

Example 1: Foregut vs. Hindgut

- 1. Find all the tags that are exclusive to the Foregut (SM107) and not in Hindgut (SM112) pg 2
- 2. Identify Tags that are up regulated in the Foregut when compared with the Hindgut pg 7
- 3. Map up regulated tags pg 9
- 4. Find GO Categories in the Hindgut library pg 13

Example 2: Gene of Interest

- 1. Retrieve tags for a gene of interest FGF8 pg 16
- 2. Determine which libraries contain identified tags pg 18

Example 1: Foregut vs. Hindgut

Task 1: find tags exclusively to hindgut and foregut libraries

1.1. Build query for Foregut (SM107)

Create a new query by clicking the New Query button. This will open the Query Builder window.

1000	ntaban	k 🔅																
₽.	Ř *	₽¥		2	ľ	D	2		Ø	Î			£	ø	GO	9		
	Type Looded Medified			Nama			Datatypa					Description						

Use the pull down menu to select GSC SAGE Libraries (Note: you must have set your user name and password to have access to these libraries. E-mail <u>discoveryspace-support@bcgsc.ca</u> if you do not have a password).

۹? New Query Definition - Query	r 2 🛛
General Query	
Query Path: Start CGAP SAGE Library	▼ <u>E</u> nd
CGAP SAGE Lit Experimental GEO SAGE Tags Id GEO SAGE Library O Name GO Terms Anchoring e GSC Clustered SAGE Tags Cell type GSC Experimental SAGE Tags GSC SAGE Libraries Human CMOST Number of tags (including linkers) Tissue Experimental CGAP Tags Taxon	
Parent Field Not Condition Value	Ī
OK Apply Close	

You can now begin to build your query. Understanding queries is crucial for successfully using DS4. It may feel overwhelming at first, but with a little practice, it will become clear.

Rew Query Definiti	on - Query	r ₂ ⊠
General Query		
Query Path: 🔫 <u>S</u> tart	GSC SAGE Libraries	▼ <u>E</u> nd
GSC SAGE Libr	ags gs	
Parent GSC SAGE Libraries	Field Not Condition Value Name = ▼	
	OK Apply Close	

Click on each item you want to set a constraint for and click the "Set Parameter" button. In the bottom window you can set your constraint by selecting a condition and setting a value.

Parameters to set:

- GSC Libraries -> Name : set to library accession (SM107)
- Experimental SAGE Tags -> Quality : set to >= 0.95 (95%)
- Experimental SAGE Tags -> Sequence -> Sequence data: set to linkers using IN clause. Click NOT box to exclude linkers from your analysis.

The final and VERY IMPORTANT step is to use the END button (on upper right hand side of the Query builder) to set what kind of data the query is going to return. In this case we set it to Sequence to return tags for our library.

Click Sequence, then click on the End button.

🛄 Discovery Space
File Window Help
? New Query Definition - Query 🗖 🗖
General Query
Query Path: 🔽 Start GSC SAGE Libraries 🗾 🗾 🔤 End
GSC SAGE Libraries
- Name
Description
Number of tags
Quality of Tags
- D Tissue
Clustered SAGE Tags
🕈 🗂 Experimental SAGE Tags
Duplicate Ditag
- 🗋 🖣 Quality
- D In Human CMOST?
CGAP Experimental Tags
COOST Experimental rags
CMOST Human mappings CMOST Mouse Mappings
GEO Experimental Tags
GSC Clustered SAGE Tags
← ☐ Ø GSC Experimental SAGE Tags
← CT MGC Virtual Tags
← □ G Normalized to short Tag Sequences
Refseq Virtual Tags
- Taxon
2 6 2 H
Parent Field Not Condition Value GSC SAGE Libraries Name □ = ▼ SM107
Experimental SAGE Tags Quality 🔲 = 🔍 0.95
Sequence Sequence data 🗹 in 🔻 [TCGGATATTAAGCCTAG, TCGGACGTACATCGTTT] 🔺
OK Apply Close

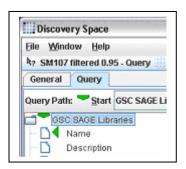
Our query is almost complete. Before clicking on "OK", make sure that you will be able to identify this query for future reference. You can do this by going to the General tab at the top of the query generator window and adding as much information as you think would be useful.

? SM107 (For	regut - TS13) - Query 🛛 🗖 🗹									
General Qu	lery									
Class:	ag Sequence									
<u>N</u> ame:	SM107 (Foregut - TS13)									
Description:	Manual - Task 1									
<u>Annotation:</u>										
Created By:	Test User									
Created On:	Jun 22, 2005, 08:19 PM PDT									
Last Modified:	Jun 22, 2005, 08:50 PM PDT									
Origin:	Created using 'New Query Definition' button									
	OK Apply Close									

The query is now complete. Click OK.

