FORWARD GENETIC APPROACHES FOR MENDELIAN DISEASE MODELING IN MICE

Exome sequencing for variant discovery module for “Big Data Skills Training for Professors BD2K”

May 24, 2017

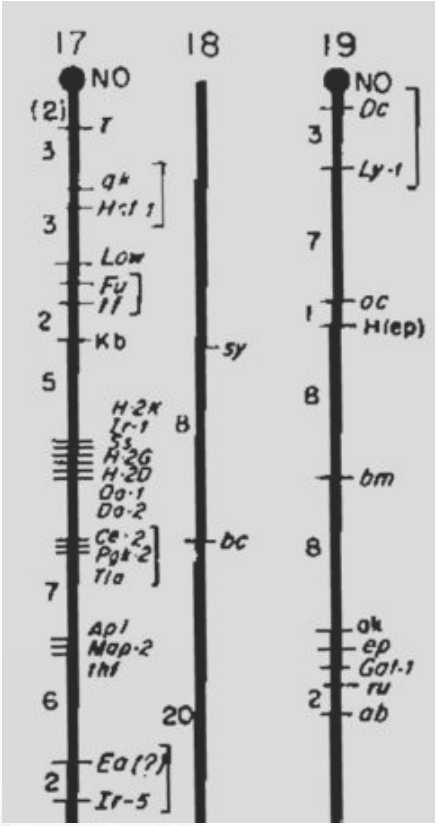
Laura Reinholdt, Ph.D.
@laura_reinholdt
laura.reinholdt@jax.org

The Jackson Laboratory
Leading the search for tomorrow's cures

Forward Genetics in Mice

ADVANTAGES

- unbiased, phenotype to gene
- selective breeding, large consanguineous pedigrees
- genetically defined inbred strain backgrounds
- mutation rates
  - spontaneous, ~6 x 10^{-6} / locus / generation



RESOURCES AT THE JACKSON LABORATORY

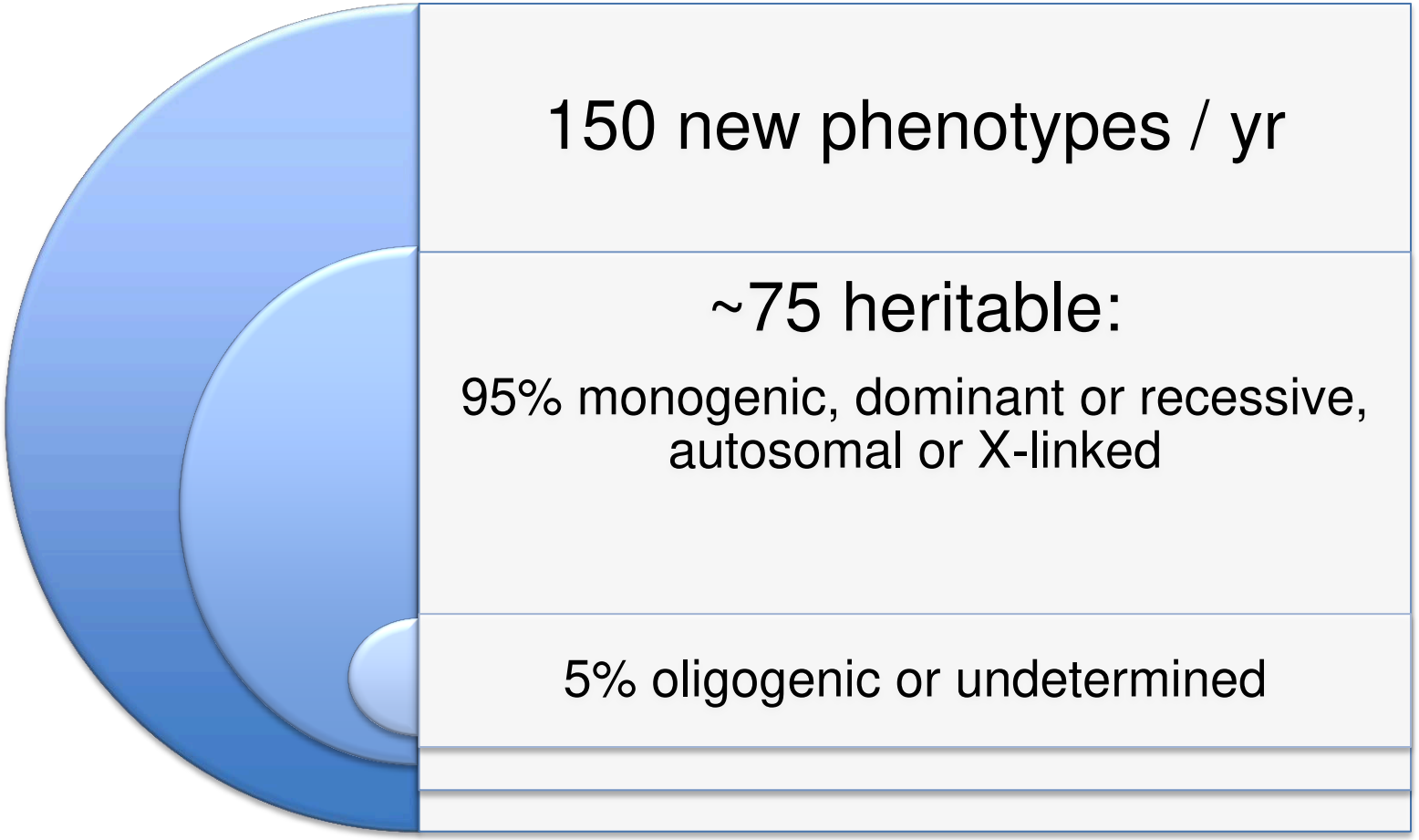
- scale
- quality control for ‘phenodeviants’

