



**University of
Zurich**^{UZH}

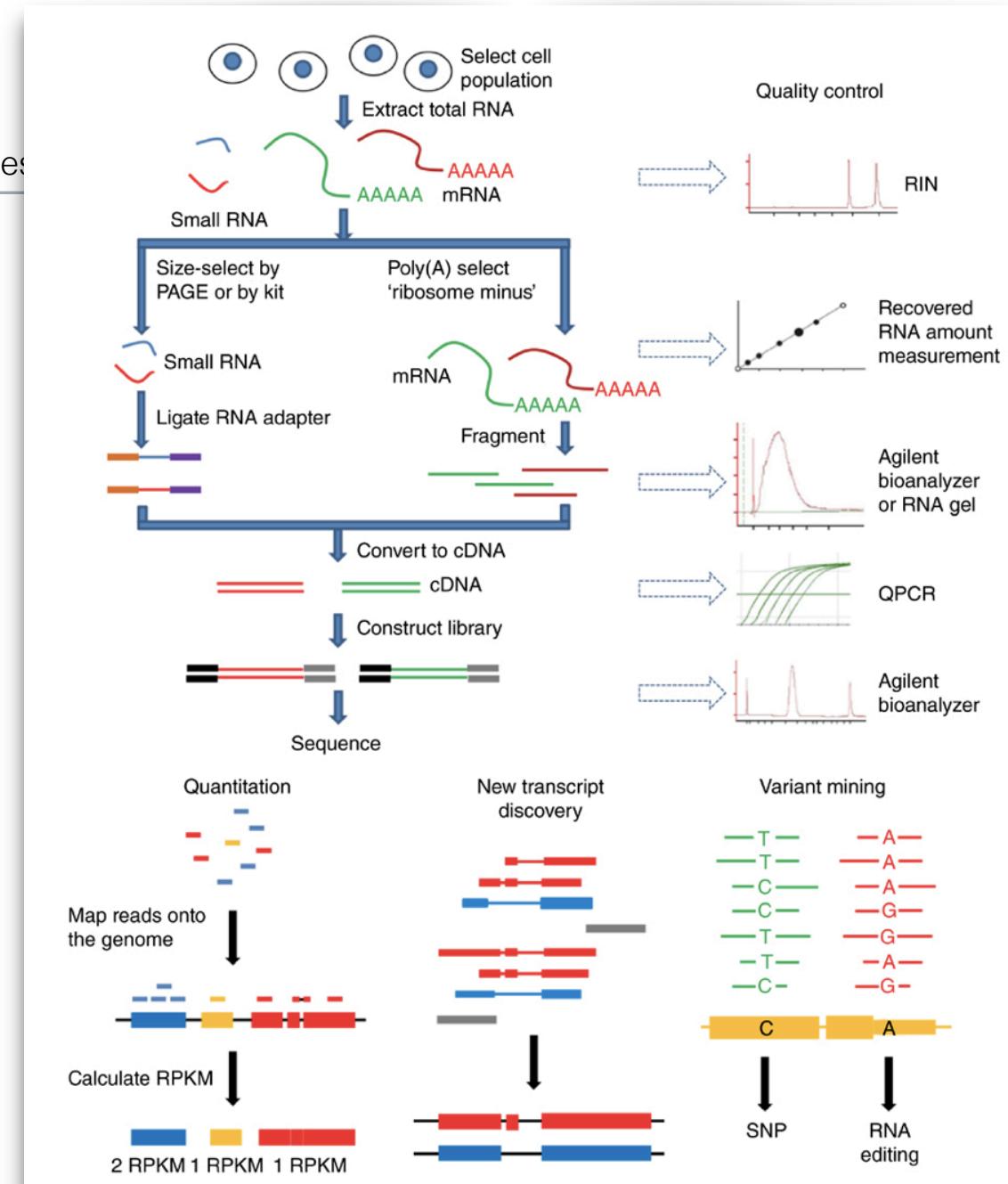
Statistical Bioinformatics // Institute of Molecular Life Sciences

Count data analysis



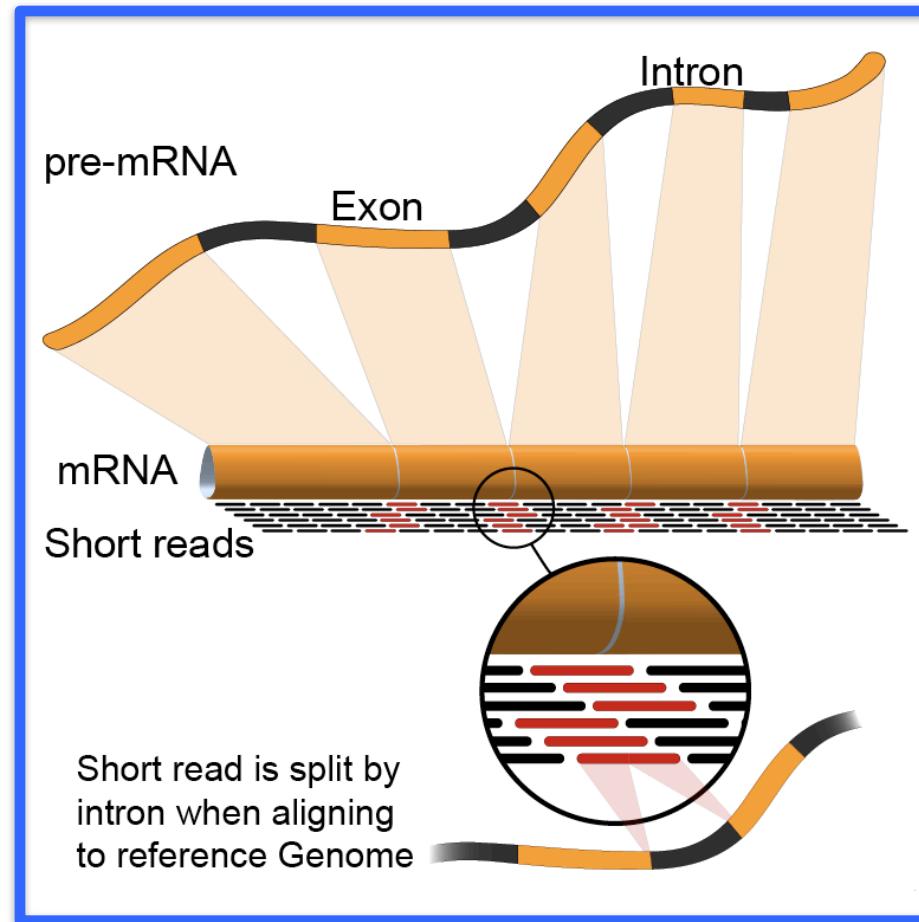
RNA-seq differential expression analyses

1. Map the reads to reference sequences
2. “Count” reads that map to genes (quantify)
3. Compute DE Statistics

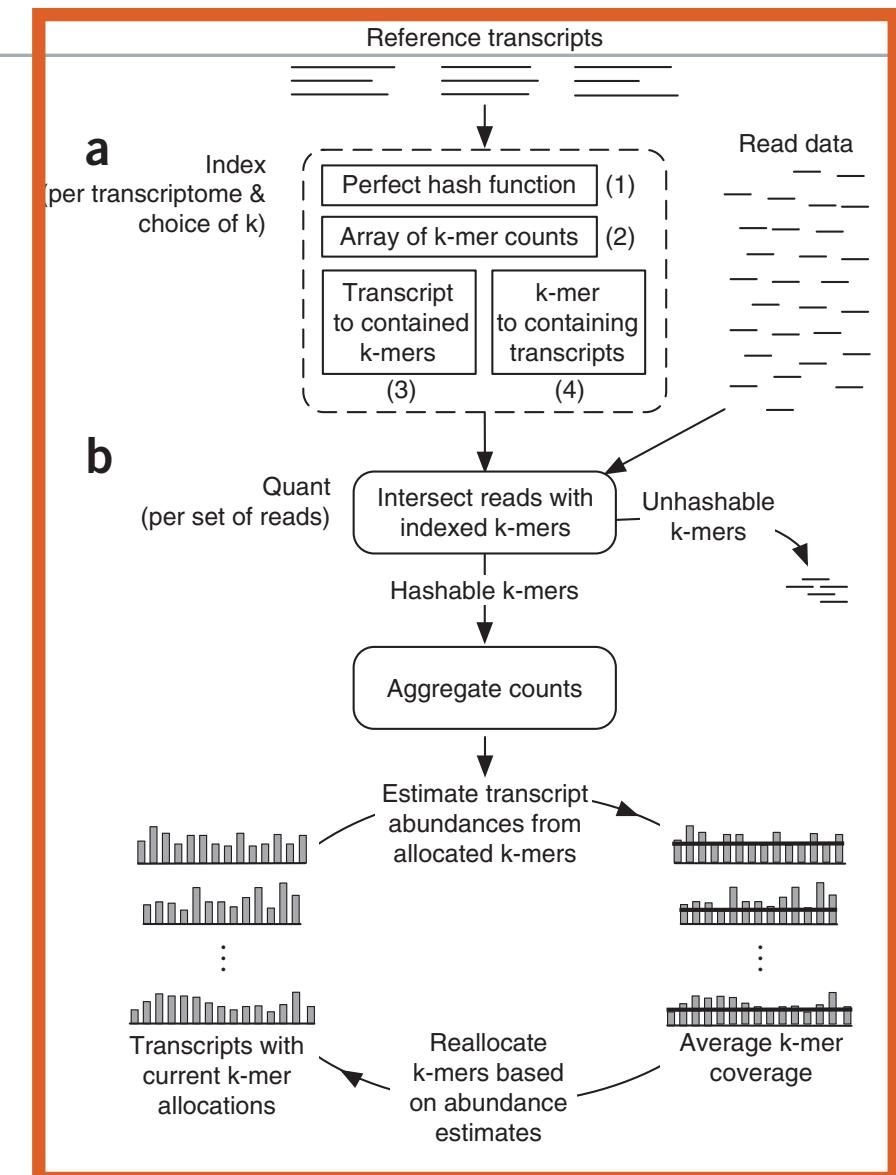


Alignment versus quasi-alignment

Statistical Bioinformatics // Institute of Molecular Life Sciences



<https://en.wikipedia.org/wiki/RNA-Seq>

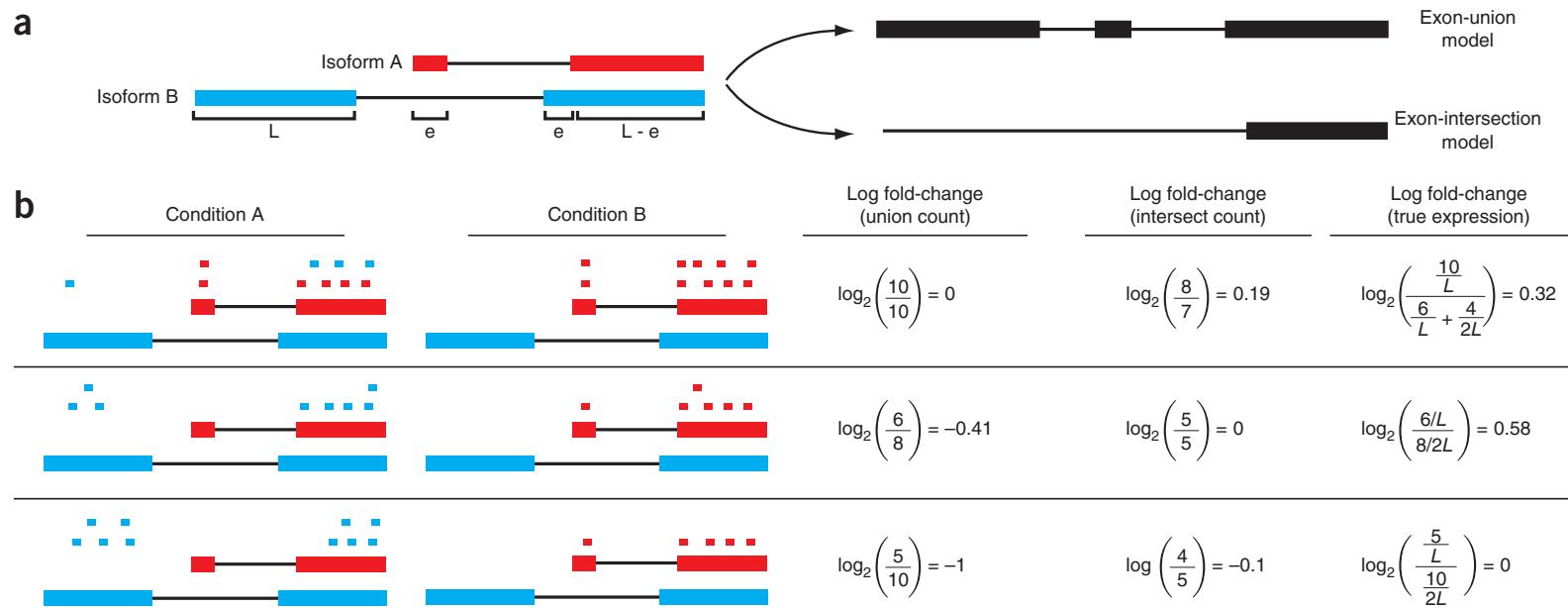


sailfish (Patro et al. 2014)