1.2. To create another query with the same parameters but for a different library, you can either repeat the previous step (1), or you can select your previous query in the databank and use the "Duplicate Selected Definitions" button on the databank to create an identical query for Hindgut (SM112).





All you have to do is change the Name parameter to SM112. The other constraints will remain unchanged.

Again, you should save this query using an appropriate name by using the General tab.

You now have two separate queries. One for Hindgut and one for Foregut.

Now go to the databank and look at the two queries. Notice they are both a query type. This indicates that if you work with this type of data you will be making a direct connection to the database. This is not necessarily a bad thing to do, but if many users are connecting in this fashion (using this data type) the queries will take a long time to return results.

🔞 Databank 🐰							ਰ ਕ
2: 1: 1:	¥ R	1			û 🔊 Go		
Туре	Loaded	Modified	Name	Datatype	Description	Origin	Last Modif
2 Ouon		~	SM107 (Foregut - TS13)	Tag Sequence	Manual - Tas	Created using 'New Query Definition' butt	Jun 22, 2005, 0{
🕜 Query 🔪		2	SM112 (Hindgut - TS13)	Tag Sequence	Manual - Tas	Duplicated from Definition 'SM107 (Foreg	Jun 22, 2005, 08

IMPORTANT: To speed up further analysis, you should create data definitions from your queries. Highlight your queries and select the Create Data Definition button. These definitions should be renamed for future reference. Click the General tab and enter identifying information. The data definitions will be used in subsequent analyses.



1.3. In the databank, select these two **data** definitions and click the Venn Diagram button.

🕲 Databank 🖉									
12: 12: 12: 12: 12: 12: 12: 12: 12: 12:	*		2	ð 1		R	GÒ	٩	
Туре					 Datatype)		Descr	iption 🔺 📃

The Venn widget will appear. You can extract tags exclusive to each library.

🔊 Tag Sequence - Venn Table		r 2 🛛
	Count 🔽 📝 1 🕂 👬	1 🗧 🕂 1 🗧
● ● ● Name Count Description ✓ ● ● SM112 (Hindgut - TS13) 110552 Manual - Task 1 ✓ ● ● SM107 (Foregut - TS13) 102985 Manual - Task 1	Item 26012 TITGGTTTAAGAAGAGG 21 21	
	TTTGGTTTAAGAAGAGG2121GCCAACAGCATCGCCAG1111GTTGTCCTTTTTCCTCT1111CACGCTCCCGGGAAGGC1010CGGTTCCACCCGAAAGG1010CTGTGCCCTCCCAGGTA1010CAAGGTGACAGGCCGGT99TTAAAGTAATGAAAAAA99TTGGTAGCCGTAGCACA99ACCAACAGGTAGCGCGT88	
All enabled Rows: 2	ACCARCAGORIGOGE 8 8 8 ACTTCAGACTCATCCAG 8 8 CGTTGACTCATCCAG 8 8 CTATTAAATTTTGCCCT 8 8 GAAGGAGTCTTTGTATG 8 8 GCCGCTCTGGTTGCTCT 8 8 CTAGAAGGAGGTCTACCG 8 8 CTAGAAGGAGGGTTCACCG 8 8 CTAGAAGGAGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGTCACCG 8 8 CTAGAAGGAGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 8 CTAGAAGGAGGGGTCACCG 8 CTAGAAGGAGGGGGTCACCG 8 CTAGAAGGAGGGGTCACCG 8 CTAGAAGGAGGGGGTCACCG 8 CTAGAAGGAGGGGGGTCACCG 8 CTAGAAGGAGGGGGGTCACCG 8 CTAGAAGGAGGGGGGTCACCG CTAGAAGGAGGGGGGTCACCG 8 CTAGAAGGAGGGGGGTCACCG 8 CTAGAAGGAGGGGGGTCACCG 8 CTAGAAGGAGGGGGGGTCACCG 8 CTAGAAGGAGGGGGGTCACCG 8 CTAGAAGGAGGGGGGTCACCG 8 CTAGAAGGAGGGGGGTCACCG CTAGAAGGAGGGGGGGTCACCG CTAGAAGGAGGGGGGGGGG	_

You can drag your tags out anywhere on the desktop. This will create a data definition for those tags. Click the General tab to rename this data for future reference.

