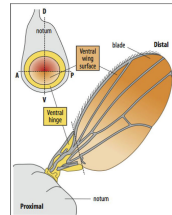
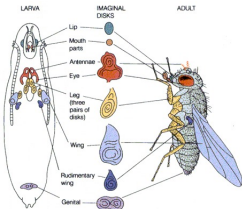


ERC *Drosophila melanogaster*

Emilio Palumbo
Amaya Abad
Silvia Perez Lluch
Alessandra Breschi
Cecilia Klein
Marina Ruiz Romero
Roderic Guigo

Lab meeting - 04/11/2015



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

- Decomposing the variation of gene expression

- Antisense

Wing compartments

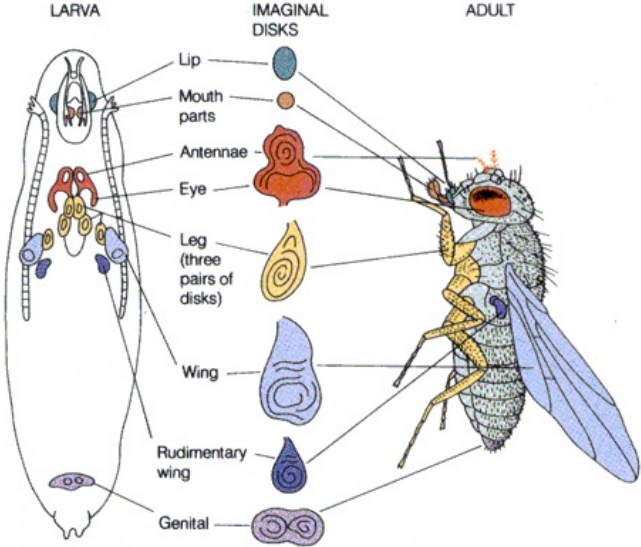
- Decomposing the variation of gene expression

- Differential gene expression - EdgeR

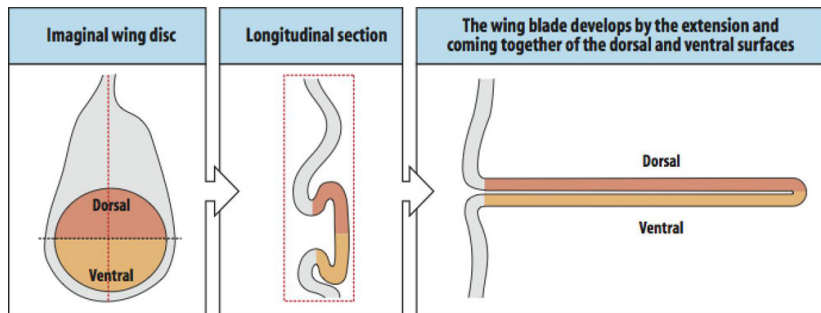
Isoform usage

Next steps

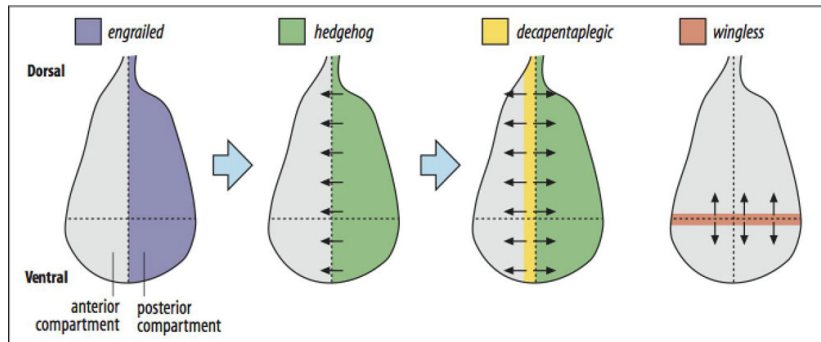
Imaginal discs



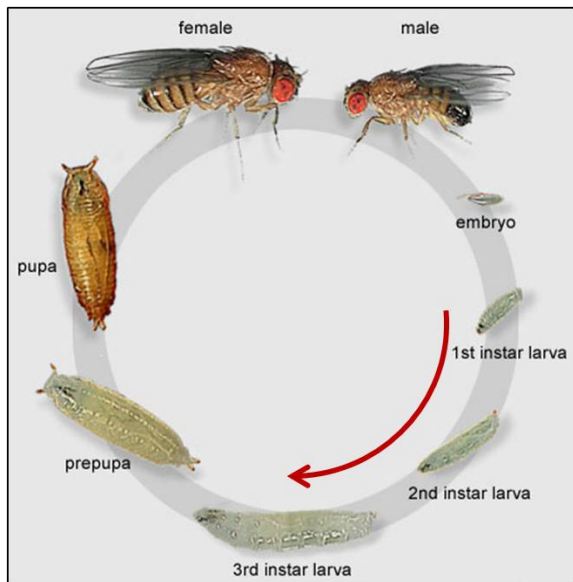
Wing imaginal disc



Wing A/P and D/V boundaries



Developmental stages



Data and goals

- ▶ RNASeq
- ▶ ChIPSeq

- ▶ Differences at transcriptome and chromatin level which distinguish the imaginal discs leading to different organs
- ▶ Temporal and spatial transcriptome and chromatin profiling during fly development
- ▶ Chromatin dynamics and its role in alternative splicing
- ▶ RNASeq: signatures of compartmentalization in wing imaginal discs

