



Functional genomics: microarray and RNAseq data

Expression Browser

Answer key

1. Exploring the available experiments

Let's use Advanced Search to explore the available data.

Click on the "Advanced Search" link in the top right corner of any VectorBase page.

Click on the "Expression" domain filter and "Experiment" subdomain filter (in the "Filter Results" section on the left side).

Question 1.1

How many experiments are available for *Aedes aegypti*?

40

Now expand the Advanced Search form and add the field "Experimental Factors" and select OrganismPart. See the screenshot below for guidance.

Click on "GO".

Search

Filter Results

Domain (Reset Filter)	Hits
Expression	77
Sub-domain (Reset Filter)	Hits
Experiment	77
Species	Hits
Anopheles gambiae	46
Aedes aegypti	25
Anopheles funestus	2
Culex quinquefasciatus	1

Enter terms

GO

Cannot find what you are looking for? Try a [global search](#)

Advanced Search

Domain/Sub-Domain

-Experiments
Add field

Experimental Factors

GrowthCondition
DiseaseState
OrganismPart

☒ Experimental Factors
☐ Species
☐ Description
☐ PubMed

1-20 OF 77 RESULTS

1
2
3
4
next >

Developmental series (Koutsos et al., 2007)

Expression > Experiment

Seven different *Anopheles gambiae* developmental stages (from embryo to adult) are profiled using (rather than instar-based). The expression...

Adult female tissues (Koutsos et al., 2007)

Expression > Experiment

Question 1.2

How many experiments are shown in the results?

12

Question 1.3

Think about the concept of "Experimental Factor". Which of the following most accurately describes the search you just performed?

	Answer
Find the experiments in which the expression of genes was compared between samples taken from different organism parts.	X
Find the experiments in which samples were dissected and only certain organism parts were used.	
Find the organism parts where genes are upregulated in expression experiments.	

Question 1.4

How many experiments in *Anopheles gambiae* are measuring gene expression differences relative to **disease state**?

12

2. Searching for differentially expressed genes

We will now search for odorant binding protein genes (more precisely, genes with symbols beginning with 'OBP...') that have high expression in certain organism parts.

Start a new Advanced Search and select the following subdomain and filter fields, as shown in the screenshot:

Domain/Sub-Domain: **Expression Statistic**

Experimental Factors: **OrganismPart**

Gene symbol: **OBP***

High expression....: **head**

The screenshot shows the 'Advanced Search' interface with the following configuration:

- Domain/Sub-Domain:** A dropdown menu showing '-Expression Statistic' with an 'Add field' button.
- Experimental Factors:** A list box containing 'GrowthCondition', 'DiseaseState', and 'OrganismPart' (which is selected).
- Gene Symbol:** A text input field containing 'OBP*'.
- High expression/upregulated in condition/tissue:** A text input field containing 'head'.

A modal window is open over the 'Experimental Factors' list, showing a checklist of available fields:

- ☒ Experimental Factors
- ☐ Species
- ☐ Gene ID
- ☐ Probe/Reporter ID
- ☒ Gene Symbol
- ☒ High expression/upregulated in condition/tissue
- ☐ Low expression/downregulated in condition/tissue
- ☐ Experiment

Question 2.1

How many different experiments do the results come from?

4

Question 2.2

Repeat the query with "midgut" instead of "head".

Note that a new experiment appears. We can exclude this from the results using an extra advanced search field:

Add field -> Experiments -> (NOT compartments)



The screenshot shows a web interface for searching experiments. There is a search bar labeled 'Experiment' with a dropdown menu. The dropdown menu is open, showing a list of options. The option '(NOT compartments)' is selected and highlighted in blue.

and click "GO" to check that works.

Now let's get the results from just one experiment: "Adult tissues (Baker et al., 2011)".

Type "Baker" into the Experiment filter box and click on "GO".

Fact 1: VectorBase expression search only returns expression statistics with P-values less than 0.05 (no multiple testing correction).

Fact 2: Expression statistics in VectorBase search are calculated for each microarray probe (or RNA-seq reference transcript). If you want to calculate the number of unique genes, click "Export results" and use an offline tool or spreadsheet to calculate the number of unique values in the 'gene_ids' column.

Now compare the results from "head" and "midgut" from the Baker et al. tissues experiment.

Which of the following statements is true?