Databank		
🏠 🎅 Tag Sequence - Venn Table		o" o" 🖂
	Count 🔽 🛃 1 🕂 🛉	
	SM107 (For Item 26012	
	TTTGGTTTAAGAAGAGG 21 21	<u></u>
📋 Top 10 tags found only in SM107 and not in S 🖬 🖾	GCCAACAGCATCGCCAG 11 11 GTTGTCCTTTTTCCTCT 11 11	
General Data	CACGCTCCCGGGAAGGC 10 10	
Class Item Quantity	CGGTTCCACCCGAAAGG 10 10	
Tag Sequence TTTGGTTTAAGAAGAGG 21	CTGTGCCCTCCCAGGTA 10 10	
Tag Sequence GCCAACAGCATCGCCAG 11	CAAGGTGACAGGCCGGT 9 9 9 TTAAAGTAATGAAAAAA 9 9	
Tag Sequence GTTGTCCTTTTTCCTCT 11		
Tag Sequence CACGCTCCCGGGAAGGC 10	ACCARCACCTACTTCCC 0 8	
Tag Sequence CGGTTCCACCCGAAAGG 10	ALLAALAGGIAGIIGGU 0 0	•
Tag Sequence CTGTGCCCTCCCAGGTA 10	Rows: 22999 Selected: 10 x 1	
Tag Sequence CAAGGTGACAGGCCGGT 9		
Tag Sequence TTAAAGTAATGAAAAAA 9		
Tag Sequence TTGGTAGCCGTAGCACA 9		
Tag Sequence ACCAACAGGTAGTTGGC 8		
OK Apply Close		

Task 2: find tags up regulated in Foregut vs. Hindgut

Highlight your two library tag Data Definitions and select the Comparison button.

🕲 Databank							
1 27 1 27 1 27	₽.	17 6			GO 🧶 🔛		
Туре	Lundrd	Modified	Name	Datatype	Description		
🗄 Data	~	~	SM112 filtered	Tag Sequence	library - for example	D	
🏥 Data	~	~	SM107 filtered 0.95	Tag Sequence	library - for example	D	
🛱 Doto	.1		Now Data Dofinition	Tog Coguoneo			

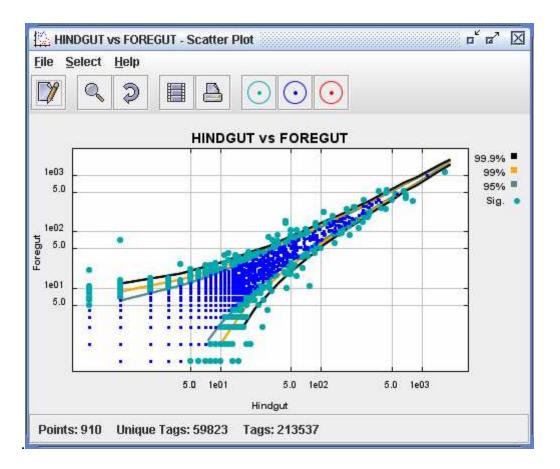
Select which library you want for each axis and rename the comparison. Close the widget.

II New Comparison Defini	tion - Com	parison 🔛	1	⊏ ⊂ ⊠				
General Comparison	Result							
Name: New Comparison De	efinition							
X		Y						
SM112 filtered	🛛 Auto	SM107 filter	ed 0.95	🖌 Auto				
Exclusive X-Y selection				, 				
Exclusive X-Y selection Name Description SM112 filtered library - for example SM107 filtered 0.95 library - for example								
ОК	Apply	Close						

Now highlight the comparison in the databank and select the Scatterplot button.

I Databank													
	*	1) 🗋 💈		ð í		1 1 1	٨	GO	9			
Туре	Loaded	Modified		Name			Datatype			Description			
II Comparison			HINDGUT VS	FORGUT		Tag S	Tag Sequence						
🏥 Data	~	~	SM112 filtere	3M112 filtered			Tag Sequence			library - for example			Dat
🏥 Data	~	~	SM107 filtered 0.95			Tag S	Tag Sequence			library - for example			Dat
H Data			Now Data Da			Tog O	Tog Coguoneo						Even

This will open the scatterplot diagram. Selected tags of interest can be dragged anywhere on the desktop to create an additional data definition. Select the up regulated tags and save them as 'UPREG'



Good job! By completing this example you have learned how to use the comparison widget and scatterplot.

Task 3: Map up regulated tags

Ensure you have dragged tags out from the scatterplot into a data definition before you start this example.

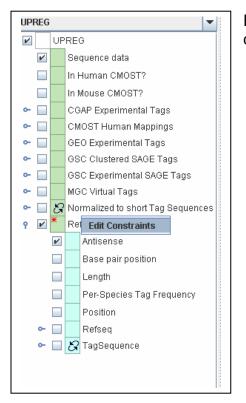
Highlight your data definition and click the explorer button.

