

## Using GUDMAP - Tutorial

This tutorial illustrates how the GUDMAP website can be used to search gene expression data to find candidate genes that may play a role in a known disease.

The starting point is a known disease – **Renal Adysplasia**

**We want to know a few things:**

- 1. What genes are associated with this disease?**
- 2. Where are genes associated with this disease expressed in the kidney?**
- 3. What other genes share similar expression patterns to these genes?**

---

### 1. Using GUDMAP Disease Resource

Navigate to *Disease > Query Disease-Genes Associations*

URL: [http://www.gudmap.org/gudmap\\_dis/Dis\\_Query.jsp](http://www.gudmap.org/gudmap_dis/Dis_Query.jsp)

Use the search box to select or type in our disease of interest. The list of disease names is taken from OMIM. Using this query we can find if there are known genes associated with our disease of interest.

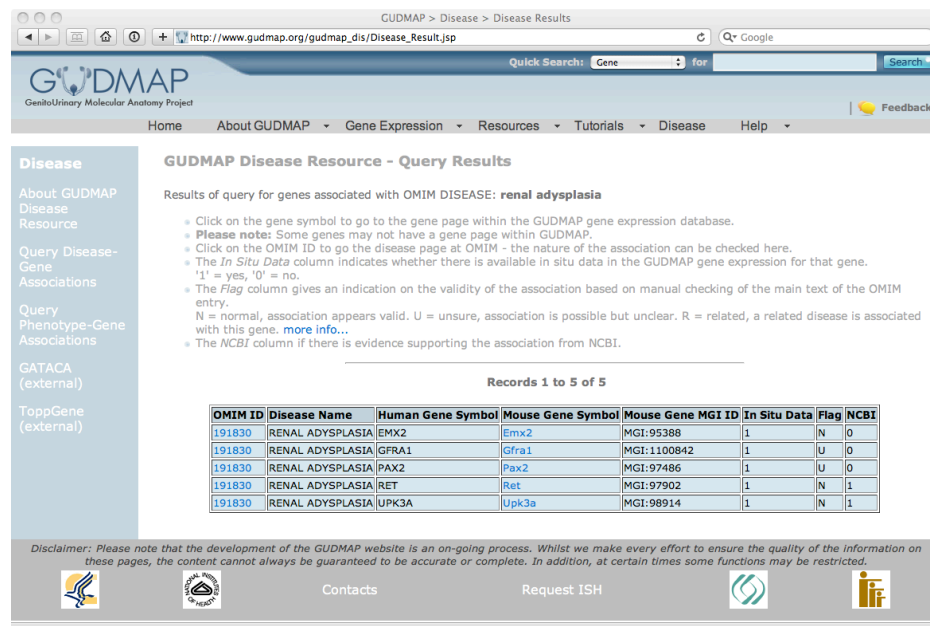
The screenshot shows a web browser window with the URL [http://www.gudmap.org/gudmap\\_dis/Dis\\_Query.jsp](http://www.gudmap.org/gudmap_dis/Dis_Query.jsp). The page title is "Querying GUDMAP Disease-Genes Associations". The main content area has a heading "How are disease-gene associations determined?" and a section "To search for genes associated with a genitourinary disease:". Below this, there is a "Select Disease:" dropdown menu with "RENAL ADYSPLASIA" selected, and an "Enter Disease:" text box with "renal adysplasia" entered. There are "Search" and "Clear" buttons. The sidebar on the left contains links for "Disease", "About GUDMAP Disease Resource", "Query Disease-Genes Associations", "Query Phenotype-Genes Associations", "GATACA (external)", and "ToppGene (external)". The top navigation bar includes "Home", "About GUDMAP", "Gene Expression", "Resources", "Tutorials", "Disease", and "Help".

Figure 1. Query Disease-Genes Associations

## 2. Select Gene from Results Page

The results page will return a list of genes that have an association to the disease. Associations flagged as being NCBI = 1 are taken from NCBI's list of disease-gene associations (mim2gene). In situ data = 1 means there is in situ data for that gene in GUDMAP.

