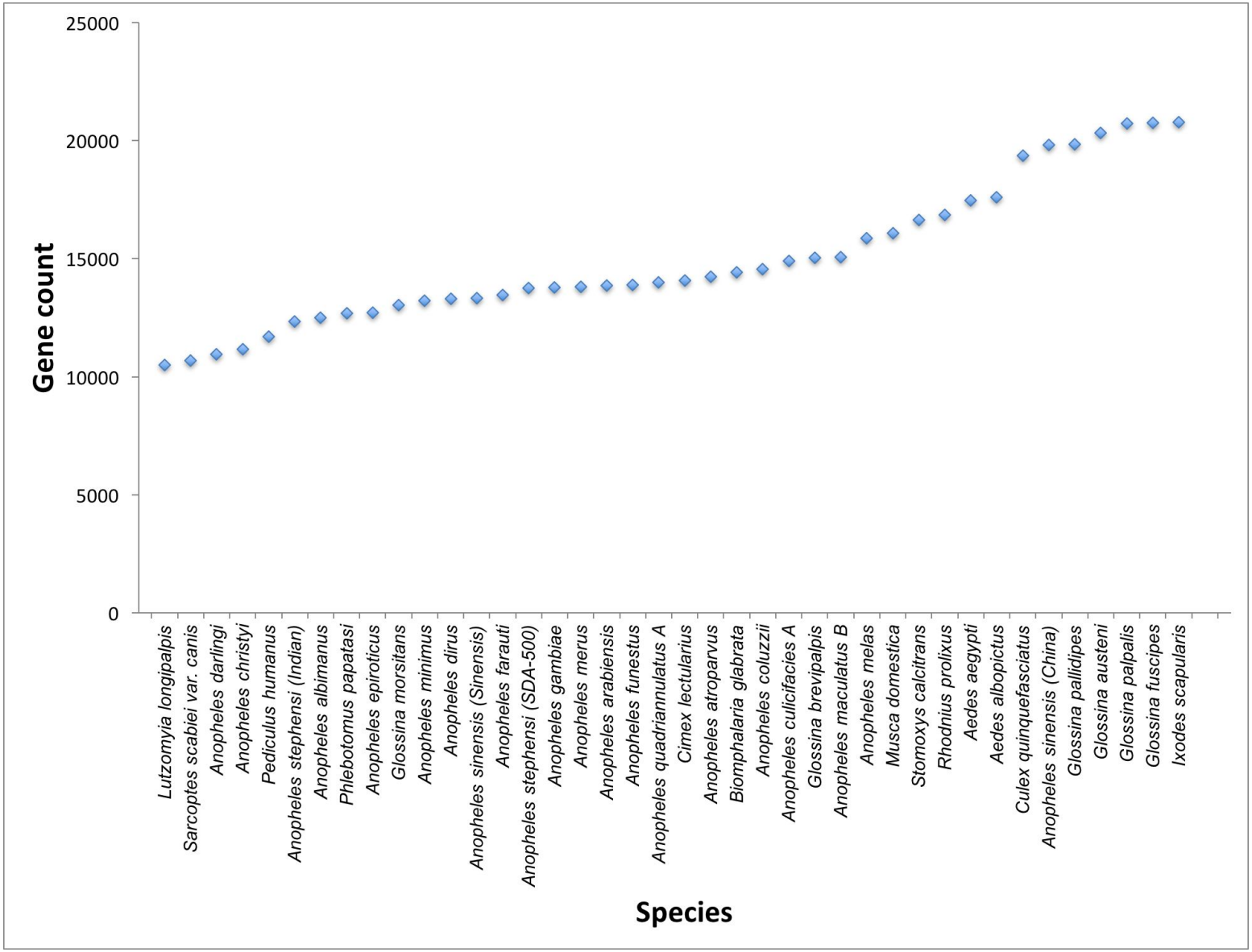
Transcript data and resources (part 1)

