

Examples of dataset construction & other queries

Question 1: Where do I find a list of animals tested for gene X?

- 1. Go to the Populations directory.
- 2. Use the colonies_to_pipelines_mapping document to determine which pipeline(s) your gene of interest was processed through.
- 3. Find your gene of interest within the correct pipeline folder and download the .csv file. Both the gene name and the colony prefix is included in the file name.

Note that for primary phenotyping data, this will be in either MGP_Pipeline_1, MGP_Pipeline_2, Mouse_GP or MGP_Select. The gene may also have been processed through one of the other screens such as Expression or a specific infection challenge, so please check these if this is the type of data you are interested in.

4. This will give you the list of mutant animals tested plus the local controls (as defined above – the controls processed during the same weeks as the mutants). The unique identifiers for each animal are the ID, Mouse Name or Mouse Barcode.

Question 2: I want to download all data for test X for gene Y, including mutants, controls and a baseline

- 1. Follow steps 1-4 for Question 1 above.
- 2. Go to the Index directory. Look in the *index_of_indexes* file to narrow down which .csv index files you will need to check for the pipeline your gene of interest was processed through.
- 3. Look in the relevant numbered index files to search for the animals in your population file within the mouseID or mouseName columns.
- 4. The location and filename columns will then direct you to where the raw data for that animal and test can be found.
- 5. Go to the Pipelines directory to access the raw data for each test. Select the pipeline stated in the filename for your population then go to your test of interest.

There may be one or more folders within the test folder if there were different data types generated as part of the test (e.g. observation values and images).



- 6. The specific numbered file containing the measurements of interest is that listed in the index file checked during step 2.
- 7. To assemble a baseline, take a note of the baseline ID specified in the population filename.

e.g. in the following file *acbd5_tm1b(eucomm)wtsi_(ID-12488_BaselineID-94795)_PMAJ_(57-ROWS).csv* the baseline ID is 94795.

- 8. The baseline files can be found within the Baselines directory. Match up the ID from your population filename with the ID in a filename here.
- 9. The raw data for the list of animals in the baseline file will need to be found in the same way as those in the population file, by following steps 2-6.

Question 3: I want to access all the wildtype/control data for test X

- 1. We suggest you use control data from a single pipeline (unless you are doing a comparison) and combining all control data is not appropriate due to differences in the pipelines over the lifetime of the Mouse Genetics Project.
- 2. Take a look at the Pipelines document in the Documentation folder to select a suitable pipeline.
- 3. The baseline files can be found within the Baselines directory. The pipeline and genetic background within the filenames will help you select the appropriate file.
- 4. This will give you the list of control animals of that genetic background tested on that pipeline. The unique identifiers for each animal are the ID, Mouse Name or Mouse Barcode.
- 5. The raw data for the list of animals in the baseline file will need to be found in the same way as detailed above in answer to Question 2, steps 2-6.

Question 4: How do I find out what was considered significant/not significant when line X was phenotyped?

- 1. Go to the Heat_Map directory.
- 2. For a visual representation, download HeatMap.xlsx



- 3. The Legend tab and the Heat_map.pdf document will help you understand the colours to determine if the results of any test performed on a line were considered significant.
- 4. The HeatMapCalls.xlsx file will detail specific parameters measured in each test deemed significant and also give the corresponding Mammalian Phenotype (MP) term(s).

Appendix: Worked example of dataset construction

Note that the following screenshots were taken using Mozilla Firefox 85.0.2 and will likely appear visually different on other browsers. The functionality should be the same, however.

Q: I want the raw haematology data for the Kcne2 line of mice plus controls.

1. Find out which adult phenotyping pipeline this line was processed through. Go to Populations and open colonies_to_pipelines_mapping.pdf

Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/ Image: Space of the sectors		111/	1		Search	Q	ס לי	•••	/ping/	resources/mouse/phe	ger.ac.uk/pub/re	ftp://ftp.sa	0	۵	G	⇒ C
Image: Size Last Modified 1801/2021 10.1000 Issuines 1901/2021 16.4600 Documentation 1002/2021 15.3200 Index Index 1002/2021 15.300 Index Index Index Index Index Index Index Index																
															ſ	
Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index Index Index Index Index Index Index Index Index Index Index Index Index Index Index Influenz, Challenge Index Index Index Index Index Influenz, Challenge Index Index_Index Inde							/	vpin	/mouse/phenc	uk/pub/resour	.sanger.ac.u	of ftp://ft	ndex			
Name Size Last Modified 18/01/2021 16/10/00 19/01/2021 16/10/00 0 Documentation 10/02/2021 15/2000 0 Documentation 10/02/2021 15/2000 0 Populations 10/02/2021 15/2000 0 Populations 10/02/2021 15/2000 0 Populations 10/02/2021 15/2000 File: ReadMepdf 7/2 KB 10/02/2021 15/2000 10/02/2021 15/2000 10/02/2021 15/2000 15/2000 File: ReadMepdf 7/2 KB 10/02/2021 15/2000 V Index 10/02/2021 15/2000 15/2000 V Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ V V V Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ V V V V V Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ V V V V V V Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ V V <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>,</td> <td>., b</td> <td>file abo, priorie</td> <td>and pais/resource</td> <td>_</td> <td></td> <td></td> <td></td> <td></td> <td></td>							,	., b	file abo, priorie	and pais/resource	_					
Baselines 18/07/2021 10/10.00 Colonies 19/07/2021 16/10.00 Documentation 10/02/2021 15/40.00 Index 10/02/2021 15/20.00 Pipelines 10/02/2021 15/10.00 Pipelines 19/01/2021 16/10.00 Pipelines 19/01/2021 16/10.00 Pipelines 19/01/2021 16/10.00 Pipelines 19/01/2021				lified	ast Mov	1.	Size				, i	-				
Colonies 19/07/2021 164600 Documentation 10/02/2021 15:2000 Heat, Map 10/02/2021 15:2000 Pipelines 10/02/2021 15:2000 Populations 10/02/2021 15:2000 File: ReadMe.pdf 10/02/2021 15:2000 T12 KB 10/02/2021 15:2000 10/02/2021 15:2000 File: ReadMe.pdf 10/02/2021 15:2000 10/02/2021 15:3600 T12 KB 10/02/2021 15:3600 10/02/2021 15:3600 10/02/2021 15:3600 Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Mane Size Last Modified 10/02/2021 15:1000 10/02/2021 15:1000 DSC.chalenge 19/01/2021 16:01:00 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Discuste_chalenge 19/01/2021 16:01:00 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Discuste_chalenge 19/01/2021 19/01/2021 16:01:00 19/01/2021 16:01							0120									
 Documentation Hest, Map 10/02/2021 15:2000 10/02/2021 15:3600 10/02/2021 15:10:00 10/02/2021																
→ C Mare 10/02/2021 15:20:00 → C Mare 10/02/2021 15:20:00 → C Mare 10/02/2021 15:20:00 → C Mare 10/02/2021 15:36:00 → C Mare C Search V V N C Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ C Search V Intervention Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/											n					
index 10/02/2021 15:20:00 ippeines 10/02/2021 15:20:00 iPopulations 10/02/2021 15:20:00 iPipeines 10/02/2021 15:20:00 iPipeines 10/02/2021 15:20:00 iPipeines 10/02/2021 15:20:00 iPipeines 10/02/2021 15:30:00 iPipeines 10/02/2021 15:30:00 iPipeines 10/02/2021 15:30:00 iPipeines 10/02/2021 15:30:00 iPipeines iPipeines iPipeines iPipeines iPipeine iPipeine iPipeines iPipeine iPipeine iPipeines iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine iPipeine<																
 Pipeline: Populations File: ReadMe.pdf → C																
 index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Index of ftp://ftp.sange																
→ C M Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ C Search L L M C Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ C Search L L M C Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Size Last Modified Indo KB DSS_											1		<u> </u>			
 							712 KB				1		_ L			
Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Name Size Last Modified File: colonies_to_pipelines_mapping.pdf 1406 KB 10/02/2021 15:10:00 0 DSS_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_DSS_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Expression_Pipeline 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2				15156166	0, 02, 2021		112100					c				
Index of ftp://ftp.sanger.ac.uk/pub/resources/mouse/phenotyping/Populations/ Name Size Last Modified File: colonies_to_pipelines_mapping.pdf 1406 KB 10/02/2021 15:10:00 0 DSS_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_DSS_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Expression_Pipeline 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2		has												^	~	
Name Size Last Modified File: colonies_to_pipelines_mapping.pdf 1406 KB 10/02/2021 15/10.00 Influenza_Challenge 19/01/2021 16/01.00 MGP_Citrobacter_Challenge 19/01/2021 16/01.00 MGP_DSS_Challenge 19/01/2021 16/01.00 MGP_Expression_Pipeline 19/01/2021 16/01.00 MGP_Expression_Pipeline 19/01/2021 16/01.00 MGP_Influenz_Challenge 19/01/2021 16/01.00 MGP_Pipeline_1 19/01/2021 16/01.00 MGP_Pipeline_2 19/01/2021 16/01.00 MGP_Salmonella_Challenge 19/01/2021 16/01.00 MGP_Select 19/01/2021 16/01.00		111.1	-		, search	~	<i>w</i>		ping/Populations/	esources/mouse/prie	jer.ac.uk/pub/re	i itp://itp.sai	•	w		- 0
Name Size Last Modified File: colonies_to_pipelines_mapping.pdf 1406 KB 10/02/2021 15:10:00 Image: DSS_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 Image: MGP_Citrobacter_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 Image: MGP_DSS_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 Image: MGP_Expression_Pipeline 19/01/2021 16:01:00 19/01/2021 16:01:00 Image: MGP_Expression_Pipeline 19/01/2021 16:01:00 19/01/2021 16:01:00 Image: MGP_Influenza_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 Image: MGP_Salmonella_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 Image: MGP_Salmonella_Challenge 19/01/2021 16:01:00 19/01/2021 16:01:00 Image: MGP_Select 19/01/2021 16:01:00 19/01/2021 16:01:00															ſ	
File: colonies_to_pipelines_mapping.pdf 1406 KB 10/02/2021 15:10:00 IDSS_Challenge 19/01/2021 16:01:00 IDMGP_ESS_Challenge 19/01/2021 16:01:00 IDMGP_Expression_Pipeline_(Legacy) 19/01/2021 16:01:00 IDMGP_Pipeline_1 19/01/2021 16:01:00 IDMGP_Pipeline_1 19/01/2021 16:01:00 IDMGP_Pipeline_1 19/01/2021 16:01:00 IDMGP_Pipeline_1 19/01/2021 16:01:00 IDMGP_Pipeline_2 19/01/2021 16:01:00 IDMGP_Salmonella_Challenge 19/01/2021 16:01:00 IDMGP_Select 19/01/2021 16:01:00					ions/	ulatio	/Ρορι	ypin	'mouse/pheno	uk/pub/resourc			_			
Implement 19/01/2021 16:01:00 Implement 19/01/2021					ions/	ulatio	/Рорі	ypin	'mouse/pheno	uk/pub/resourd			_	ļ		
Implement 19/01/202 16:01:00				ified				ypin	'mouse/pheno	uk/pub/resourd		o higher level	🖺 Up			
MGP_Citrobacter_Challenge 19/01/2021 16:01:00 MGP_DSS_Challenge 19/01/2021 16:01:00 MGP_Expression_Pipeline 19/01/2021 16:01:00 MGP_Expression_Pipeline_(Legacy) 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2021 16:01:00 MGP_Select 19/01/2021 16:01:00					ast Mod	La	Size	ypin	'mouse/pheno		lirectory	o higher level ame	🖺 Up	ļ		
MGP_DSS_Challenge 19/01/2021 16:01:00 MGP_Expression_Pipeline 19/01/2021 16:01:00 MGP_Expression_Pipeline_(Legacy) 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2021 16:01:00 MGP_Select 19/01/2021 16:01:00				15:10:00	.ast Moc 0/02/2021	La 10,	Size	ypin	'mouse/pheno		lirectory pipelines_mapping.	o higher level ame e: colonies_to	1) Up	1		
MGP_Expression_Pipeline 19/01/2021 16:01:00 MGP_Expression_Pipeline_(Legacy) 19/01/2021 16:01:00 MGP_Influenza_Challenge 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2021 16:01:00 MGP_Select 19/01/2021 16:01:00				15:10:00 16:01:00	.ast Moc 0/02/2021 9/01/2021	La 10, 19,	Size	ypin	'mouse/pheno		lirectory pipelines_mapping.	o higher level ame e: colonies_to DSS_Challeng	۹ Up	1		
MGP_Expression_Pipeline_(Legacy) 19/01/2021 16:01:00 MGP_Influenza_Challenge 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2021 16:01:00 MGP_Select 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00	ast Moc 0/02/2021 9/01/2021 9/01/2021	La 10, 19, 19,	Size	ypin	'mouse/pheno		lirectory ipelines_mapping. lenge	o higher level ame e: colonies_to DSS_Challeng Influenza_Cha	۹۵ Up	Ī		