Caveat: simple gene-level counting not perfect, but good first approximation

Trapnell et al. 2013 Nat Biotech



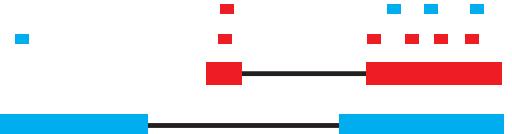
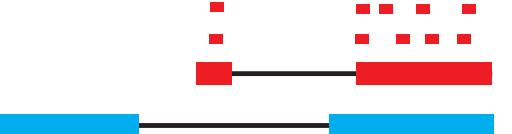
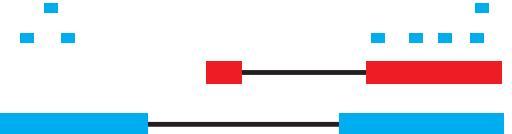
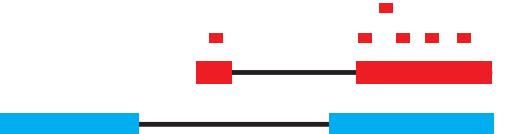
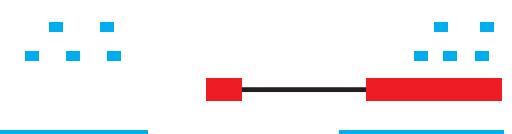
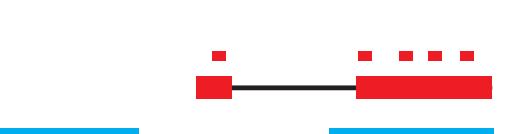
Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene

Mar González-Porta¹, Adam Frankish², Johan Rung¹, Jennifer Harrow² and Alvis Brazma^{1*}

Counting/Quantification

union counters → simple sum of all reads
transcript counters → sum of length-normalized reads
(often unknown which reads map to which transcript → portioning)

b

	Condition A	Condition B	Log fold-change (union count)	Log fold-change (intersect count)	Log fold-change (true expression)
			$\log_2\left(\frac{10}{10}\right) = 0$	$\log_2\left(\frac{8}{7}\right) = 0.19$	$\log_2\left(\frac{\frac{10}{L}}{\frac{6}{L} + \frac{4}{2L}}\right) = 0.32$
			$\log_2\left(\frac{6}{8}\right) = -0.41$	$\log_2\left(\frac{5}{5}\right) = 0$	$\log_2\left(\frac{6/L}{8/2L}\right) = 0.58$
			$\log_2\left(\frac{5}{10}\right) = -1$	$\log\left(\frac{4}{5}\right) = -0.1$	$\log_2\left(\frac{\frac{5}{L}}{\frac{10}{2L}}\right) = 0$



How do all these methods of counting affect DE analyses?

You've been doing your RNA-Seq all wrong

Posted by: RNA-Seq Blog in Expression and Quantification November 12, 2015 13,162 Views

In recent years, RNA-seq is emerging as a powerful technology in estimation of gene and/or transcript expression, and RPKM (Reads Per Kilobase per Million reads) is widely used to represent the relative abundance of mRNAs for a gene. In general, the methods for gene quantification can be largely divided into two categories: transcript-based approach and 'union exon'-based approach. Transcript-based approach is intrinsically more difficult because different isoforms of the gene typically have a high proportion of genomic overlap. On the other hand, 'union exon'-based approach method is much simpler and thus widely used in RNA-seq gene quantification. Biologically, a gene is expressed in one or more transcript isoforms. Therefore, transcript-based approach is logically more meaningful than 'union exon'-based approach. Despite the fact that gene quantification is a fundamental task in most RNA-seq studies, however, it remains unclear whether 'union exon'-based approach for RNA-seq gene quantification is a good practice or not.

Researchers at [Pfizer Worldwide Research & Development](#) carried out a side-by-side comparison of 'union exon'-based approach and transcript-based method in RNA-seq gene quantification. It was found that the

F1000Research

F1000Research 2016, 4:1521 Last updated: 05 APR 2016



METHOD ARTICLE

REVISED Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]

Charlotte Soneson^{1,2}, Michael I. Love^{3,4}, Mark D. Robinson^{1,2}

¹Institute for Molecular Life Sciences, University of Zurich, Zurich, 8057, Switzerland

²SIB Swiss Institute of Bioinformatics, University of Zurich, Zurich, 8057, Switzerland

³Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, 02210, USA

⁴Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA, 02115, USA

v2 First published: 30 Dec 2015, 4:1521 (doi: [10.12688/f1000research.7563.1](https://doi.org/10.12688/f1000research.7563.1))

Latest published: 29 Feb 2016, 4:1521 (doi: [10.12688/f1000research.7563.2](https://doi.org/10.12688/f1000research.7563.2))

Open Peer Review





Differential expression: why not use methods developed for microarrays?

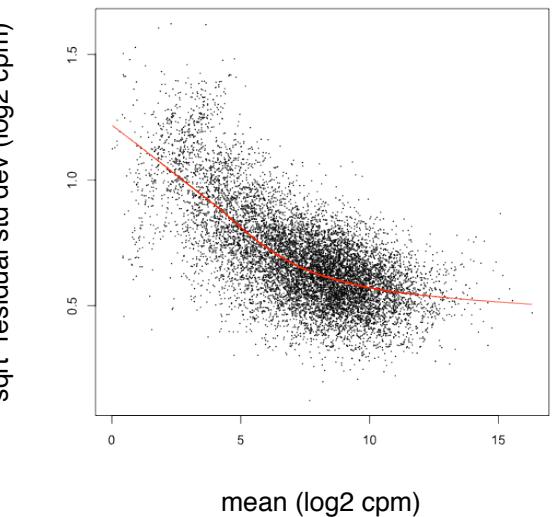
Count data is discrete, not continuous.

Methods designed for microarrays are not directly applicable and suboptimal (**more on this later**)

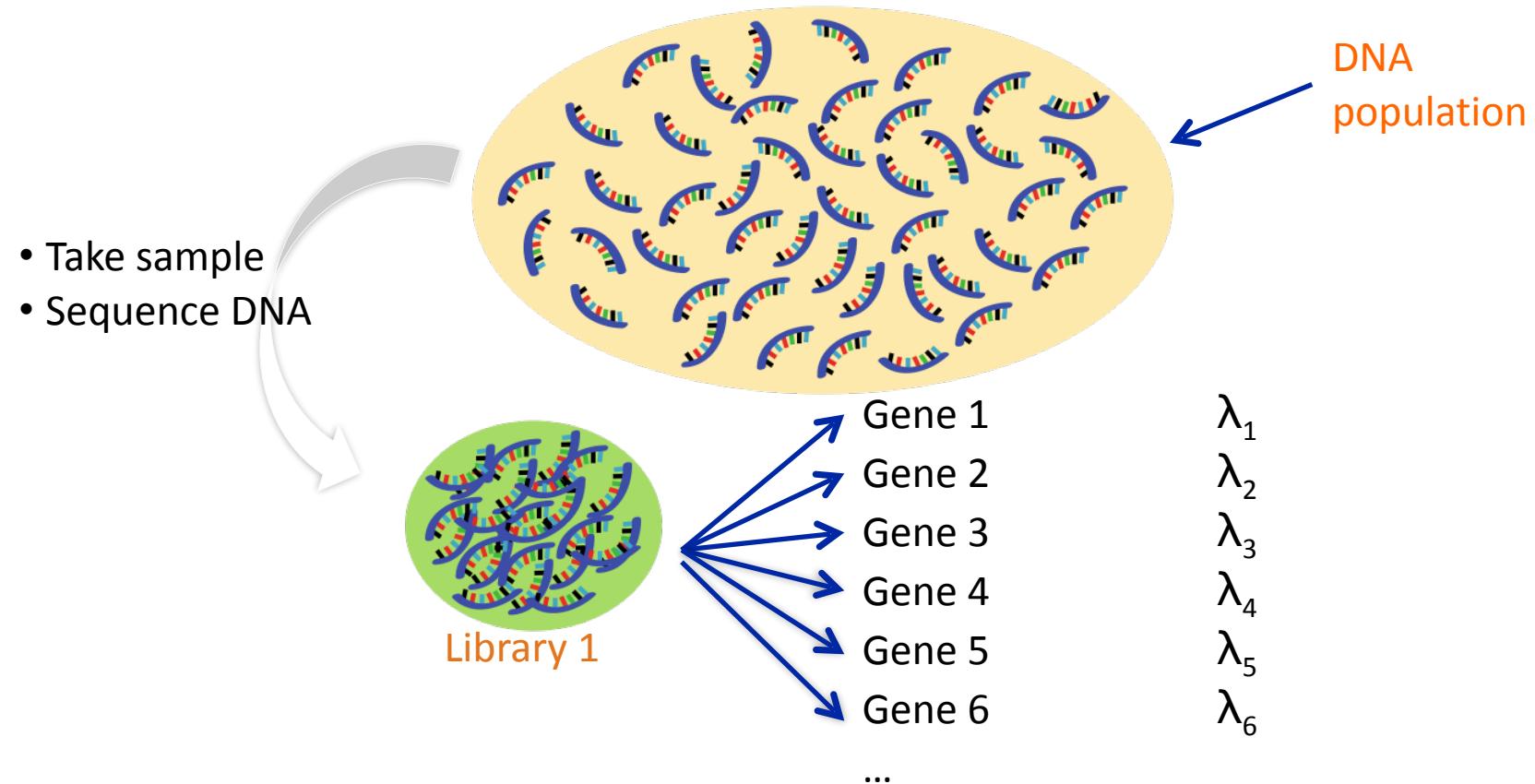
Two options:

Transform count data and apply standard methodology

Analyze using models for count data



Sampling reads from population of DNA fragments is multinomial





For a single gene, it's a coin toss, i.e. Binomial



$$Y_i \sim \text{Binomial}(M, \lambda_i)$$

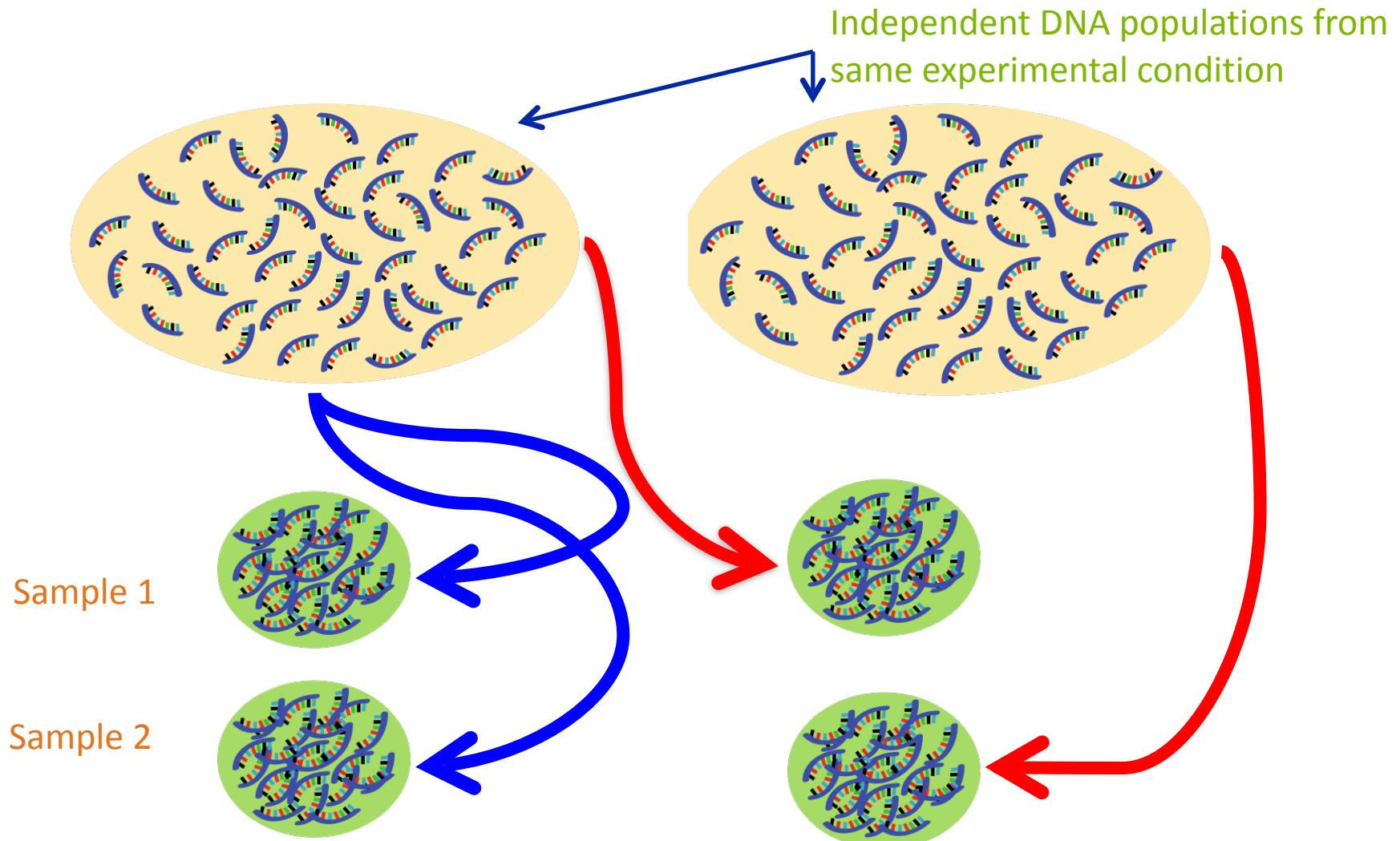
Y_i - observed number of reads for gene i

M - total number of sequences

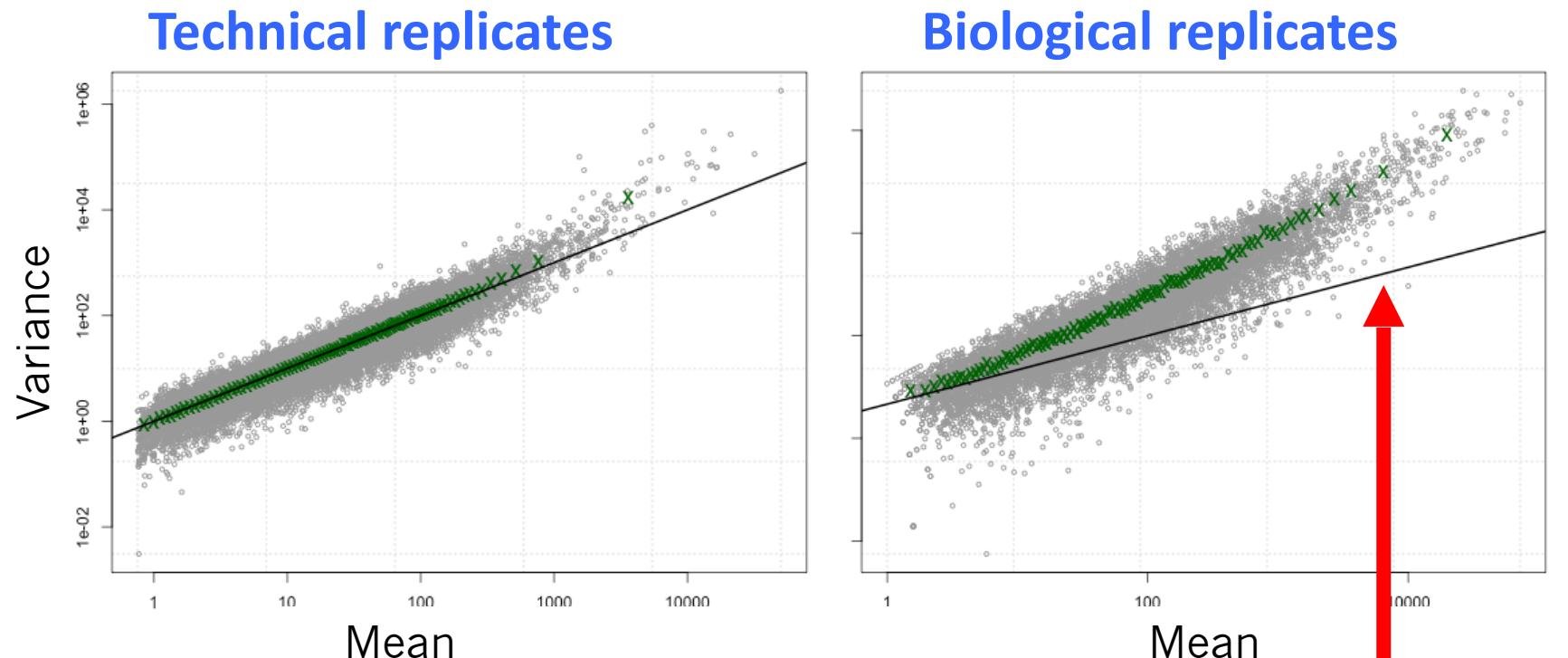
λ_i - proportion

Large M , small $\lambda_i \rightarrow$ approximated well by Poisson($\mu_i = M \cdot \lambda_i$)

Technical replication versus biological replication



Mean-Variance plots: What we see in real data



Data from Marioni et al. *Genome Research* 2008

Data from Parikh et al.
Genome Biology 2010

mean=variance
(Poisson assumption)



Count data modeling assumptions

Poisson adequately describes technical variation

$$Y_i \sim \text{Pois}(M * \lambda_i)$$

$$\text{mean}(Y_i) = \text{variance}(Y_i) = M * \lambda_i$$

Negative binomial (gamma-Poisson) model is a natural extension that allows **biological** variability:

$$Y_i \sim \text{NB}(\mu_i = M * \lambda_i, \phi_i)$$

Same mean, variance is quadratic in the mean:

$$\text{variance}(Y_i) = \mu_i (1 + \mu_i \phi_i)$$

M = library size

λ_i = relative contribution of gene i



Similar interpretation

$$Y_i \sim NB(\mu_i = N_i * \lambda_i, \phi_i)$$

$$E(y_{gi}) = \mu_{gi} = N_i \pi_{gi}.$$

(Coefficient of variation = standard deviation/mean)

$$\text{var}(y_{gi}) = E_\pi[\text{var}(y|\pi)] + \text{var}_\pi[E(y|\pi)] = \mu_{gi} + \phi_g \mu_{gi}^2.$$

Dividing both sides by μ_{gi}^2 gives

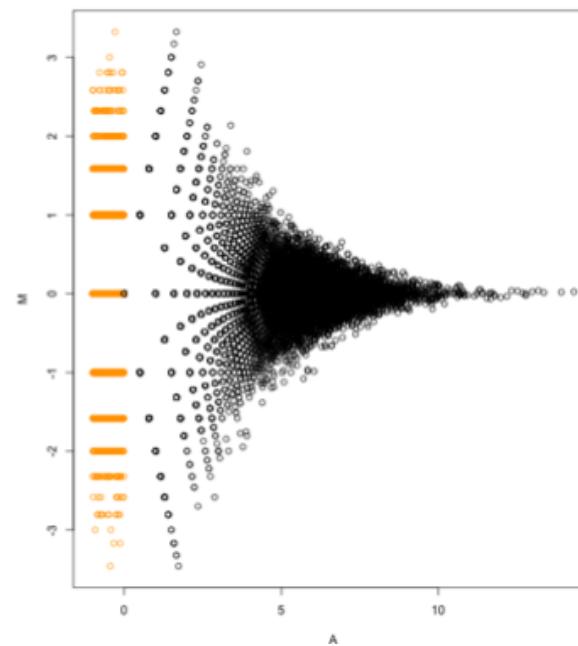
$$CV^2(y_{gi}) = 1/\mu_{gi} + \phi_g.$$



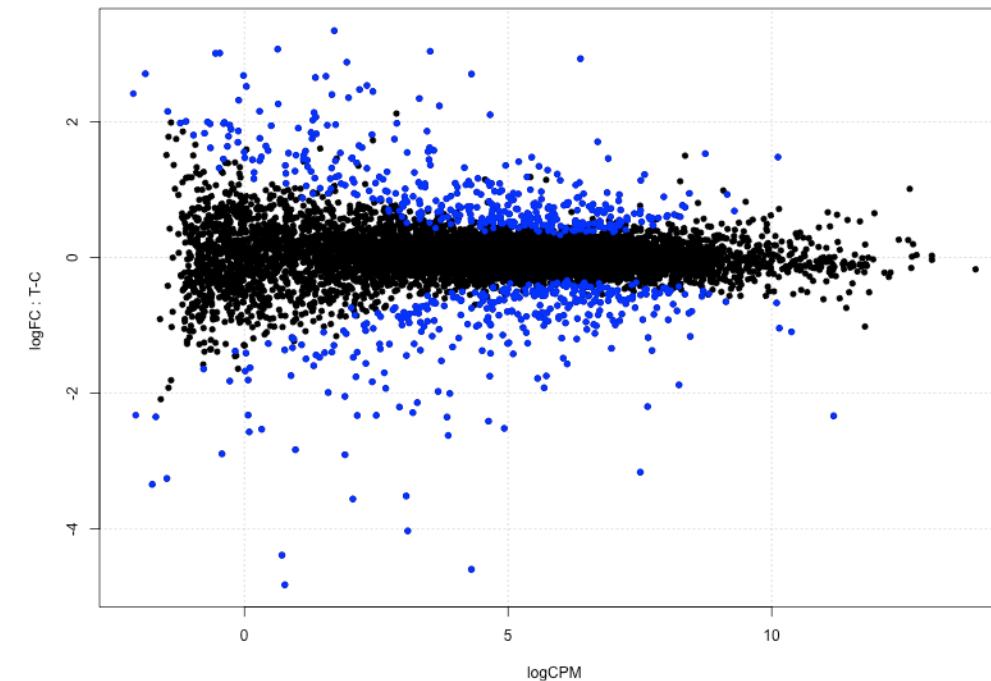
$$CV^2(\gamma_{gi}) = 1/\mu_{gi} + \phi_g.$$

A confirmation of what the theory states

Technical replicates
(~Poisson)



Biological replicates





Differential expression, small sample inference —> **with counts**

- Table of data (e.g., microarray gene expression data with replicates of each of condition A, condition B)
 - rows = features (e.g., genes), columns = experimental units (samples)
- Most common problem in statistical bioinformatics: want to infer whether there is a change in the response
—> a statistical test for each row of the table.

What test might you use? Why is this hard? What issues arise? How much statistical power is there [1] ?

> head(y)	group0	group0	group0	group1	group1	group1
gene1	-0.1874854	0.2584037	-0.05550717	-0.4617966	-0.3563024	-0.03271432
gene2	-3.5418798	-2.4540999	0.11750996	-4.3270442	-5.3462622	-5.54049106
gene3	-0.1226303	0.9354707	-1.10537767	-0.1037990	0.5221678	-1.72360854
gene4	-2.3394536	-0.3495697	-3.47742610	-3.2287093	6.1376670	-2.23871974
gene5	-3.7978820	1.4545702	-7.14796503	-4.0500796	4.7235714	10.00033769
gene6	1.4627078	-0.3096070	-0.26230124	-0.7903434	0.8398769	-0.96822312



What was successful with microarray data: classical/moderated/ shrunken t-tests

$$t_g = \frac{\bar{y}_{\text{mu}} - \bar{y}_{\text{wt}}}{s_g c}$$

Feature-specific

$$\tilde{t}_g = \frac{\bar{y}_{\text{mu}} - \bar{y}_{\text{wt}}}{\tilde{s}_g u}$$

Moderated

$$t_{g,\text{pooled}} = \frac{\bar{y}_{\text{mu}} - \bar{y}_{\text{wt}}}{s_0 c}$$

Common



Let's try the same strategy with counts

At one extreme, assume all genes have same dispersion (too strong)

At other extreme, estimate dispersion separately/independently for each gene (poor estimates)

Shrink individual estimates toward common/trend (how?)

