KOMP-DCC / I-DCC Report

Martin Ringwald The Jackson Laboratory

KOMP-DCC: Objectives

- Serve the KOMP research network laboratories as a central information resource regarding publicly available null and conditional mutants
- Provide query and display tools to support prioritizing new mouse genes for knockout experiments.
- Collect information generated by the KOMP, track progress of the knockout mutant production pipelines, and support coordination of work
- Provide web-based query and display tools for KOMP data
- Export data to other relevant community databases such as Ensembl, the UCSC Genome Browser, NCBI, Mouse Genome Informatics (MGI), and IMSR
- \Rightarrow We developed KOMP-DCC Database and Web Site (www.knockoutmouse.org)

NIH Knockout Mouse Project (KOMP)

The IKMC Web Portal

www.knockoutmouse.org

- Extension of KOMP-DCC web site to serve as the common web portal for all IKMC projects:
 - KOMP
 - EUCOMM
 - NorCOMM
 - TIGM
- Site developed and maintained by the KOMP-DCC and I-DCC

• Supported by NIH and the European Union





Objectives

- Provide data/tools internal to IKMC
 - Generate and maintain common gene list
 - Choose and prioritize target genes
 - Coordinate work among IKMC projects
 - Track progress
- Provide IKMC information to the public
 - Central public web site for IKMC data
 - Export IKMC data to other resources

Gene List + Annotations



Genome Annotations (Build 37)

Administrative Web Interface

(1							Welcome	, Read Only User (<u>logout</u>) Read only access
	Admin Home Lists	CSD Peperts - Pegener	n Penerte -		Find chang	ios hu Data				
	Admin Home Lists *	CSD Reports * Regenere	on Reports *	EUCOMM Reports *	rind chang	es by Date				
Γ	Filter Search Upload Do	wnload Saved Filters								
	Currently applied filters are displayed Adding a filter restricts the current wo Currently applied filters:	below. To add a new filter, click on the colu rking set	umn header of the colur	nn that displays the value by	which you wish to f	ilter.				Sorted by: Symbol
	No filters currently defined (viewing all genes)									
	Refine your search by adding or	removing multiple filters at once								
	Page: 1 of 1542 Next > Last >>									
#	Gene Y	Other IDs T	Biotype	Y	DCC Status 🍸	CSD 7	Regeneron	Repository Status	EUCOMM / NorCOMM	Other Status Y
1	0610005C13Rik MGI:1918911 Chr7:52823165-52830697(-) Secondary MGI IDs (show) Old Swebals (show)	Filter by Other IDs Ensembl: Don't care		Gene (miscRNA) bessed_transcript) N_processed_transcript)	Master	Un-Assigned	Un-Assigned			Gene traps (MGI): <u>30</u> Gene traps (TIGM): 78 IMSR: <u>79</u> IMSR ES Cells: <u>79</u> Non-IKMC alleles: 30
2	0610006L08Rik MGI:1923503 Chr7:81968501-81998696(-)	Don't care Vega: Don't care	\$	Gene (protein-coding)		Un-Assigned	Un-Assigned			
3	0610007C21Rik MG:1918918 Chr:531350865-31357006(+) Secondary MGI IDs (<u>show</u>) Old Symbols (<u>show</u>)	CCDS: Has CCDS ID(s) Cancel Submit CCDS19170.1	•	Gene (protein-coding) ein_coding) N_protein_coding)	Target	Un-Assigned	Un-Assigned		EUCOMM: Project ID: <u>72859</u> ES Cells - Targeting Confirmed Effective: 2010-06-06	Gene traps (IGTC): 1 Gene traps (IGI): 7 Gene traps (IGIM): 1 IMSR: 3 IMSR ES Cells: 2 IMSR Mice: 1 EUCOMM alleles: 2 IKMC alleles: 2 Non-IKMC alleles: 7
4	0610007L01Rik MGI:1918817 Chr5:130695614- 130719635(+) Secondary MGI IDs (<u>show</u>) Old Symbols (<u>show</u>)	Ensembl ID: <u>ENSMUSG00000053094</u> NCBI ID: <u>71667</u> CCDS ID: <u>CCDS39295.1</u>	Entrez Gene (NCBI	I) Gene (protein-coding) tein_coding)	Target	Un-Assigned	Assigned Project ID: VG13171 ES cell colonies screened / QC positives Effective: 2009-07-28	Regeneron Vector Available Regeneron ES Cell Available	EUCOMM: Project ID: <u>69211</u> Vector - Initial Attempt Unsuccessful Effective: 2010-03-08	Gene traps (IGTC): 12 Gene traps (MGI): 26 Gene traps (TIGM): 6 IMSR: 12 IMSR Mice: 1 IMSR Mice: 1 IKMC alleles: 1 Non-IKMC alleles: 26 All other alleles: 1 Ocer ther (MCH): 1