Outline

Introduction

Overview of RNA-seq samples

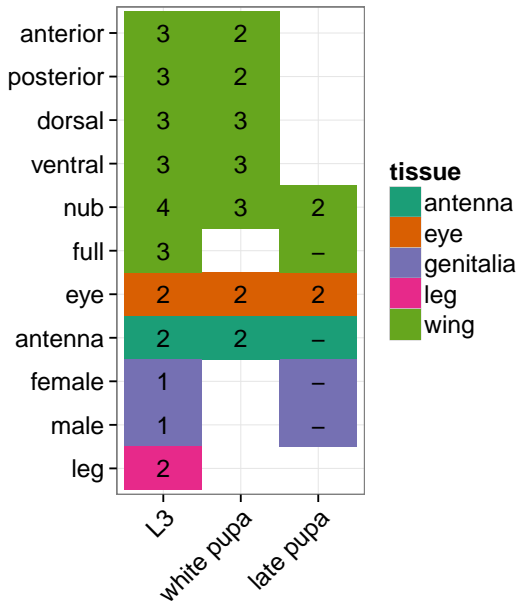
Wing and eye

Wing compartments

Isoform usage

Next steps

Overview of processed RNA-seq samples



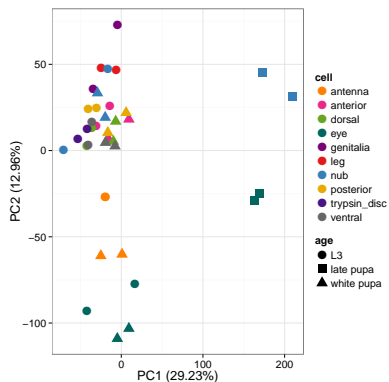
RNA-seq pipeline

- ▶ Assembly: dm6
- ▶ Annotation: FlyBase r6.05

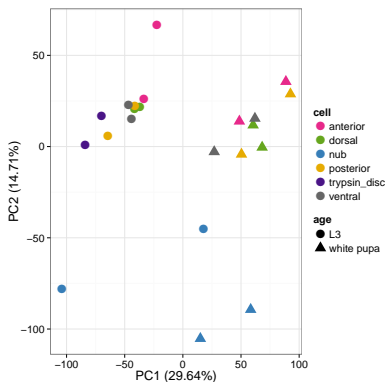
- ▶ Grape pipeline - STAR+RSEM

PCA - selected replicates - $IDR \leq 0.1$

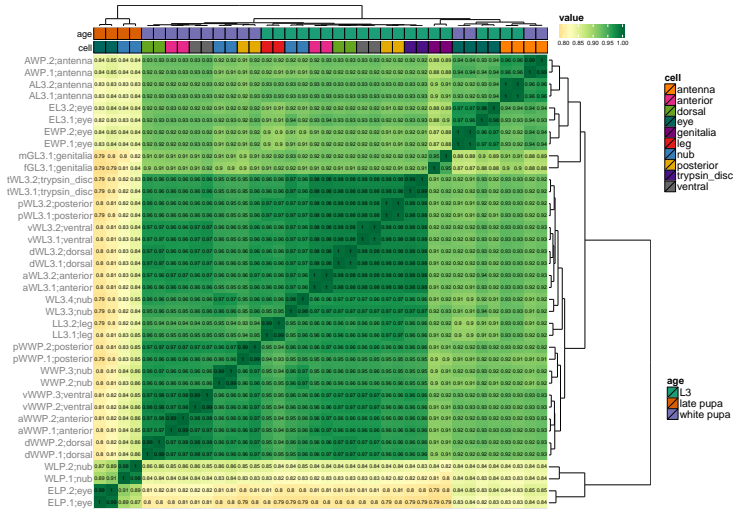
17158 genes



wing compartments



Clustering by gene expression - $IDR \leq 0.1$ (Spearman)



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Decomposing the variation of gene expression

Genes with high variation across space

Genes with high variation across time

Antisense

Wing compartments

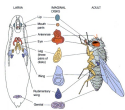
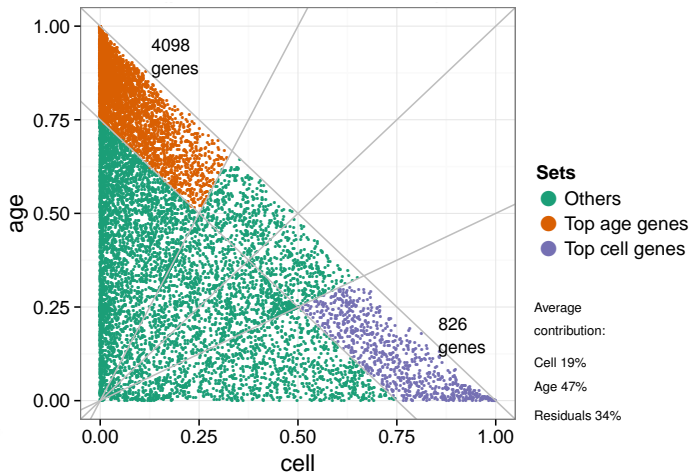
Isoform usage

Next steps

RNA-seq samples



Decomposing the variation of gene expression across time and space



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Decomposing the variation of gene expression

Genes with high variation across space

Genes with high variation across time

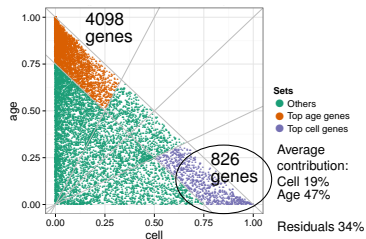
Antisense

Wing compartments

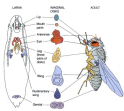
Isoform usage

Next steps

Genes with high variation across space



Biotype	Cluster1
mRNA	250
ncRNA	26
pseudogene	3
Biotype	Cluster2
mRNA	529
ncRNA	16
pseudogene	2



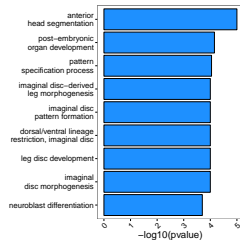
Variance decomposition

topCellGenes = 826 genes

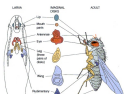
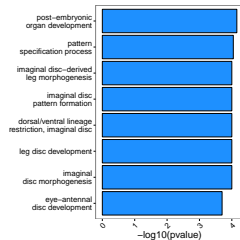
cluster 1 = 279 genes



Pvalue	Term - Cluster1
--------	-----------------



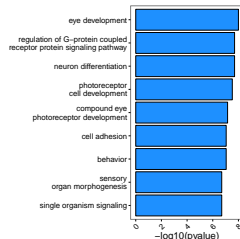
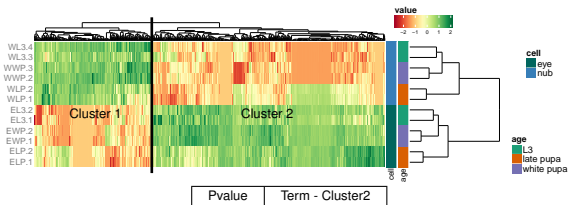
wing, eye, imaginal, disc, pattern, development, signaling



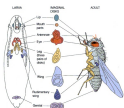
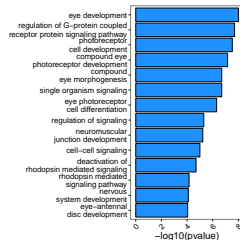
Variance decomposition

topCellGenes = 826 genes

cluster 2 = 547 genes



wing, eye, imaginal, disc, pattern, development, signaling



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Decomposing the variation of gene expression

Genes with high variation across space

Genes with high variation across time

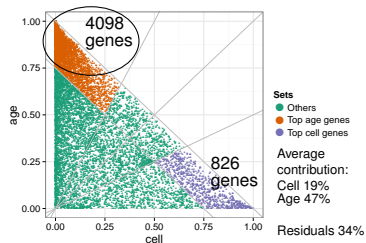
Antisense

Wing compartments

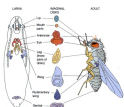
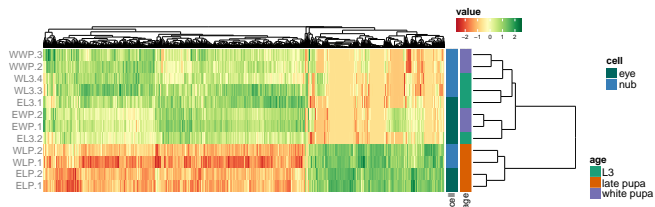
Isoform usage

Next steps

Genes with high variation across time



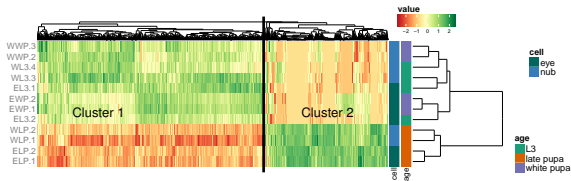
Biotype	Cluster1
mRNA	2584
ncRNA	82
pseudogene	16
rRNA	1
Biotype	Cluster2
mRNA	1329
ncRNA	75
pseudogene	9
rRNA	2



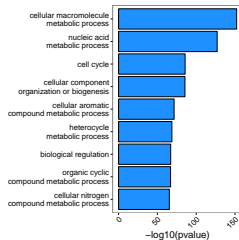
Variance decomposition

topAgeGenes = 4098 genes

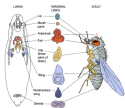
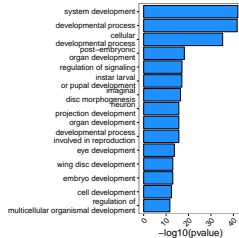
cluster 1 = 2683 genes