KEY TECHNOLOGIES

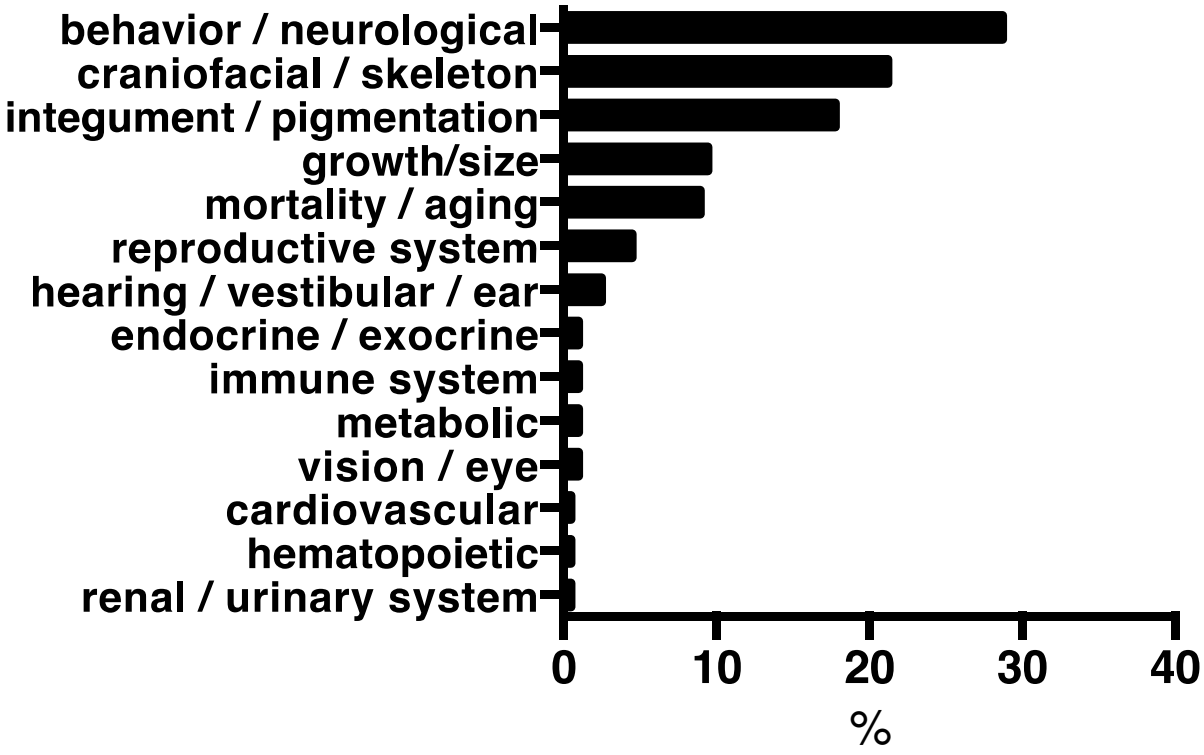
- High-throughput sequencing
- Common inbred strain reference genomes (Sanger Mouse Genomes Project)