	Answer
More genes are highly expressed in head than in midgut according to the Baker et al experiment.	
More OBP genes are highly expressed in head than in midgut according to the Baker et al experiment.	X
More OBP genes are highly expressed in midgut than in head according to the Baker et al experiment.	

3. Understanding expression reports

Change your previous search from Question 2.2 back to **High expression in 'head'**. The first hit should be for a probe corresponding to gene **OBP9** (OBP20). If not scroll down to find it, and click on it.

You will land on a page entitled:

Probe Set **Ag.X.4.0_CDS_at** Expression in Experiment **Adult tissues (Baker et al., 2011)**

Note: A "Probe Set" is roughly equivalent to a "Probe". Probe Sets are only used on microarrays manufactured by Affymetrix. We will usually refer to Probe Sets as Probes for simplicity.

Take some time to find the following items on this page:

- Which gene is associated with this probe (it should be OBP9!)
- A link to the genome browser showing where the probe sequence aligns to the genome
- A link to an information page for the experiment
- The "Expression summary"
- An expression profile plot

Did you know...? Probe-level expression summaries can be downloaded using BioMart (Expression database). In BioMart, probes are called "Reporters".

Question 3.1

What does this page show? Tick the most appropriate answer.

	Answer
The levels of mRNA detected by one probe in one experiment	X
A summary of OBP9 gene expression	
A summary of the Adult tissues (Baker et al., 2011) experiment.	

Question 3.2


Examine the expression plot. Which of the following statements are true about the mRNA expression levels measured by this probe?

	True	False
Error bars are +/- one standard deviation.		X
Expression in heads (male or female) is significantly higher than other individual body parts at the 95% level.	X	
Female head expression is significantly higher than male head expression at the 95% level.		X
Expression in whole males is significantly higher than whole females at the 95% level.	X	

Question 3.3

In the Probe Information section, right-click on the links under "Probe ID" and "Associated gene(s)" to open **two new browser tabs**.

One page is a gene-based summary, the other is a probe-based summary. Look at both carefully and mark the following statements as true or false.

	True	False
Probe expression reports contain fewer experiments because probes are smaller than genes.		X
Gene expression reports contain more experiments because they show the data from <i>all</i> probes from <i>all</i> microarrays that correspond to the gene of interest, and because experiments can be performed with different microarrays.	X	
Both probe and gene expression reports list the experiments in ascending P-value order.	X	
The two male vs female experiments (look for the  icon on its own) are in disagreement with each other.		X

4. When a gene has more than one probe

Some microarrays are designed with several probes per gene. The probes can correspond to different exons and VectorBase's expression reports can provide information about the expression of different isoforms.

Start a new search for the gene GSTD1 in *Anopheles gambiae*.

Fine the gene page and locate the "Expression report" link on the left hand side and click on it.

Anopheles gambiae (AgamP4) Location: 2R:50,997,215-51,001,446 Gene: GSTD1

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Sequence
- External references
- Regulation
- Expression report
- Ontologies
 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process
- Pathways
- PubMed (1)
- Comparative Genomics
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Phenotype
- Genetic Variation
 - Variant table
 - Structural variants
 - Variant image
 - External data

Gene: GSTD1 AGAP004164

Description glutathione S-transferase delta class 1 [Source:VB Community Annotation;Acc:AGAP004164]

Location [Chromosome 2R: 50,997,215-51,001,446](#) forward strand.
AgamP4:CM000357.1

About this gene This gene has 4 transcripts ([splice variants](#)), [61 orthologues](#) and [24 paralogues](#).

Transcripts [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Translation ID	Biotype	UniProt	RefSeq	Flags
GSTD1	AGAP004164-RA	902	209aa	AGAP004164-PA	Protein coding	Q93112	XM_313049 XP_313049	
GSTD1	AGAP004164-RB	898	209aa	AGAP004164-PB	Protein coding	Q93113	XM_313050 XP_313050	
GSTD1	AGAP004164-RD	818	186aa	AGAP004164-PD	Protein coding	O77462	XM_562680 XP_562680	
GSTD1	AGAP004164-RC	805	216aa	AGAP004164-PC	Protein coding	O77473	XM_313048 XP_313048	

Summary

The gene expression report lists a lot of experiments, but the "Embryonic development" experiment will be listed near the top of the table. Click on the "Plots and data" link.

Examine the large plot and the four smaller plots.

Question 4.1

Which of the following statements are true?