🔞 Databank 📎												
1 : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:	*	1	۵ 🏾 (Î		£	ø	GO	٩		
Туре	Loaded	Modified		Name		Da	tatype			Desc	ription	
🏥 Data	~	~	UPREG	IPREG Tag Sequence								E
1 Comparison		1	HINDGUT VS E	ORGUT		IOUNO2 NCT	nce					0

This will open the Explorer widget which will allow you to map your tags. This widget is central to DS4.

UPREG - Discovery Space Explorer		c c 🛛
		Ħ
UPREG 🗸 🗸	UPREG	
UPREG	987 TTCATTATAATCTCAAA	
	920 CAAGGTGACAGGCCGCT	-
Sequence data	884 ATAATACATAAAAAAAA	=
In Human CMOST?	634 CAAAAATAAAAGCCGCA	
In Mouse CMOST?	490 AGAGCGAAGTGGCGGAA	
🗢 🔲 🗖 CGAP Experimental Tags	482 CTGAACATCTCCCCCTT	
	432 GGGGAAATCGCCAGCTT	
🗢 📃 🔤 CMOST Human Mappings	414 CACGCTCCCGGGAAGGG	
🗢 🔲 🛛 CMOST Mouse Mappings	329 CCCTACTTCATCCTTTG 326 GTGAAACTAAAAAAAA	
🗢 🔲 🗾 GEO Experimental Tags	319 GCAGAGTGCGCCTGCTG	
GSC Clustered SAGE Tags	280 CTAATAAAGCCACTGTG	
	255 TCTTCTCACAAGACTTT	
🗢 🔲 GSC Experimental SAGE Tags	228 TTCCAGCTGCTTGACCC	
🗢 🔲 🛛 MGC Virtual Tags	211 TGTACTCAATAAACGAT	
🗣 🔲 🌠 Normalized to short Tag Sequences	210 ATCCGAAAGATGAAGCT	
Refseq Virtual Tags	205 TTCAGCTCGAGCGCCAA	
- Reisey viitual Lays	195 CATTITCTGGCAAAATC	
	185 ААААААААААААААА	
	182 ATCAACACCGCAACCTT	
	172 GAAATTTAAAGCAGGTT	
	167 GTGTTAACCAGCTGAAA	
	162 TCTGGACGCGGCAAGCA 156 GTGGCTAACAACCATCT	
	153 CACCACCACAGGATCAA	
	152 CAAACCTCCATAGACCG	
	137 CAATAAACTGAAAAGAG	-
	Finished query for specified UPREG. 29	6 rows returned.

In order to map our tags, we have to set some constraints similar to those set with the Query Builder.



Right click on RefSeq. This will allow you to open the constraints window.

You might want to exclude antisense mappings and only look at unambiguous matches.

Constraint		$\overline{\mathbf{X}}$
● <u>A</u> N <u>N</u> ew Filter ○ <u>O</u> F		
Refseq Virtual Tags	Antisense 👻 🖢 =	▼ false + -
Refseq Virtual Tags	Per-Species Tag Frequency 🔻 💆 =	▼ 1 + -
4	OK Cancel	

You also might want to only look at mouse matches. Click New Filter. Click the RefSeq Virtual Tags button.

This will open the Select a Node window.

🛄 Select A Node 🛛 🔀
Refseq Virtual Tags
♀ ○ □ Refseq ► ○ □ GO Terms
O PA-GOSUB Analysis
 O Protein localization predictions
► ○ C Virtual Tags
 C □ TagSequence
OK Cancel

By selecting the Taxon item you can narrow your mappings down by species.

In our case, we set the Taxon accession to 10090 for mouse.

Constraint					X
	AND Offset: 0 OR Limit: 150,000 1	41	Order <u>b</u> y:		r v Descending
Refseq Virtual Tags	Antisense	-	Fog =	•	false +
Refseq Virtual Tags	Per-Species Tag Frequency	-	tog =	-	1 + =
Taxon	Accession	-	toz =	-	10090
•					
	ОК		Cancel		