The Jackson Laboratory
The Jackson Laboratory, 2013

Annual recovery of “phenodeviants” from a production scale vivarium

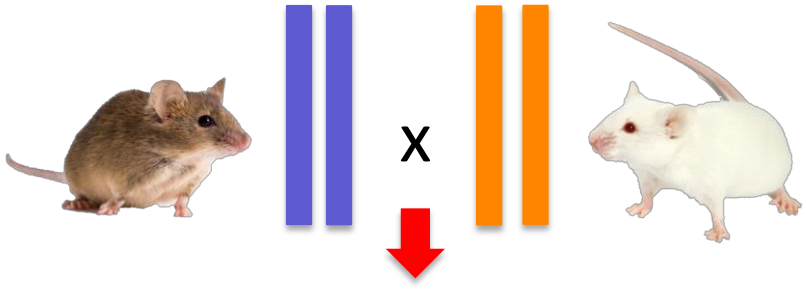


Heritable Mendelian phenotypes at the The Jackson Laboratory mouse “clinic”

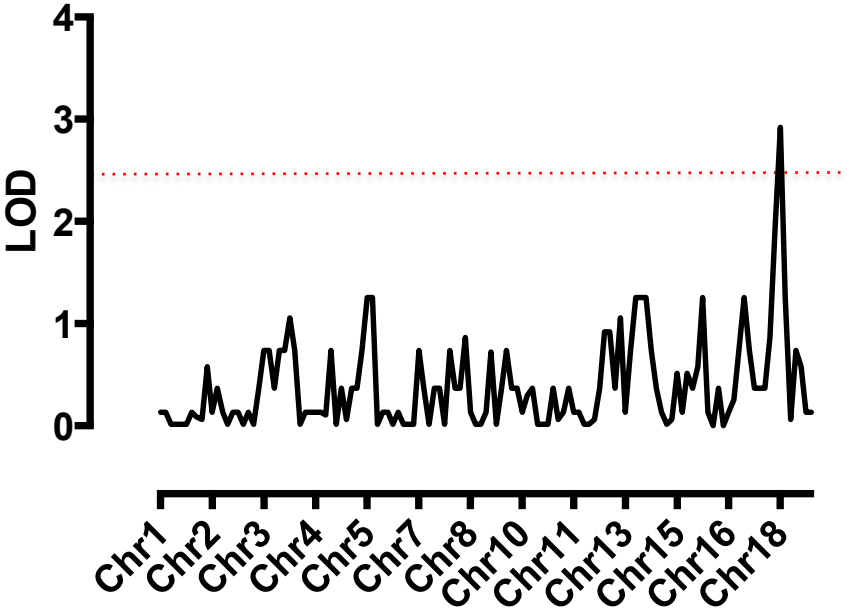


Mammalian Phenotype Ontology: Smith CL and Eppig JT., Mamm Genome. 2012 Oct;23(9-10):653-68