Pvalue	Term - Cluster1
1.5542469751877e-34	Ribosome
1.56907308620959e-13	Spliceosome
7.41109730641896e-13	RNA transport
1.61890438245164e-09	mRNA surveillance pathway



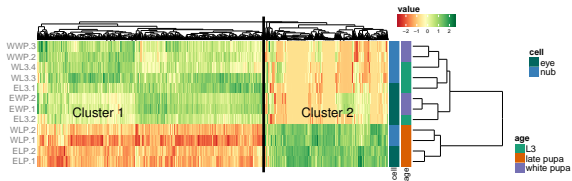
wing, eye, imaginal, disc, pattern, development, signaling



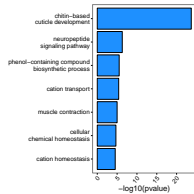
Variance decomposition

topAgeGenes = 4098 genes

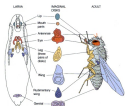
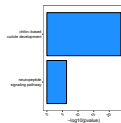
cluster 2 = 1415 genes



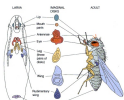
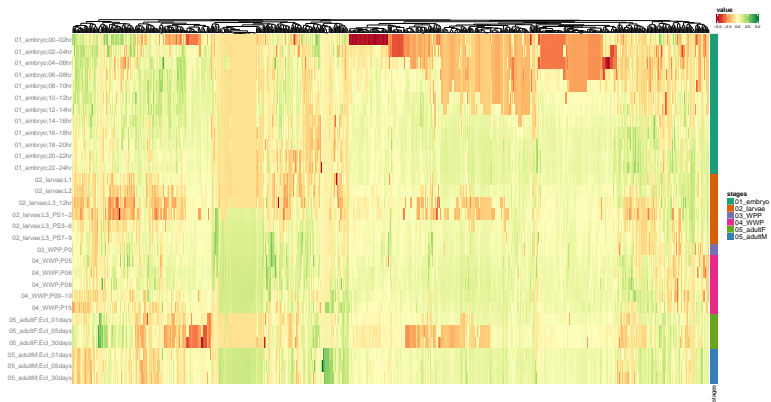
Pvalue	Term - Cluster2
1.98209815298486e-08	Metabolic pathways



wing, eye, imaginal, disc, pattern, development, signaling



modENCODE - gene expression cell genes



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Decomposing the variation of gene expression

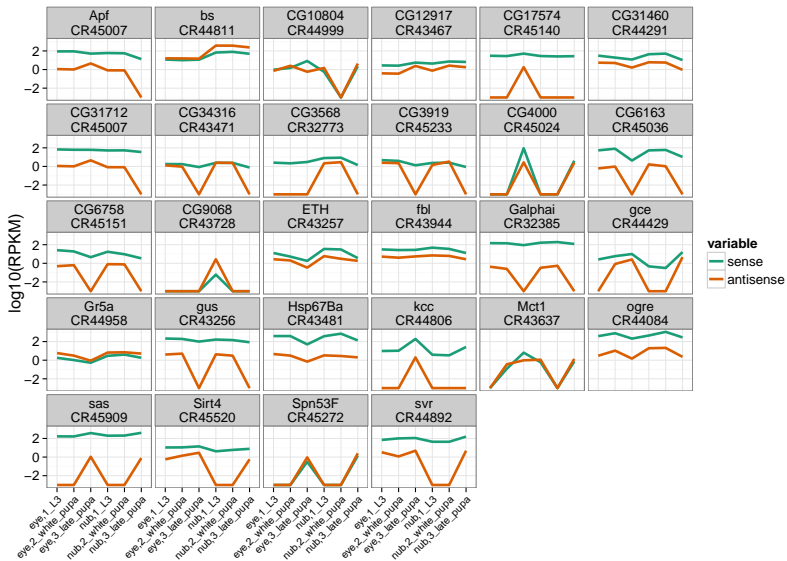
Antisense

Wing compartments

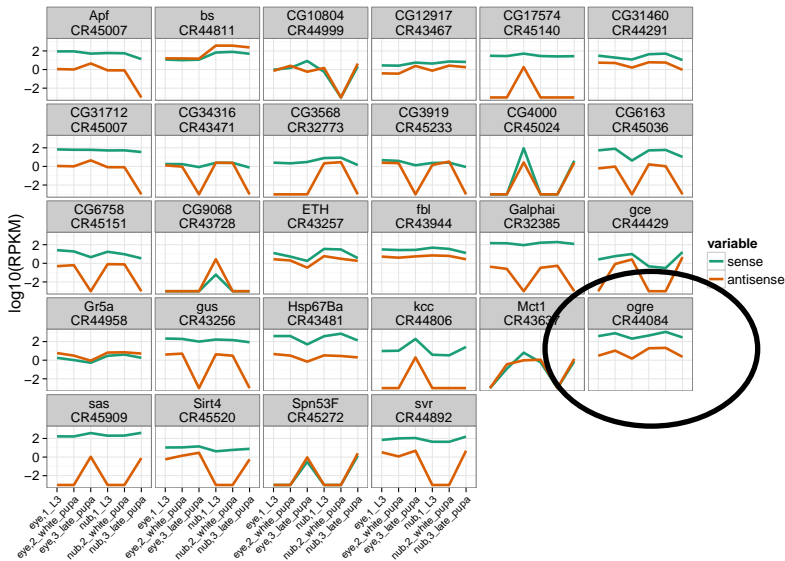
Isoform usage

Next steps

antisense - positively correlated



antisense - positively correlated

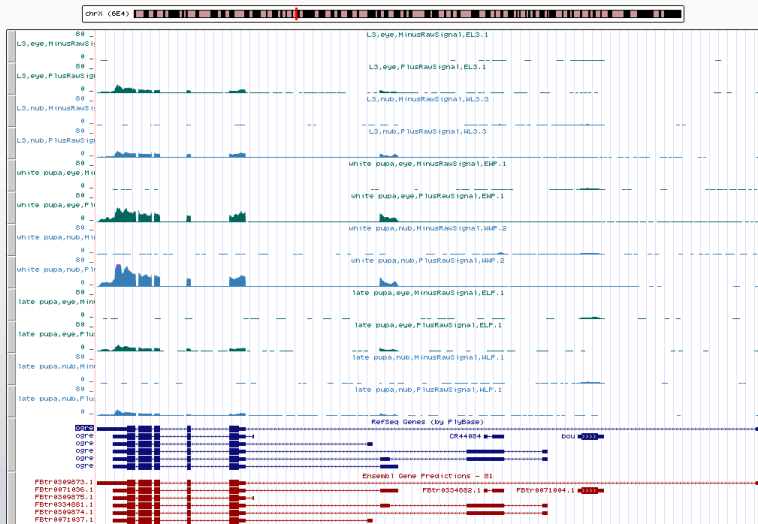


antisense - positively correlated

UCSC Genome Browser on *D. melanogaster* Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6) Asse

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

<:6,973,620-6,991,318 17,699 bp. enter position, gene symbol or search terms go [hg38 replaces hg19 as default huma](#)