	True	False
The four small plots show the expression of neighbouring genes		X

The four small plots show the expression of mRNA as measured by four different probes present on the microarray used in this experiment.	X	
There are two qualitatively different expression patterns in the four small plots.		X
The large plot shows the expression values averaged over all four probes.	X	
Clicking on a small plot opens a single probe-experiment report page	X	

Question 4.2

Now examine the gene-experiment pages for this gene in **Adult tissues (Baker et al., 2011)** and **Blood meal time series (Marinotti et al., 2006)** and compare with each other and the Embryonic development experiment.

Which of the following statements is the most accurate reflection of what the data tells you?

	Answer
All probes report approximately similar expression patterns	
Each experiment shows two different patterns of expression, always involving the same two pairs of probes: Ag.2R.71.5_s_at/Ag.2R.71.2_CDS_at and Ag.2R.71.0_CDS_at/Ag.2R.71.4_CDS_a_at	
Probes Ag.2R.71.5_s_at and Ag.2R.71.2_CDS_at show mostly similar expression patterns in all three experiments, but there are some differences in tissue expression. Probes Ag.2R.71.0_CDS_at and Ag.2R.71.4_CDS_a_at have roughly similar patterns in tissues and blood feeding but very different expression in embryonic development.	X

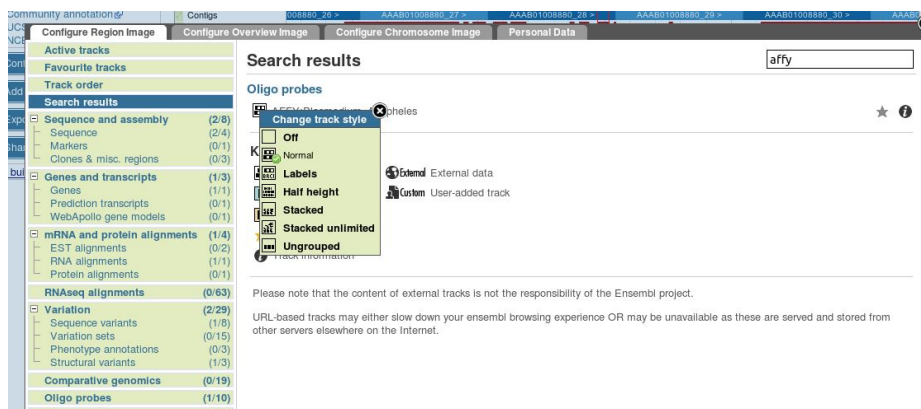
Question 4.3

Go back to the genome browser gene page for AGAP004164 and click on the **Location** tab or link.

Click on the **gear icon** in the bottom image and switch on the track indicating where the probe sequences on the Affymetrix array align to the genome as follows:

- Type the word "affy" into the popup window's search box, top right and press **enter**.
This locates the track configuration for the Affymetrix array probes.

- Click on the square/checkbox and choose "Normal" from the "Change track style" menu, as shown below.



- Close the popup window by clicking outside it or on the black tick icon, top right.

You should now see a graphic showing the location of the probes with respect to the different exons of this gene.

Click on the green probe features to bring up information popups. Now you can identify which probes are which.

Which probe measures the expression of all four alternative transcripts?

	Answer
Ag.2R.71.5_s_at	X
Ag.2R.71.0_CDS_at	
Ag.2R.71.4_CDS_a_at	
Ag.2R.71.2_CDS_at	