y	Ē	È	∧ ∨ (▶)		9 🖩 🏟							Į
REG						25		UPRE	G			
_	-			11111	Sequence data		W.		Refseq Vi	rtual Tags		
		GSC	C Experimental SAGE Ta	at the			Antisense	Per-Specie			fseq	
		MGC Virtual Tags		A CONTRACTOR OF CONT	~			Accession	Definition	Taxon		
	R	Nor	malized to short Tag Seg			0					Accession	and when the two property in the set of the
	*	100000	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	a statement	TTCATTATAATCTCAAA		false	1	NM_008972	Mus musc	10090	Mus musculus
		Refseq Virtual Tags Antisense	seq Virtual Tags	* Interested	CAAGGTGACAGGCCGCT							
	V	1	Antisense	-	ΑΤΑΑΤΑCΑΤΑΑΑΑΑΑΑ			Kana K				¥
	-	-	Daga poly position		CAAAAATAAAAGCCGCA	12						<u> </u>
	-		Base pair position	· · · · · · · · · · · · · · · · · · ·	AGAGCGAAGTGGCGGAA	100	false	1	NM_018860	Mus musc	10090	Mus musculus
			Length		CTGAACATCTCCCCCTT		¥/////////////////////////////////////		A			¥
3	17	-		2	GGGGAAATCGCCAGCTT	- 4	¥/////////////////////////////////////		A			¥
	V	Per-Species Tag Freque	rei-opecies rag rieque		CACGCTCCCGGGAAGGG	11					10000	<u> </u>
		1	Position	- Internation	CCCTACTTCATCCTTTG		false	1	NR_001592	Mus musc	10090	Mus musculus
9	V	-	Refseq	· · · · · · · · · · · · · · · · · · ·	GTGAAACTAAAAAAAAA							¥ ////////////////////////////////////
			and second se	* Statesteates	GCAGAGTGCGCCTGCTG	- 12	false		NUL 007000	1//////	40000	X/////////////////////////////////////
		V	Accession -	and the second second second	CTAATAAAGCCACTGTG	-	false	1	NM_007990		10090	Mus musculus
		V	Definition	* Construction	TCTTCTCACAAGACTTT		false		NM_026055 NM 023418		10090	Mus musculus
			and a second sec	· · · · · · · · · · · · · · · · · · ·	TTCCAGCTGCTTGACCC TGTACTCAATAAACGAT	100	laise		NIM_023418	wus musc	10090	Mus musculus
			Comment	-	ATCCGAAAGATGAAGCT		¥ ////////////////////////////////////		X////////////////////////////////////			
			Genbank		TTCAGCTCGAGCGCCAA	- 44	false	1	NM 011300	Muo muoo	10090	Mus musculus
				· ·	CATTTTCTGGCAAAATC		false	1	NM 010514		10090	Mus musculus
			Gene Names	1 married	AAAAAAAAAAAAAAAAAAAAAA	17	(//////////////////////////////////////	kinniki	_010314	Mus muse	1//////////////////////////////////////	mus musculus
			Location	and the second second	ATCAACACCGCAACCTT	1	¥ ////////////////////////////////////		X////////			X
			Due due to blow a c	and the second second	GAAATTTAAAGCAGGTT		¥ ////////////////////////////////////		X////////			X
_			Product Names	and the second second	GTGTTAACCAGCTGAAA	1	¥ ////////////////////////////////////		X////////////////////////////////////			\$ \\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\

Hit the run button and you will map your tags!!! As you can see, the tag counts are in the same table as the mappings.

Task 4: Find GO Categories in Hindgut Library

In order to use the GO Browser, we must retrieve data that can be inputted into GO. As you may know, GO accepts RefSeq, MGC, and Ensembl accession numbers. Therefore, we do not want tag data, we want gene data.

Since we want to retrieve gene data for the Hindgut library, we can duplicate our query that retrieved the tag data for Hindgut and modify it to retrieve gene data.

4.1. Duplicate Hindgut query as in task 1. You must use your query definition NOT your data definition. The data definition only contains your tag data, you need the actual query so you can modify it.

4.2. Edit the query to retrieve RefSeq genes.

SM112 filtered -	Query ຕໍ່ ຜ້າ	\mathbf{X}
General Query		
Query Path: 🔽 <u>S</u> tar	t GSC SAGE Libraries 🗾 🖬 End	1
P 🗂 _ Re	fseq Virtual Tags	
	Antisense	
	Base pair position	
	Length	
	Per-Species Tag Frequency	
	T Usition	
	Refseq	
	Accession	
	Definition	
	Comment	
	Genbank	
	Gene Names	
	Location	
	Product Names	
	Version	
	GO Terms	
• •	Locuslinks	
	PA-GOSUB Analysis	=
	Protein localization predictions	
¶	Taxon	
	- Accession	
	Scientifc name	
	Common names	
	Genbank common name	
	Rank	
►	$\Box \mathcal{B}$ Parent Taxon	
1. 1. 1	Ē	i
Parent	Field Not Condition Value	
Refseq Virtual Tags		
Refseq Virtual Tags Taxon	Per-Species Tag Frequency = 1 Accession = 10090	
Taxoff		

Since we are doing GO Analysis, we need to retrieve gene data. Therefore, set the END button to RefSeq.

We need to narrow down the genes we return just as in Task 3. Therefore, we set the following parameters:

RefSeq Virtual Tags -> Antisense : set to "false" RefSeq Virtual Tags -> Per-Species Tag Frequency : set to "1" RefSeq Virtual Tags -> RefSeq -> Taxon -> Accession : set to 10090

This will return mouse genes that unambiguously map our tags. Make sure you rename this new query.