Gloria I. Giraldo-Calderón
October 2016

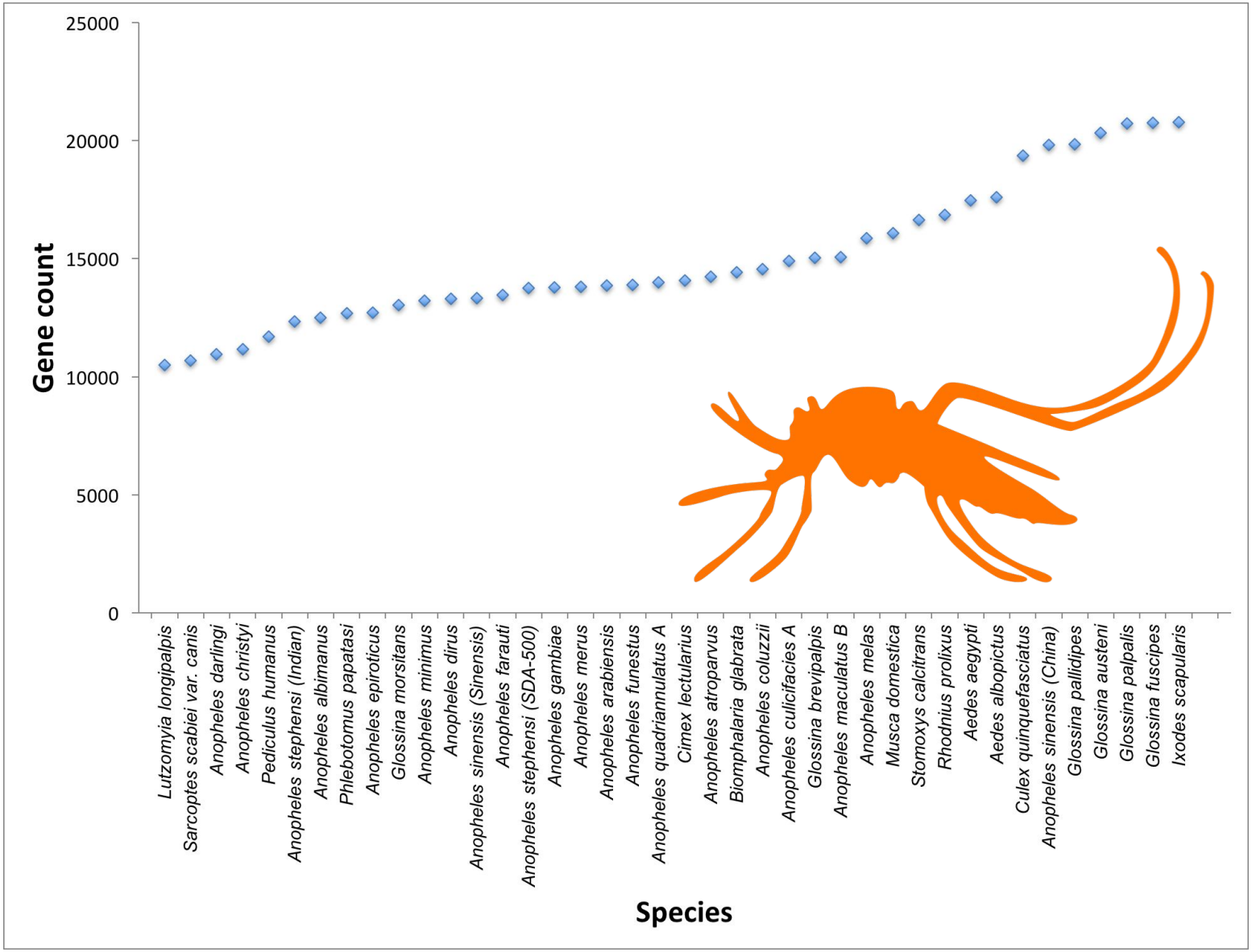
VectorBase

Bioinformatics Resource for Invertebrate Vectors of Human Pathogens

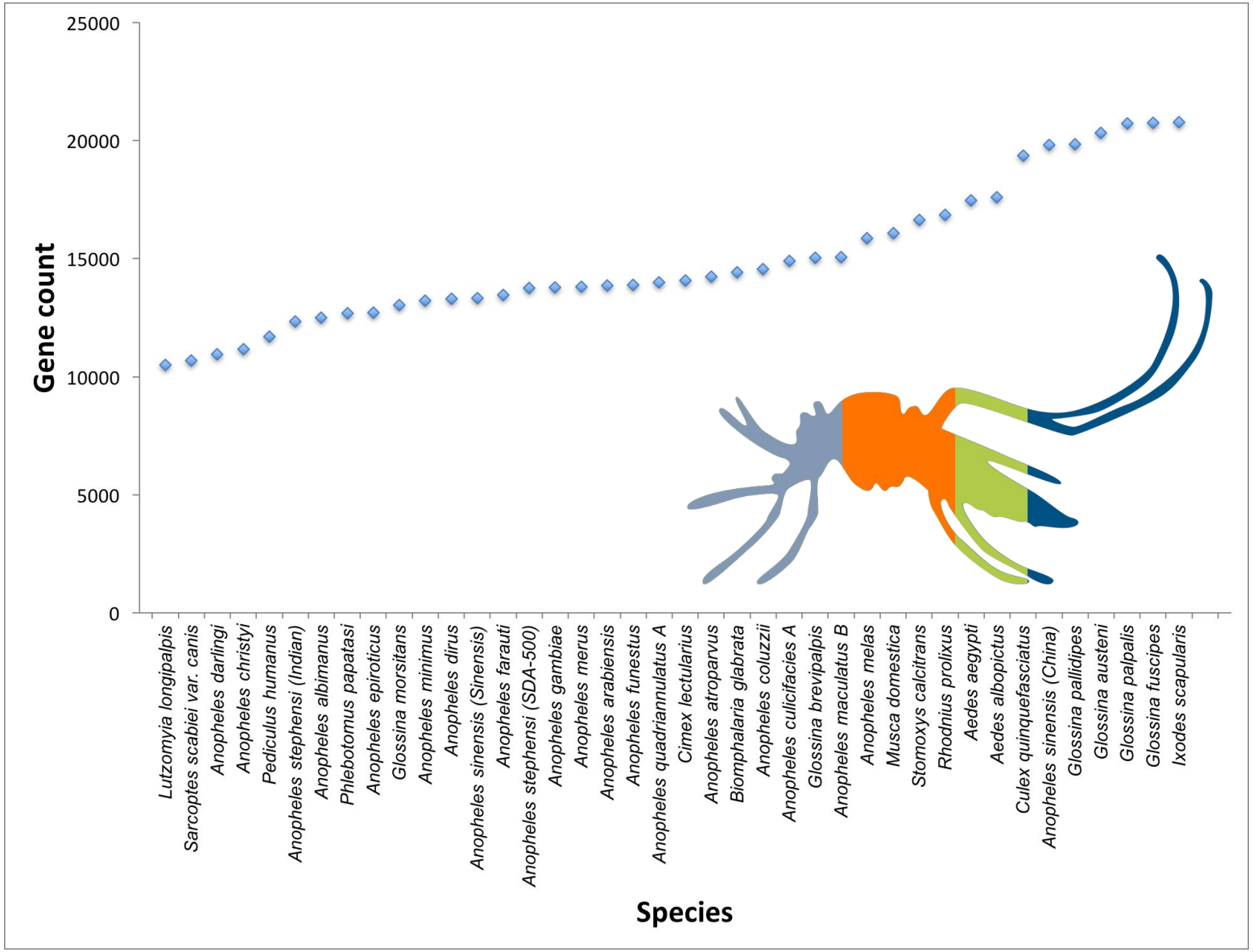
VectorBase Genomes



VectorBase Genomes



VectorBase Genomes



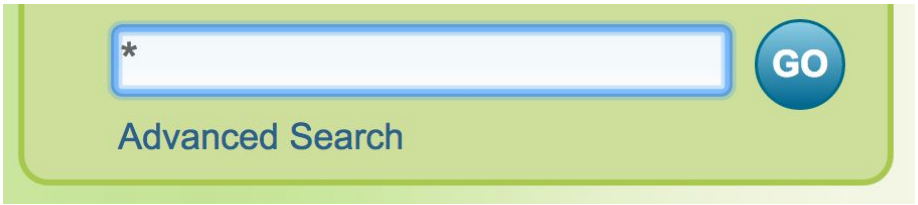
Outline

1. Available data
2. The basics about microarrays and RNAseq
3. What can you use the transcript expression data for
4. Browse gene expression data for a gene of interest
5. Assess gene expression data critically
6. Submit data

1. Available Data

Published and about-to-be-published
microarray and RNAseq data

1. Available Data



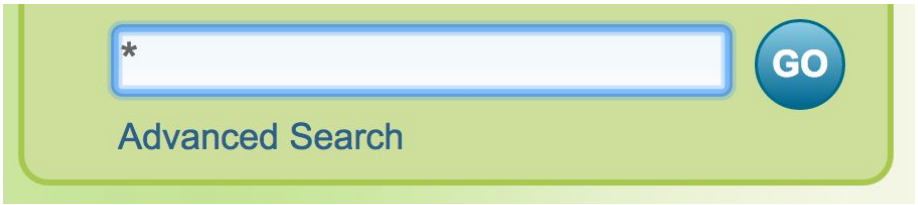
Filter Results

Domain (Reset Filter)	Hits
Expression	84

Sub-domain (Reset Filter)	Hits
Experiment	84

Species	Hits
Anopheles gambiae	47
Aedes aegypti	29
Anopheles funestus	2
Aedes albopictus	1
Culex quinquefasciatus	1
Musca domestica	1

1. Available Data



RNAseq tracks and track groups are not covered in this tutorial

Search

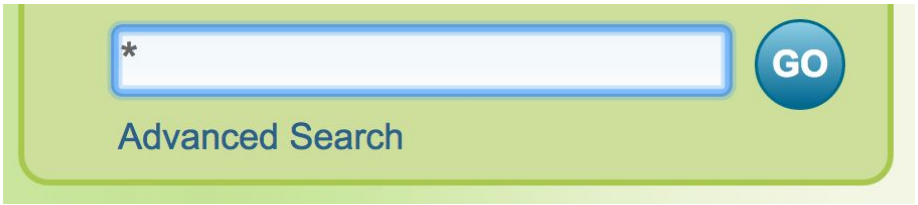
Filter Results

Domain (Reset Filter)	Hits
Expression	2469479

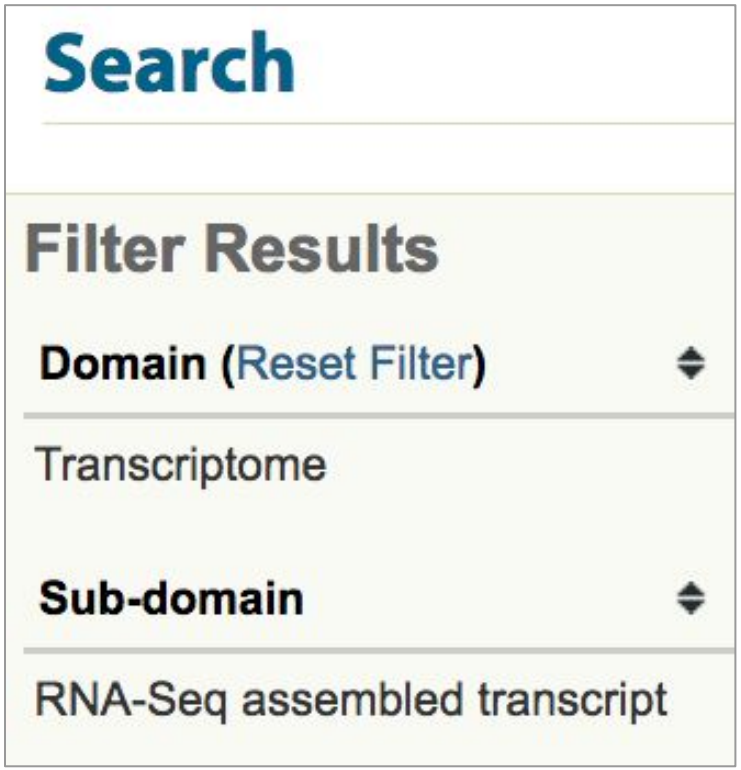
Sub-domain	Hits
Probe	1339156
Expression statistic	1128602
Sample	1332
RNA-seq tracks	223
Experiment	84
RNA-seq track groups	44
Platform	38