Mapping cross, recessive trait



F1



F2

LOD scores

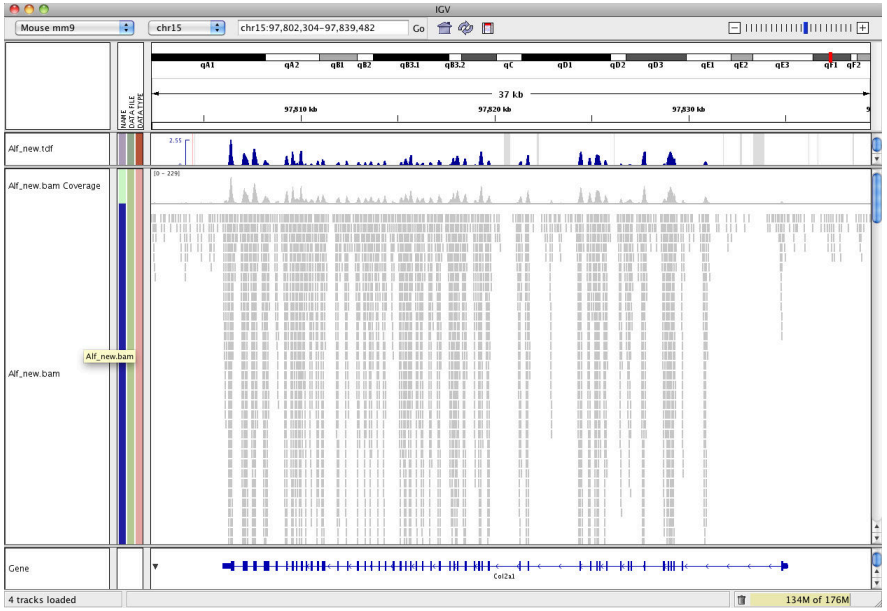
0.1
0.3
1.2
0.9
3.0

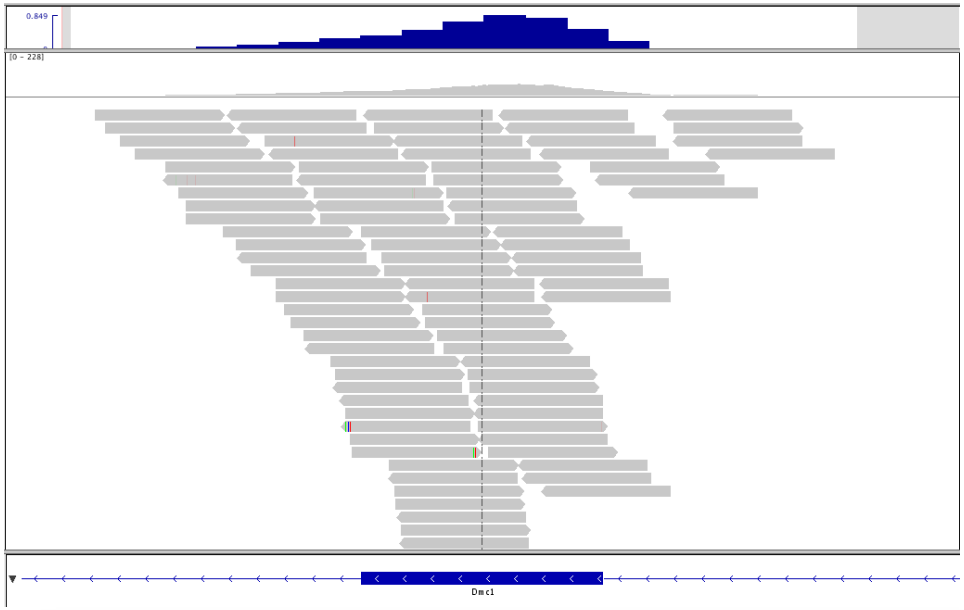
Chr_Pos

$$LOD = \log10 \frac{(1-\theta)^{NR} \times \theta^{R}}{0.5^{(NR+R)}}$$

$$\theta = R/(NR+R)$$

New technologies, resources, impact





- Based on Mouse Gene Catalog (curated by Mouse Genome Database), non-redundant gene predictions from NCBI, Ensembl and Vega
- 203,225 exonic regions ~33,000 genes and miRNAs

Fairfield *et al. Genome Biology* 2011, **12**:R86
http://genomebiology.com/2011/12/9/R86

Highly accessed

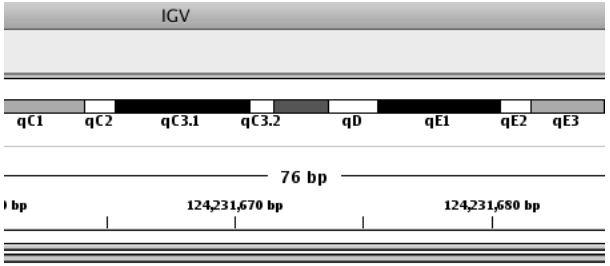
Genome Biology

METHOD Open Access

Mutation discovery in mice by whole exome sequencing

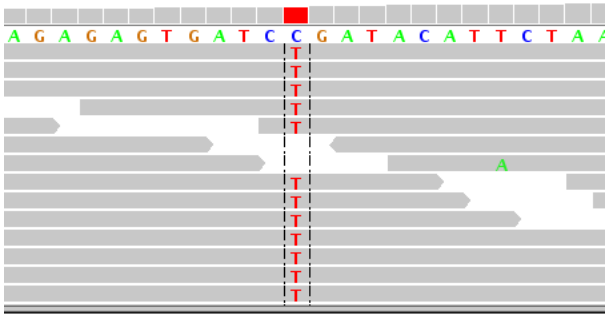
Heather Fairfield[1], Griffith J Gilbert[1], Mary Barter[1], Rebecca R Corrigan[2], Michelle Curtain[1], Yueming Ding[3], Mark D'Ascenzo[4], Daniel J Gerhardt[4], Chao He[5], Wenhui Huang[6], Todd Richmond[6], Lucy Rowe[1], Frank J Probst[2], David E Bergstrom[1], Stephen A Murray[1], Carol Bult[1], Joel Richardson[1], Benjamin T Kile[7], Ivo Gut[8], Jorg Hager[8], Snaevar Sigurdsson[9], Evan Mauceli[9], Federica Di Palma[9], Kerstin Lindblad-Toh[9], Michael L Cunningham[10], Timothy C Cox[10], Monica J Justice[2], Mona S Spector[5], Scott W Lowe[5], Thomas Albert[4], Leah Rae Donahue[1], Jeffrey Jeddeloh[6], Jay Shendure[10] and Laura G Reinholdt[1*]

