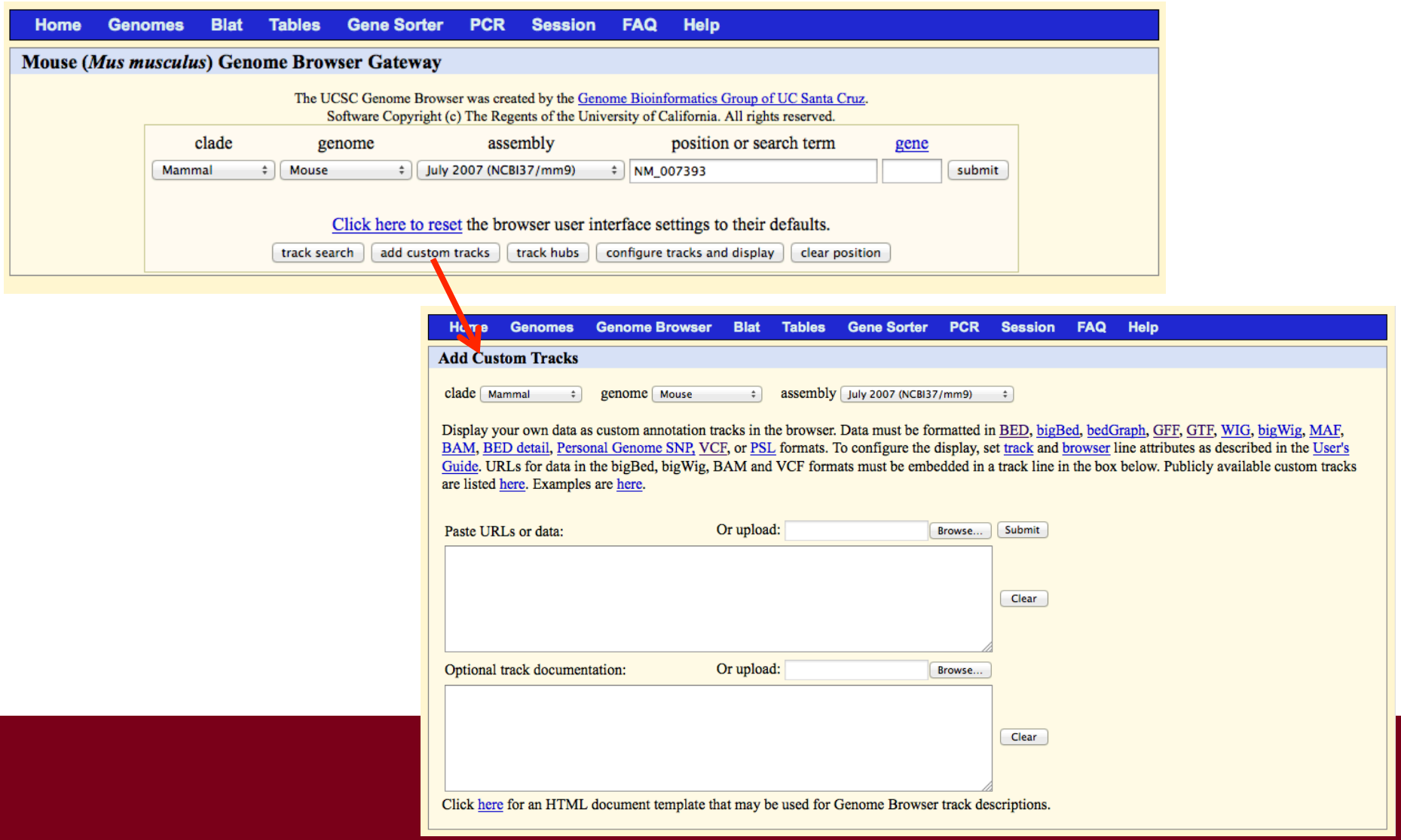
- Direct visualization of “positive controls”:
  - IGV viewer
  - UCSC Genome Browser
- Statistical checks of data structure:
  - PCA: principle component analysis
  - MDS: multi-dimension scaling
  - Unsupervised clustering and Heatmap
- System-level Analysis:
  - IPA: ingenuity pathway analysis

Integrative Genomics Viewer (IGV)

- Fast genome browser
- Supports array-based and next-generation sequence data, and genomic annotations
- Free Java program
- Launch:
  - From Galaxy
  - From Desktop: allocate enough memory 