We are now ready to use the GO Browser!

4.3. Highlight your new gene query and click the GO button.

🕼 Da	atabani	k 🤤													
₽.	Ě ×	₽.	R *	2	ľ	ē ,	2	٩	Î	Ł	ß	ø	GO	۲	

This will open the GO Browser. Please refer to the GO Browser primer for detailed instructions of its use.

You will notice two things different in DS4.

- 1) Tag counts are imported into the GO Brower (left-most column)
- 2) There is an option to "Score using Counts". This will calculate percentages based on tag counts. If you deselect this box, the percentages will be calculated based on number of genes.

8			Select funct	ion 🔻 🗹 Score using Counts
of i	mported recor	ds		
	Count	Туре	Identifier	Description
				Mus musculus ribosomal protein, large, P1 (Rplp1), mRNA.
	385 Refse	q Gene	NM_024277	Mus musculus ribosomal protein S27a (Rps27a), mRNA.
				Mus musculus prothymosin alpha (Ptma), mRNA.
	323 Refse	q Gene	NM_009098	Mus musculus ribosomal protein S8 (Rps8), mRNA.
				Mus musculus ribosomal protein, large P2 (Rplp2), mRNA.
				Mus musculus actin, beta, cytoplasmic (Actb), mRNA.
				Mus musculus lactate dehydrogenase 1, A chain (Ldh1), mRNA.
			-	Mus musculus ribosomal protein 10 (RpI10), mRNA.
				Mus musculus guanine nucleotide binding protein, beta 2, related sequence 1 (Gnb2-rs1), mRNA.
	221 Refse	q Gene	NM_013765	Mus musculus ribosomal protein S26 (Rps26), mRNA.
				Mus musculus ribosomal protein L9 (RpI9), mRNA.
				Mus musculus ribosomal protein L37a (RpI37a), mRNA.
				Mus musculus embigin (Emb), mRNA.
				Mus musculus ribosomal protein L41 (Rpl41), mRNA.
				Mus musculus cofilin 1, non-muscle (Cfi1), mRNA.
			-	Mus musculus ribosomal protein L35 (RpI35), mRNA.
			-	Mus musculus ribosomal protein S27 (Rps27), mRNA.
				Mus musculus ribosomal protein S29 (Rps29), mRNA.
				Mus musculus heat shock protein 1, beta (Hspcb), mRNA.
				Mus musculus ribosomal protein L23 (Rpl23), mRNA.
				Mus musculus ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit (Atp5b), nuclear gene enc
	149 Refse	eq Gene	NM_172086	Mus musculus ribosomal protein L32 (RpI32), mRNA.

4.4. Select "Show associated terms" from the drop down menu.

GO SM112 filtered - genes - GO Analysis										
<>) =	¢ 🖺 🗛 🕻	Select function	Score using Counts							
Set of import	ed records	Select function								
Cour	nt Type	Show associated terms	Description							
4	75 Refseq Gene	Show only directly associated terms	1 (Rplp1), mRNA.							
3	85 Refseq Gene	Show only records with terms	(ps27a), mRNA.							
3	81 Refseq Gene	Show only records without terms	, mRNA.							
	23 Refseq Gene	Show product of records & terms	8), mRNA.							
	64 Refseq Gene	Show product of records & only direct terms	2 (Rplp2), mRNA.							
2	61 Refseq Gene	אַש אָרָעָט אָא אָאָא אָאָא אָאָא אָאָא אָאָא אָאָא אָאָא אָאַראָא אָאָא אָאַראָ אָאַא אָאַא אָאַאַ אַ	ctb), mRNA.							
2	52 Refseq Gene	NM_010699 Mus musculus lactate dehydrogenase 1, .	A chain (Ldh1), mRNA.							

This will display the GO categories and the percentage of transcripts in each category.

IMPORTANT:

Direct Association: the lowest branch a gene is categorized in

Ancestral Association: gene is in this category, but has also been put into lower branches

For example: a transcription factor (Direct Association) is also in DNA binding (Ancestral Association).