Variant features



Reference (REF)

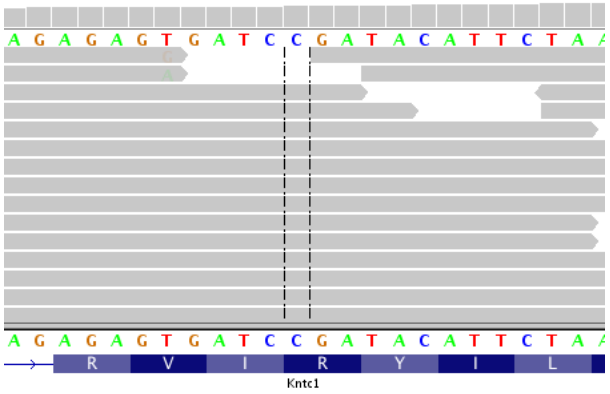
depth = 12



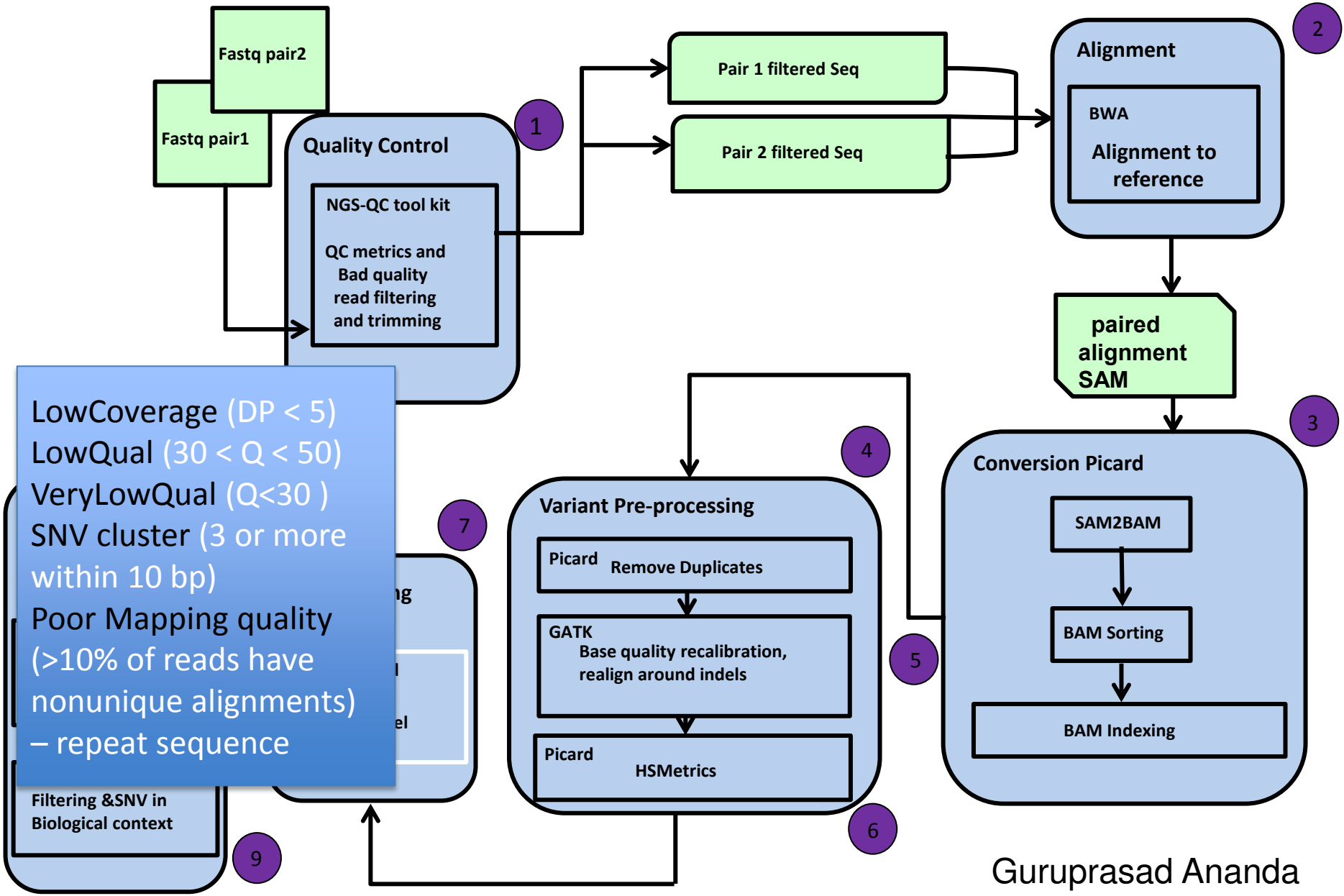
allele ratio (REF/ALT) = 0/12