http://www.broadinstitute.org/igv/home

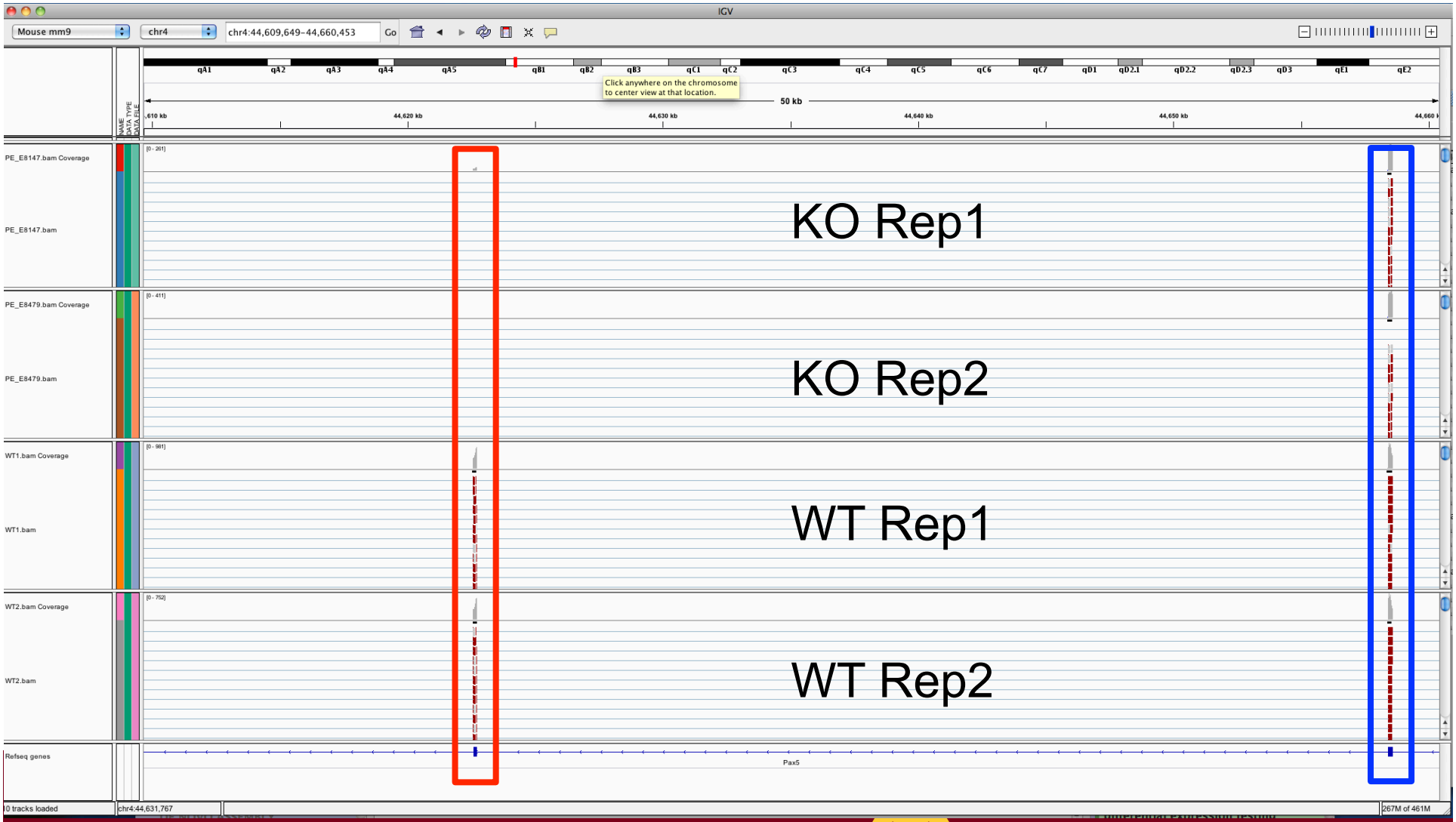
UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway)



No. 1 in your Check-List

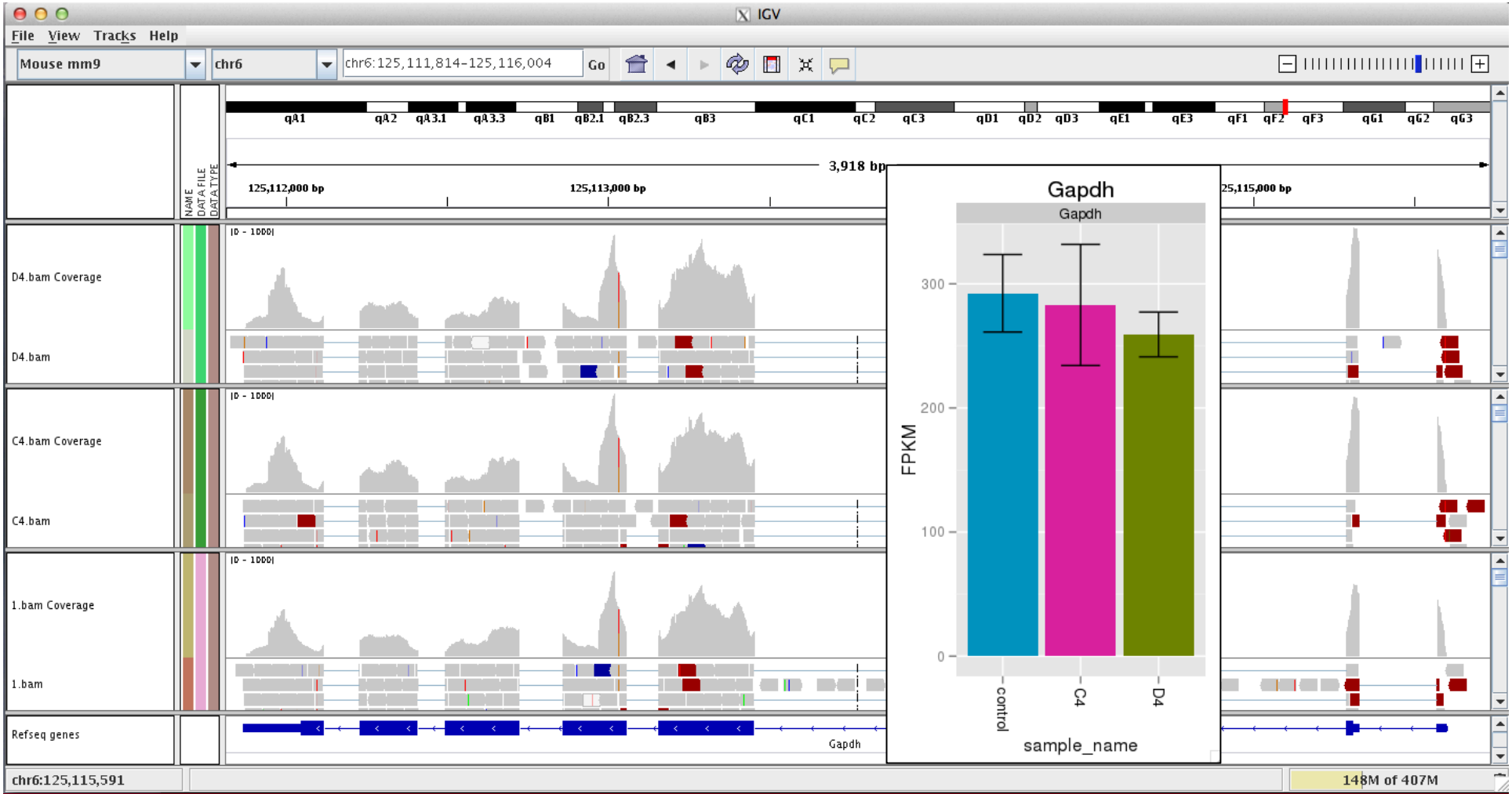
"Are my data behaving as expected?"