We want to look at a gene that has in situ data so that we can see where it is expressed in the kidney. In this result set both RET and **UPK3A** have mutations that cause **Renal Adysplasia** and there is in situ data in GUDMAP for these genes. So we select one of these (**Upk3a**) to investigate further. Click on the Upk3a link in the Mouse Gene Symbol column.



Results of query for genes associated with OMIM DISEASE: renal adysplasia

- Click on the gene symbol to go to the gene page within the GUDMAP gene expression database.
- Please note:** Some genes may not have a gene page within GUDMAP.
- Click on the OMIM ID to go the disease page at OMIM - the nature of the association can be checked here.
- The *In Situ Data* column indicates whether there is available in situ data in the GUDMAP gene expression for that gene. '1' = yes; '0' = no.
- The *Flag* column gives an indication on the validity of the association based on manual checking of the main text of the OMIM entry.  
N = normal, association appears valid. U = unsure, association is possible but unclear. R = related, a related disease is associated with this gene. [more info...](#)
- The *NCBI* column if there is evidence supporting the association from NCBI.

Records 1 to 5 of 5

OMIM ID	Disease Name	Human Gene Symbol	Mouse Gene Symbol	Mouse Gene MGI ID	In Situ Data	Flag	NCBI
<a href="#">191830</a>	RENAL ADYSPLASIA	<a href="#">EMX2</a>	<a href="#">Emx2</a>	MGI:95388	1	N	0
<a href="#">191830</a>	RENAL ADYSPLASIA	<a href="#">GFRA1</a>	<a href="#">Gfra1</a>	MGI:1100842	1	U	0
<a href="#">191830</a>	RENAL ADYSPLASIA	<a href="#">PAX2</a>	<a href="#">Pax2</a>	MGI:97486	1	U	0
<a href="#">191830</a>	RENAL ADYSPLASIA	<a href="#">RET</a>	<a href="#">Ret</a>	MGI:97902	1	N	1
<a href="#">191830</a>	RENAL ADYSPLASIA	<a href="#">UPK3A</a>	<a href="#">Upk3a</a>	MGI:98914	1	N	1

Disclaimer: Please note that the development of the GUDMAP website is an on-going process. Whilst we make every effort to ensure the quality of the information on these pages, the content cannot always be guaranteed to be accurate or complete. In addition, at certain times some functions may be restricted.

Contacts Request ISH

Figure 2. Disease-Gene Results Page

### 3. Upk3a – Gene Page – access to expression data

All information about Upk3a can be viewed on its gene page, including details of what entries have been submitted to the database for Upk3a

The screenshot shows the GUDMAP Gene Expression Database interface for the gene Upk3a. The page is titled "Expression Database" and "Gene Details". It displays the following information:

- Symbol:** Upk3a
- Name:** uropodin 3A
- Synonyms:** 1110017C07Rik, Upk3
- Chromosome:** Chromosome: 15, Start: 84847650, End: 84852864, Genome Build: Ensembl, December 2010
- Gene Disease:** OMIM(1) TS17-28
- Gene Expression:** A bar chart showing expression levels across different tissues.
- In situ expression profile:** A bar chart showing expression levels across different tissues.
- In situ expression images:** A small image showing a kidney section.
- Microarray expression profile:** A bar chart showing expression levels across different tissues.
- Genesets:** A list of genesets associated with the gene.
- Items in my Genes:** 0
- Probes Linked to Gene:** MGI:1900588
- Links:** MGI: MGI:98914, Ensembl: ENSMUSG0000022435, UCSC: View probes for this gene in UCSC Browser, OMIM: Search OMIM, Entrez: Search entrez, GeneCards: Search GeneCards, GO: MGI GO Annotations
- In-Situ:** A table of in-situ hybridisation (ISH) entries.