- apply data filters to define gene set
- assign gene set to production center

Apply multiple filters at once, and save and share filter sets





New data filters for allele information and IMSR data

- IKMC alleles
- KOMP alleles
- EUCOMM alleles
- NorCOMM alleles
- Non-IKMC targeted alleles
- IMSR ES cells
- IMSR mice

Addition of new filter capabilities



Search for Languishing Genes

Genes for which the most advanced pipeline status, excluding terminal and finished statuses, has not changed since the specified number of days "Languishing Genes" Report Utility available to all IKMC projects

T EUCOMM	▼ NorCOMM
EUCOMM Assignment Status Un-Assigned Assigned	Un-Assigned
EUCOMM Pipeline Status Multiple select allowed On Hold Withdrawn From Pipeline Design Requested Redesign Requested Alternate Design Requested VEGA Annotation Requested Weak of the pipeline status Search for Languishing Genes	Un-Assigned Un-Assigned
Days since last pipeline status update: (excludes terminal and finished statuses)	Un-Assigned

Filters for genes with nominations and for high-priority genes

Page: 1	of 1721	Next >	Last >>		
Y	DCC Status 🔻	CSD		V	Regeneron
scRNA) anscript)	Filter by DCC	Status			ssigned
sed_transcript)	Select which lis Multiple select all	st owed			
	KOMP Target List IKMC Target List IKMC Master List Other (Not on any list)				
known)	Is High Priority Target			ssigned	
	Cancel Submit				
otein-coding) ig) _coding)	Target	Un-Assigne	d		Un-Assigned

Objectives

- Provide data/tools internal to IKMC
 - Generate and maintain common gene list
 - Choose and prioritize target genes
 - Coordinate work among IKMC projects
 - Track progress
- Provide IKMC information to the public
 - Central public web site for IKMC data
 - Export IKMC data to other resources





IKMC Gene Progress Summary 🥹

Targeted alleles					
Total Gapes	к	OMP	EUCOMM	NerCOMM	
Total Genes	CSD	Regeneron	EUCOMMTools	Norecomm	
Vectors available	6597	4733	7956	839	
ES cells available	4789	3430	6486	569	
Mutant mice available	349	281	573	4	

View details and project goals View details about the acronyms used

Gene trap alleles

Total Genes	TIGM	EUCOMM	NorCOMM
ES cells available	9404	4418	4593
Mutant mice available	81		

NOTE: Not all gene traps are in C57BL/6 ES cells. View strain information

Status Report: updated on a daily basis



Status Report: updated on a daily basis

NorCOMM

NorCOMM

4593

839

569

4



alleles have been generated.





IKMC Gene Progress Summary 99

largeted alleles							
Total Conor	к	OMP	EUCOMM	NorCOM			
Total Genes	CSD	Regeneron	EUCOMMTools				
Vectors available	6597	4733	7956	83			
ES cells available	4789	3430	6486	56			
Mutant mice available	349	281	573				
Many datalla and analast av	and a Manual da	Andles always they	a second second				

View details and project goals View details about the acronyms used

Gene trap alleles

Total Genes	TIGM	EUCOMM	NorCOMM
ES cells available	9404	4418	4593
Mutant mice available	81		