Species	Hits
Aedes aegypti	1493846
Anopheles gambiae	768626

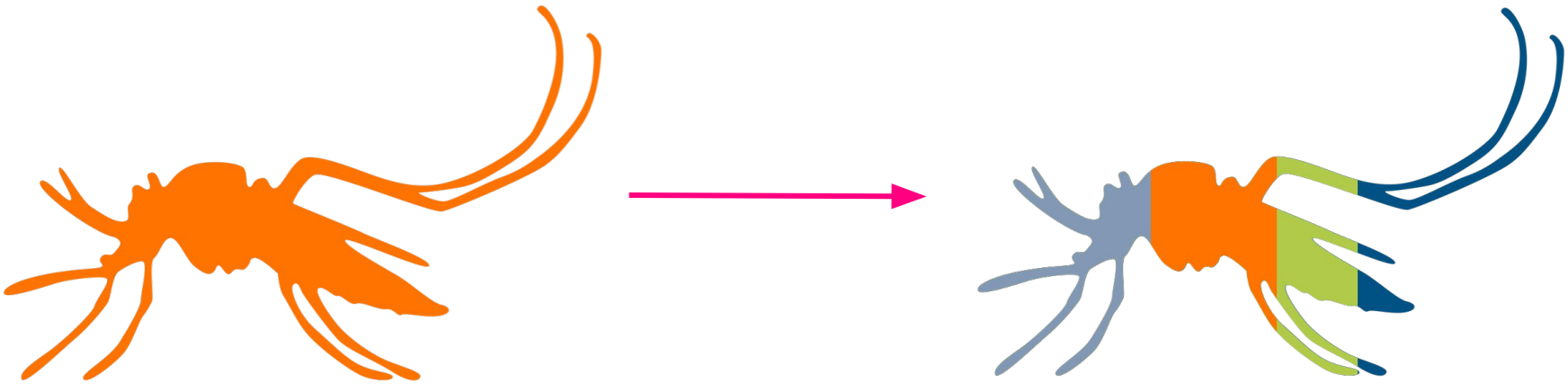
1. Available Data



not covered in this tutorial



2. The basics about microarrays and RNAseq



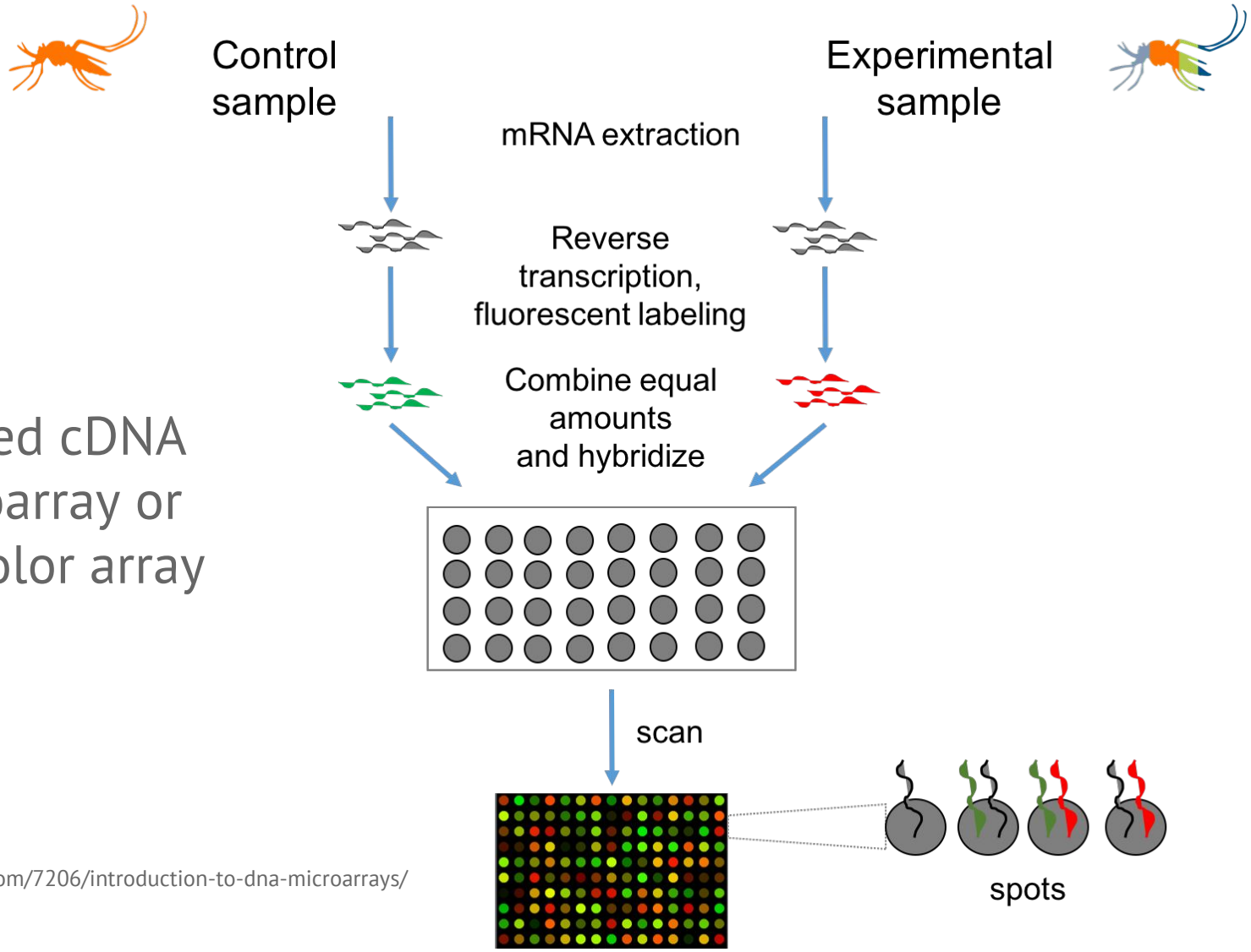
2. Basics about microarrays

- VectorBase has data for the two most popular microarrays:
  - spotted cDNA microarray or two-color array
  - oligonucleotide chips (from Genechip and Affymetrix)

- Despite of the technology, the basic principle in microarrays is hybridization to complementary sequences