No hierarchical model (e.g. limma) to do this —> **approximations,
weighted likelihood**

No t-distribution theory to formulate statistical tests.



Count data modeling assumptions

Poisson adequately describes technical variation

$$Y_i \sim \text{Pois}(M * \lambda_i)$$

$$\text{mean}(Y_i) = \text{variance}(Y_i) = M * \lambda_i$$

Negative binomial (gamma-Poisson) model is a natural extension that allows **biological** variability:

$$Y_i \sim \text{NB}(\mu_i = M * \lambda_i, \phi_i)$$

Same mean, variance is quadratic in the mean:

$$\text{variance}(Y_i) = \mu_i (1 + \mu_i \phi_i)$$

M = library size

λ_i = relative contribution of gene i

Second challenge: Moderate dispersion estimate

Weighted likelihood -- individual log-likelihood plus a weighted version of the common log-likelihood:

$$WL(\phi_g) = l_g(\phi_g) + \alpha l_C(\phi_g)$$

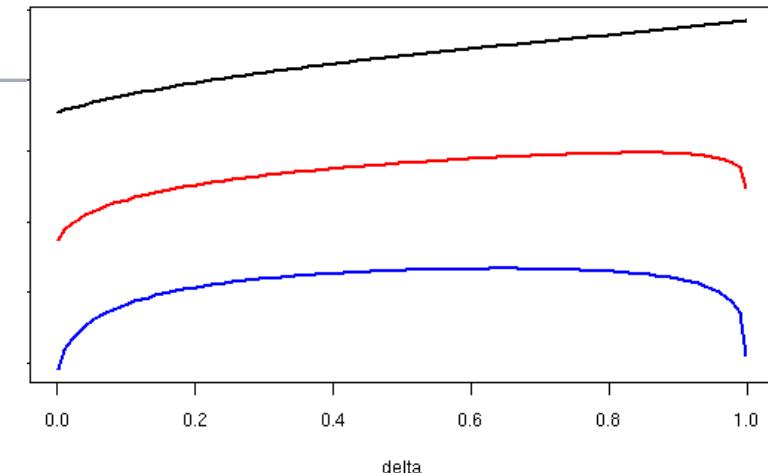
l_g - quantile-adjusted conditional likelihood

Black: single tag

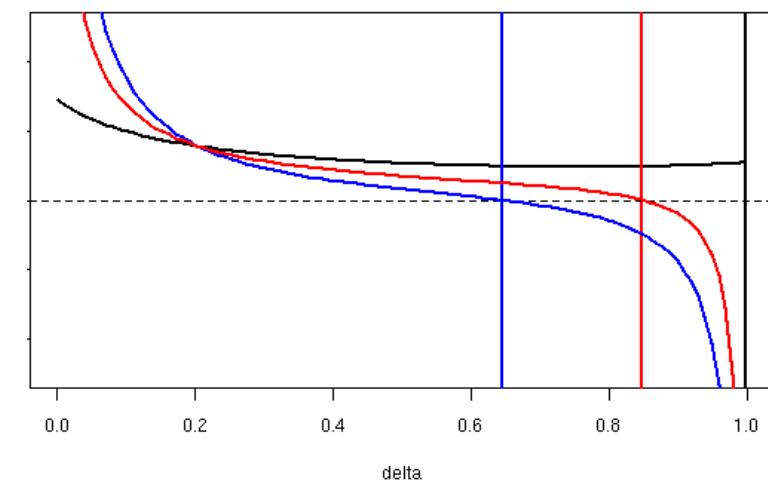
Blue: common dispersion

Red: Linear combination of the two

Log-Likelihood



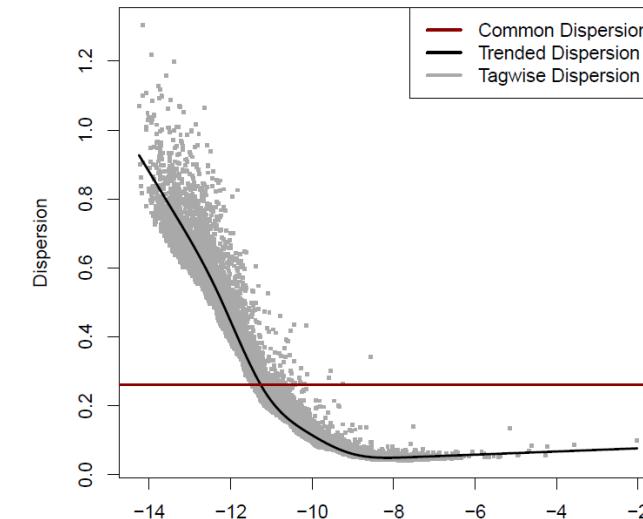
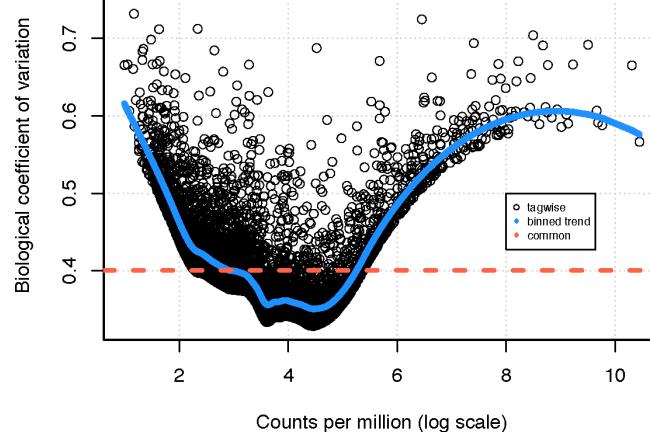
Score (1st derivative of LL)



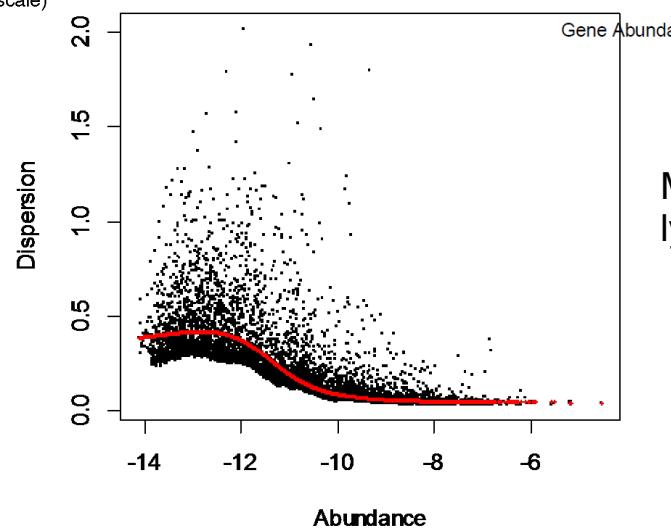
$$\delta = \frac{\phi}{\phi+1}$$

Dispersion varies with mean: moderate dispersion towards **trend**

Data:
Tuch et al.,
2008



Mouse hemopoietic
stem cells



Mouse
lymphomas

Advantage: genes are allowed to have their own variance.



Nature Reviews Genetics | AOP, published online 18 November 2008; doi:10.1038/nrg2484

INNOVATION

RNA-Seq: a revolutionary tool for transcriptomics

Zhong Wang, Mark Gerstein and Michael Snyder

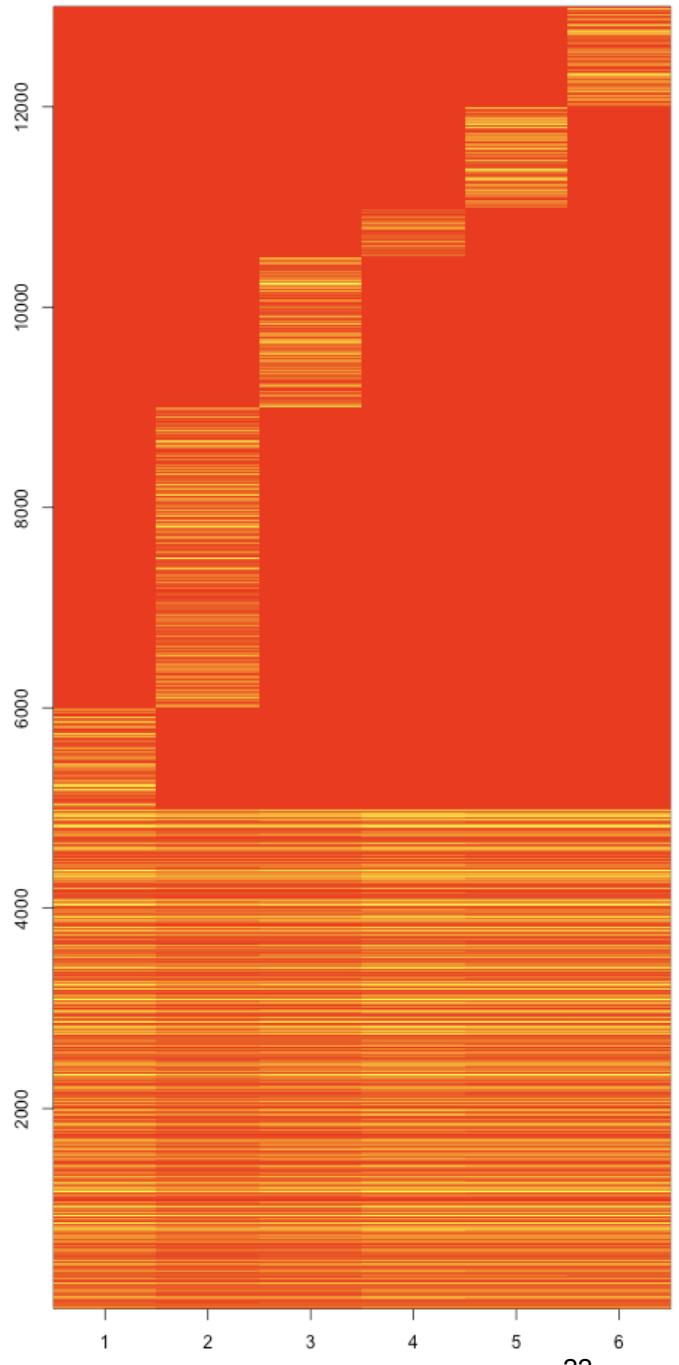
One particularly powerful advantage of RNA-Seq is that it can capture transcriptome dynamics across different tissues or conditions without sophisticated normalization of data sets^{19,20,22}.