GUDMAP ID	TS	Sex	System	Section	MGI
GUDMAP:10025	TS17	unknown	genitourinary system	wholmount	MGI:1900588
GUDMAP:10390	TS21	male	male reproductive system	wholmount	MGI:1900588
GUDMAP:10391	TS21	female	female reproductive system	wholmount	MGI:1900588
GUDMAP:10392	TS21	female	urinary system	wholmount	MGI:1900588
GUDMAP:10393	TS21	female	urinary system	wholmount	MGI:1900588
GUDMAP:9904	TS23	unknown	metanephros	section	MGI:1900588
GUDMAP:9905	TS23	unknown	ureter	section	MGI:1900588
GUDMAP:13627	TS23	male	urethra of male...	section	MGI:1900588
GUDMAP:13631	TS23	male	male reproductive system	section	MGI:1900588
GUDMAP:13628	TS27	male	bladder	section	MGI:1900588


Figure 3. Upk3a Gene Page

The *In Situ* section of the page shows how many in situ hybridisation (ISH) entries GUDMAP holds for this gene – in this case 5 wholmount ISH and 5 section ISH entries. Each in situ can be viewed by clicking on the GUDMAP ID in the **In-Situ** section (see Fig. 3).

We want to know where Upk3a is expressed in the kidney and use this information to search for other genes that share a similar expression pattern. So we want to look at a section ISH entry for the metanephros – in this case GUDMAP:9904. Click on the ID number to view the entry page.

## 4. View Expression Mapping

Expression Mapping

View annotated components as a tree 




















ComponentID	Component	Expression	Patterns/Locations
EMAP:8233	collecting duct	 present	
EMAP:28508	developing capillary loop stage nephron GROUP	 not detected	
EMAP:8239	early nephron GROUP	 not detected	
EMAP:30864	immature loop of Henle	 not detected	
EMAP:28529	late tubule	 not detected	
EMAP:28487	maturing nephron	 not detected	
EMAP:28063	medullary collecting duct 	 present (strong)	
EMAP:8226	metanephros	 present	 regional
EMAP:27724	nephrogenic zone	 not detected	
EMAP:28121	pelvic smooth muscle	 not detected	
EMAP:28079	pelvic urothelial lining 	 present (strong)	
EMAP:8232	pelvis	 present	 regional
EMAP:28085	perihilar interstitium	 not detected	
EMAP:8236	renal cortex	 not detected	
EMAP:28520	renal interstitium group	 not detected	
EMAP:28043	renal medulla	 present	 regional
EMAP:28158	renal medullary vasculature	 not detected	
EMAP:28460	renal vasculature	 not detected	
EMAP:8217	urinary system	 present	

Figure 4. Expression Mapping of GUDMAP:9904. Annotations viewed as a list.

On the entry page for GUDMAP:9904 there is a section called 'Expression Mapping'. This gives details of all the annotation for that entry. By default they are displayed as an ontology tree. The annotation can be viewed as a simple list (Figure 4) by clicking on the appropriate tab. Now we have a discrete list of annotations for GUDMAP:9904.

We can use these annotations to perform a search of the database to find genes that share a similar expression pattern. We do this using the Boolean Query. The annotations for GUDMAP:9904 should be noted down.

From the annotations of Upk3a we can see it is expressed (present) in the collecting duct, medullary collecting duct and renal medulla and that it is not expressed (not detected) in the nephrogenic zone, renal cortex and maturing nephron. So, we'll use this information to perform a Boolean Query to find other genes with similar expression.

## 5. Using the Boolean Query

Navigate to [http://www.gudmap.org/gudmap/pages/boolean\\_test.html](http://www.gudmap.org/gudmap/pages/boolean_test.html)

The user-interface that helps build a Boolean query is limited to only 3 components. Therefore, we'll manually construct the query syntax we require. Help on using the Boolean Query syntax is found at

[http://www.gudmap.org/Help/Boolean\\_Syntax\\_Help.html](http://www.gudmap.org/Help/Boolean_Syntax_Help.html)

### Expression Database > Boolean Anatomy Search

Search gene expression patterns in selected anatomical components.