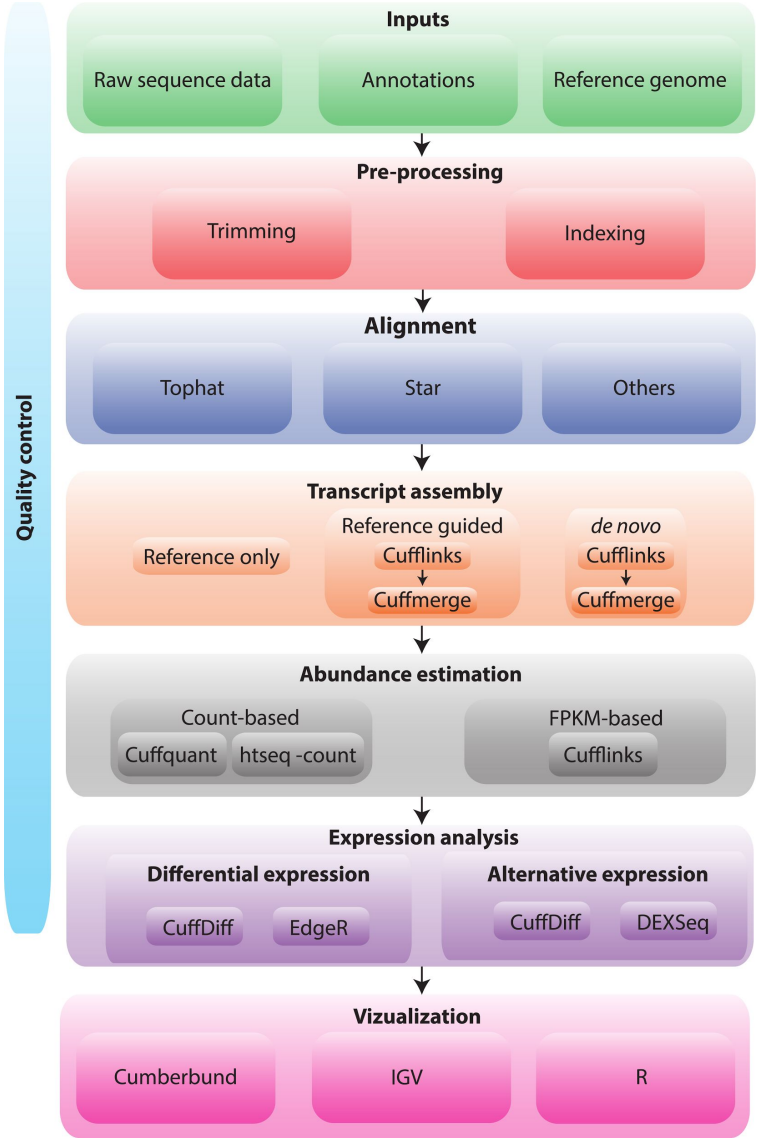
2. Basics about microarrays

spotted cDNA microarray or two-color array



Modified from
http://bitesizebio.com/7206/introduction-to-dna-microarrays/

2. Basics about RNA sequencing (RNAseq)



Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL (2015) Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. PLoS Comput Biol 11(8): e1004393. doi:10.1371/journal.pcbi.1004393

3. What can you use the transcript expression data for?

Microarrays:

- transcript expression data

RNAseq:

- transcript expression data (less expensive and more sensitive)
- other applications

3. What can you use the transcript expression data for?

Gene AGAP001212 Expression Report

Gene links:

Anopheles gambiae

Browse Genome

Expression Map

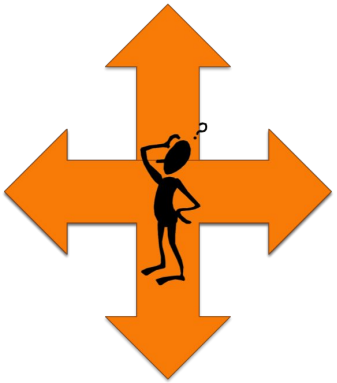
Probe Information show details:

Expression summary:

Experiment	P-value	Test	Experimental factor	Summary
Adult tissues (Baker et al., 2011) Microarray experiment info Plots and data	0.0	ANOVA	Organism part and sex 	Significant **differential expression** ↑ midgut:male ↓ ovaries:female
Circadian rhythm: heads, light-dark (Rund et al., 2011) Microarray experiment info Plots and data	1.6e-12	ANOVA	Time 	Significant **differential expression** ↑ 36.0 h ↓ 28.0 h
Embryonic development (Goltsev et al., 2009) Microarray experiment info	2.2e-09	ANOVA	Age 	Significant **differential expression** ↑ 43h

4. Browse gene expression data for a gene of interest

Demo



Keywords to look for genes with Search, Advanced Search and the Expression Browser Search:

- ✔ VectorBase gene ID
- ✘ Gene metadata (name and description) - Not recommended
- ✔ Terms from ontologies to describe experiments or samples

For batch queries (many genes or complete genomes) use BioMart.

4. Browse gene expression data for a gene of interest

- Independent of the route you chose, you will arrive to the **Expression Browser**. The difference is that this tool has different landing pages:
    - Expression report
    - Expression in experiment
- In either of these pages, how do you analyse the obtained results?

4. Browse gene expression data for a gene of interest

Gene AAEL006498 Expression Report

Gene links:

Aedes aegypti

Browse Genome

Expression Map

Probe Information show details:

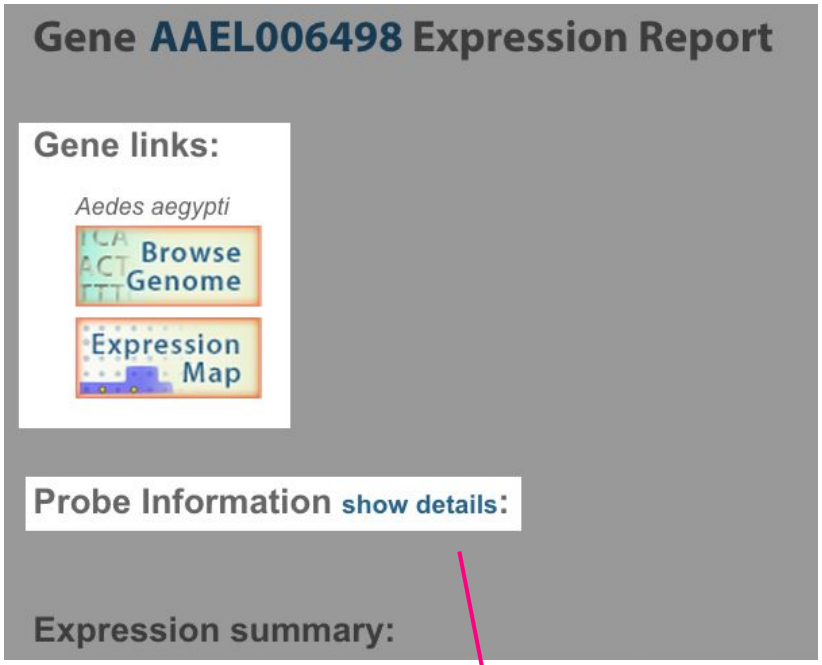
Expression summary:

Experiment	P-value	Test	Experimental factor	Summary
Hemocyte vs. carcass with bacterial infections (Choi et al., 2012) Microarray experiment info Plots and data	5.4e-05	t-test	Disease state and organism part	Significant **68.0-fold down-regulation** with respect to E. coli infected:hemolymph v carcass details
Click here for more results from Hemocyte vs. carcass with bacterial infections (Choi et al., 2012)				
Male vs. female (Dissanayake et al., 2010) Microarray experiment info Plots and data	0.00069	ANOVA	Sex	Significant **differential expression** ↑ male ↓ female details
Blood fed vs. Sugar Fed (Bonizzoni et al., 2011) RNA-Seq experiment info Plots and data	0.026	ANOVA	Growth condition	Significant **differential expression** ↑ sugar fed ↓ blood fed details

4. Browse gene expression data for a gene of interest

Gene links

Probe information



Probe Information hide details:

Uniquely associated with AAEL006498:

Probe ID	Probe annotation
AAEL006498-RAP00012	Probe sequence aligns to 2 genomic locations, 1 gene and 1 transcript.
AAEL006498-RAP00327	Probe sequence aligns to 2 genomic locations, 1 gene and 1 transcript.
AAEL006498-RAP00585	Probe sequence aligns to 2 genomic locations, 1 gene and 1 transcript.
AEG_V1.7272	Probe sequence aligns to 2 genomic locations, 1 gene and 1 transcript.
AaegL3.1_AAEL006498-RA	RNAseq reference transcript
AaegL3.3_AAEL006498-RA	RNAseq reference transcript