< > ≡	Itered - genes - GO Analysis			Score using Counts		⁶ ם ם •
Terms associ	ated with the selected records					
Accession	Name	Term Type	Depth		Direct Associa An	cestral Ass
11	all	universal	0		0.000	85.870
0:0003674	molecular_function	function	1		0.000	76.069
O:0008150	biological_process	process	1		0.000	69.912
0:0005575	cellular_component	component	1		0.000	72.279
bsolete_fun	obsolete_function	function	1		0.000	0.022
0:0000004	biological_process unknown	process	2		3.209	3.209
0:0007610	behavior	process	2		0.024	0.259
0:0005554	molecular_function unknown	function	2		2.759	2.759
0:0030528	transcription regulator activity	function	2		0.186	3.466
0:0007275	development	process	2		0.820	5.503
0:0031012	extracellular matrix	component	2		0.000	0.739
0:0045182	translation regulator activity	function	2		0.002	2.453
0:0003774	motor activity	function	2		1.016	1.035
0:0003824	catalytic activity	function	2		1.569	25.053
0:0016032	viral life cycle	process	2		0.091	0.140
0:0005198	structural molecule activity	function	2		1.834	15.657
0:0050789	regulation of biological process	process	2		0.000	10.935
0:0043234	protein complex	component	2		0.000	25.694
0:0009987	cellular process	process	2		0.000	32.188
0:0005488	binding	function	2		0.862	44.768
0:0004871	signal transducer activity	function	2		0.458	5.094
0:0003767	co-chaperone activity	function	2		0.022	0.022
0:0005576	extracellular	component	2		0.512	8.286
0:0016209	antioxidant activity	function	2		0.756	0.935
0:0030188	chaperone regulator activity	function	2		0.015	0.016
0:0007582	physiological process	process	2		0.120	64.382
0.0001002	ontyme regulator activity	function	2		0.022	2 4 4 1

Example 2: Gene of Interest

Task 1: Retrieve tags for a gene of interest – FGF8

1.1 Build a query to return tags for a given gene. This time we will use the Template builder.

Click the Template button to create a Template.



You should be comfortable with queries after completing Example 1. In this example we want to return tags from a given gene. We will use RefSeq Gene Sequence to retrieve all longSAGE mouse unambiguous tags. Notice the END function is on the TagSequence item.

Click the query tab and create this query:

Discovery Space	
Eile <u>W</u> indow <u>H</u> elp	
🚯 New Template Definition - Template	rt 🛛
General Query Template	n fredd dawrae
	10000
Query Path: Start Refseq Gene Sequence	<u>E</u> nd
🗂 🌄 Refseq Gene Sequence	
- C ## Accession	
Comment	
- 🗋 Genbank	
- 🗋 Gene Names	
- D Location	
Product Names	
- 🗋 Version	
🗢 🗂 Entrez Gene	
GO Terms	
• 🚍 Locuslinks	
PA-GOSUB Analysis	
Protein localization predictions Taxon	
- Clession	
- Common names	
- Common name	
← 🖬 🛱 Parent Taxon	
Y Carl aren raxon	
Antisense	
Base pair position	
Length	
Per-Species Tag Frequency	
- N Position	
- C Refseq	
🗠 🗂 🛌 TagSequence	
	uaranananan
10. 10 12 I	
Parent Field Not Condition Value	
Taxon Accession □ = ▼ 10090 Virtual Tags Antisense □ = ▼ false	
Virtual Tags Length	
Virtual Tags Per-Species Tag Frequency 🔲 = 🔽 1	
OK Apply Close	

🚯 New Temp	ate Definition - Template		• • • 2
General Q	uery Template		
Name <u>P</u> refix	gene		Copy Description
Name <u>S</u> uffix			Copy Annotation
Example	gene FGF8		\frown
Variable	/ Refseq Gene Sequence / Accession		#
a		ſ	Example Usage Steps
5050	Values for: Accession		
FGF8			
Cha	nge Variable Property		
Select	a property as new variable:		
	Refseq Gene Sequence	▲	
	Accession		
	Definition		
	Comment		Outionally abanyo the
	Genbank		. Optionaly, change the /ariable property above
	🗰 Gene Names		andwe property above
	Location	2	. Provide approprate
	Product Names		alues for the variable
	Vereion	<u>-</u> u	ising the table to the left
/ Refs	eq Gene Sequence / Gene Names		. Click the Create
	OK Cancel		Queries button below

We have to set the variable. This allows us to run this query as many times as we like to create an identical query for different gene names. You can then input FGF8 into Values for: Gene Names column.

You can set Name Prefix and Name Suffix to help label your queries. In this example, the query will be named 'gene FGF8'.

NOTE: We can now reuse this query over and over to retrieve tags for different genes!

Click Apply. Click Create queries.

This creates your specified queries in the Databank. Highlight your new query and select the Explorer button.

When the query is run, it will return tags.