NOTE: Not all gene traps are in C57BL/6 ES cells. View strain information

Query Results Summary

(Internatio	help me location Search 20681, Chr13:22210730-22311689				
	Home About IKMC MartSearch	h Download I	Nominate gene FAQ Order Products	Contact	IKMC	
F	Results: 1 - 1 of 1 for <i>Fgf10</i>		Generated allele	types: 0	Conditional-Ready N No	n-Conditional Deletion
	Corro (9)	IKMC Knockout Attempts 🥹			Other Becourses	
	Gene 🤍	Project	Status		Availability	Other Resources S
	Fgf10 Express interest	KOMP-Regeneron	ES cell colonies screened / QC positives Order vector » Order ES cell »		Order vector » Order ES cell »	IMSR: 4
Symbol: Fgf10 Chr13:119458598-119580922(+) (MGI,UCSC,Ensembl)		EUCOMM	ES Cells - Targeting Confirmed		Order ES cell » Order vector »	Targeted Mutations (MGI): 8 Other Mutations (MGI): 7
I	■ Show Other Ids	TIGM Gene traps 3 Gene traps available			Order gene trap ES cell »	
F	Results: 1 - 1 of 1 for Fgf10					











Mutagenesis Predictions

This gene has 3 wild type transcripts, of which 1 are protein-coding. Following removal of the floxed region, 1 transcript is **predicted** to produce a truncated protein product of which 0 may be subject to non-sense mediated decay (NMD). The original allele for this mutation is of type 'Knockout-First - Reporter Tagged Insertion'. The table below shows the **predicted** structure of the gene transcripts after application of Flp and Cre (forming a 'Knockout-First, Post-Flp and Cre - Deletion, No Reporter' allele - more information on IKMC alleles can be found here). Click the 'view' button for each transcript to see the full prediction for that transcript.

Ensembl transcript id	Ensembl Biotype	Floxed transcript description	Details
ENSMUST0000022246	protein_coding	Residual N-terminal, novel C-terminal product	

Predicted translation:

MMKWILTRICASAFPHLPGCCCCFILLFLV3SFPVTCQALGQDMVSQFATNCSSSSSSFSSPSSAGRHVRSYNHLQGDVRW RRLFSFTKYFLTFIEKNGKV3GTKNEDCPYKRV*

Ensembl Exon ID	Pfam Domains	WildType	Floxed			
ENSMUSE00000389351		UTR + start codon + CDS	UTR + start codon + CDS			
ENSMUSE00000120353		CDS	Deleted			
ENSMUSE00000354860		CDS + stop codon + UTR	CDS + stop codon + UTR			
Legend: in frame deleted frameshifted						

show/hide more transcripts









IKMC Gene Progress Summary 99

Та	arge	eted	alle	les

Total Gapes	ко	MP	EUCOMW	NorCOMM
Total Genes	CSD	Regeneron	EUCOMMTools	Norcomin
Vectors available	6597	4733	7956	839
ES cells available	4789	3430	6486	569
Mutant mice available	349	281	573	4

View details and project goals View details about the acronyms used

Gene trap alleles

Total Genes	TIGM	EUCOMM	NorCOMM
ES cells available	9404	4418	4593
Mutant mice available	81		

NOTE: Not all gene traps are in C57BL/6 ES cells. View strain information

IKMC MartSearch

The **BioMarts**

This portal integrates information on IKMC mouse knockout resources with numerous other relevant datasets, including Ensembl, Europhenome, EurExpress and EMMA. For more information about this portal and the way in which it unites and searches the data, please see the about page. If you wish to access any of these BioMarts directly, click on the names of the datasets in the image below and you will be taken to the standard MartView interface.



Objectives

- Provide data/tools internal to IKMC
 - Generate and maintain common gene list
 - Choose and prioritize target genes
 - Coordinate work among IKMC projects
 - Track progress
- Provide IKMC information to the public
 - Central public web site for IKMC data
 - Export IKMC data to other resources

IKMC tracks in Genome Browsers

- Ensembl
- UCSC: Mouse and Human
- MGI Mouse Gbrowse

Tracks link to entries in IKMC web portal

Representation of IKMC alleles: registration and nomenclature

- Developed for all types of targeted alleles automated procedures to upload information into MGI and to assign official nomenclature, and pertinent QC
- All KOMP, EUCOMM, and NorCOMM alleles are now represented in MGI, allele information is updated on a weekly basis
- Allele reports available for download from IKMC web portal and from MGI ftp site