Exploring results using Empirical Cases – Example I: no reads mapped at knock-out site



Data Courtesy of Dr. Mike Farrar and Dr. Lynn Harris (unpublished data)

Example II: Housekeeping genes should behave similarity across multiple samples



Data Courtesy of Dr. David Bernlohr and Dr. Ann Hertzel (unpublished data)

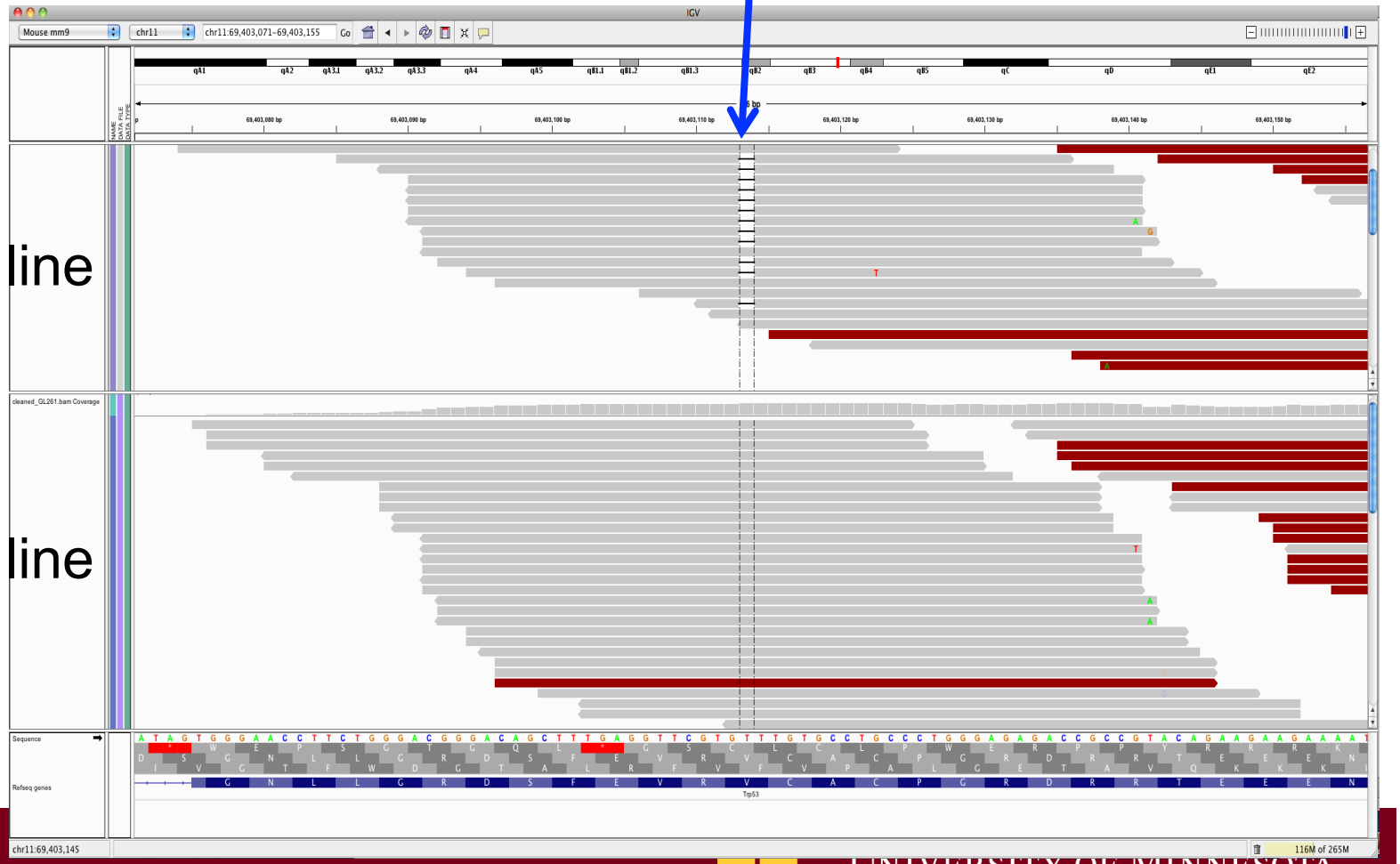
UNIVERSITY OF MINNESOTA
Driven to Discover℠