MGP_Influenza_Challenge 19/01/2021 16:01:00 MGP_Pipeline_1 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2021 16:01:00 MGP_Select 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 9/01/2021 9/01/2021 9/01/2021	La 10, 19, 19, 19,	Size	ypin	'mouse/pheno		directory pipelines_mapping. lenge ter_Challenge	o higher level ame e: colonies_to DSS_Challeng Influenza_Cha MGP_Citroba	top Up	1		
MGP_Pipeline_1 19/01/2021 16:01:00 MGP_Pipeline_2 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2021 16:01:00 MGP_Select 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moo 0/02/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021	La 10, 19, 19, 19, 19,	Size	ypin	'mouse/pheno		directory pipelines_mapping. lenge ter_Challenge llenge	o higher level ame colonies_to, DSS_Challeng Influenza_Cha MGP_Citroba MGP_DSS_Ch	עף ני	Ī		
MGP_Pipeline_2 19/01/2021 16:01:00 MGP_Salmonella_Challenge 19/01/2021 16:01:00 MGP_Select 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021	La 10, 19, 19, 19, 19,	Size	ypin	'mouse/pheno	g.pdf	directory ipelines_mapping. lenge ter_Challenge llenge un_Pipeline	o higher level ame colonies_to, DSS_Challeng Influenza_Chal MGP_Citroba MGP_DSS_Challeng MGP_Express	L Op	1		
MGP_Salmonella_Challenge 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021	Lz 10, 19, 19, 19, 19, 19, 19,	Size	ypin	'mouse/pheno	g.pdf	tirectory ipelines_mapping. lenge ter_Challenge llenge In_Pipeline n_Pipeline_(Legac	o higher level ame colonies_to, DSS_Challeng Influenza_Chi MGP_Citroba MGP_DSS_Ch MGP_Express MGP_Express	L Up	<u> </u>		
MGP_Select 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021	La 10, 19, 19, 19, 19, 19, 19,	Size	ypin	'mouse/pheno	g.pdf	lirectory ipelines_mapping, enge tengc terg_Challenge llenge on_Pipeline on_Pipeline_(Legac _Challenge	o higher level ame colonies_to, DSS_Challeng Influenza_Chi MGP_Citroba MGP_DSS_Ch MGP_Express MGP_Express MGP_Influen:	L Op	1		
				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021	Lz 10, 19, 19, 19, 19, 19, 19, 19,	Size	ypin	'mouse/pheno	g.pdf	tirectory ipelines_mapping. lenge ter_Challenge llenge nn_Pipeline nn_Pipeline _Challenge 1	e: colonies_to, DSS_Challeng Influenza_Chi MGP_Citroba MGP_DSS_Ch MGP_Express MGP_Express MGP_Influen: MGP_Pipelind	Up [1		
Mouse_GP 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021	Lz 10, 19, 19, 19, 19, 19, 19, 19, 19,	Size	ypin	'mouse/pheno	g.pdf	tirectory ipelines_mapping. lenge ter_Challenge llenge n_Pipeline n_Pipeline _Challenge 1 2	o higher level ame e: colonies_to, DSS_challeng Influenza_Ch. MGP_Citroba MGP_DSS_ch MGP_Express MGP_Express MGP_Influen: MGP_Pipelind MGP_Pipelind	Up [1		
				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021	Lz 10, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19	Size	ypin	'mouse/pheno	g.pdf	tirectory ipelines_mapping. lenge ter_Challenge llenge n_Pipeline n_Pipeline _Challenge 1 2	o higher level ame e: colonies_to, DSS_Challeng Influenz_Chi MGP_Dirroba MGP_Express MGP_Pripelini MGP_Pipelini MGP_Pipelini MGP_Salmon		1		
T27_Pipeline 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021 0/01/2021	Lz 10,/ 19,/ 19,/ 19,/ 19,/ 19,/ 19,/ 19,/ 19	Size	ypin	'mouse/pheno	g.pdf	tirectory ipelines_mapping. lenge ter_Challenge llenge n_Pipeline n_Pipeline _Challenge 1 2	o higher level ame e: colonies_to, DSS_Challeng Influenza_Chi MGP_Dirobas MGP_DSS_Chi MGP_Express MGP_Pipelini MGP_Pipelini MGP_Salmon MGP_Select	עף ני	1		
Trichuris_challenge 19/01/2021 16:01:00				15:10:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00 16:01:00	ast Moc 0/02/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021 9/01/2021	Lz 10, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19	Size	ypin	'mouse/pheno	g.pdf	tirectory ipelines_mapping. lenge ter_Challenge llenge n_Pipeline n_Pipeline _Challenge 1 2	o higher level ame e: colonies_to, DSS_Challeng Influenza_Ch. MGP_Diroba MGP_DSS_Ch MGP_Express MGP_Pipelini MGP_Pipelini MGP_Pipelini MGP_Select Mouse_GP	Upped to the second	1		