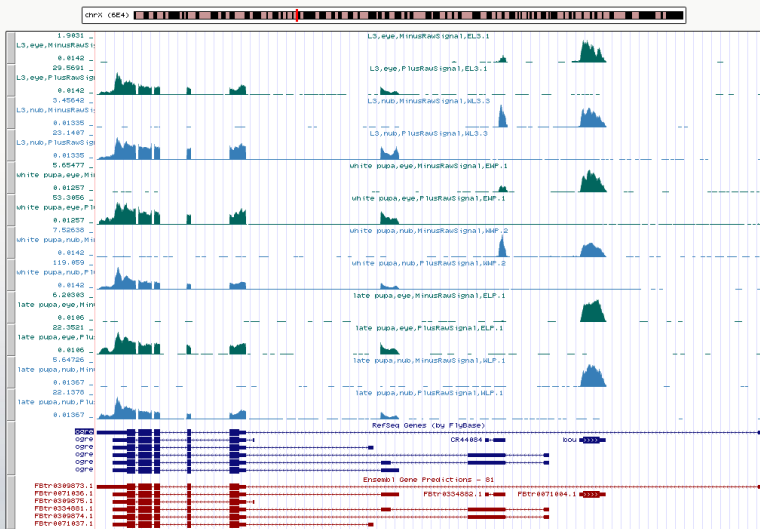


antisense - positively correlated

UCSC Genome Browser on *D. melanogaster* Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6) Asse

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

X:6,973,620-6,991,318 17,699 bp. enter position, gene symbol or search terms go **hg38 replaces hg19 as default hum**



antisense - positively correlated

General Information			
Symbol	Dmellogre	Species	<i>D. melanogaster</i>
Name	optic ganglion reduced	Annotation symbol	CG3039
Feature type	protein_coding_gene	FlyBase ID	FBgn0004646
Gene Model Status	Current	Stock availability	17 publicly available
Families, Domains and Molecular Function			
Gene Group Membership (FlyBase)	INNEXINS		
Protein Family (UniProt, Sequence Similarities)	Belongs to the pannexin family. (ECO:0000255 PROSITE-ProRule:PRU000351). (P27716)		
Protein Domains/Motifs	UniProt (Sequence Similarities) - InterPro Innexin		
Molecular Function (see GO section for details)	Experimental Evidence - Predictions/Assertions gap junction channel activity		
Gene Ontology (GO): (9 terms)			
<input checked="" type="checkbox"/> Molecular Function (1 term)			
Terms Based on Experimental Evidence (0 terms)			
Terms Based on Predictions or Assertions (1 term)			
CV term	Evidence	References	
gap junction channel activity	inferred by curator from GO:0005921	<i>(Bohmann and Zimmermann, 2008)</i>	
<input checked="" type="checkbox"/> Biological Process (4 terms)			
Terms Based on Experimental Evidence (1 term)			
CV term	Evidence	References	
phototransduction	inferred from mutant phenotype	<i>(Curtin et al., 2002)</i>	
Terms Based on Predictions or Assertions (3 terms)			
CV term	Evidence	References	
intercellular transport	inferred from sequence or structural similarity with Inx3	<i>(Curators, 2008--)</i>	
nervous system development	non-traceable author statement	<i>(Swiss-Prot Project Members, 1992.8.1)</i>	
visual behavior	inferred from sequence or structural similarity	<i>(Phelan et al., 1998)</i>	

antisense - positively correlated

modENCODE Temporal Expression Data

([modENCODE_mRNA-Seq_U](#))

Styles

- Linear
- Logarithmic
- Heatmap

Scales

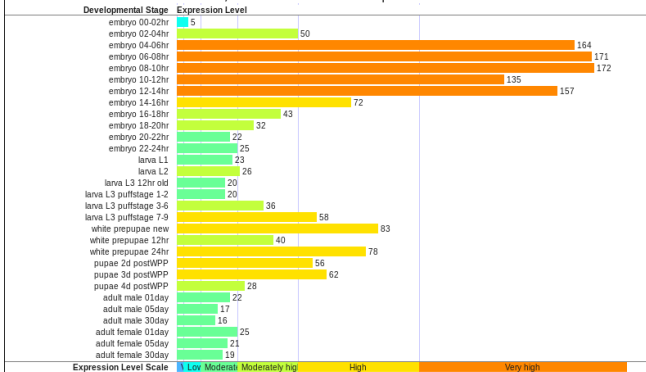
- gene max expression
- Moderate expression bin max
- High expression bin max
- Extremely high expression bin max

[\[download data \(TSV\)\]](#)

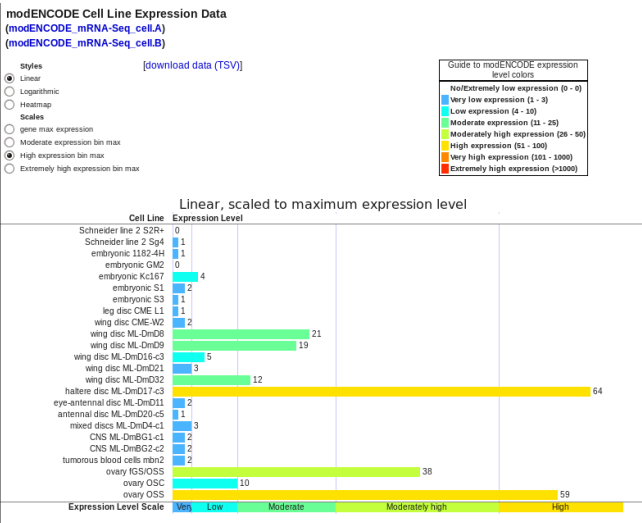
Guide to modENCODE expression level colors

No/Extremely low expression (0 - 0)
Very low expression (1 - 3)
Low expression (4 - 10)
Moderate expression (11 - 25)
Moderately high expression (26 - 50)
High expression (51 - 100)
Very high expression (101 - 1000)
Extremely high expression (>1000)

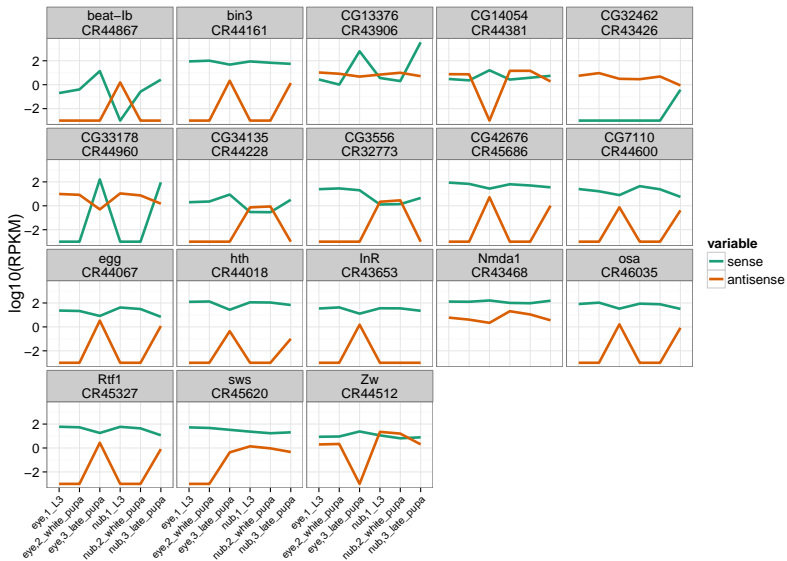
Linear, scaled to maximum expression level



antisense - positively correlated



antisense - negatively correlated



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Wing compartments

Decomposing the variation of gene expression

Genes with high variation across space

Genes with high variation across time

modENCODE - chromatin marks

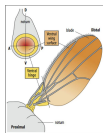
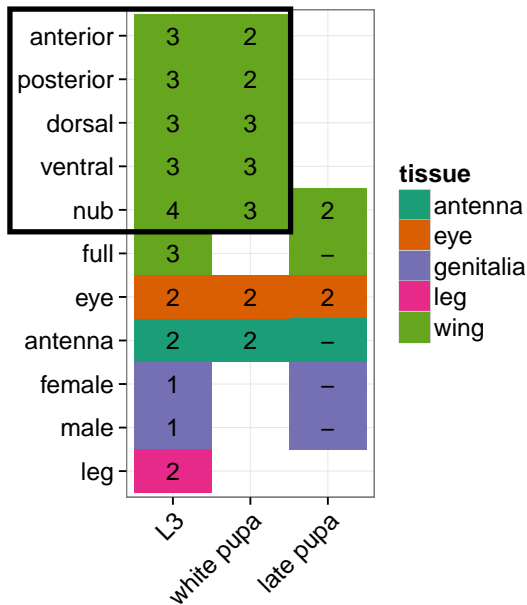
Differential gene expression - EdgeR