Example III: review of known biomarkers, for example, known SNP and indel

Heterozygous deletion of ‘T’ with 46% penetrance

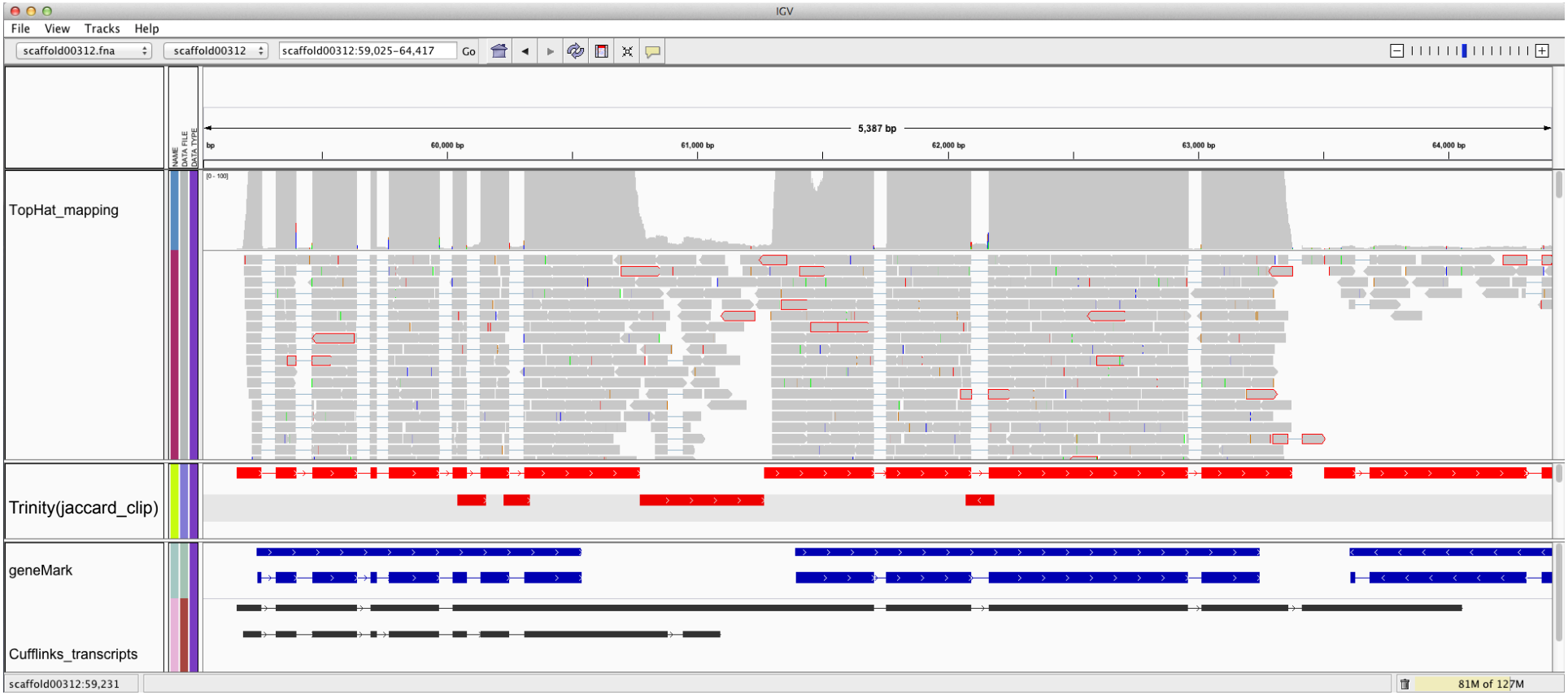
Cancer cell line

Control cell line



Data Courtesy of Dr. John Ohlfest and Dr. Flavia Popescu (unpublished data)

Example IV: detect the caveat of programs



Data courtesy of Dr. Steve Gantt and Dr. Karen Tang (unpublished data)

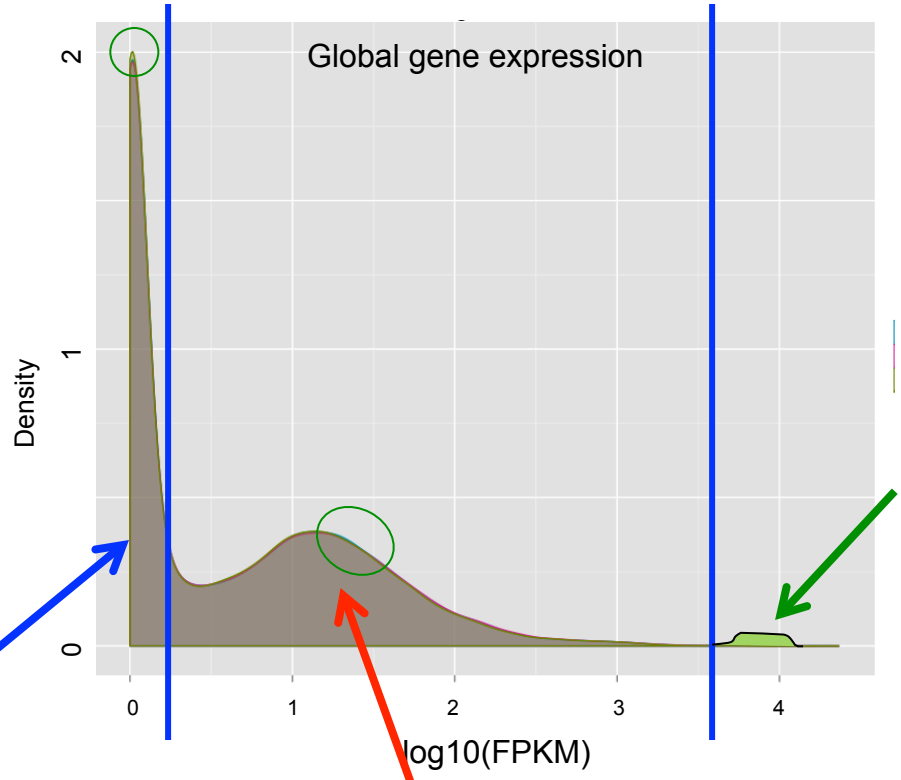
Specific Notes for Prokaryotes' samples

- Cufflinks developer: 

"We don't recommend assembling bacteria transcripts using Cufflinks at first. If you are working on a new bacteria genome, consider a computational gene finding application such as Glimmer."

- So for bacteria transcriptome:
  - If the genome is available, do genome annotation first then reconstruct the transcriptome.
  - If the genome is not available, try *de novo* assembly of the transcriptome, followed by gene annotation.