Registration of IKMC alleles in MGI: IDs and nomenclature

MGI About Halo EAO		Homo	Phonotypes Eve	ression Pecombinas	Eusction Dathways	Symbols, or IDs	Quick Search
Search - Download	d 👻 Mo	re Resources	Submit Data	Find Mice (IMSR)	Analysis Tools	Contact Us	
Abca2tm2a(KOMP)Wtsi Your Input Welcome Targeted Allele Detail							
Nomenclature	Symbol: Name: MGI ID: Gene:	Abca2 ^{tm2a} (ATP-binding ca MGI:4451500 Abca2 Locatio	KOMP) Wtsi ssette, sub-family A (AB	C1), member 2; targeted	I mutation 2a, Wellcome Tr Senetic Position: Chr2, 17.2	rust Sanger Institute 5 cM, cytoband A2-B	e k
Mutation origin	Mut Germline Pa S	tant Cell Lines: Transmission: arent Cell Line: Strain of Origin:	EPD0514_3_A01, EP EPD0514_3_G02 (W Unknown JM8A3.N1 (ES Cell) C57BL/6N-A ^{tm1Brd}	D0514_3_E01, EPD0 ellcome Trust Sanger Ins	514_3_E03, EPD0514_: titute)	3_F04, EPD0514_	3_G01,
Mutation description	Allele Typ Mutatic	Illele Type: Targeted (Floxed/Frt) Mutation: Insertion Vector: L1L2_Bact_P The L1L2_Bact_P cassette was inserted at position 25288175 of Chromosome 2 upstream of the critical exon(s) (Build 37). The cassette is composed of an FRT site followed by lacZ sequence and a loxP site. This first loxP site is followed by neomycin under the control of the human beta-actin promoter, SV40 polyA, a second FRT site and a second loxP site. A third loxP site is inserted downstream of the targeted exon(s) at position 25291409. The critical exon(s) is/are thus flanked by loxP sites. A "conditional ready" (floxed) allele can be created by flp recombinase expression in mice carrying this allele. Subsequent cre expression results in a knockout mouse. If cre expression occurs without flp expression, a reporter knockout mouse will be created. Further information on targeting strategies used for this and other KOMP alleles can be found at http://www.knockoutmouse.org/aboutkompstrategies. (<i>J: 148605</i>)					
Find Mice (IMSR)	Mouse strains and cell lines available from the International Mouse Strain Resource (IMSR) Carrying this Mutation: Mouse Strains: 0 strains available Cell Lines: 0 lines available Carrying any Abca2 Mutation: 2 strains or lines available						
References	Original: All:	J:148605 Well MGI Direct Dat 1 reference(s)	come Trust Sanger Inst a Submission 2009;():	itute, "Alleles produced fo	or the KOMP project by the	Wellcome Trust San	ger Institute"

Targeted IKMC alleles represented in MGI (as of 9/26/11)

KOMP:	84,373 ES cell lines	12,282 alleles	8,137 genes
EUCOMM:	96,597 ES cell lines	12,420 alleles	6,483 genes
NorCOMM:	3,160 ES cell lines	580 alleles	569 genes
Total:	184,130 ES cell lines	25,282 alleles	13,145 genes
Cond-Ready:	91,564 ES cell lines	10,932 alleles	10,031 genes

The IKMC Web Portal Team

The Jackson Laboratory

Jeremy Mason Kevin Stone Hamsa Tadepally Jim Kadin Carol Bult Janan Eppig Martin Ringwald The Sanger Institute

Vivek Iyer Darren Oakley Bill Skarnes

Supported by NIH/NHGRI and the European Union

Nucleic Acids Research Advance Access published October 6, 2010

Nucleic Acids Research, 2010, 1–7 doi:10.1093/nar/gkq879

The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium

Martin Ringwald^{1,*}, Vivek Iyer², Jeremy C. Mason¹, Kevin R. Stone¹, Hamsa D. Tadepally¹, James A. Kadin¹, Carol J. Bult¹, Janan T. Eppig¹, Darren J. Oakley², Sebastien Briois³, Elia Stupka⁴, Vincenza Maselli⁴, Damian Smedley⁵, Songyan Liu⁶, Jens Hansen⁷, Richard Baldock⁸, Geoff G. Hicks⁶ and William C. Skarnes^{2,*}