Isoform usage

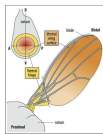
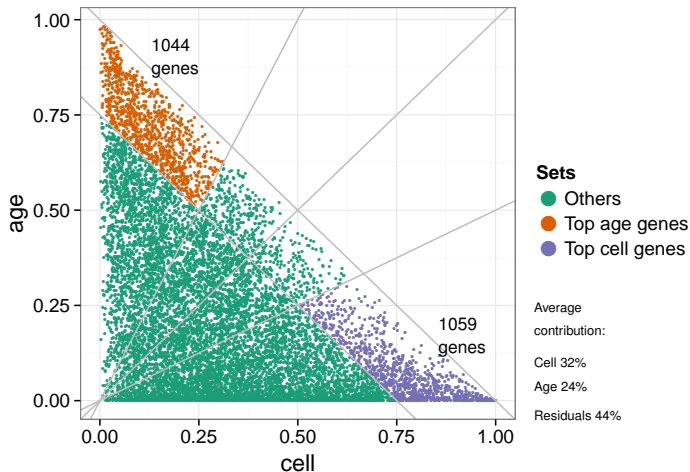
Next steps

RNA-seq samples

Wing compartments



Decomposing the variation of gene expression across time and space



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Wing compartments

Decomposing the variation of gene expression

Genes with high variation across space

Genes with high variation across time

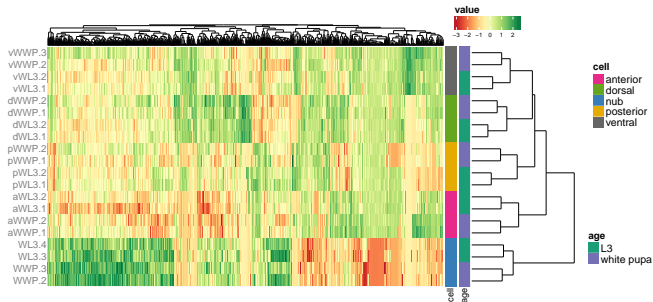
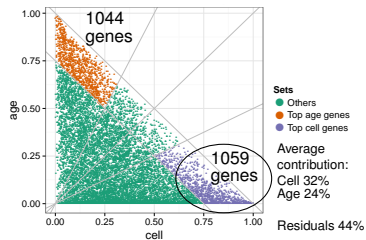
modENCODE - chromatin marks

Differential gene expression - EdgeR

Isoform usage

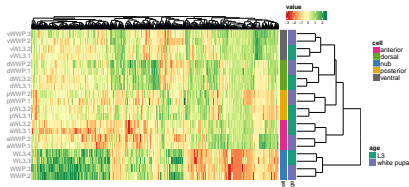
Next steps

Genes with high variation across space

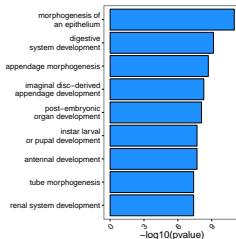


Decomposing the variation of gene expression across space

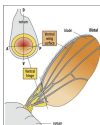
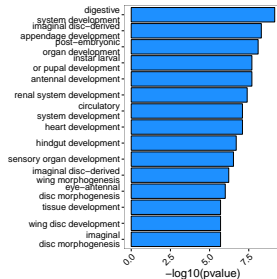
topCellGenes = 608 genes



Pvalue	Term
2.32846460522083e-06	Oxidative phosphorylation
2.51047792269406e-05	Proteasome



wing, eye, imaginal, disc, pattern, development, signaling



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Wing compartments

Decomposing the variation of gene expression

Genes with high variation across space

Genes with high variation across time

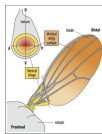
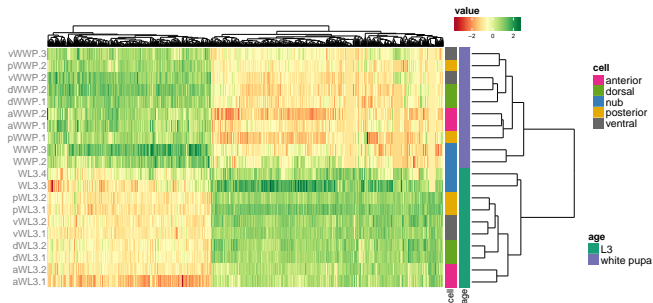
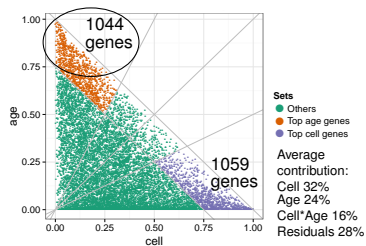
modENCODE - chromatin marks

Differential gene expression - EdgeR

Isoform usage

Next steps

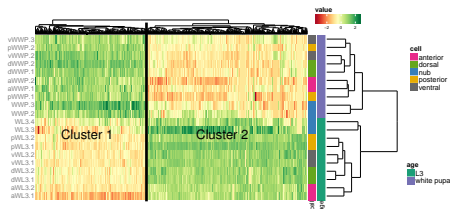
Genes with high variation across time



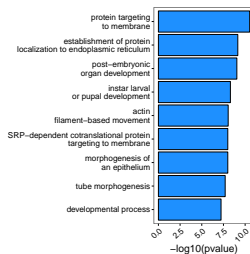
Variance decomposition

topAgeGenes = 1044 genes

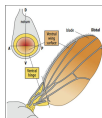
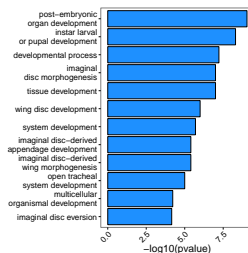
cluster 1 = 435 genes



Pvalue	Term - Cluster1
1.00401087099388e-09	Protein export
2.43046134848088e-09	Protein processing in endoplasmic reticulum



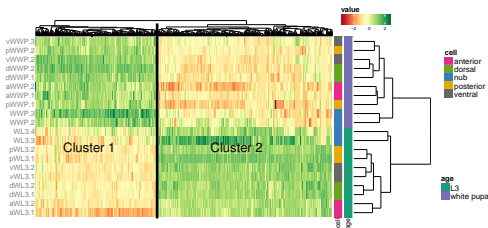
wing, eye, imaginal, disc, pattern, development, signaling



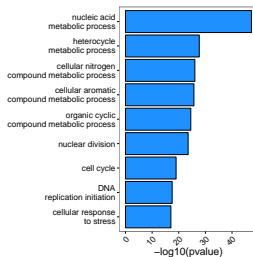
Variance decomposition

topAgeGenes = 1044 genes

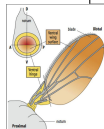
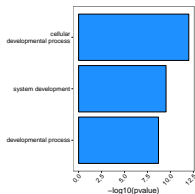
cluster 2 = 624 genes