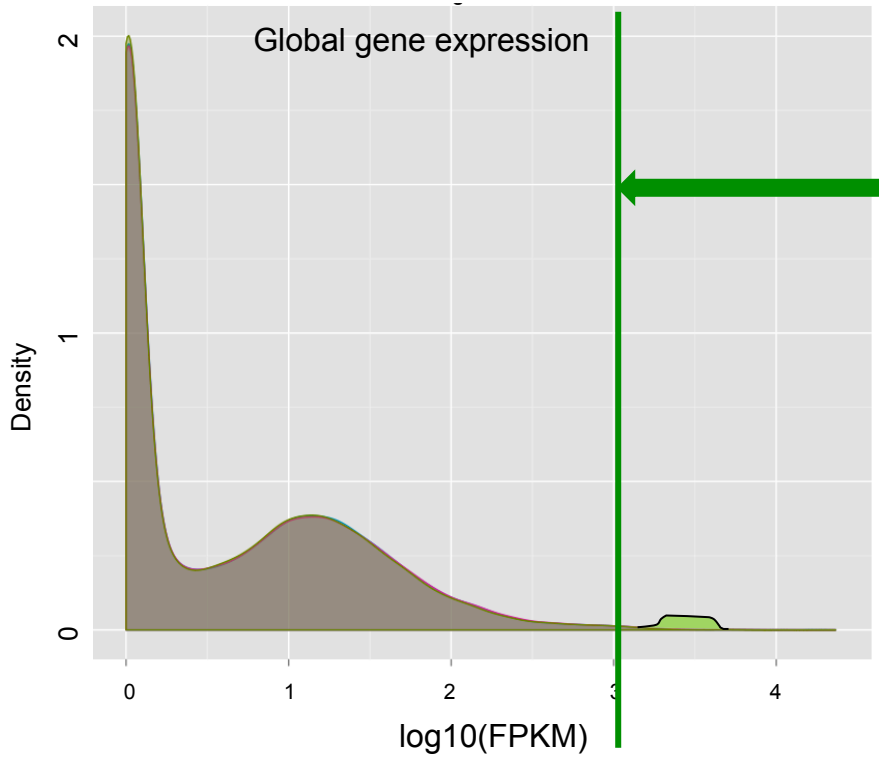
Explore the global distribution of data



Many genes will have little or no expression.

A set of genes have a high expression.

Very few genes have an usually high expression.



Exclude the highly-expressed genes for highly-unbalanced expression between conditions.
Set “yes” to “Perform quartile normalization”.

Perform quartile normalization:

Yes

Removes top 25% of genes from FPKM denominator to improve accuracy of differential expression calls for low abundance transcripts.

Example: red cell blood compared to other tissue

Warning: don't throw the baby with the bathwater…

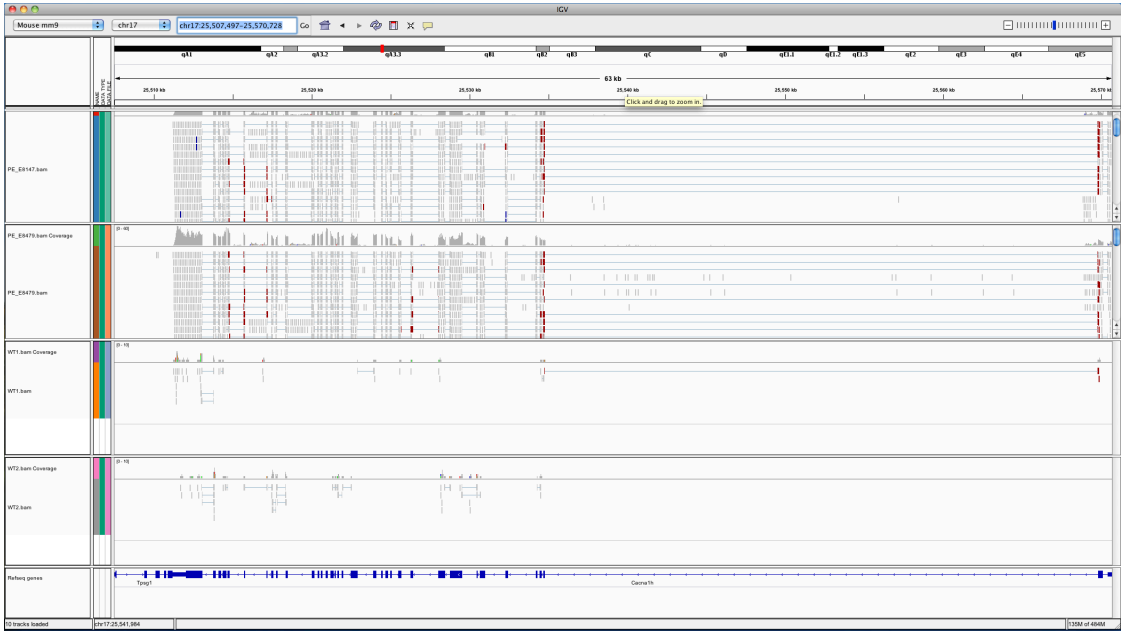
Cuffdiff: "Min Alignment Count" must be satisfied in **all** samples – too high a value will remove genes not expressed in one condition but strongly expressed in another!