Question 4.4

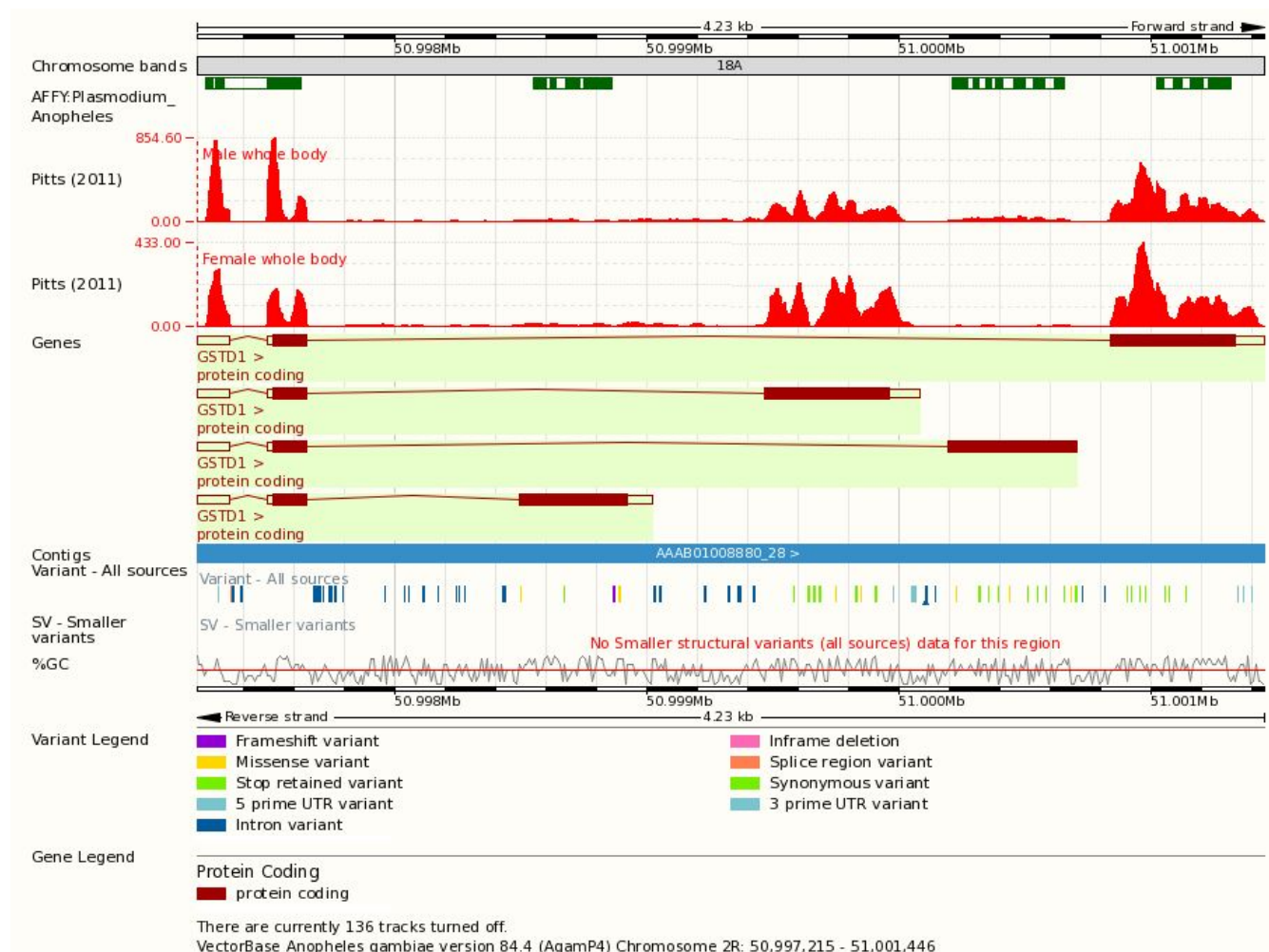
Which transcript's expression is measured by only one probe on the Affymetrix microarray?

	Answer
AGAP004164-RA	X
AGAP004164-RB	

AGAP004164-RC	
AGAP004164-RD	

Optional exercise: RNA-seq

Use the same gear icon "configure this page" link to add "wiggle plot" **RNA-seq tracks** for "Male whole body" and "Female whole body" - you should see something like the image below. Which transcript isoforms are most highly expressed? **AGAP004164-RB**, **AGAP004164-RA** Do any of the other RNA-seq tracks demonstrate expression of the lowly expressed isoforms?



Optional exercise: Expression Map

Go to the *Anopheles gambiae* expression map from the AGAP010489/OBP4 gene expression report page, using the icon at the top of the page:



Each gray circle is a cluster of genes with similar expression. Larger circles indicate larger clusters.

The yellow bar indicates in which cluster contains OBP4. Click on a cluster to get more information.

Which other OBPs have very similar expression?

AGAP001409 (OBP3)

AGAP001556 (OBP7)

AGAP005208 (OBP20)

AGAP007287 (OBP47)

Genes in neighbouring clusters also have similar expression patterns.

To find more OBP genes, search in the orange box with **GO:0005549**

Different regions of the map represent different expression patterns. You can find them with the red, green and blue menus and sliders.

How many "broad expression profile classes" do the OBPs fall into?

3/4