“Composition” or “Diversity” can affect read depth

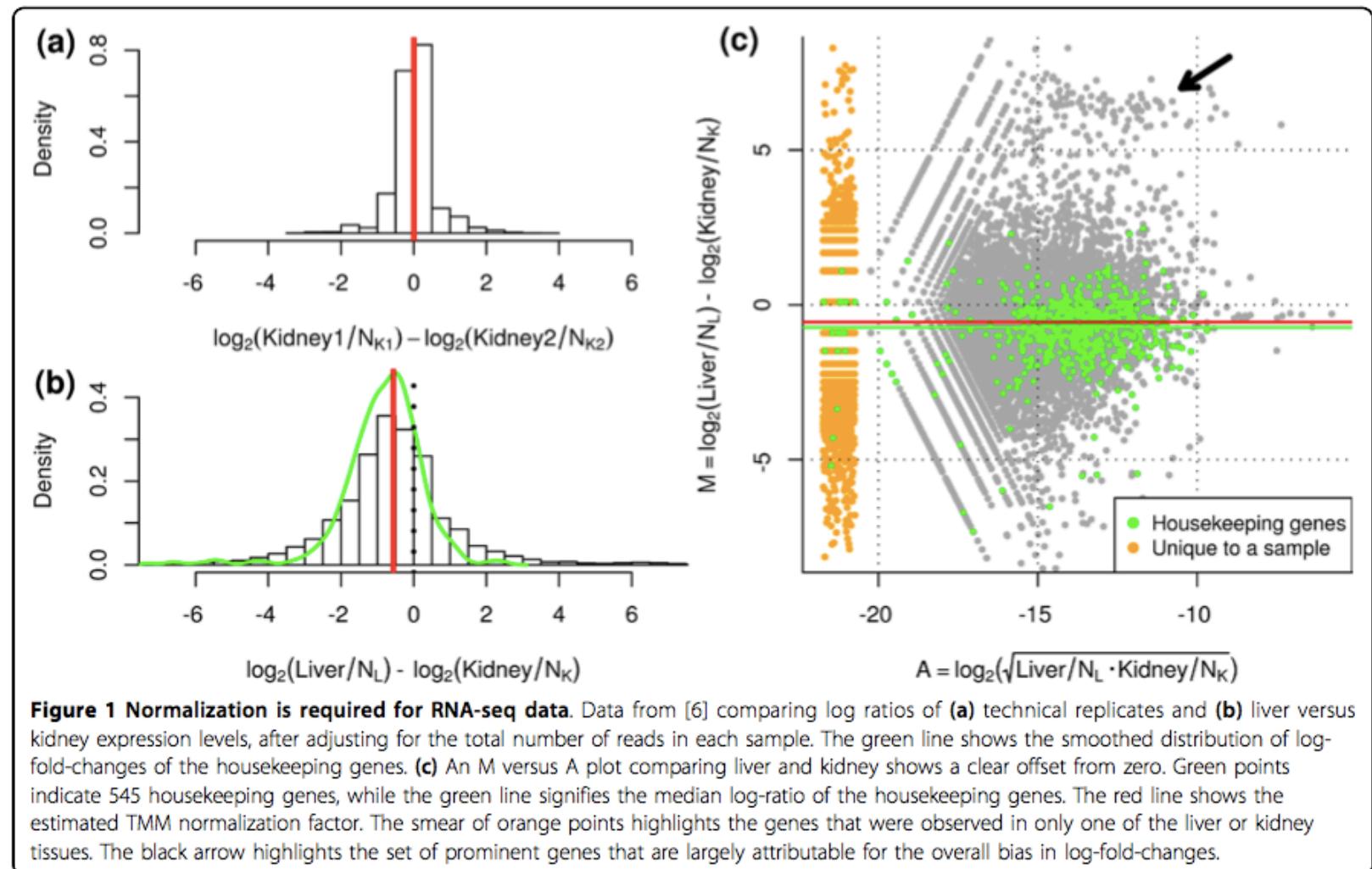
- Hypothetical example: Sequence 6 libraries to the same depth, with varying levels of unique-to-sample counts
- Read depth is affected not only by expression (and length), but also expression levels of other genes
- Composition can induce (sometimes significant) differences in counts

Red=low, goldenyellow=high





Kidney and Liver RNA have very different composition





Use scaling factor (“offset”) in statistical model

Assumption: core set of genes/loci that do not change in expression.

Our Pick a reference sample, compute a weighted trimmed mean of M-values (TMM) to reference

Adjustment to statistical analysis:

- Use “effective” library size (edgeR)
- Use additional offset (GLM)

Note: count data is not modified



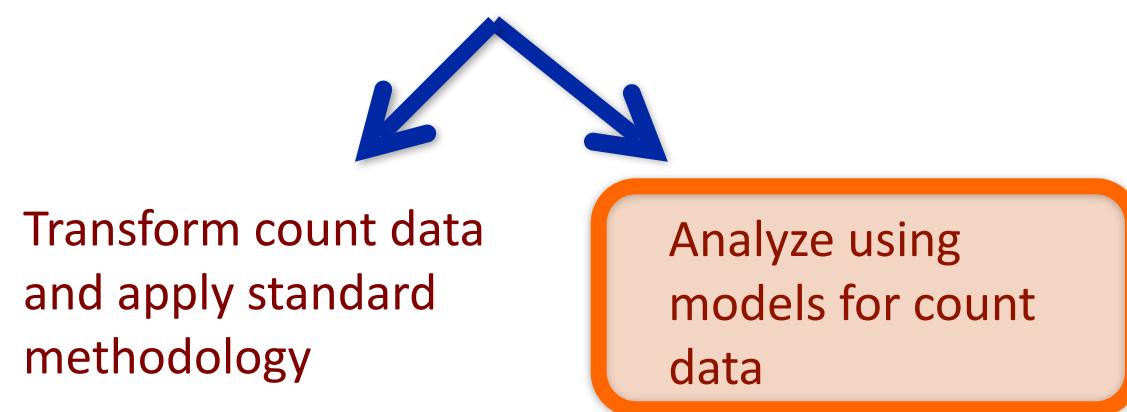
Differential expression: why not use methods developed for microarrays?

Count data is discrete, not continuous.

Methods designed for microarrays are not directly applicable and suboptimal

Transforming count data with logs, with some special treatment, can give very good results

Two options:





What does transformation do to M-V relationship?

For Poisson data, square-root should stabilize

Logarithm is too strong – variance decreases to asymptote (Neg Bin) or 0 (Poisson)

How to pick? Doesn't matter —> voom

voom: mean-variance modeling at the observational level

voom

package:limma

R Documentation

Transform RNA-Seq Data Ready for Linear Modelling

Description:

Transform count data to log2-counts per million, estimate the mean-variance relationship and use this to compute appropriate observational-level weights. The data are then ready for linear modeling.



Model log counts per million

log counts per million:

$$z_{gi} = \log_2 \left(1e6 \frac{\text{count}_{gi} + 0.5}{\text{libsize}_{gi} + 1.0} \right) = \log_2 \left(1e6 \frac{y_{gi} + 0.5}{M_{gi} + 1.0} \right)$$

normalize libsize in advance or normalize z_{gi} as for microarrays.

Linear modelling:

$$E(z_{gi}) = \mu_{gi} = x_i^T \beta_g$$

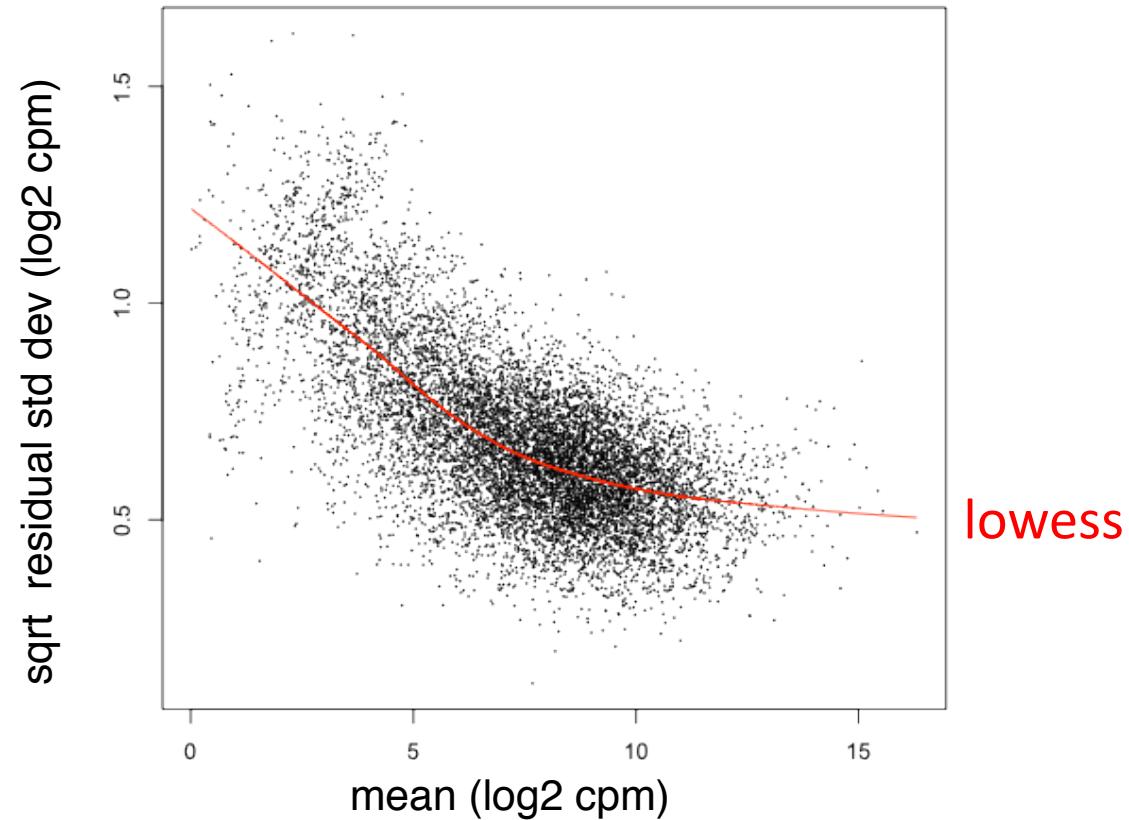
$$\text{var}(z_{gi}) = s(\mu_{gi}) \sigma_g^2$$



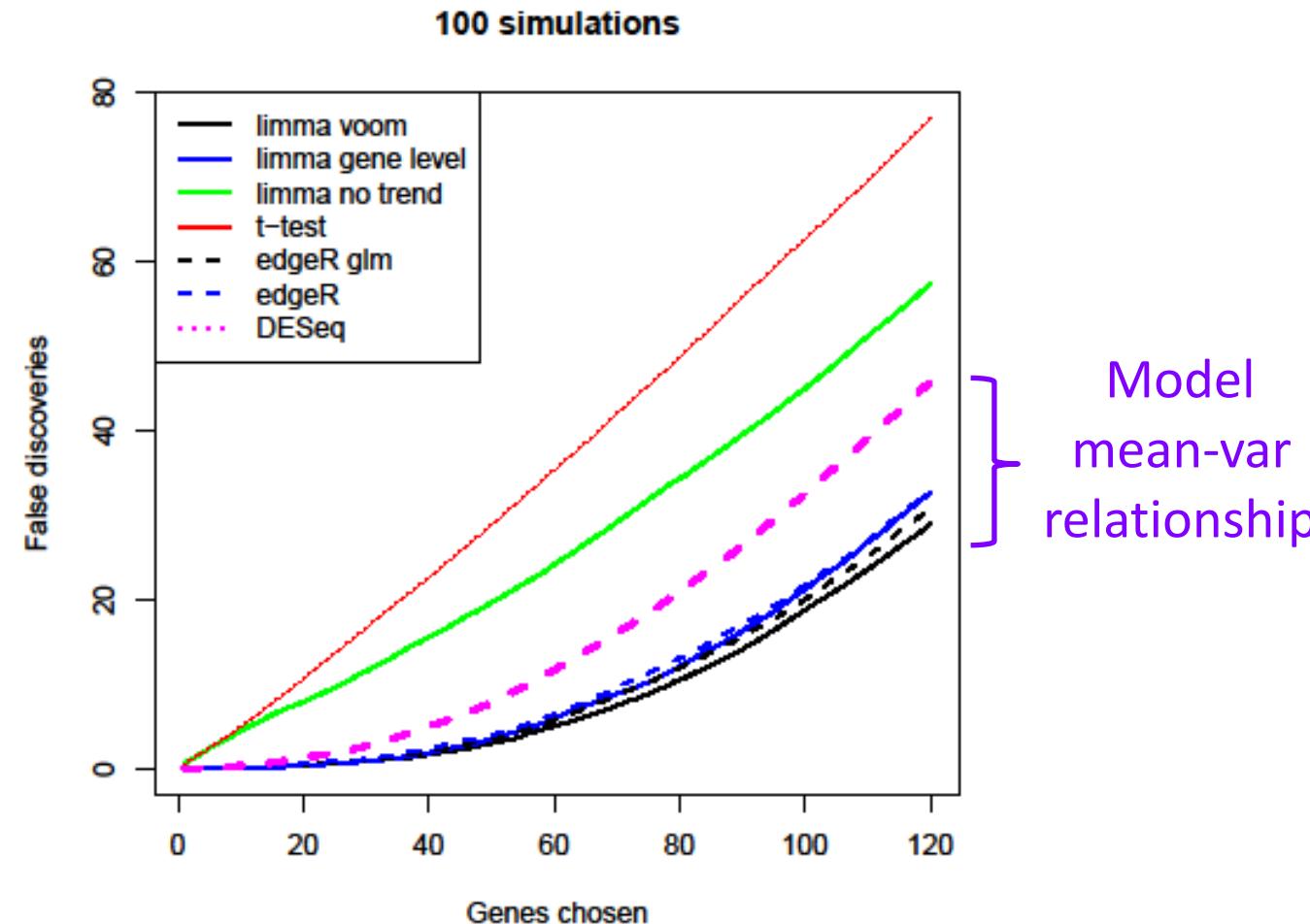
Smooth function of mean



voom fits a lowess trend to the mean-variance relationship ...