Mut Rep 1

Mut Rep 2

Wt Rep 1

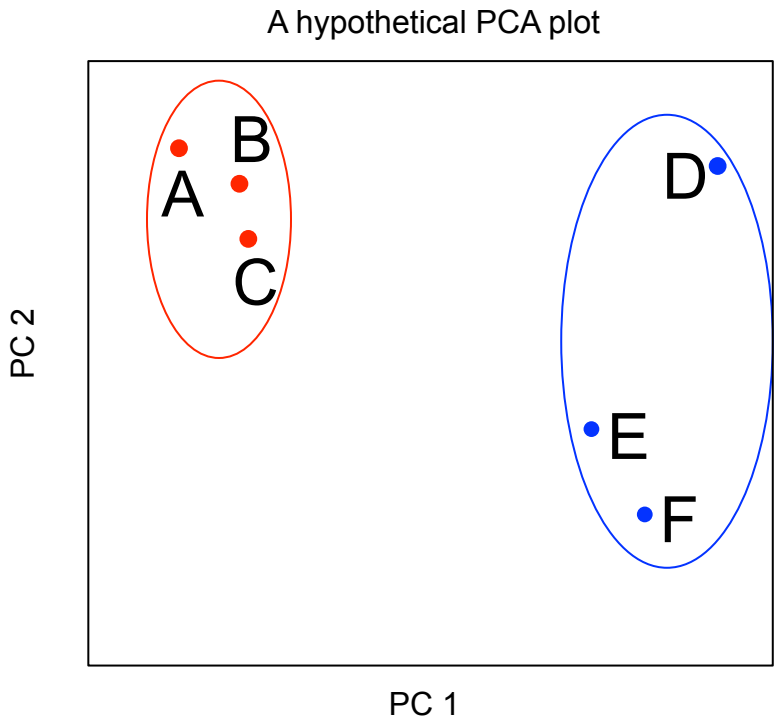
Wt Rep 2

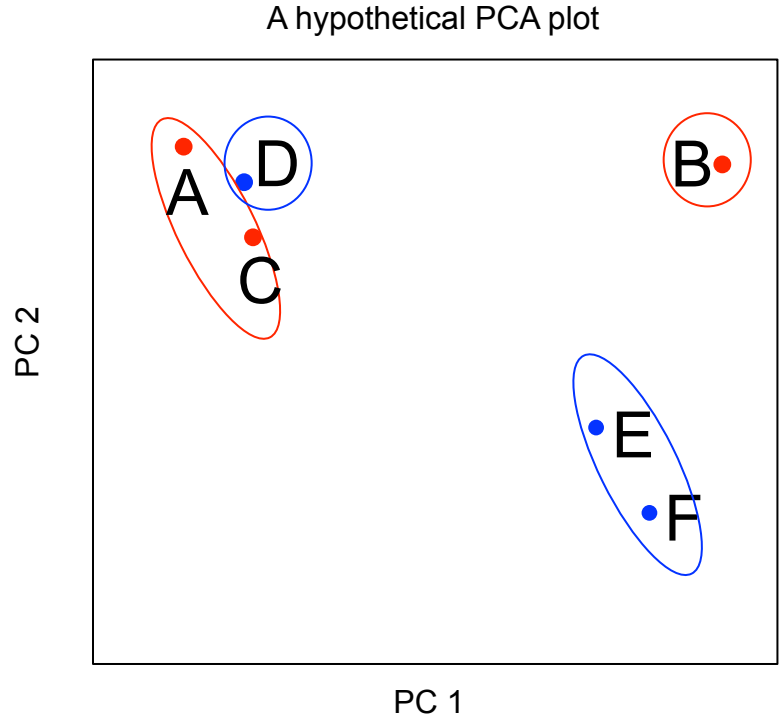


This gene was reported as DE with "Min Alignment Count" = 10, but not with 100.

Statistical Checks of data structure – Multi-Variable Analysis

- Biological replicates should show grouping behavior in multi-variable analysis:
  - innate consistence between samples





Within-group variation: non-biological variations

- Source of non-biological variation:
  - Batch effect
    - How were the samples collected and processed? Were the samples processed as groups, and if so what was the grouping?
  - Non-synchronized cell cultures
    - Were all the cells from the same genetic backgrounds and growth phase?
  - Use technical replicates rather than biological replicates

How to check for data variation?

- Principle Component Analysis (PCA)
  - Uses an orthogonal transformation
  - The first principle component has the largest possible variance
- Multi-Dimensional Scaling (MDS)
  - Computes euclidean distances among all pairs of samples
- Unsupervised Clustering / heatmap
  - Identify the hidden structure in “unlabeled” data
- Tools:
  - Galaxy
  - Statistical Package: R, SPSS, MatLab
  - Partek and Genedata Expressionist

Steps in PCA analysis

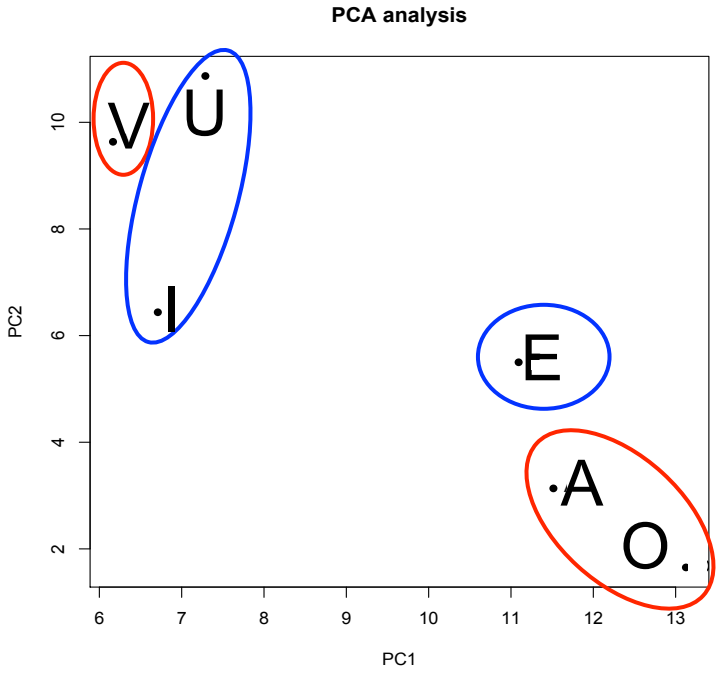
1. Construct the multiple variable matrix → 2. Run PCA analysis and explore the result

e.g. tables of FPKM values

transcript	Sample A	Sample V	Sample O	Sample E	Sample I	Sample U
gene1	6.18	6.64	6.46	6.30	6.58	6.54
gene2	5.48	0.11	1.00	0.24	0.02	0.68
gene3	20.53	18.93	18.79	18.51	18.00	18.26
gene4	55.47	52.71	50.39	54.66	49.15	44.68
gene5	7.28	8.09	8.57	7.82	8.29	9.38
gene6	14.65	13.88	13.48	13.98	14.72	12.47
gene7	16.41	13.80	14.99	17.20	14.39	13.50
gene8	6.17	6.79	7.20	6.70	8.42	7.26
gene9	25.83	24.24	25.63	27.09	22.18	23.09
gene10	38.04	30.39	35.53	37.42	28.72	27.28
gene11	195.06	179.88	178.18	208.25	179.01	155.15
gene12	32.82	32.04	31.84	33.62	31.06	29.46
gene13	18.41	16.75	16.72	17.33	16.32	16.87
gene14	24.00	21.05	22.68	22.72	22.08	22.45
			.....			