Pvalue	Term - Cluster2
4.12201666158476e-21	DNA replication
7.50087633662912e-12	Mismatch repair
8.1618297773066e-08	Nucleotide excision repair
1.66795771880413e-06	Ribosome biogenesis in eukaryotes
7.98022928115518e-05	Pyrimidine metabolism



wing, eye, imaginal, disc, pattern, development, signaling



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Wing compartments

Decomposing the variation of gene expression

Genes with high variation across space

Genes with high variation across time

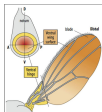
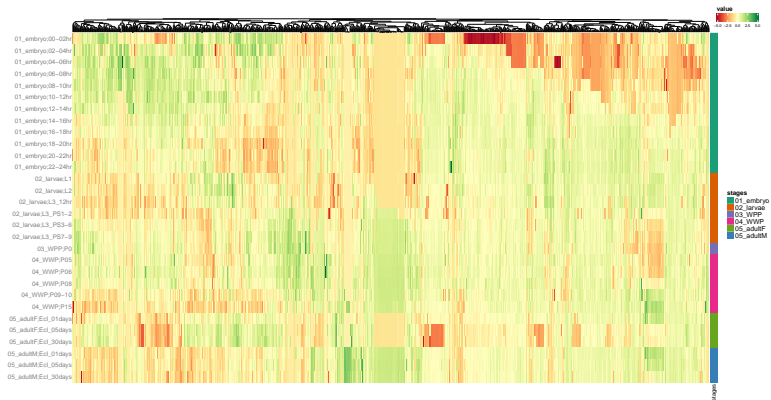
modENCODE - chromatin marks

Differential gene expression - EdgeR

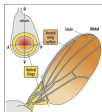
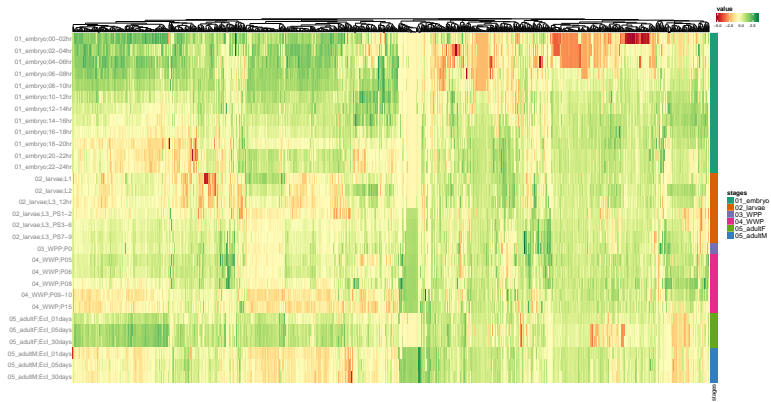
Isoform usage

Next steps

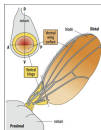
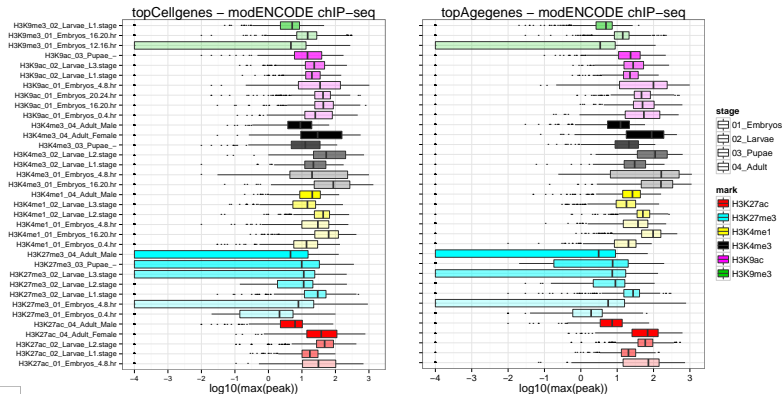
modENCODE - gene expression cell genes



modENCODE - gene expression age genes



The highest peak (measured as the log₁₀ of the signal profiles reported by modENCODE) within the gene body



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Wing compartments

Decomposing the variation of gene expression

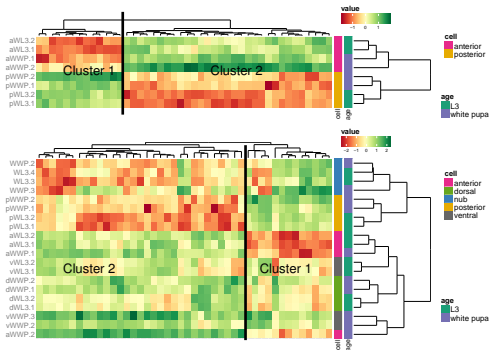
Differential gene expression - EdgeR

Space-specific genes - pairwise comparison of wing compartments

Isoform usage

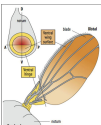
Next steps

anterior vs posterior - space-specific genes = 44

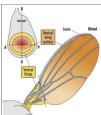
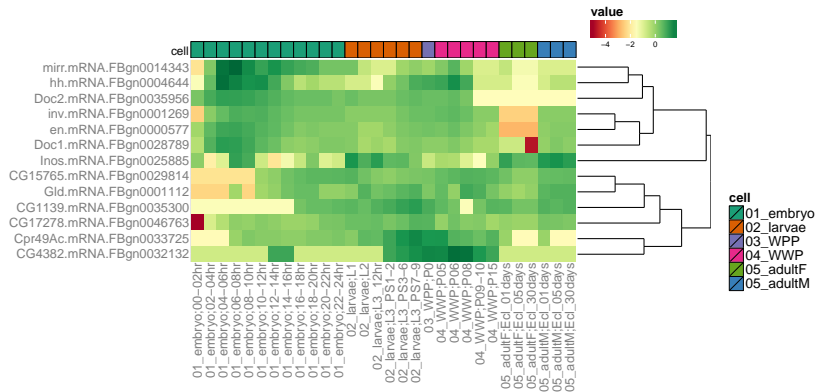


Pvalue	Term BP - Cluster1
2e-07	wing disc anterior/posterior pattern formation
1e-06	anterior commissure morphogenesis
1e-06	forebrain development
8e-06	anterior/posterior lineage restriction, imaginal disc
2e-05	compartment pattern specification
2e-05	analia development
2e-05	central nervous system neuron axonogenesis
3e-05	imaginal disc-derived wing vein specification
3e-05	central nervous system neuron differentiation
Pvalue	Term BP - Cluster2
2e-07	cell adhesion
4e-07	cell morphogenesis involved in differentiation
2e-06	multicellular organismal development
5e-06	generation of neurons
6e-06	neuron development
1e-05	neuron projection guidance
2e-05	cell projection morphogenesis
3e-05	axonogenesis
4e-05	cellular component morphogenesis