—> Use weights ($1/\text{var}$) in limma analysis .. i.e., **heteroscedastic regression**





Linear Models (microarray setting)

In general, need to specify:

- Dependent variable
- Explanatory variables (experimental design, covariates, etc.)

More generally:

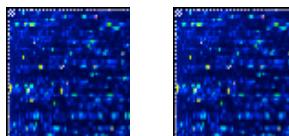
$$y = X\beta + \epsilon$$

vector of observed data design matrix Vector of parameters to estimate

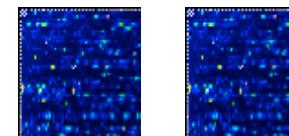
The diagram illustrates the components of the linear model equation $y = X\beta + \epsilon$. A blue arrow points to the variable y , indicating it is the vector of observed data. A red arrow points to the term $X\beta$, indicating it is the design matrix. An orange arrow points to the term ϵ , indicating it is the vector of parameters to estimate.

Analysis of Variance → Linear model

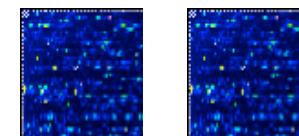
WT x 2



Cond A x 2



Cond B x 2



$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

α_1 = wt log-expression

α_2 = Cond A - wt

α_3 = Cond B - wt

$$E[y_1] = E[y_2] = \alpha_1$$

$$E[y_3] = E[y_4] = \alpha_1 + \alpha_2$$

$$E[y_5] = E[y_6] = \alpha_1 + \alpha_3$$

Applications: paired designs, multi-factor designs, interactions

→ This particular model only valid for continuous response



Generalized linear models: a more general framework

Gaussian (normal) distributed response —> various other (common) types.

Three components:

1. Probability distribution of response (in exponential family)
2. Linear predictor (covariates; design matrix)
3. Link function (link mean to linear predictor)



Link function and linear predictor

$$E(Y_i) = \mu_i$$

$$g(\mu_i) = \eta_i$$

Link function

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

Linear predictor (covariates)

$$\text{var}(Y_i) = \phi V(\mu)$$

Provides a way to link the mean of response to a linear predictor.

Data is not transformed.

Variance is a function of mean.



Common distributions, “Canonical” link functions

Common distributions with typical uses and canonical link functions

Distribution	Support of distribution	Typical uses	Link name	Link function	Mean function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\beta = \mu$	$\mu = \mathbf{X}\beta$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Inverse	$\mathbf{X}\beta = \mu^{-1}$	$\mu = (\mathbf{X}\beta)^{-1}$
Gamma					
Inverse Gaussian	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\beta = \mu^{-2}$	$\mu = (\mathbf{X}\beta)^{-1/2}$
Poisson	integer: $[0, +\infty)$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\beta = \ln(\mu)$	$\mu = \exp(\mathbf{X}\beta)$
Bernoulli	integer: $[0, 1]$	outcome of single yes/no occurrence			
Binomial	integer: $[0, N]$	count of # of "yes" occurrences out of N yes/no occurrences			
Categorical	integer: $[0, K]$	outcome of single K-way occurrence	Logit	$\mathbf{X}\beta = \ln\left(\frac{\mu}{1 - \mu}\right)$	$\mu = \frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)} = \frac{1}{1 + \exp(-\mathbf{X}\beta)}$
	K-vector of integer: $[0, 1]$, where exactly one element in the vector has the value 1				
Multinomial	K-vector of integer: $[0, N]$	count of occurrences of different types (1 .. K) out of N total K-way occurrences			

http://en.wikipedia.org/wiki/Generalized_linear_model



Large sample theory – Result 3 (likelihood ratio test)

$$\begin{aligned} D &= -2 \ln \left(\frac{\text{likelihood for null model}}{\text{likelihood for alternative model}} \right) \\ &= -2 \ln(\text{likelihood for null model}) + 2 \ln(\text{likelihood for alternative model}) \end{aligned}$$

http://en.wikipedia.org/wiki/Likelihood-ratio_test

General form (exponential family)

$$-2 \log \lambda = 2 \sum_{i=1}^n \frac{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)}{a_i(\phi)}$$

edgeR::glmLRT()

Again, large sample theory says this is approx. χ^2 with degrees of freedom according to the difference in the number of parameters between null and alternative (assuming they are nested).



Some interesting generalizations of NB modeling for RNA-seq (3)

$$LRT_k = 2(\ell_k(\hat{\mu}_k|\mathbf{y}_k) - \ell_k(\tilde{\mu}_k|\mathbf{y}_k)) \longrightarrow LRT_k \sim \Phi_k \chi_q^2 + O_p(n^{-1/2})$$

$$\hat{\Phi}_k = \frac{2(\ell_k(\mathbf{y}_k|\mathbf{y}_k) - \ell_k(\hat{\mu}_k|\mathbf{y}_k))}{n - p}$$

$$F_{QL} = \frac{LRT_k/q}{\hat{\Phi}_k}$$

Accounting for the uncertainty in
estimating dispersion

`edgeR::glmQLF`

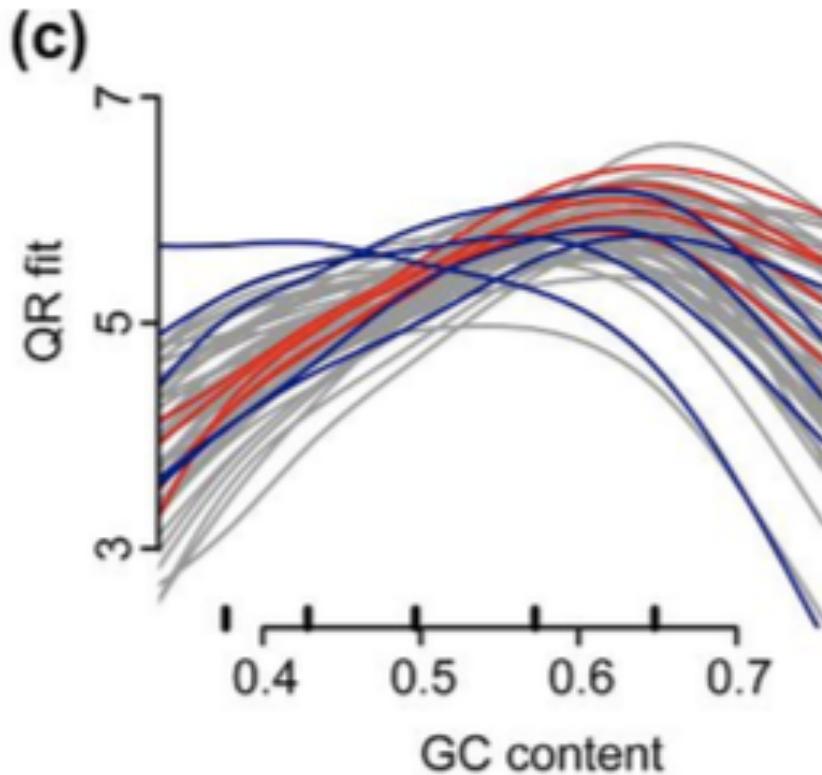
Lund et al., SAGMB 2012; 11(5):8



$$E[Y] = \mu = g^{-1}(\eta) = g^{-1}(X\beta + \xi)$$



Some interesting generalizations of NB modeling for RNA-seq (4)



Integrate sample-specific normalization via offset

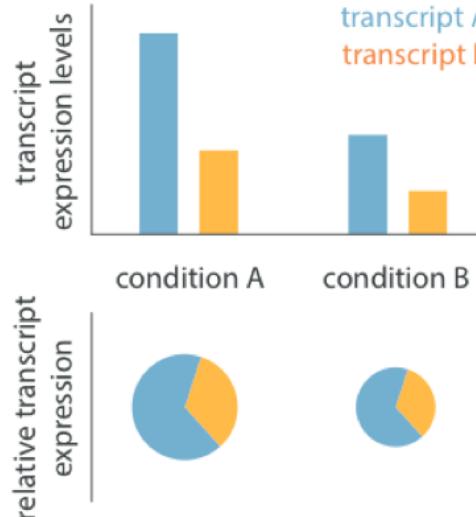
Profiles vary from sample to sample:
GC content
Gene length

DOES NOT change data, use offsets to modify expected mean

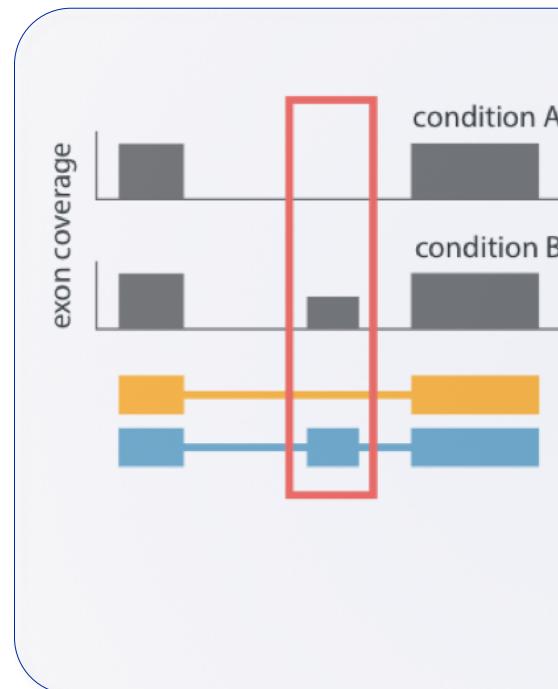
Give a sample (or gene)-specific offset to edgeR/DESeq2

Some terms: DTE, DEU, DTU

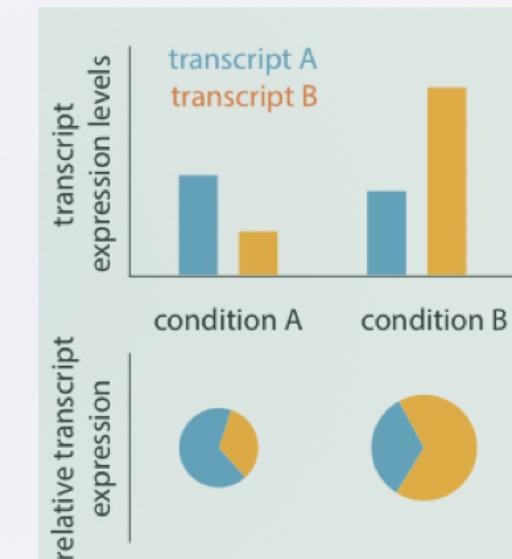
Differential transcript expression (DTE)



Differential exon usage (DEU)