Group 1 (A,V,O)

Group 2 (E,I,U)



Heatmap: Unsupervised clustering

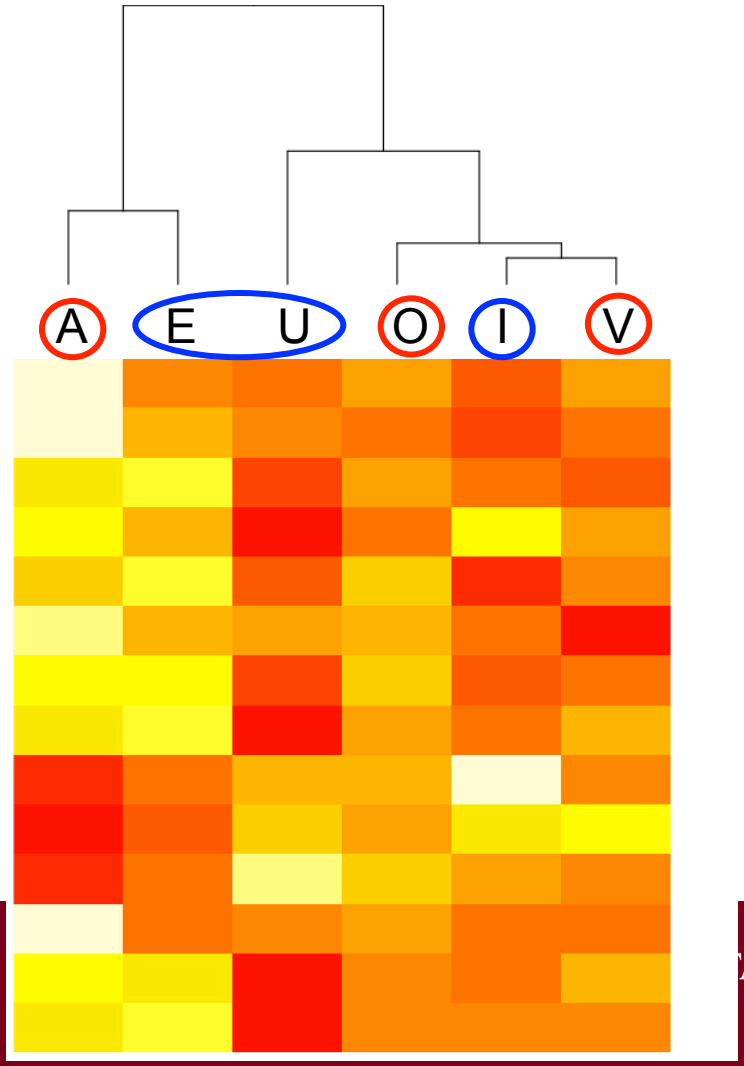
1. Construct the multiple variable matrix → 2. Run Unsupervised Clustering and generate Heatmap

e.g. tables of FPKM values

transcript	Sample A	Sample V	Sample O	Sample E	Sample I	Sample U
gene1	6.18	6.64	6.46	6.30	6.58	6.54
gene2	5.48	0.11	1.00	0.24	0.02	0.68
gene3	20.53	18.93	18.79	18.51	18.00	18.26
gene4	55.47	52.71	50.39	54.66	49.15	44.68
gene5	7.28	8.09	8.57	7.82	8.29	9.38
gene6	14.65	13.88	13.48	13.98	14.72	12.47
gene7	16.41	13.80	14.99	17.20	14.39	13.50
gene8	6.17	6.79	7.20	6.70	8.42	7.26
gene9	25.83	24.24	25.63	27.09	22.18	23.09
gene10	38.04	30.39	35.53	37.42	28.72	27.28
gene11	195.06	179.88	178.18	208.25	179.01	155.15
gene12	32.82	32.04	31.84	33.62	31.06	29.46
gene13	18.41	16.75	16.72	17.33	16.32	16.87
gene14	24.00	21.05	22.68	22.72	22.08	22.45
			.....			