2. By searching this document, we can see that there is data for this line from the Mouse GP pipeline.

Note: for general details about this pipeline, we can look in the Documentation subdirectory.

Ush1c <tm1a(komp)wtsi></tm1a(komp)wtsi>	B6N	MCSB			MGP Select
Anp32e <tm1e(komp)wtsi></tm1e(komp)wtsi>	B6N	MCSC		Mouse GP	
Cfh <tm1a(eucomm)wtsi></tm1a(eucomm)wtsi>	B6N	MCSE		Mouse GP	
Rhot2 <tm1(komp)wtsi></tm1(komp)wtsi>	B6N	MCSF		Mouse GP	
3110001I22Rik <tm1a(eucomm)wtsi></tm1a(eucomm)wtsi>	B6N	MCSH		Mouse GP	
Kcne2 <tm1a(eucomm)wtsi></tm1a(eucomm)wtsi>	B6N	MCSJ		Mouse GP	
Rab29 <tm1a(eucomm)wtsi></tm1a(eucomm)wtsi>	B6N	MCSP		Mouse GP	
Pld5 <tm1b(komp)wtsi></tm1b(komp)wtsi>	B6Brd;B6Dnk;B6N-Ty	MCSQ			MGP Select
Spns2 <tm1b(komp)wtsi></tm1b(komp)wtsi>	B6Brd;B6Dnk;B6N-Ty	MCSR		Mouse GP	MGP Select
Sar1b <tm1a(eucomm)wtsi></tm1a(eucomm)wtsi>	B6N	MCSS		Mouse GP	

3. Still in the Populations subdirectory, go to the Mouse_GP folder and search for either the gene name (Kcne2) or colony prefix (MCSJ) to locate the correct file to download.

File: jari	id2_(ID-3106_BaselineID-65784)_MAEF_(28-ROWS).csv	13 KB	18/01/2021	16:37:00
File: kaz	zn_(ID-3906_BaselineID-46808)_MCFF_(62-ROWS).csv	26 KB	18/01/2021	16:39:00
File: kcr	ne2_(ID-4030_BaselineID-46808)_MCSJ_(47-ROWS).csv	20 KB	18/01/2021	16:48:00
File: kdr	m4b_(ID-1974_BaselineID-46785)_MBCQ_(63-ROWS).csv	32 KB	18/01/2021	16:33:00
File: kdr	m4c_(ID-1763_BaselineID-46808)_MBXV_(28-ROWS).csv	12 KB	18/01/2021	16:42:00

- 4. In the below screenshot showing a subset of the .csv file indicated above, we can see columns which will be important later:
 - a. Unique animal identifiers: any of ID, Mouse Name, Mouse Barcode (highlighted blue).
 - b. Cohort type (highlighted green): Control or Subject and Genotype (highlighted yellow) which shows that for this line the subjects were homozygous.
 - c. Sex (highlighted orange) to help us split the data for analysis appropriately.