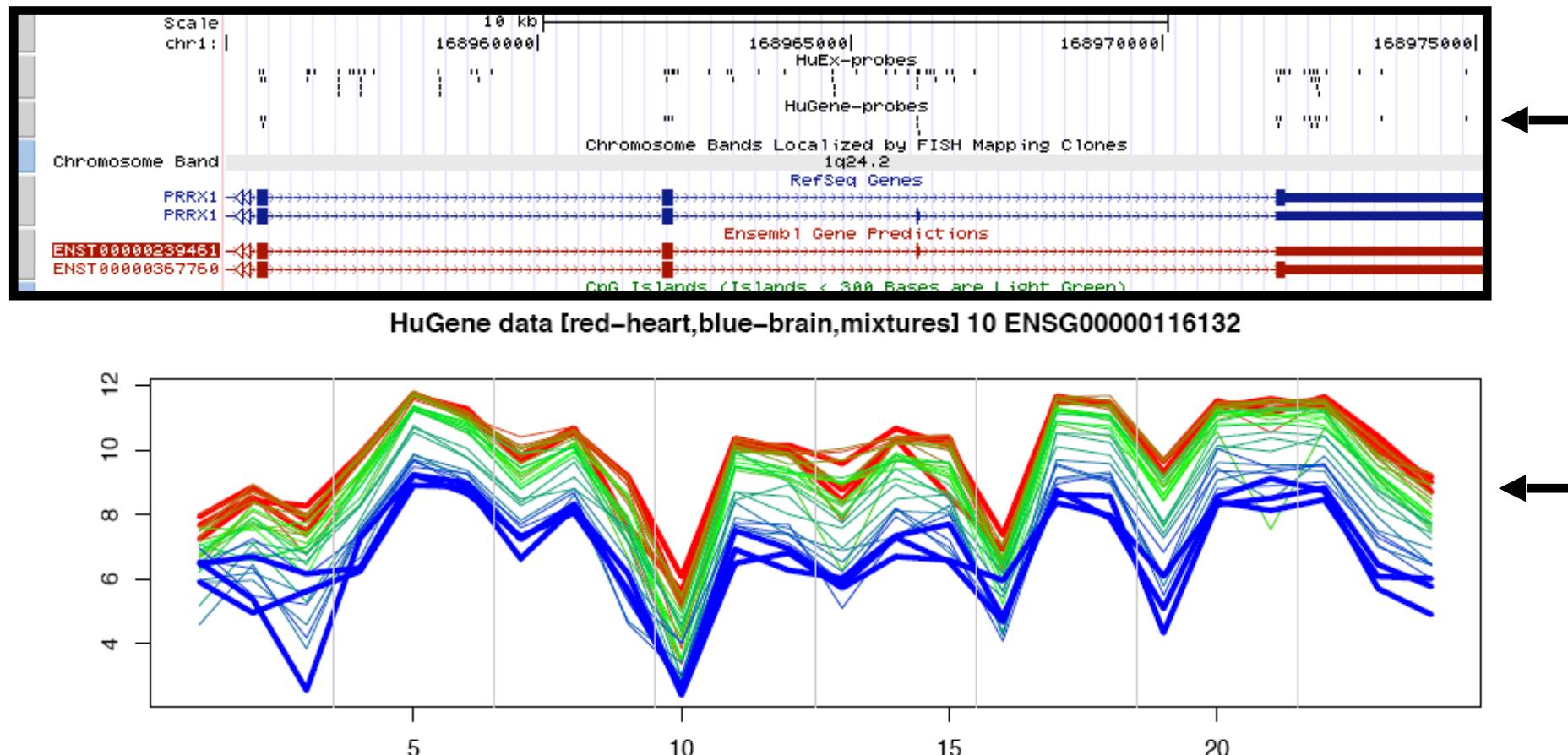
Differential transcript usage (DTU)



differential splicing

Digression 1/3: The nature of Affymetrix Probe Level Data

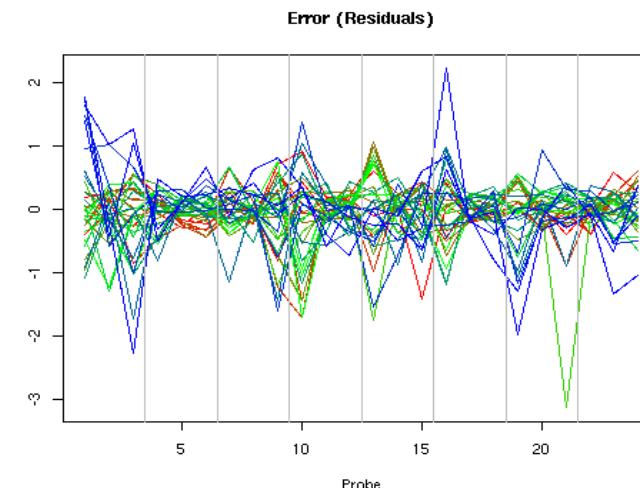
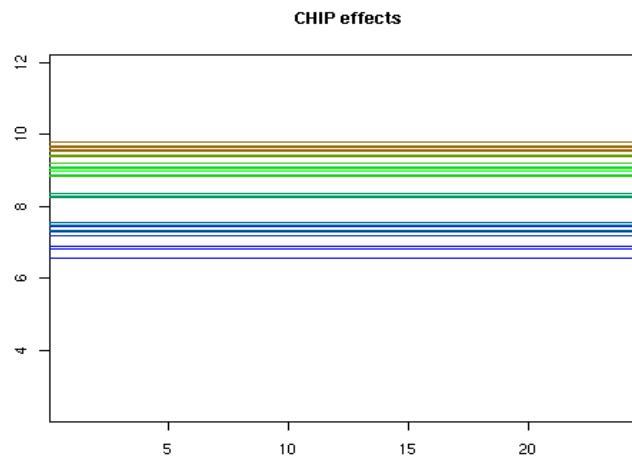
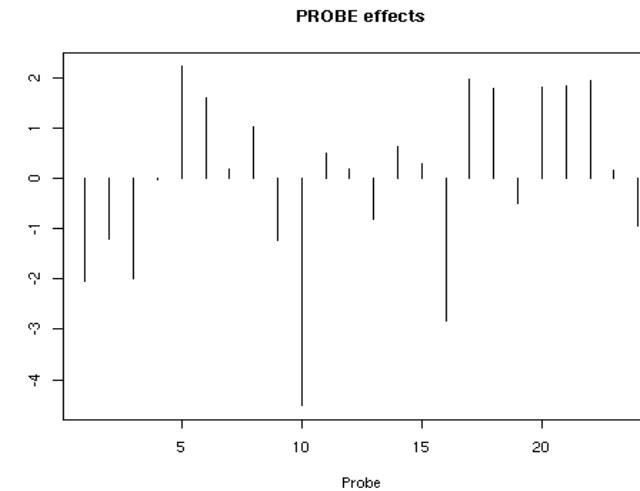
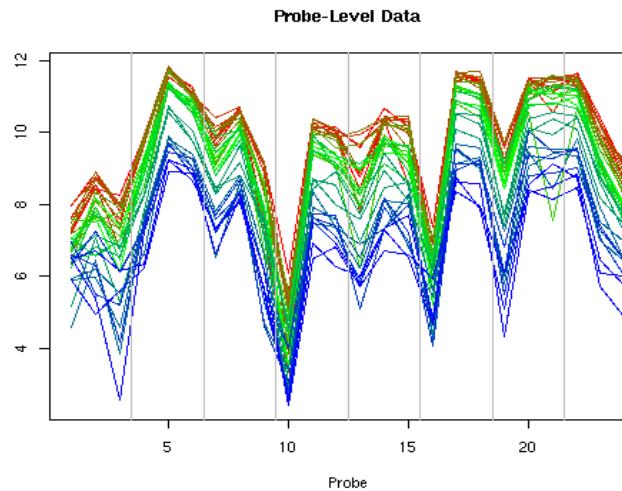
Statistical Bioinformatics // Institute of Molecular Life Sciences



- Data for gene that is DE between heart (red=100% heart) and brain (blue=100% brain).
- 11 mixtures x 3 replicates = 33 samples (33 lines)
- Note the parallelism: probes have different affinities



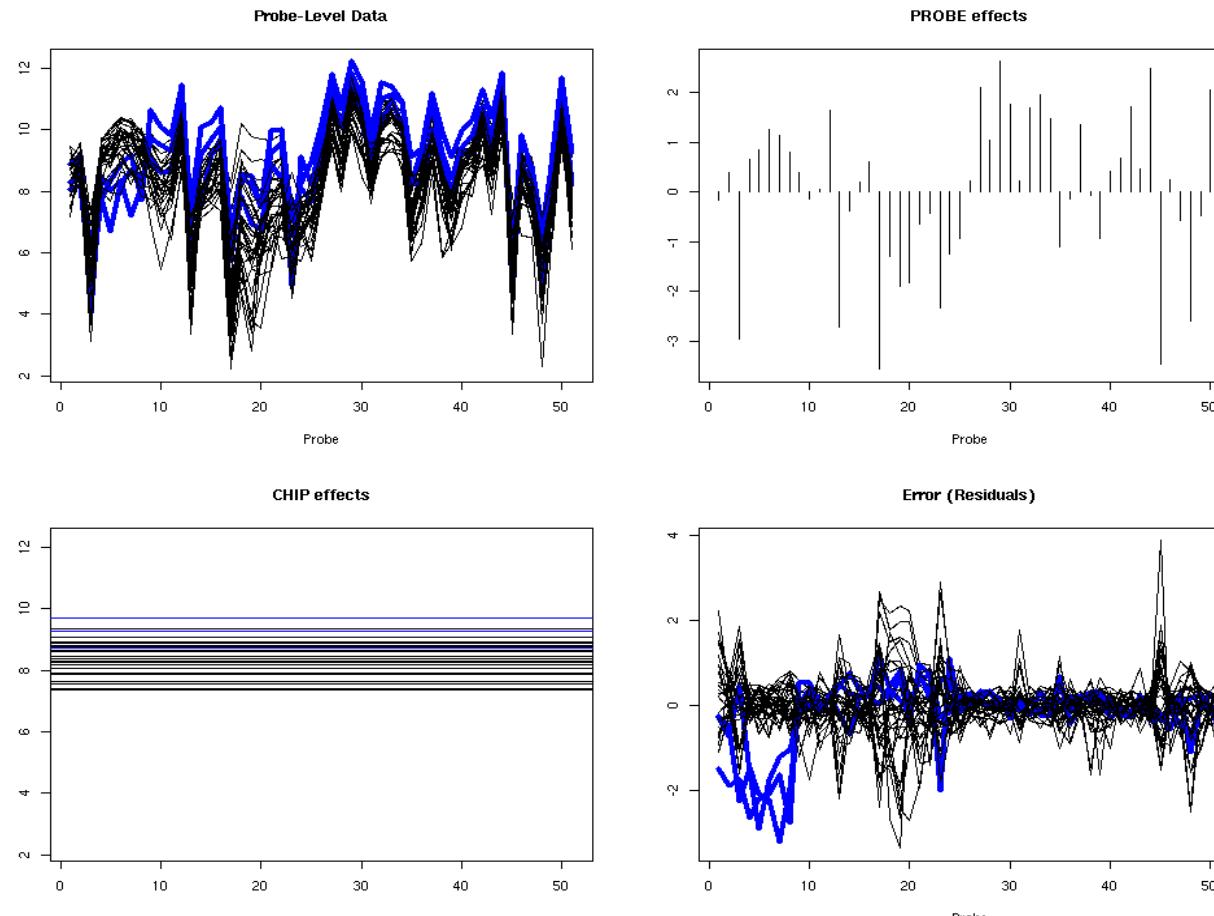
(Digression 2/3) Differential expression: Affy microarrays



$$y_{ik} = g_i + p_k + e_{ik}$$



Digression 3/3: “Differential splicing” or “Differential isoform usage”: Affy microarrays



$$y_{ik} = g_i + p_k + e_{ik}$$



(back to RNA-seq) Beyond differential expression: differential splicing

Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments

Hugues Richard^{1,*}, Marcel H. Schulz^{1,2}, Marc Sultan³, Asja Nürnberg³,
Sabine Schrinner³, Daniela Balzereit³, Emilie Dagand³, Axel Rasche³, Hans Lehrach³,
Martin Vingron¹, Stefan A. Haas¹ and Marie-Laure Yaspo³