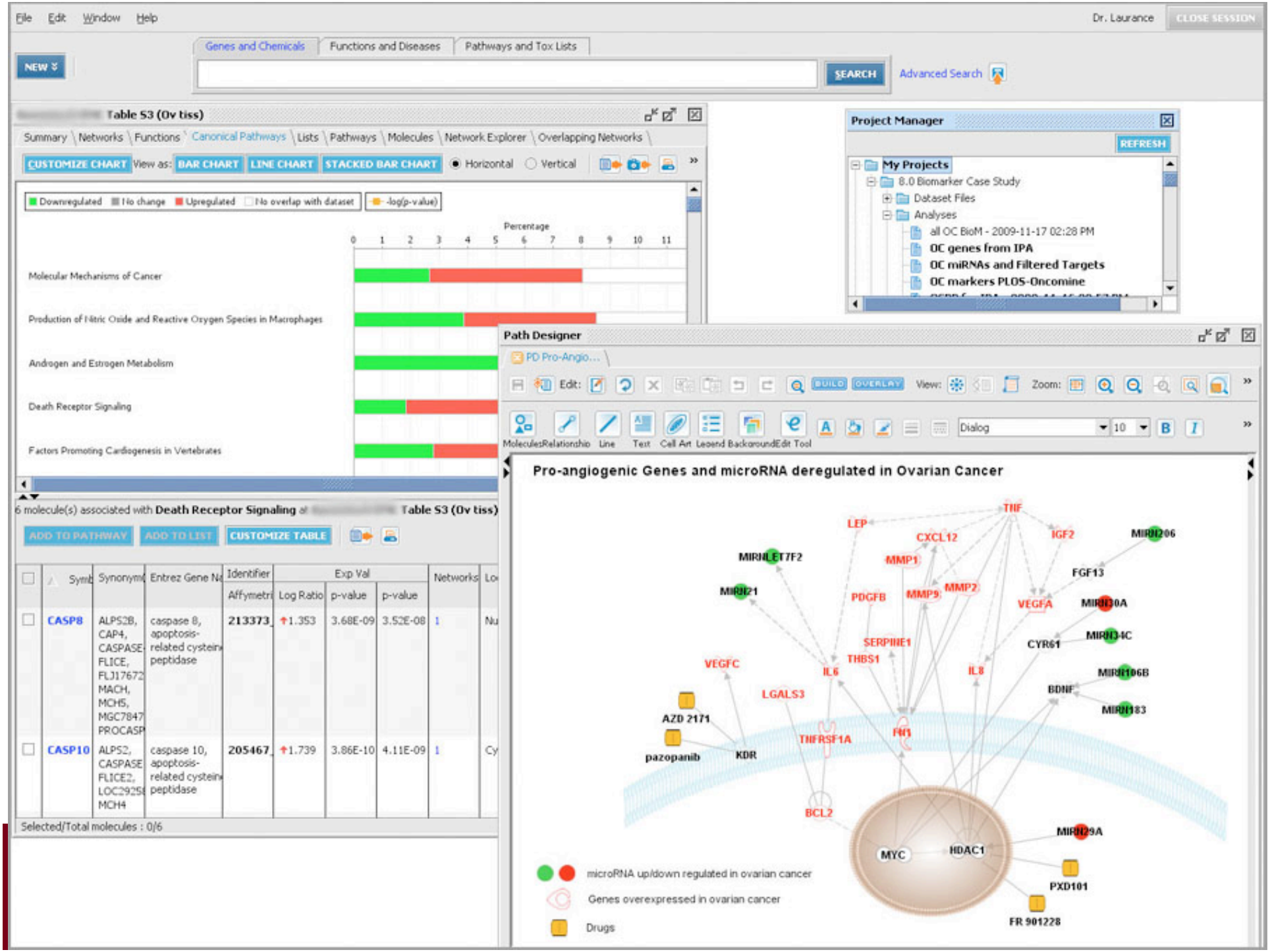
Group 1 (A,V,O)

Group 2 (E,I,U)



Exploring data at system-level: Ingenuity Pathway analysis

- Using the differentially expressed genes
- Connecting the genes with known knowledge
- Testing for the significance of the identified network
- Check the details at:
  - http://ingenuity.com/products/pathways_analysis.html


File Edit Window Help
Dr. Laurance
CLOSE SESSION
Genes and Chemicals
Functions and Diseases
Pathways and Tox Lists
NEW
SEARCH
Advanced Search
Table S3 (Ov tiss)
Summary Networks Functions Canonical Pathways Lists Pathways Molecules Network Explorer Overlapping Networks
CUSTOMIZE CHART View as: BAR CHART LINE CHART STACKED BAR CHART Horizontal Vertical
Downregulated No change Upregulated No overlap with dataset -log(p-value)
Percentage
0 1 2 3 4 5 6 7 8 9 10 11
Molecular Mechanisms of Cancer
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages
Androgen and Estrogen Metabolism
Death Receptor Signaling
Factors Promoting Cardiogenesis in Vertebrates
6 molecule(s) associated with Death Receptor Signaling
Table S3 (Ov tiss)
ADD TO PATHWAY
ADD TO LIST
CUSTOMIZE TABLE
Synb
Synonym(
Entrez Gene Na
Identifier
Affymetri
Exp Val
Log Ratio
p-value
p-value
Networks
Lo
CASP8
ALPS2B, CAP4, CASPASE-FLICE, FLJ17672 MACH, MCH5, MGC7847 PROCASP
caspase 8, apoptosis-related cystein peptidase
213373_
↑1.353
3.68E-09
3.52E-08
1
Nu
CASP10
ALPS2, CASPASE FLICE2, LOC2925 MCH4
caspase 10, apoptosis-related cystein peptidase
205467_
↑1.739
3.86E-10
4.11E-09
1
Cy
Selected/Total molecules : 0/6
Project Manager
REFRESH
My Projects
8.0 Biomarker Case Study
Dataset Files
Analyses
all OC BioM - 2009-11-17 02:28 PM
OC genes from IPA
OC miRNAs and Filtered Targets
OC markers PLOS-Oncomine
Path Designer
PD Pro-Angio...
Edit:
BUILD
OVERLAY
View:
Zoom:
Molecules
Relationship
Line
Text
Cell Art
Legend
Background
Edit Tool
Dialog
10
Pro-angiogenic Genes and microRNA deregulated in Ovarian Cancer
LEP
TNF
CXCL12
IGF2
MIRN206
MIRNLET7F2
MMP1
FGF13
MIRN21
PDGFB
MMP9
MMP2
VEGFA
MIRN130A
SERPINE1
CYR61
MIRN34C
VEGFC
THBS1
IL6
IL8
MIRN106B
LGALS3
BDNF
MIRN183
AZD 2171
TNFRSF1A
FN1
pazopanib
KDR
BCL2
MIRN29A
MYC
HDAC1
PXD101
FR 901228
microRNA up/down regulated in ovarian cancer
Genes overexpressed in ovarian cancer
Drugs

Discussion and Questions?

- Get Support at MSI:
  - Email: help@msi.umn.edu
  - General Questions:
    - Subject line: “RISS:…”
  - Galaxy Questions:
    - Subject line: “Galaxy:…”