allele frequency (var %) (for non reference allele) = 1.0

depth = 13



Mouse exome CIVET pipeline



Guruprasad Ananda

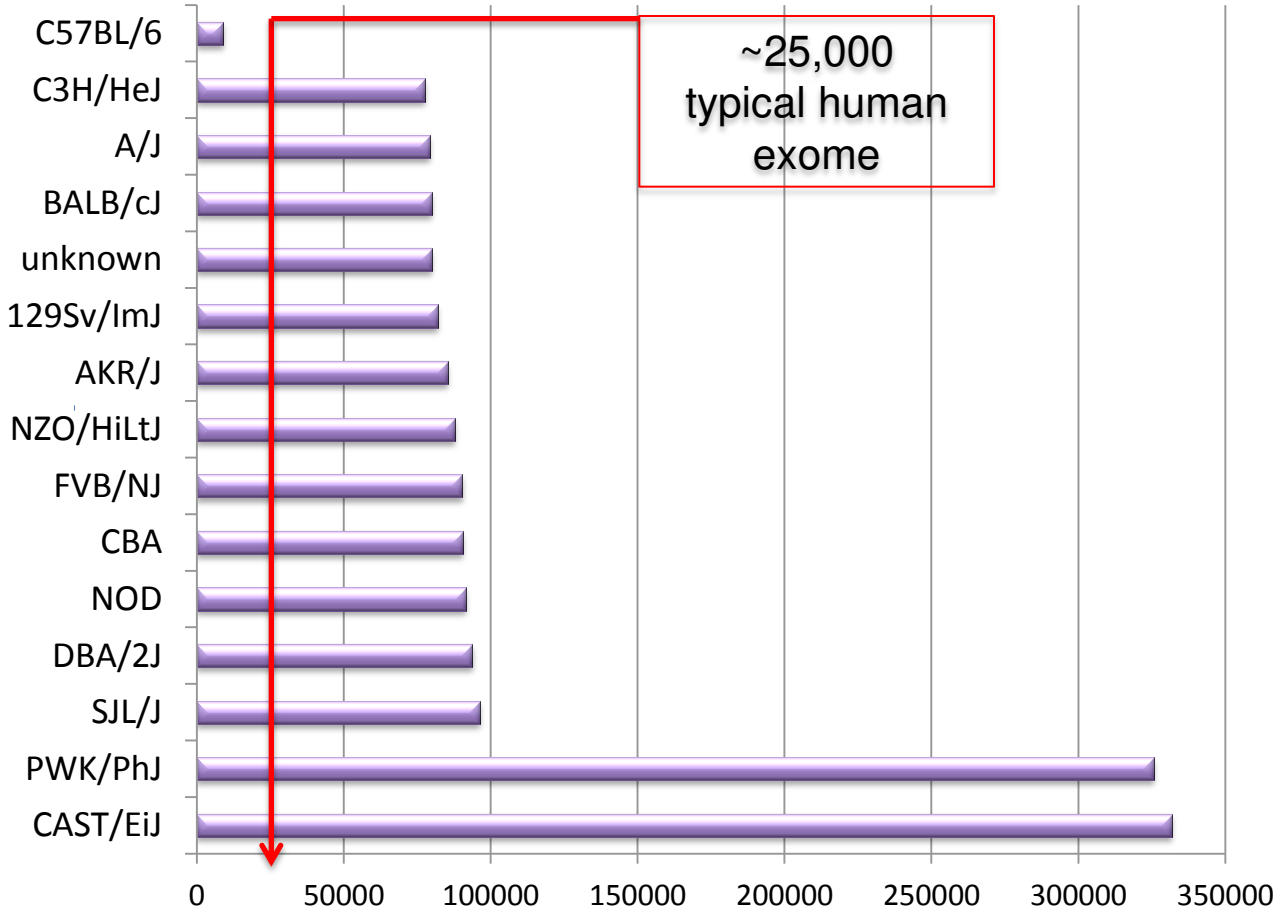
Variant caller format (VCF)