¹Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestr. 73,

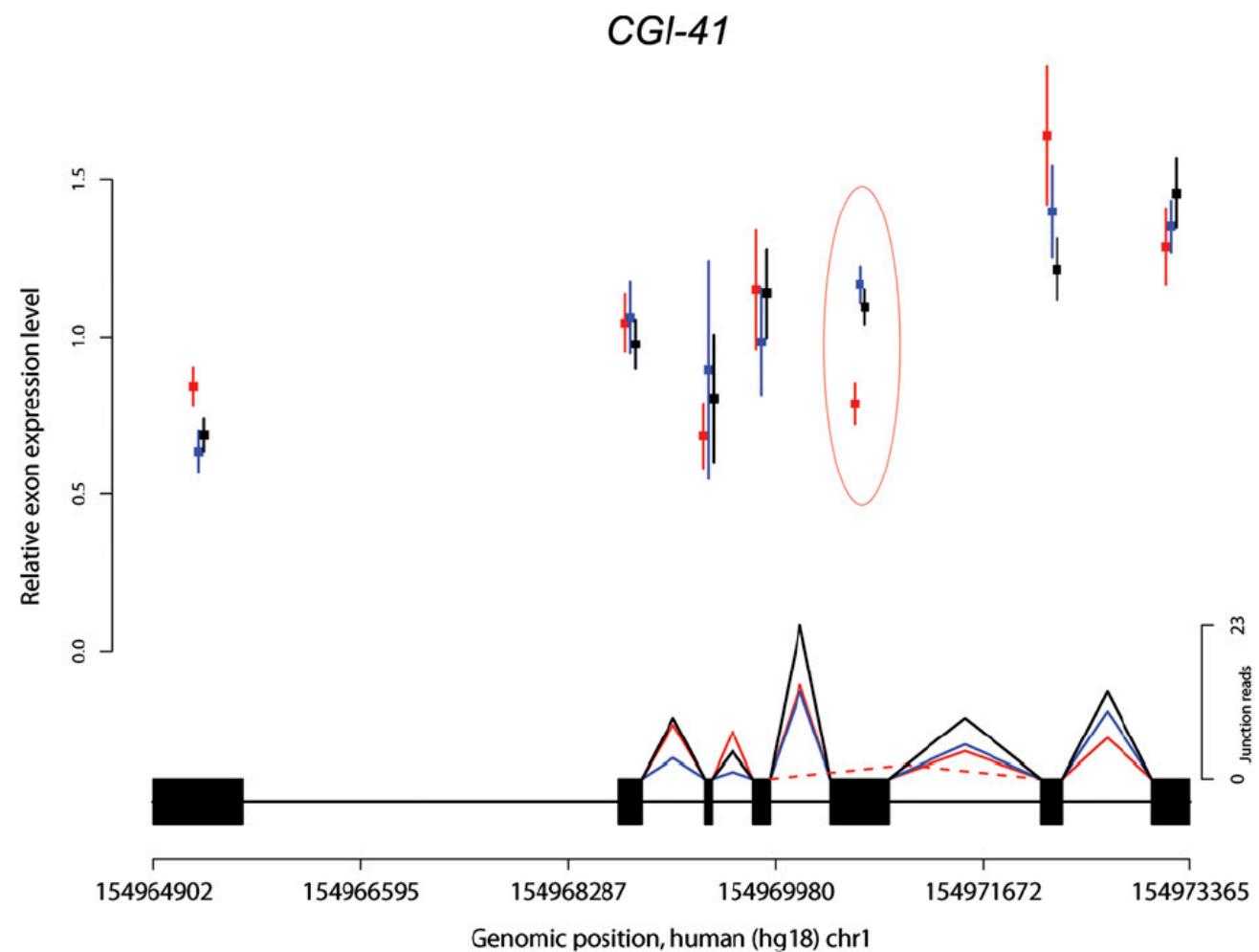
²International Max Planck Research School for Computational Biology and Scientific Computing, and

³Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 73, 14195 Berlin,
Germany

Sex-specific and lineage-specific alternative splicing in primates

Ran Blekhman,^{1,4,5} John C. Marioni,^{1,4,5} Paul Zumbo,² Matthew Stephens,^{1,3,5}
and Yoav Gilad^{1,5}

¹Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA; ²Keck Biotechnology Laboratory, New Haven,
Connecticut 06511, USA; ³Department of Statistics, University of Chicago, Chicago, Illinois 60637, USA



Counting: a few considerations (exon-level)

All the downstream statistical methods start with a count table.

How to get one?

- annotation-based? What about novel genes?
- gene-level versus transcript-level? versus exon-level?
- ambiguities
- junctions?

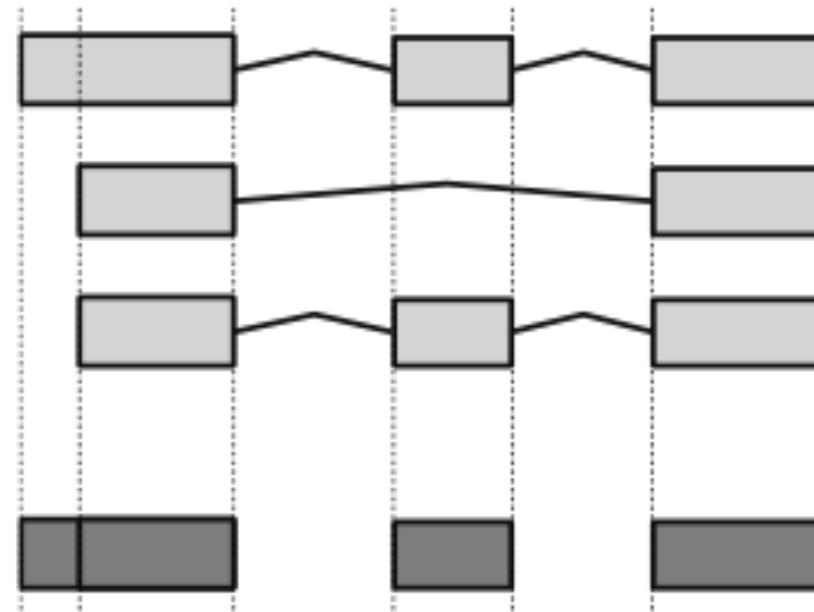


Figure 1. Flattening of gene models: This (fictional) gene has three annotated transcripts involving three exons (light shading), one of which has alternative boundaries. We form counting bins (dark shaded boxes) from the exons as depicted; the exon of variable length gets split into two bins.



Detecting differential usage of exons from RNA-seq data

Simon Anders,^{1,2} Alejandro Reyes,¹ and Wolfgang Huber

European Molecular Biology Laboratory, 69111 Heidelberg, Germany

Transcript inventory versus differential expression

Shotgun RNA-seq data can be used both for identification of transcripts and for differential expression analysis. In the former, one annotates the regions of the genome that can be expressed, i.e., the exons, and how the pre-mRNAs are spliced into transcripts. In differential expression analysis, one aims to study the regulation of these processes across different conditions. For the method described here, we assume that a transcript inventory has already been defined, and focus on differential expression.



DEXSeq – general structure: exon-level models

We use generalized linear models (GLMs) (McCullagh and Nelder 1989) to model read counts. Specifically, we assume K_{ijl} to follow a negative binomial (NB) distribution:

$$K_{ijl} \sim NB\left(\text{mean} = s_j \mu_{ijl}, \text{dispersion} = \alpha_{il}\right), \quad (1)$$

where α_{il} is the dispersion parameter (a measure of the distribution's spread; see below) for counting bin (i, l) , and the mean is predicted via a log-linear model as

$$\log \mu_{ijl} = \beta_i^G + \beta_{il}^E + \beta_{ip_j}^C + \beta_{ip_j l}^{EC}. \quad (2)$$

i – gene

j – sample ... ρ_j is condition (categorical)

l – bin

β^G – baseline “expression strength”

β^E – “exon” (bin) effect

β^C – condition effect

β^{EC} – condition x “exon” interaction

DEXSeq: sig. interaction terms = differential exon usage

(DEXSeq
vignette)

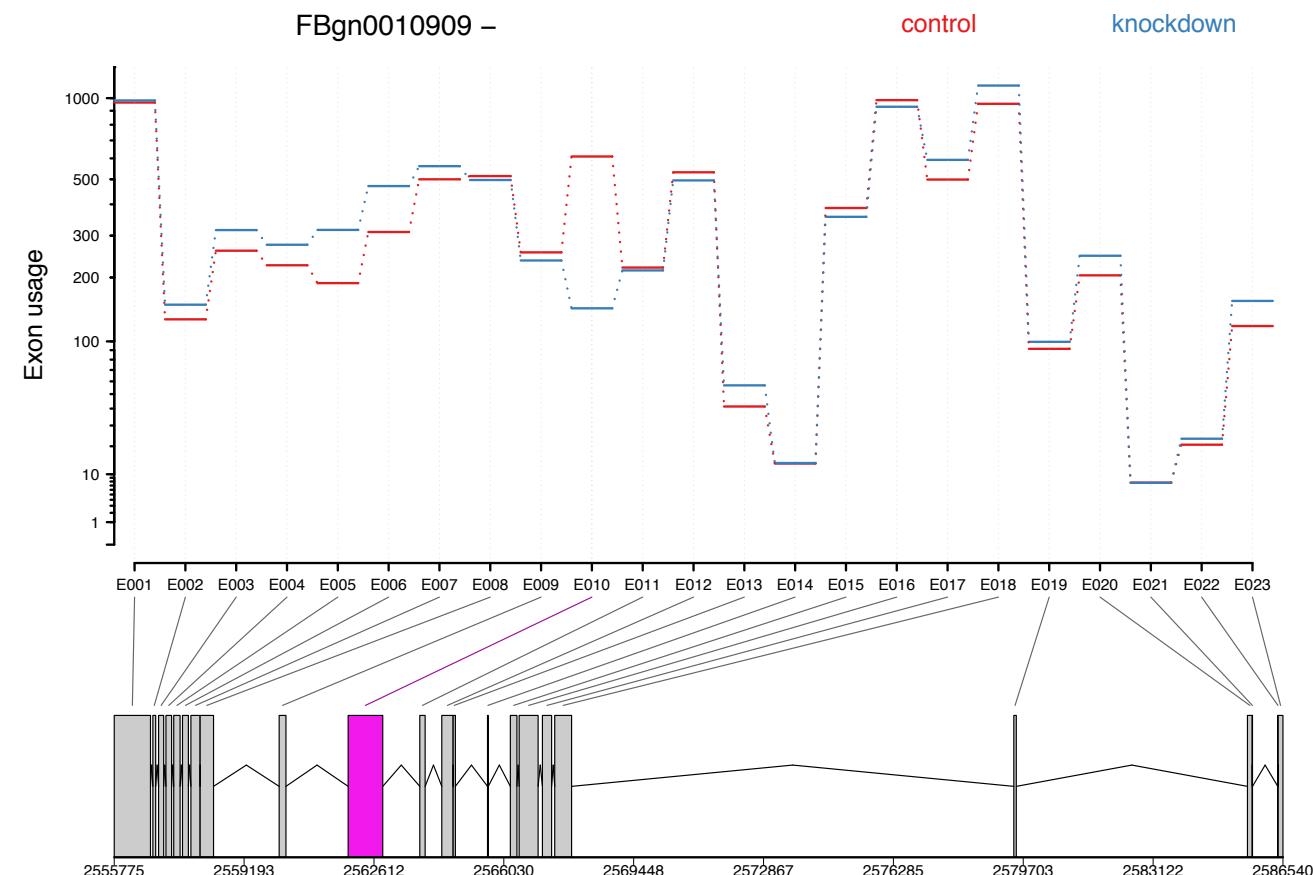
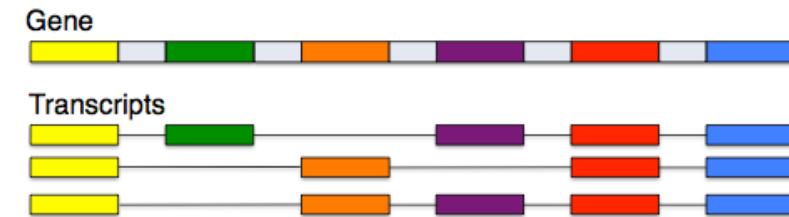


Figure 6: Fitted splicing

The plot represents the estimated effects, as in Figure 3, but after subtraction of overall changes in gene expression.



DTU → dirichlet-multinomial distribution

Estimated:

- transcript ratios

$$\Pi = (\pi_1, \pi_2, \pi_3)$$

Observed:

- transcript counts
- gene expression

$$Y = (y_1, y_2, y_3)$$

$$n = \sum_{j=1}^k y_j$$

Multinomial: $P(\mathbf{Y} = \mathbf{y} | \Pi = \pi) = \binom{n}{\mathbf{y}} \prod_{j=1}^k \pi_j^{y_j}$

Dirichlet: $P(\Pi = \pi) = \frac{\Gamma(\gamma_+)}{\prod_{j=1}^k \Gamma(\gamma_j)} \prod_{j=1}^k \pi_j^{\gamma_j - 1}, \gamma_+ = \sum_{j=1}^k \gamma_j$

Dirichlet-multinomial: $P(\mathbf{Y} = \mathbf{y}) = \binom{n}{\mathbf{y}} \frac{\Gamma(\gamma_+)}{\Gamma(n + \gamma_+)} \prod_{j=1}^k \frac{\Gamma(y_j + \gamma_j)}{\Gamma(\gamma_j)}, \gamma_j = \pi_j \gamma_+$