Current anatomy display is for stage range TS17 to TS28.

The screenshot shows the Boolean Anatomy Search interface. On the left is a tree of anatomical components, with 'collecting duct' selected. The tree includes components like 'immature loop of Henle (TS22-TS24)', 'medullary collecting duct (TS22-TS24)', 'renal medullary interstitium (TS22-TS27)', 'developing vasculature of medulla (TS22-TS22)', 'outer medulla (TS25-TS28)', 'loop of Henle, outer medullary portion (TS25-TS28)', 'outer medullary interstitium (TS25-TS28)', 'outer stripe of outer medulla (TS28-TS28)', 'inner stripe of outer medulla (TS28-TS28)', 'outer medullary collecting duct (TS25-TS28)', 'inner medulla (TS25-TS28)', 'loop of Henle, inner medullary portion (TS25-TS28)', 'inner medullary interstitium (TS25-TS28)', 'inner medullary collecting duct (TS25-TS28)', 'renal medullary vasculature (TS23-TS28)', 'renal medullary lymphatics (TS23-TS28)', 'renal medullary nerves (TS23-TS28)', 'G Loop of Henle group (TS25-TS28)', 'calyx (TS23-TS28)', and 'pelvis (TS21-TS28)'. On the right is a search panel with three query components. The first component has 'genes' in the search box, 'present' selected for expression, and 'medullary collecting duct' in the 'in' dropdown. The second component has 'present' and 'not detected' selected for expression, and 'TS17' to 'TS28' in the 'in' dropdown. The third component has 'present' selected for expression, and 'TS17' to 'TS28' in the 'in' dropdown. A red arrow points from the first component to the search box. Two yellow callout boxes provide instructions: '1. The user-interface only allows for three components' and '2. So enter boolean query syntax directly into the search box. Click on the 'i' to find help using the syntax.'

Figure 5. Using the Boolean Query to find genes with similar expression

The syntax of the query we'll use is:

GENE: p{in "collecting duct" TS17..TS28} AND p{in "medullary collecting duct" TS17..TS28} AND p{in "renal medulla" TS17..TS28} AND nd{in "nephrogenic zone" TS17..TS28} AND nd{in "renal cortex" TS17..TS28} AND nd{in "maturing nephron" TS17..TS28}

A total of 9 genes are returned by the query – including Upk3a (as expected).

## 6. Use Gene Expression Summaries to compare genes.

Each gene returned by the Boolean Query is displayed as a gene expression summary. This makes it possible to quickly view the expression of each gene and spot any similarities or differences.



**Figure 6. Gene Expression Summaries. (A) Microarray Expression Profiles. (B) In situ Images. (C) Tick boxes to select genes for collections. (D) Collection features to build and modify user collections.**

The Microarray Expression Profiles (Fig. 6A) can be compared for each gene to that of Upk3a (final row).

The in situ images for each gene can be viewed (Fig. 6B) by clicking on the 'In situ expression images' thumbnail.

Genes of interests can then be built into user collections by selecting genes (Fig. 6C) and using the collections features at the foot of the page (Fig. 6D). In order to be able to save collections you need to be logged in to the site. Logins can be obtained by e-mailing the GUDMAP Editorial Office ([gudmap-editors@gudmap.org](mailto:gudmap-editors@gudmap.org)).

By manually viewing expression profile, in situ images and in situ annotations of these genes we can further refine our list of genes.

Looking in more detail at the microarray expression profiles we can probably refine our list further. By clicking on the microarray expression profile for Upk3a we link to a more detailed microarray heatmap. If we select the Developing Kidney heatmap at the top of the page and click 'Ok' we can see where probes for Upk3a are expressed in components of the kidney. Upk3a shows expression in urothelium and medullary collecting duct from its microarray expression profile (Fig. 7).