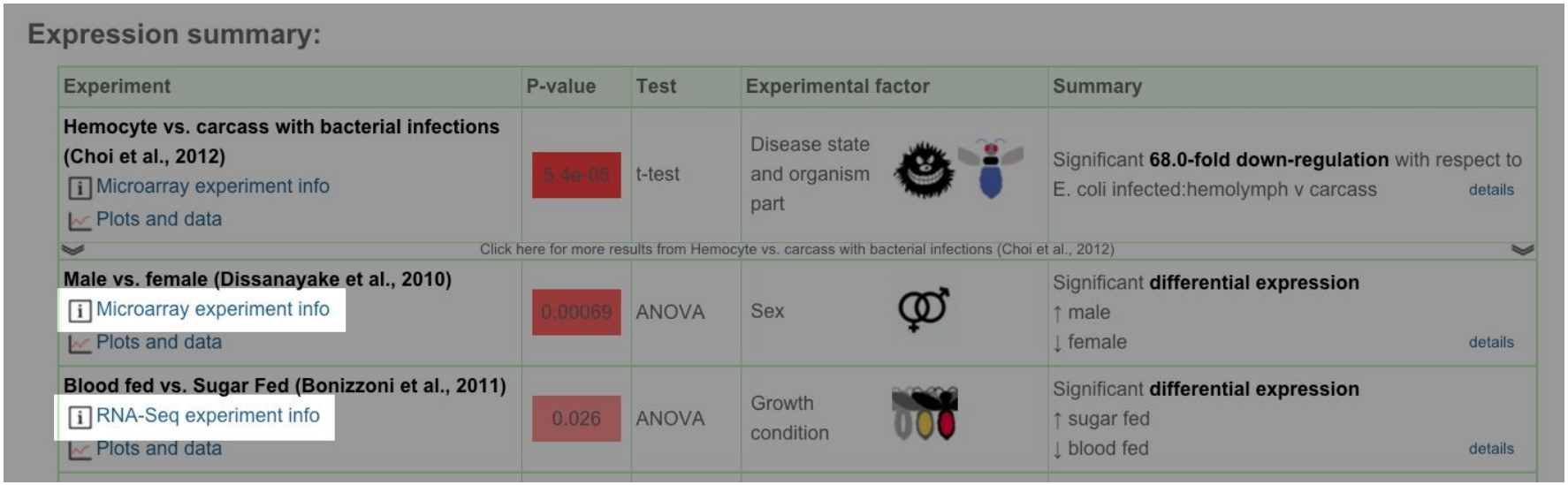
4. Browse gene expression data for a gene of interest

One-line experiment summary

Experiment	P-value	Test	Experimental factor	Summary
Hemocyte vs. carcass with bacterial infections (Choi et al., 2012) Microarray experiment info Plots and data	5.4e-05	t-test	Disease state and organism part 	Significant **68.0-f** E. coli infected:h

- **Experiment**
- **P-value**
- **Test**
- **Experimental factor**
- **Summary**

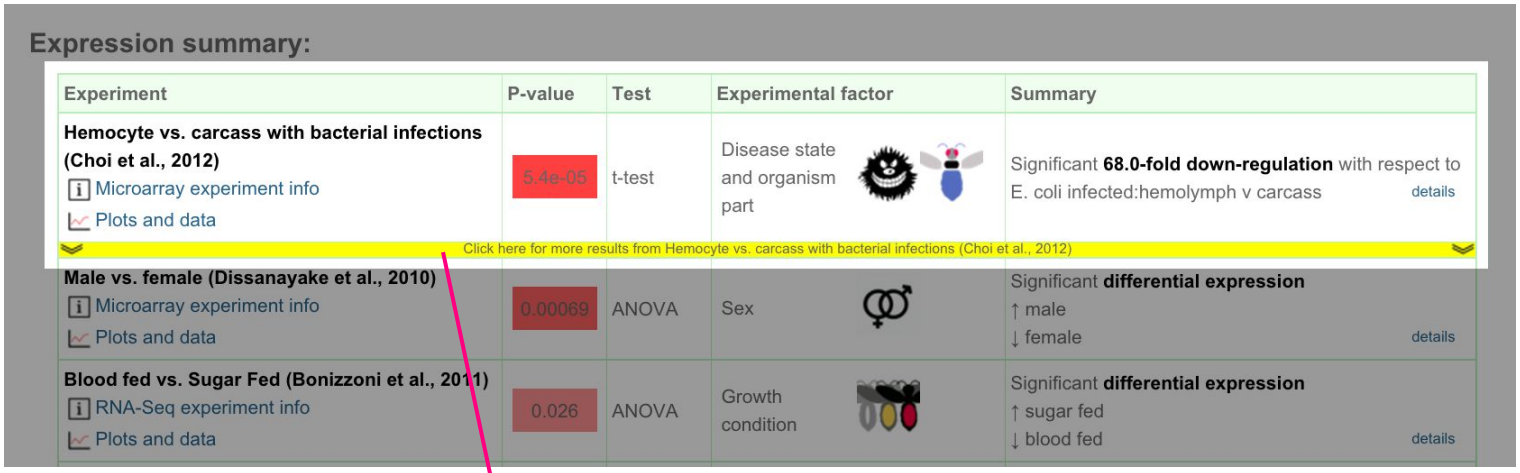
4. Browse gene expression data for a gene of interest

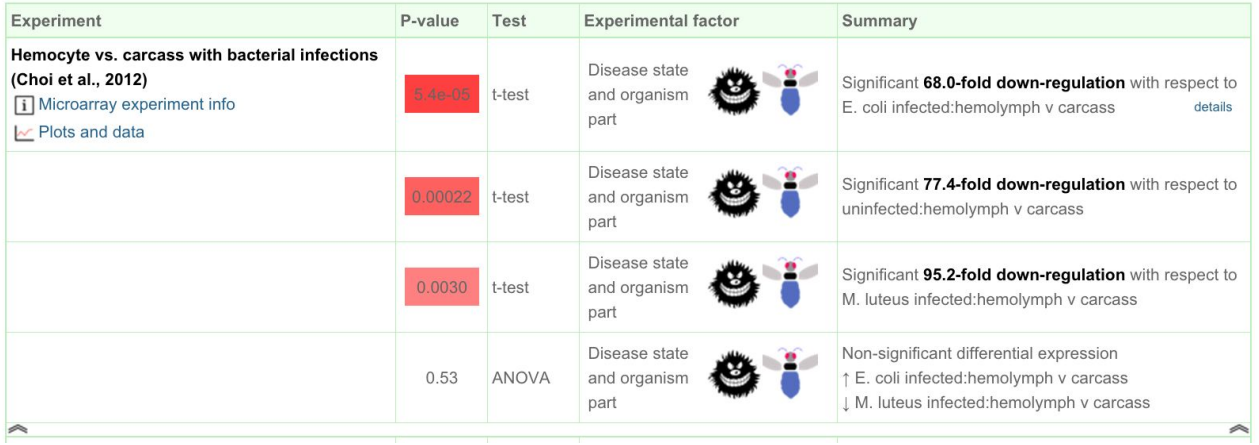


Expression summary:

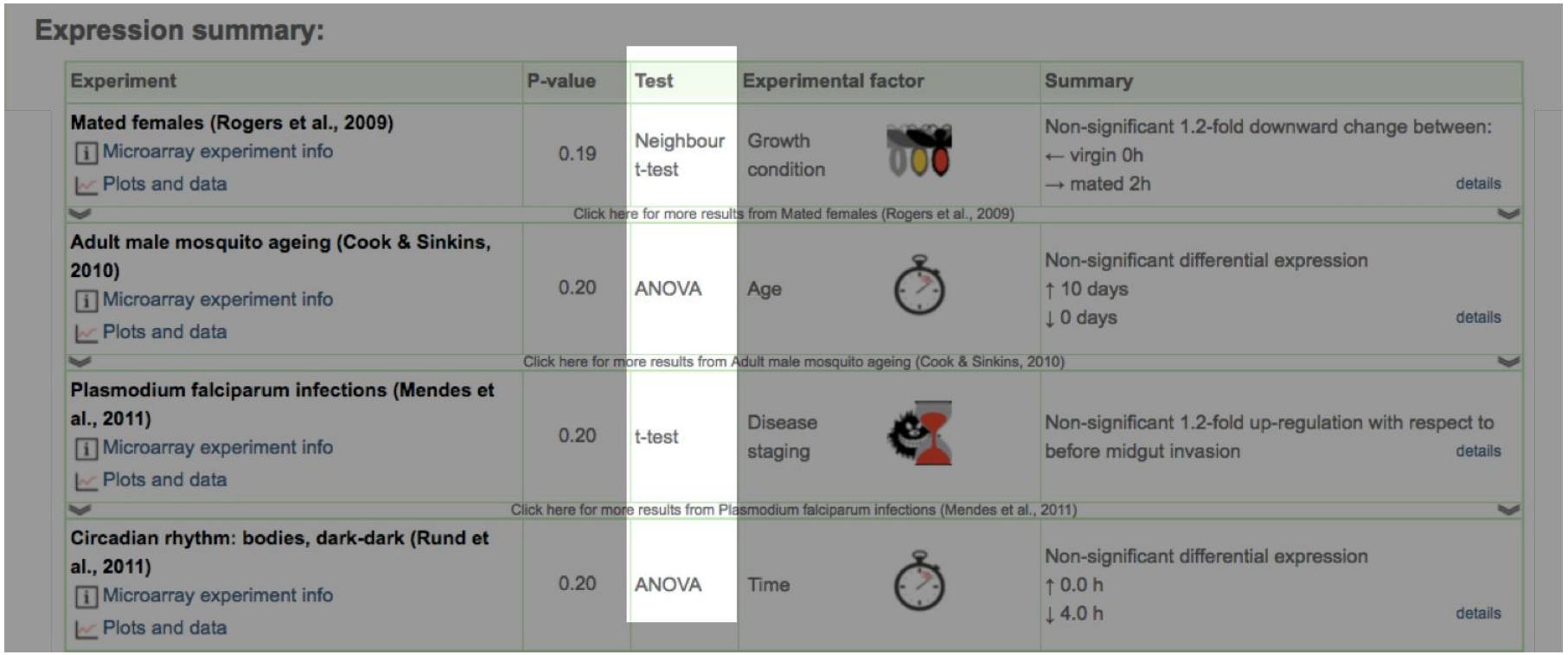
Experiment	P-value	Test	Experimental factor	Summary
Hemocyte vs. carcass with bacterial infections (Choi et al., 2012) Microarray experiment info Plots and data	5.4e-05	t-test	Disease state and organism part	Significant **68.0-fold down-regulation** with respect to E. coli infected:hemolymph v carcass details
Click here for more results from Hemocyte vs. carcass with bacterial infections (Choi et al., 2012)				
Male vs. female (Dissanayake et al., 2010) Microarray experiment info Plots and data	0.00069	ANOVA	Sex	Significant **differential expression** ↑ male ↓ female details
Blood fed vs. Sugar Fed (Bonizzoni et al., 2011) RNA-Seq experiment info Plots and data	0.026	ANOVA	Growth condition	Significant **differential expression** ↑ sugar fed ↓ blood fed details

4. Browse gene expression data for a gene of interest





4. Browse gene expression data for a gene of interest



Expression summary:

Experiment	P-value	Test	Experimental factor	Summary
Mated females (Rogers et al., 2009) Microarray experiment info Plots and data	0.19	Neighbour t-test	Growth condition	Non-significant 1.2-fold downward change between: ← virgin 0h → mated 2h details
Click here for more results from Mated females (Rogers et al., 2009)				
Adult male mosquito ageing (Cook & Sinkins, 2010) Microarray experiment info Plots and data	0.20	ANOVA	Age	Non-significant differential expression ↑ 10 days ↓ 0 days details
Click here for more results from Adult male mosquito ageing (Cook & Sinkins, 2010)				
Plasmodium falciparum infections (Mendes et al., 2011) Microarray experiment info Plots and data	0.20	t-test	Disease staging	Non-significant 1.2-fold up-regulation with respect to before midgut invasion details
Click here for more results from Plasmodium falciparum infections (Mendes et al., 2011)				
Circadian rhythm: bodies, dark-dark (Rund et al., 2011) Microarray experiment info Plots and data	0.20	ANOVA	Time	Non-significant differential expression ↑ 0.0 h ↓ 4.0 h details

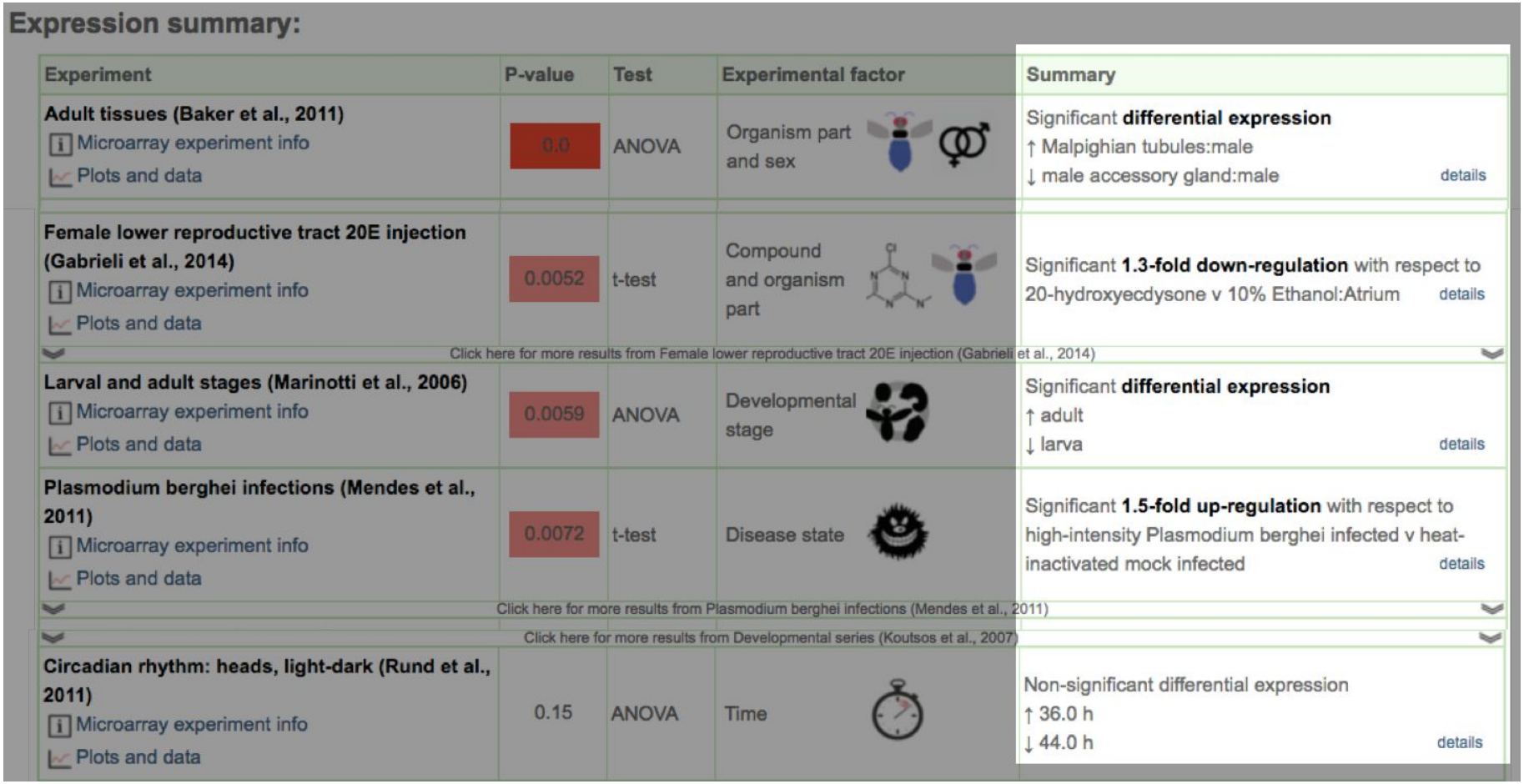
4. Browse gene expression data for a gene of interest

Expression summary:

Experiment	P-value	Test	Experimental factor	Summary
Adult tissues (Baker et al., 2011) Microarray experiment info Plots and data	0.0	ANOVA	Organism part and sex 	Significant **differential expression** ↑ Malpighian tubules:male ↓ male accessory gland:male details
MR4 cell-lines (AVCL consortium, 2014) RNA-Seq experiment info Plots and data	4.4e-12	ANOVA	Strain or line 	Significant **differential expression** ↑ MRA-920 L35 ↓ MRA-918 4a-3A details
Blood meal time series (Marinotti et al., 2006) Microarray experiment info Plots and data	3.8e-09	ANOVA	Growth condition 	Significant **differential expression** ↑ blood-fed 15d ↓ blood_fed 72h details
Click here for more results from Blood meal time series (Marinotti et al., 2006)				
Embryonic development (Goltsev et al., 2009) Microarray experiment info Plots and data	3.7e-08	ANOVA	Age 	Significant **differential expression** ↑ 43h ↓ 7h details
Click here for more results from Embryonic development (Goltsev et al., 2009)				
Female lower reproductive tract post-mating time-series (Gabrieli et al., 2014) Microarray experiment info Plots and data	2.1e-07	ANOVA	Developmental stage 	Significant **differential expression** ↑ virgin adult ↓ mated adult details