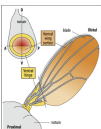
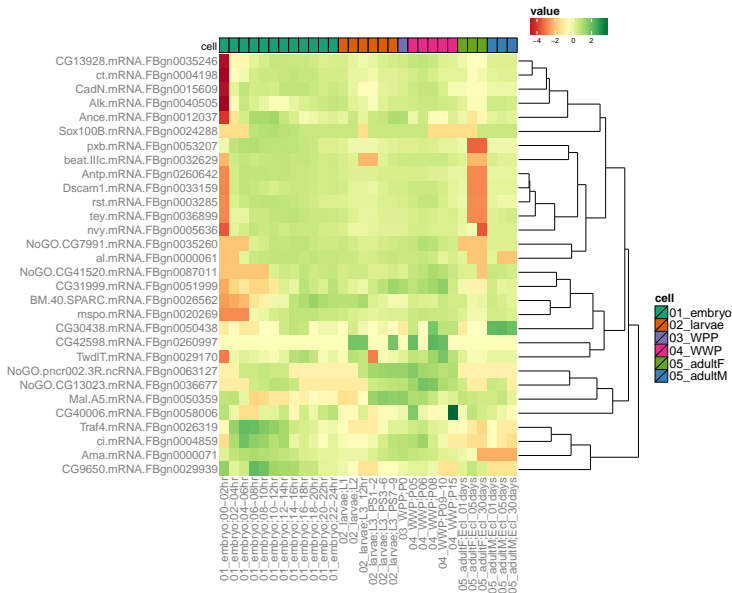
Biotype	Cluster1	Cluster2
mRNA	13	30
ncRNA	0	1
GO terms	13	27
GO terms (wing, anterior, posterior, imaginal disc, development)	9	14



modEncode for cluster1



modEncode for cluster2



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

Wing compartments

Decomposing the variation of gene expression

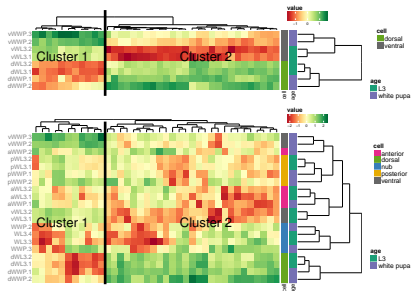
Differential gene expression - EdgeR

Space-specific genes - pairwise comparison of wing compartments

Isoform usage

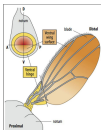
Next steps

dorsal vs ventral - space-specific genes = 38

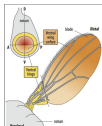
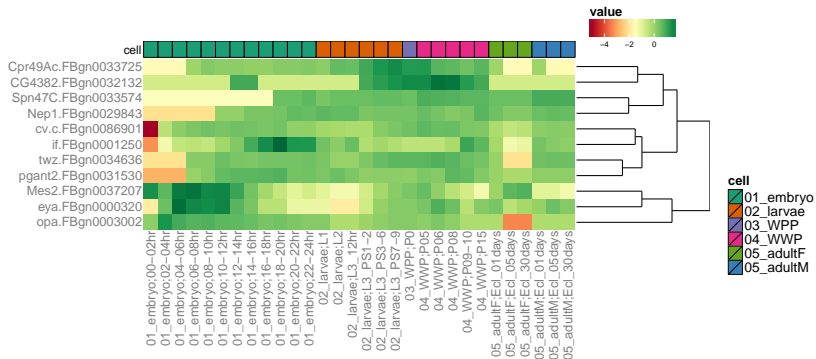


Pvalue	Term BP - Cluster1
4e-06	midgut development
7e-05	maintenance of epithelial integrity, open tracheal system
1e-04	embryo development
1e-04	cell motility
1e-04	digestive system development
3e-04	imaginal disc development
4e-04	post-embryonic development
7e-04	tissue homeostasis
7e-04	eye-antennal disc morphogenesis
Pvalue	Term BP - Cluster2
3e-06	myoblast migration
3e-06	fibroblast growth factor receptor signaling pathway
3e-06	response to fibroblast growth factor
6e-05	regulation of crystal cell differentiation
7e-05	cellular response to growth factor stimulus
1e-04	larval visceral muscle development
2e-04	glial cell proliferation
3e-04	larval somatic muscle development
5e-04	enzyme linked receptor protein signaling pathway

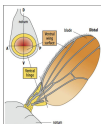
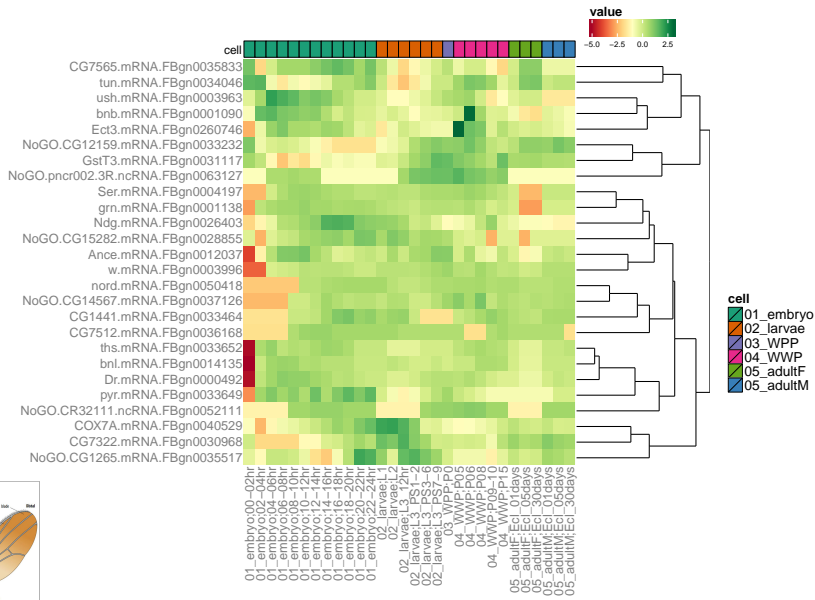
Biotype	Cluster1	Cluster2
mRNA	11	25
ncRNA	0	2
GO terms	11	21
GO terms (wing, dorsal, ventral, imaginal disc, development)	6	9