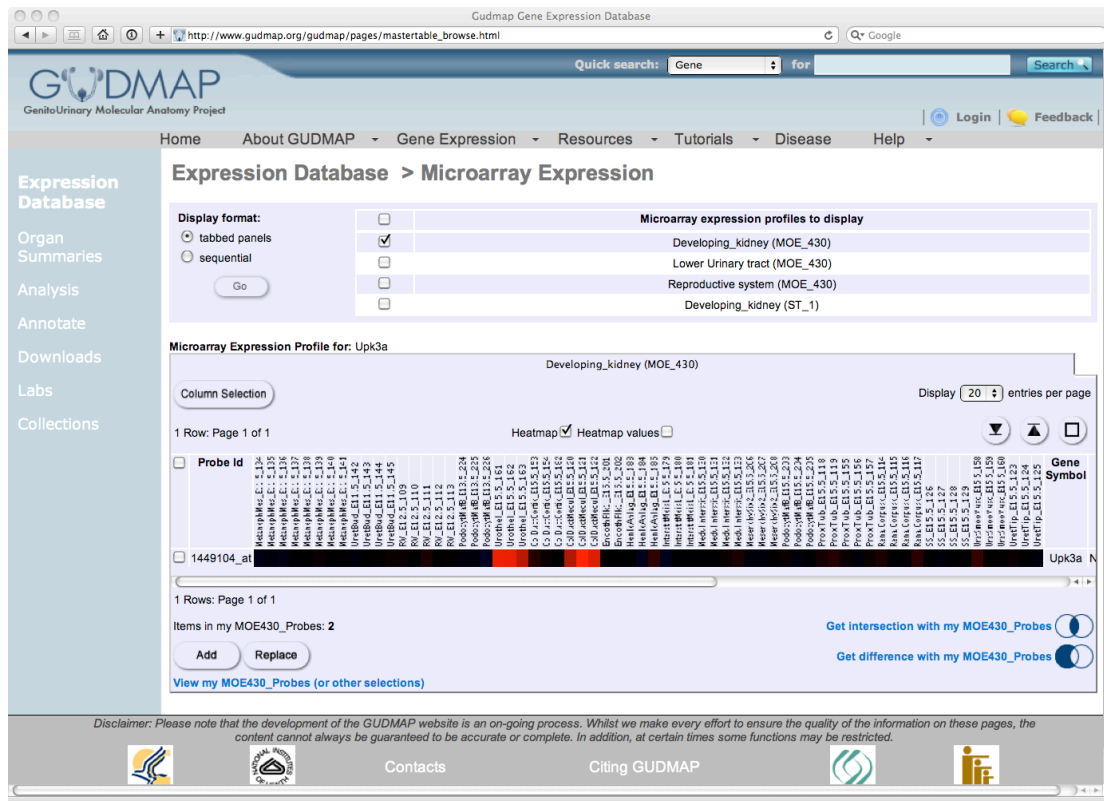


Figure 7. Microarray Expression for Upk3a in Developing Kidney

If we now look in more detail at the other genes in our list we can select those that share this type of expression. On this basis we can pick out *Al836003*, *Foxa1*, *Gsdmc2* and *Gsdmc4* as sharing microarray expression profiles similar to *Upk3a* based on elevated expression in both the urothelium and medullary collecting duct compared to other components of the developing kidney.

This set of genes represented potential disease candidate genes for Renal Adysplasia – generated through the use of the GUDMAP website to investigate gene expression patterns.

## Summary

By using the GUDMAP Disease Resource we have identified a gene associated with a disease of interest. We have then been able to look at in situ expression entries in the GUDMAP database for that gene to see where it is expressed. Using this information we have performed a Boolean Query of the database to find other genes that share a similar expression pattern to our gene of interest. From the results of this search we have then been able to further refine our results by looking in more detail at the microarray expression profiles of the gene.

This gives us a working set of potential disease candidate genes – simply by searching the gene expression data stored in GUDMAP.