##contig=<ID=chrY,length=91744698>
##reference=file:///data/shared/genomes/Mus_musculus/UCSC/mm10_3-18-2013/Sequence/WholeGenomeFasta/genome.fa
##source_20140024.1=vcf-annotate(r797) -a /data/shared/mmr/dbSNP.mm10.tab.gz -c CHROM,FROM,TO,ID

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SJL									
chr10	3120209	rs29378454	A	G	2004.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=77	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	3120231	rs29358647	A	G	2049.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=85	Dels=0.00	FS=3.740	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	3120311	rs29371433	C	T	1125.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=43	Dels=0.00	FS=6.453	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	3125031	rs47392805	G	A	625.77	PASS	AC=2	AF=1.00	AN=2	DP=25	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=57.17	MQ0=0	QD=25.03
chr10	3125055	rs48707627	G	C	509.77	PASS	AC=2	AF=1.00	AN=2	DP=17	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=55.80	MQ0=0	QD=29.99
chr10	3125193	rs50658327	C	T	234.78	PASS	AC=2	AF=1.00	AN=2	DP=8	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0	QD=29.35
chr10	3268447	rs29364458	A	G	674.77	PASS	AC=2	AF=1.00	AN=2	DP=24	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=53.32	MQ0=0	QD=28.12
chr10	3268450	rs29346449	G	A	614.77	PASS	AC=2	AF=1.00	AN=2	DP=22	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=53.62	MQ0=0	QD=27.94
chr10	3268460	.	TG	T	668.73	PASS	AC=2	AF=1.00	AN=2	DP=19	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=54.01	MQ0=0	QD=35.20	RPA=3,2
chr10	3268474	rs29368948	C	T	620.77	PASS	AC=2	AF=1.00	AN=2	DP=21	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=54.68	MQ0=0	QD=29.56
chr10	3545733	rs29315555	A	G	845.81	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=35	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	3554546	rs29376248	T	C	450.77	PASS	AC=2	AF=1.00	AN=2	DP=15	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0	QD=30.05
chr10	3554631	rs29383399	C	T	497.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=20	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	3872653	rs33849715	T	C	637.77	PASS	AC=2	AF=1.00	AN=2	DP=21	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=57.77	MQ0=0	QD=30.37
chr10	3872657	rs29364945	G	A	711.77	PASS	AC=2	AF=1.00	AN=2	DP=25	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=58.13	MQ0=0	QD=28.47
chr10	3933583	rs29358987	C	T	1011.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=40	Dels=0.00	FS=2.636	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	3978708	rs29362746	rs29320259	T	C	582.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=21	Dels=0.00	FS=3.222	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=57.70
chr10	3978709	rs29362746	G	A	593.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=21	Dels=0.00	FS=3.222	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=57.70	MQ0=0
chr10	3978749	.	T	TTTTG	467.75	PASS	AC=1	AF=0.500	AN=2	BaseQRankSu	DP=14	FS=6.990	HaplotypeSc	MLEAC=1	MLEAF=0.50	MQ=56.95	MQ0=0	MQRankSum
chr10	4007800	rs29378605	rs222352237	T	C	684.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=28	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00
chr10	4032345	rs29322875	T	C	525.28	PASS	AC=1	AF=0.500	AN=2	BaseQRankSu	DP=26	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=1	MLEAF=0.50	MQ=60.00	MQ0=0
chr10	4041894	rs47215848	G	T	941.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=34	Dels=0.00	FS=4.523	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	4048178	rs51676436	C	T	1479.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=60	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	4103364	rs29333656	C	A	1495.77	PASS	AC=2	AF=1.00	AN=2	BaseQRankSu	DP=61	Dels=0.00	FS=10.685	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0
chr10	4155568	.	CTT	C	556.73	PASS	AC=2	AF=1.00	AN=2	DP=13	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=57.23	MQ0=0	QD=42.83	RPA=3,1
chr10	4155580	rs29320956	A	G	396.77	PASS	AC=2	AF=1.00	AN=2	DP=14	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=55.69	MQ0=0	QD=28.34
chr10	4166804	rs29357681	C	T	104.03	LowCoverage	AC=2	AF=1.00	AN=2	DP=4	Dels=0.00	FS=0.000	HaplotypeSc	MLEAC=2	MLEAF=1.00	MQ=60.00	MQ0=0	QD=26.01
chr10	4368518	rs51768590	A	C	177.77	PASS	AC=1	AF=0.500	AN=2	BaseQRankSu	DP=101	Dels=0.00	FS=4.671	HaplotypeSc	MLEAC=1	MLEAF=0.50	MQ=55.20	MQ0=0
chr10	4368528	rs227191183	A	G	287.77	PASS	AC=1	AF=0.500	AN=2	BaseQRankSu	DP=129	Dels=0.00	FS=11.550	HaplotypeSc	MLEAC=1	MLEAF=0.50	MQ=55.92	MQ0=0
chr10	4368577	rs258468563	G	T	845.77	SnpCluster	AC=1	AF=0.500	AN=2	BaseQRankSu	DP=211	Dels=0.00	FS=12.259	HaplotypeSc	MLEAC=1	MLEAF=0.50	MQ=55.48	MQ0=0