	Α	В	С	D	E	F	G	н	1	J	К
1	ID	Mouse Name	Mouse Barcode	Cohort Type	Coat	D.O.B.	D.O.W.	D.O.D.	Sex	Genotype	G.Backgro
2	835901	MALM41.1d	M00835901	Control	Black	02-Jun-11	22-Jun-11	22-Sep-11	Female	WT	C57BL/6N;
З	835902	MALM41.1e	M00835902	Control	Black	02-Jun-11	22-Jun-11	22-Sep-11	Female	WT	C57BL/6N;
4	835903	MALM41.1f	M00835903	Control	Black	02-Jun-11	22-Jun-11	22-Sep-11	Female	WT	C57BL/6N;
5	837334	MAQG63.2g	M00837334	Control	Black	04-Jun-11	24-Jun-11	22-Sep-11	Female	WT	C57BL/6N;
6	837335	MAQG63.2h	M00837335	Control	Black	04-Jun-11	24-Jun-11	22-Sep-11	Female	WT	C57BL/6N;
7	842157	MAQG63.2i	M00842157	Control	Black	04-Jun-11	24-Jun-11	22-Sep-11	Female	WT	C57BL/6N;
8	835853	MAQG64.2d	M00835853	Control	Black	02-Jun-11	22-Jun-11	22-Sep-11	Female	WT	C57BL/6N;
9	835854	MAQG64.2e	M00835854	Control	Black	02-Jun-11	22-Jun-11	22-Sep-11	Female	WT	C57BL/6N;
10	863483	MAQG71.1g	M00863483	Control	Black	05-Jul-11	26-Jul-11	24-Oct-11	Female	WT	C57BL/6N;
11	863484	MAQG71.1h	M00863484	Control	Black	05-Jul-11	26-Jul-11	24-Oct-11	Female	WT	C57BL/6N;
12	863485	MAQG71.1i	M00863485	Control	Black	05-Jul-11	26-Jul-11	24-Oct-11	Female	WT	C57BL/6N;
13	863486	MAQG71.1j	M00863486	Control	Black	05-Jul-11	26-Jul-11	24-Oct-11	Female	WT	C57BL/6N;
14	832395	MCSJ23.3a	M00832395	Subject	Black	24-May-11	15-Jun-11	14-Sep-11	Male	Kcne2:Hom	C57BL/6N;
15	832396	MCSJ23.3b	M00832396	Subject	Black	24-May-11	15-Jun-11	14-Sep-11	Male	Kcne2:Hom	C57BL/6N;
16	832398	MCSJ23.3d	M00832398	Subject	Black	24-May-11	15-Jun-11	14-Sep-11	Male	Kcne2:Hom	C57BL/6N;
17	828310	MCSJ26.2b	M00828310	Subject	Black	21-May-11	15-Jun-11	14-Sep-11	Male	Kcne2:Hom	C57BL/6N;
18	828314	MCSJ26.2f	M00828314	Subject	Black	21-May-11	15-Jun-11	14-Sep-11	Female	Kcne2:Hom	C57BL/6N;



5. Now we have our list of animals, we can use these to find the raw data for haematology. At the top level, go to the Index directory and open index_of_indexes.pdf

Index of ftp://ftp.sanger.ac.uk/pub/resourc	es/mouse/prienotyping/		-
Name	Size	Last Modified	
Baselines		18/01/2021 10:10:00	
		19/01/2021 16:46:00	
Documentation		10/02/2021 15:40:00	
Heat Map		10/02/2021 15:20:00	
📕 Index		10/02/2021 15:20:00	
Pipelines		10/02/2021 15:10:00	
Populations		10/02/2021 15:20:00	
File: ReadMe.pdf	712 KB	10/02/2021 15:36:00	
File: index_42_(30000_rows).csv	5213 KB	21/01/2021 14:07:00	
File: index_43_(30000_rows).csv	4980 KB	21/01/2021 14:07:00	
File: index_44_(30000_rows).csv	5855 KB	21/01/2021 14:07:00	
File: index_45_(30000_rows).csv	6493 KB	21/01/2021 14:07:00	
File: index_46_(30000_rows).csv	6132 KB	21/01/2021 14:07:00	
File: index_47_(30000_rows).csv	6591 KB	21/01/2021 14:07:00	
File: index_48_(30000_rows).csv	8064 KB	21/01/2021 14:07:00	
File: index_49_(30000_rows).csv	5611 KB	21/01/2021 14:07:00	
File: index_50_(30000_rows).csv	4255 KB	21/01/2021 14:07:00	
File: index_51_(7681_rows).csv	1289 KB	21/01/2021 14:07:00	
File: index_of_indexes.pdf	410 KB	10/02/2021 15:12:00	

- 6. In this file we can see that the indexes for data from Mouse GP are in index files 35 50 so we only need to search 16 files, rather than all 51.
- 7. The relevant index files will then need to be searched for where:
 - a. The mouseId or MouseName (from the index) matches an ID or Mouse Name in the population file
 - AND
 - b. The sopId in the index file matches your test of interest (in this case Haematology)
- 8. The rows within the index files where the above two conditions are met should be extracted and combined to create a dataset of haematology data for the *Kcne2* line.