All 14 experimental factors can be found here:
http://goo.gl/24nG8r

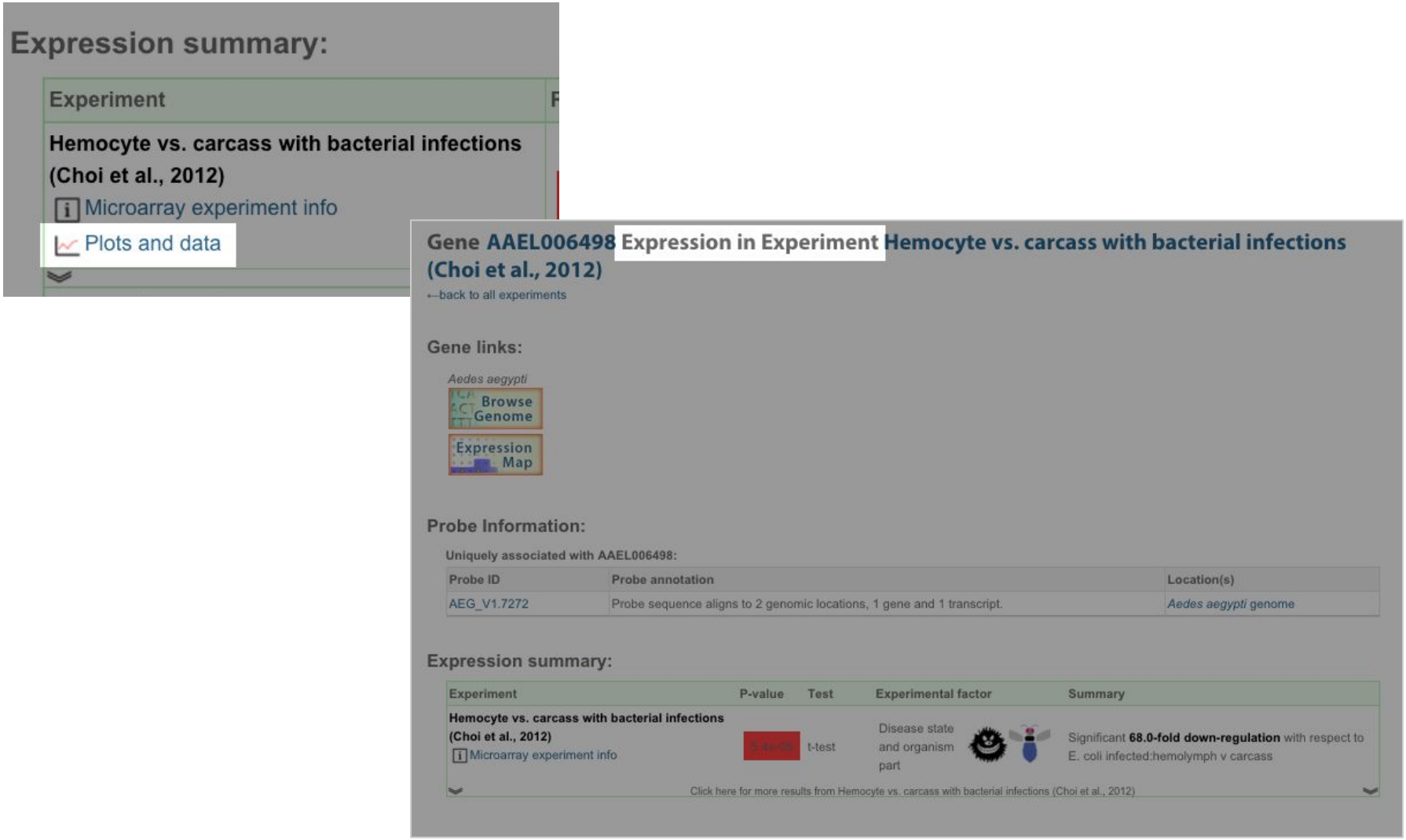
4. Browse gene expression data for a gene of interest



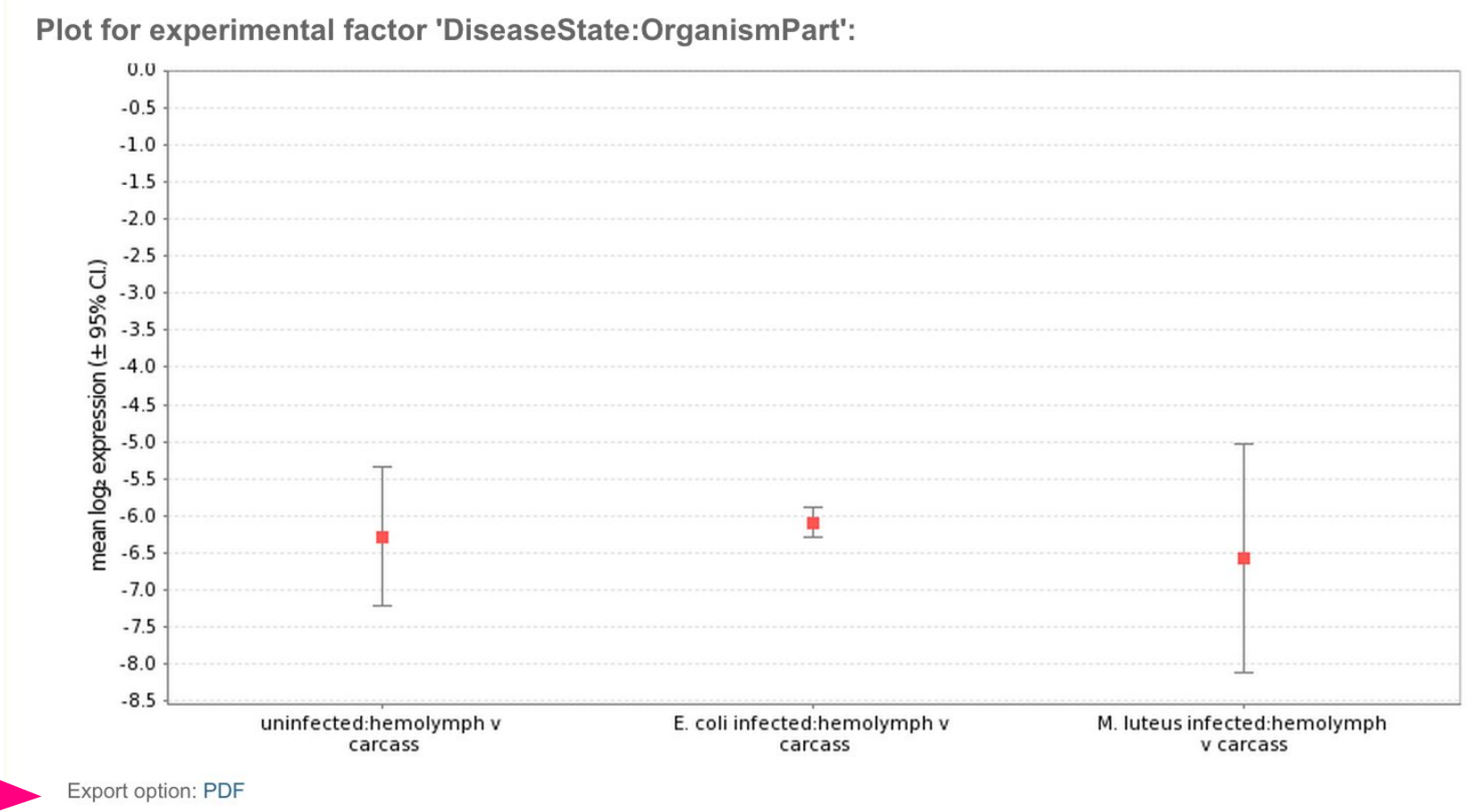
Expression summary:

Experiment	P-value	Test	Experimental factor	Summary
Adult tissues (Baker et al., 2011) Microarray experiment info Plots and data	0.0	ANOVA	Organism part and sex	Significant **differential expression** ↑ Malpighian tubules:male ↓ male accessory gland:male details
Female lower reproductive tract 20E injection (Gabrieli et al., 2014) Microarray experiment info Plots and data	0.0052	t-test	Compound and organism part	Significant **1.3-fold down-regulation** with respect to 20-hydroxyecdysone v 10% Ethanol:Atrium details
Click here for more results from Female lower reproductive tract 20E injection (Gabrieli et al., 2014)				
Larval and adult stages (Marinotti et al., 2006) Microarray experiment info Plots and data	0.0059	ANOVA	Developmental stage	Significant **differential expression** ↑ adult ↓ larva details
Plasmodium berghei infections (Mendes et al., 2011) Microarray experiment info Plots and data	0.0072	t-test	Disease state	Significant **1.5-fold up-regulation** with respect to high-intensity Plasmodium berghei infected v heat-inactivated mock infected details
Click here for more results from Plasmodium berghei infections (Mendes et al., 2011)				
Click here for more results from Developmental series (Koutsos et al., 2007)				
Circadian rhythm: heads, light-dark (Rund et al., 2011) Microarray experiment info Plots and data	0.15	ANOVA	Time	Non-significant differential expression ↑ 36.0 h ↓ 44.0 h details

4. Browse gene expression data for a gene of interest



4. Browse gene expression data for a gene of interest



4. Browse gene expression data for a gene of interest

Data tables for **microarrays or RNAseq**:

Plot data in tabular form:

DiseaseState:OrganismPart	mean	stdev	min	lower 95% C.I.	median	upper 95% C.I.	max	n
uninfected:hemolymph v carcass	-6.27	0.59	-6.76	-7.21	-6.38	-5.34	-5.57	4
E. coli infected:hemolymph v carcass	-6.09	0.08	-6.14	-6.28	-6.13	-5.9	-6.00	3
M. luteus infected:hemolymph v carcass	-6.57	0.62	-7.08	-8.12	-6.76	-5.03	-5.88	3

Export options: CSV | XML

Only for **microarrays**:

Spot data:

20 items found, displaying all items.

Hybridisation	DiseaseState:OrganismPart	Probe / Probe Set	position	value
hemocyte-naive-cy3-rep4_Cy3_carcass-naive-cy5-rep4_Cy5	uninfected:hemolymph v carcass	AEG_V1.7272	23364	-5.55
hemocyte-naive-cy3-rep4_Cy3_carcass-naive-cy5-rep4_Cy5	uninfected:hemolymph v carcass	AEG_V1.7272	2064	-5.58
hemocyte-mlut-cy3-rep3_Cy3_carcass-mlut-cy5-rep3_Cy5	M. luteus infected:hemolymph v carcass	AEG_V1.7272	23364	-5.82

Only for **RNAseq**:

RNA-Seq data:

4 items found, displaying all items.

Sequencing Run	GrowthCondition	Ref Transcript	position	value
LSF_A.amplification	sugar fed	AaegL3.1_AAEL006498-RA	880	11.81
LSF_B.amplification	sugar fed	AaegL3.1_AAEL006498-RA	880	11.79
LBF_B.amplification	blood fed	AaegL3.1_AAEL006498-RA	880	10.98
LBF_A.amplification	blood fed	AaegL3.1_AAEL006498-RA	880	10.65

4 items found, displaying all items. Export options: CSV | XML

5. Assess gene expression data critically

Microarrays:

- only provide an indirect estimate for the expression of a gene
- probes may give unexpected results for a number of reasons

RNAseq:

- results are affected by various experimental/bench factors
- critical considerations: number of replicates, statistical analyses and technique standardization

5. Assess gene expression data critically

- Wide error bars or poor P-values
- Transcript abundance
- population, organism or strain differences

5. Assess gene expression data critically

Demo

- Data from multiple reporters is usually pooled and averaged in VectorBase

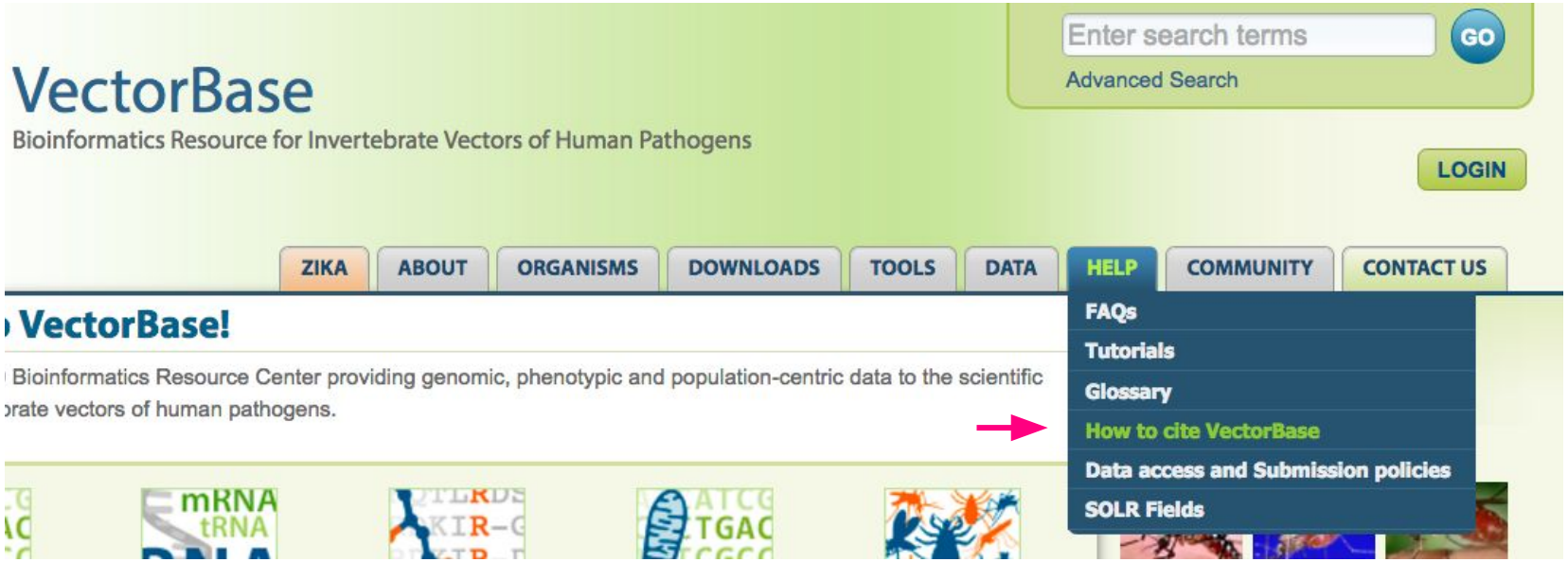
- If one reporter gives markedly different results, you may want to find why

- Navigate to the genome browser and use ‘Configure this page’ to check the location of the reporters

6. Submit data

Making your data available to the community and submission policy for VectorBase and archival repositories.

Citations


Enter search terms
GO
Advanced Search
VectorBase
Bioinformatics Resource for Invertebrate Vectors of Human Pathogens
LOGIN
ZIKA
ABOUT
ORGANISMS
DOWNLOADS
TOOLS
DATA
HELP
COMMUNITY
CONTACT US
VectorBase!
Bioinformatics Resource Center providing genomic, phenotypic and population-centric data to the scientific
rate vectors of human pathogens.
FAQs
Tutorials
Glossary
How to cite VectorBase
Data access and Submission policies
SOLR Fields
mRNA
tRNA

How to search for more information or help?

E-mail us at
info@vectorbase.org

Thank you!