New Query Definition - Discovery Space Explo														َ <mark>ت</mark>	a" [2	3
	1			 ••••• 					Ø							
FGF	FGF8 example 👻						New		ery Defi					 		٦
~	FGF8 example				1	AACA	AGA		I <mark>UENCE</mark> GGAAGC		1					
	Sequence data				GGCAGCCCCGCTCCGC GCCTTTACCCGCAAGGG TGAGGGAGCAGAGCCTG											
												-				
	In Mouse CMOST?				1				CGCGGG	-						
~	CGAP Experimental Tags			an a	1	GCAG	AAG	GACG	GAGACO	:C						
~				an a												
•				ana an												
~	🗢 🔲 🗖 GSC Clustered SAGE Tags			and the second												
•	🕶 🔲 🖉 GSC Experimental SAGE Tags			an a												
•	🕶 🔲 🛛 MGC Virtual Tags															
•	🗢 🔲 🛃 Normalized to short Tag Sequences															
•			Refseq Virtual Tags	ana an												
				ana an												
				18												

Task 2: Determine which libraries contain these tags

2.1. Create query of all available libraries

We need an additional query to return available libraries. This query will be used to open the Search For Tags in Libraries widget.

For our example, let's look at all longSAGE mouse libraries in the GSC Libraries.

Let's also remove the contaminated libraries.

Discovery Space								
<u>File Window H</u> elp								
? New Query Definit	tion - Query					ᄚᅜ	\boxtimes	
General Query								
Query Path: 🔽 <u>S</u> tart	GSC SAGE L	ibraries			•	<u>E</u> n	d	
GSC SAGE Libi	raries							
Name								
- 🔤 🗧 Description								
Number of 1	tags							
	Project							
Protocol								
	Quality of Tags							
	Tissue							
	Clustered SAGE Tags							
	- Scientifc name							
	n names							
	k common na	ame						
← 🗖 🖓 Parent Taxon								
A . T		·····						
						Ē	i	
Parent	Field	Not	Condition		Valu			
GSC SAGE Libraries	Description Protocol	🖌 coi	ntains T	CONT longS/		IATION		
Taxon	Accession		•	10090				
	ок	Apply	Close					

2.2. Highlight the GSC SAGE Libraries query and click the Search For Tags in Libraries button.



Now highlight the tags from your Tag Sequence query and drag them into the Search for Tags widget.

Click the libraries on the left column.

wailable			-	Res	ults										
					Absolute 💌 Quality Cutoff						0.9	0.99 +			
Name	Description	Protocol		X	X	X	X	X	X	X			1		
SM096	Large Intesti	the state is a second se	1		22	60	0	69	US S	5					
SM095	xxxxxx	LongSAGE			PAC	1993	50	8.AG	000	586					
SM099	X0000X	LongSAGE		ΠĻ	G.B.(000	6.8(603	55	661	Name	Description	Total Tags		
SM102	X0000X	LongSAGE			ACG	TGC	60.9	CCC	000	866	Name	Description	Totar rago		
SM091	X0000X	LongSAGE			sсаваавасвеаваесс 🗙	AAGCGCCTGCCGGGGG 🗙	текееекескекесте 🗙	эсстттасссесальеее 🗙	GGCAGCCCCCGGCTCCGC	AACAAGGGGGGAAGCT					
SM103	X0000X	LongSAGE			AG.	260	580	Đ	5C.B	\$C.A.					
SM104	X0000X	LongSAGE			ŭ			ŭ	ŭ	Å.	014440	(440500		
SM100	200000	LongSAGE			_	12	-	-	-	-	SM112	X0000X	110586		
SM106	200000	LongSAGE			_	2					SM107	X0000X	103008		
SM107	X0000X	LongSAGE													
SM108	X0000X	LongSAGE													
SM109	X0000X	LongSAGE													
SM105	X0000X	LongSAGE													
SM114	X0000X	LongSAGE													
SM112	X0000X	LongSAGE													
SM106a	X0000X	LongSAGE													
SM117	200000	LongSAGE													
SM116	X0000X	LongSAGE													
SM113	200000	LongSAGE													
SM118	200000	LongSAGE													
SM122	X0000X	LongSAGE													
SM124	X0000X	LongSAGE													
SM116a	X0000X	LongSAGE													
SM104_f30	200000	LongSAGE													
SM104_122	200000	LongSAGE													
SM119	200000	LongSAGE													
SM120	200000	LongSAGE													
	200000	LongSAGE													
SM133	200000	LongSAGE	=												
SM133 SM139		LongSAGE													
	X0000X	Leongen or													
SM139	x0000X	LongSAGE													
SM139 SM140															
SM139 SM140 SM137	X0000X	LongSAGE													

You might want to search public libraries for your tags. You can create a query for GEO libraries and repeat these steps.

৫ মি? New Query Definition - Query
General Query
Query Path: 🔽 Start GEO SAGE Library 🔽 🔄 End
GEO SAGE Library Name Description Long description Number of tags Protocol Target Experimental Tags Taxon Accession Scientifc name Common names Genbank common name Rank Parent Taxon
Parent Field Not Condition Value Accession Accession ■ ■ 10090
OK Apply Close