modEncode for cluster1



modEncode for cluster2



Outline

Introduction

Overview of RNA-seq samples

Wing and eye

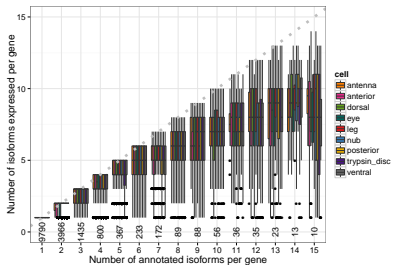
Wing compartments

Isoform usage

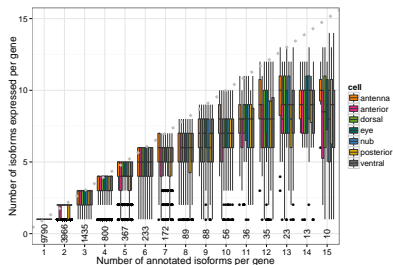
Next steps

Number of isoforms expressed vs annotated

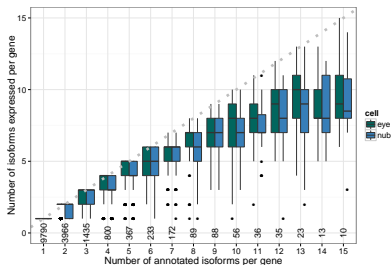
L3



white pupa

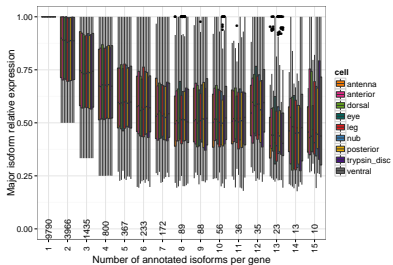


late pupa

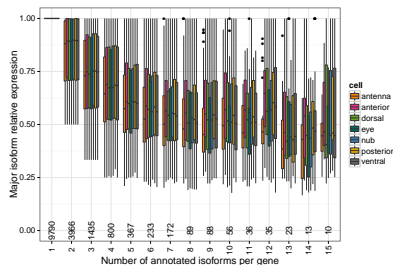


Expression of the major isoform

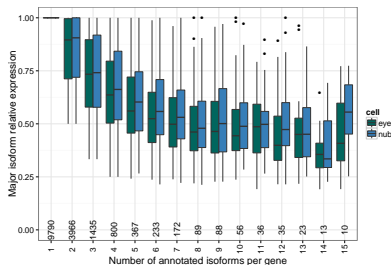
L3



white pupa

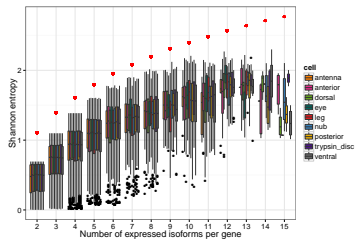


late pupa

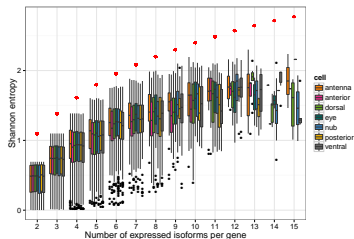


Shannon entropy of isoform expression

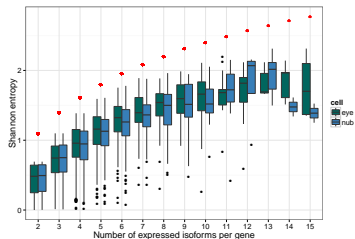
L3



white pupa

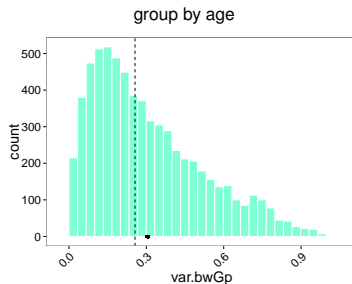
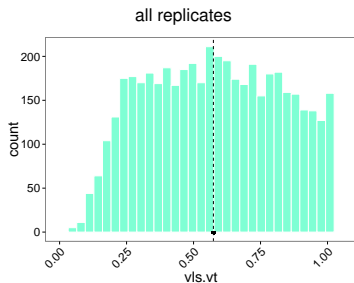


late pupa

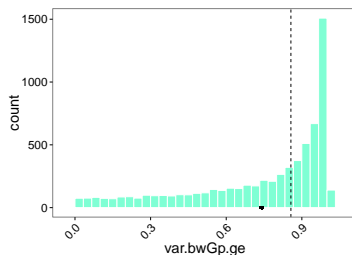


$$H = - \sum_{i=1}^n p(x_i) \log p(x_i)$$

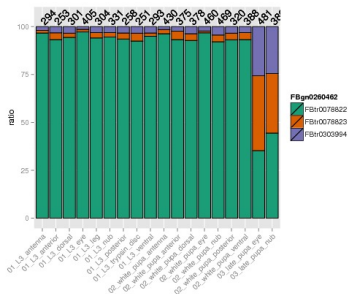
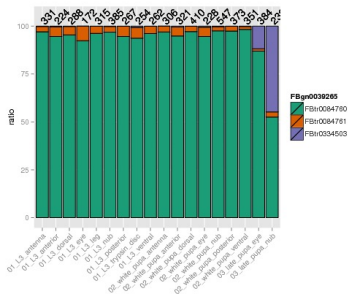
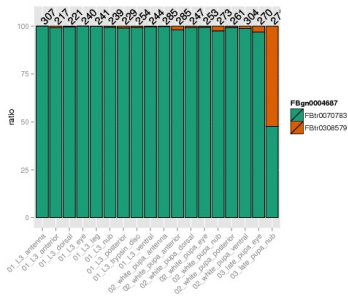
higher entropy \rightarrow isoforms of a given gene are more uniformly expressed



- ▶ vls.vt (i.e. gene expression contribution estimate)
- ▶ var.bwGp (i.e. proportion of variance explained by group classification)
Here variance means variance in transcript expression.
- ▶ var.bwGp.ge (i.e. contribution of gene expression in this variance that is explained by the groups).



vlsvt - selected



UCSC Genome Browser on *D. melanogaster* Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6) Assembly

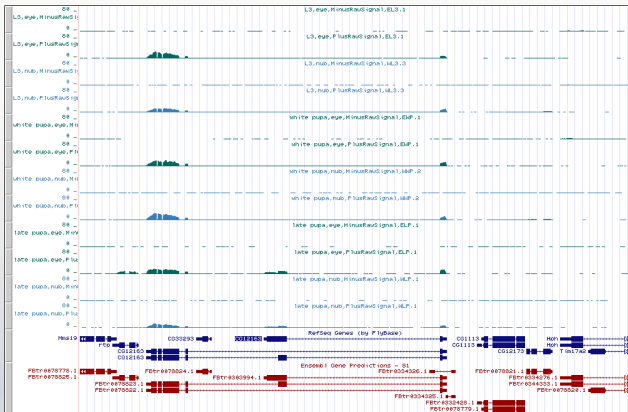
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr3R:5,234,558-5,260,339 25,782 bp. enter position, gene symbol or search terms

go

[hg38 replaces hg19 as default human ass.](#)

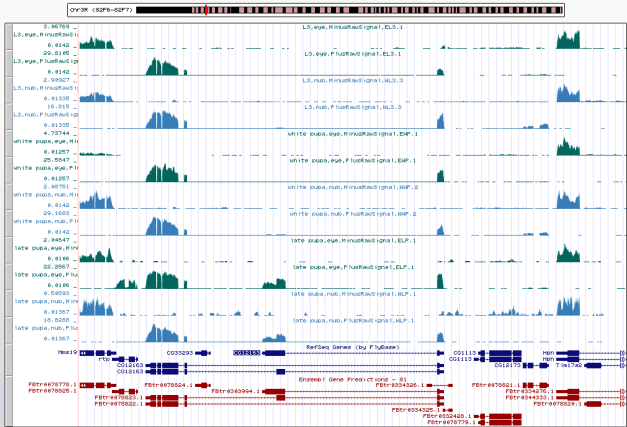
ov3R (62K-52P7)



UCSC Genome Browser on *D. melanogaster* Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr3R:5,234,558-5,260,339 25,782 bp. enter position, gene symbol or search terms go **hg38 replaces hg19 as default human ass**



Outline

Introduction

Overview of RNA-seq samples

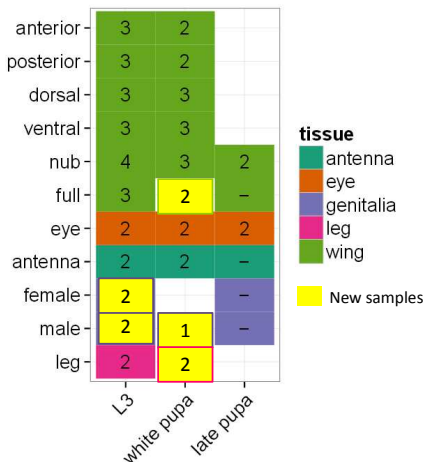
Wing and eye

Wing compartments

Isoform usage

Next steps

Just sequenced RNA-seq samples



Analysis

- ▶ Further characterize AP DV boundaries
- ▶ Differences in the relative abundance of isoforms (Jean Monlong - multiGroupSplicingComp)
- ▶ AS events - IPSA, AStalavista...
- ▶ ...
- ▶ Suggestions?



obrigado

Dank U

Merci

mahalo

Köszi

спасибо

Grazie

Thank
you

mauruuru

Takk

Gracias

Dziękuję

Děkuju

danke

Kiitos