option for soft filter

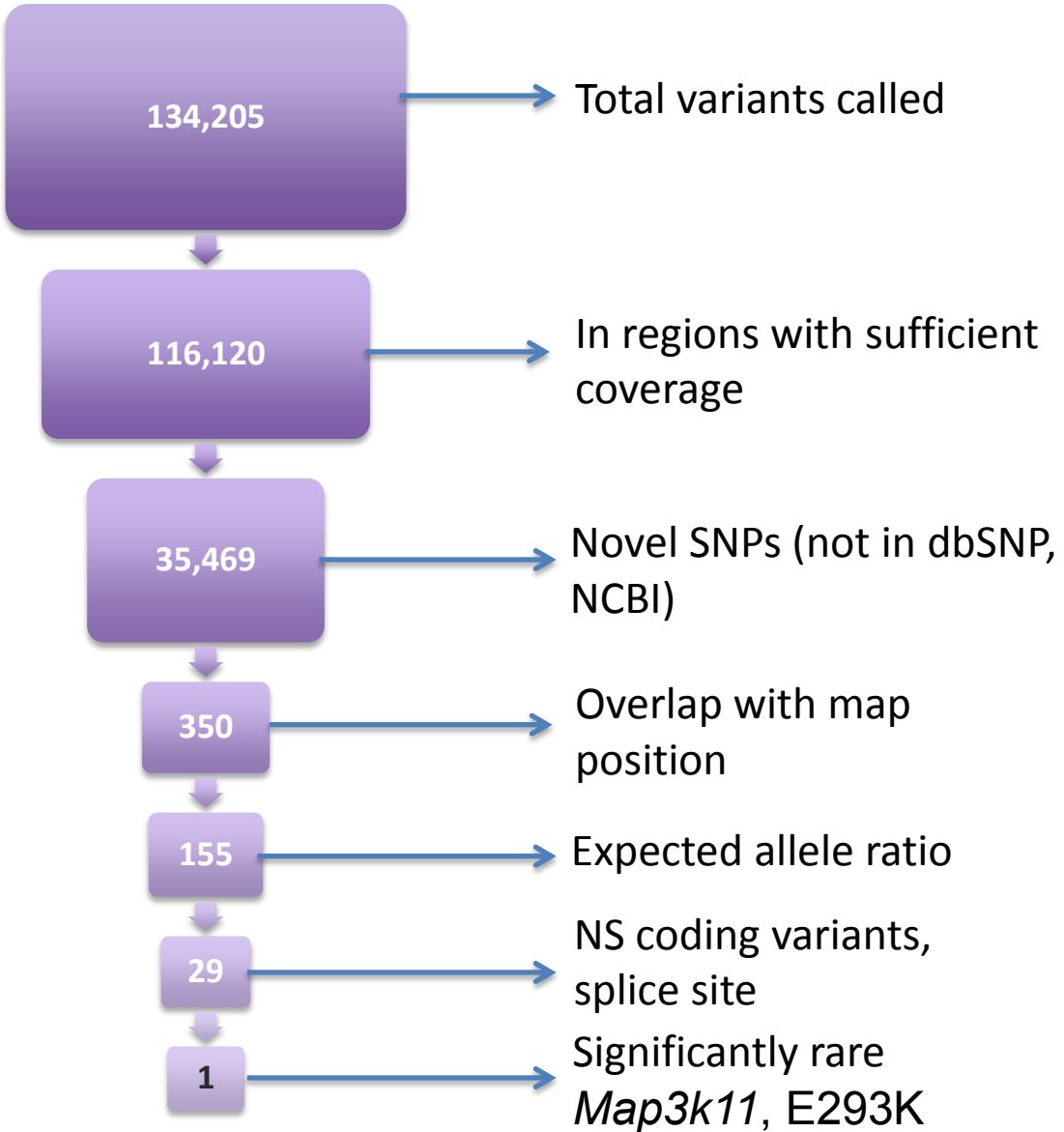
Total exome variants by strain background



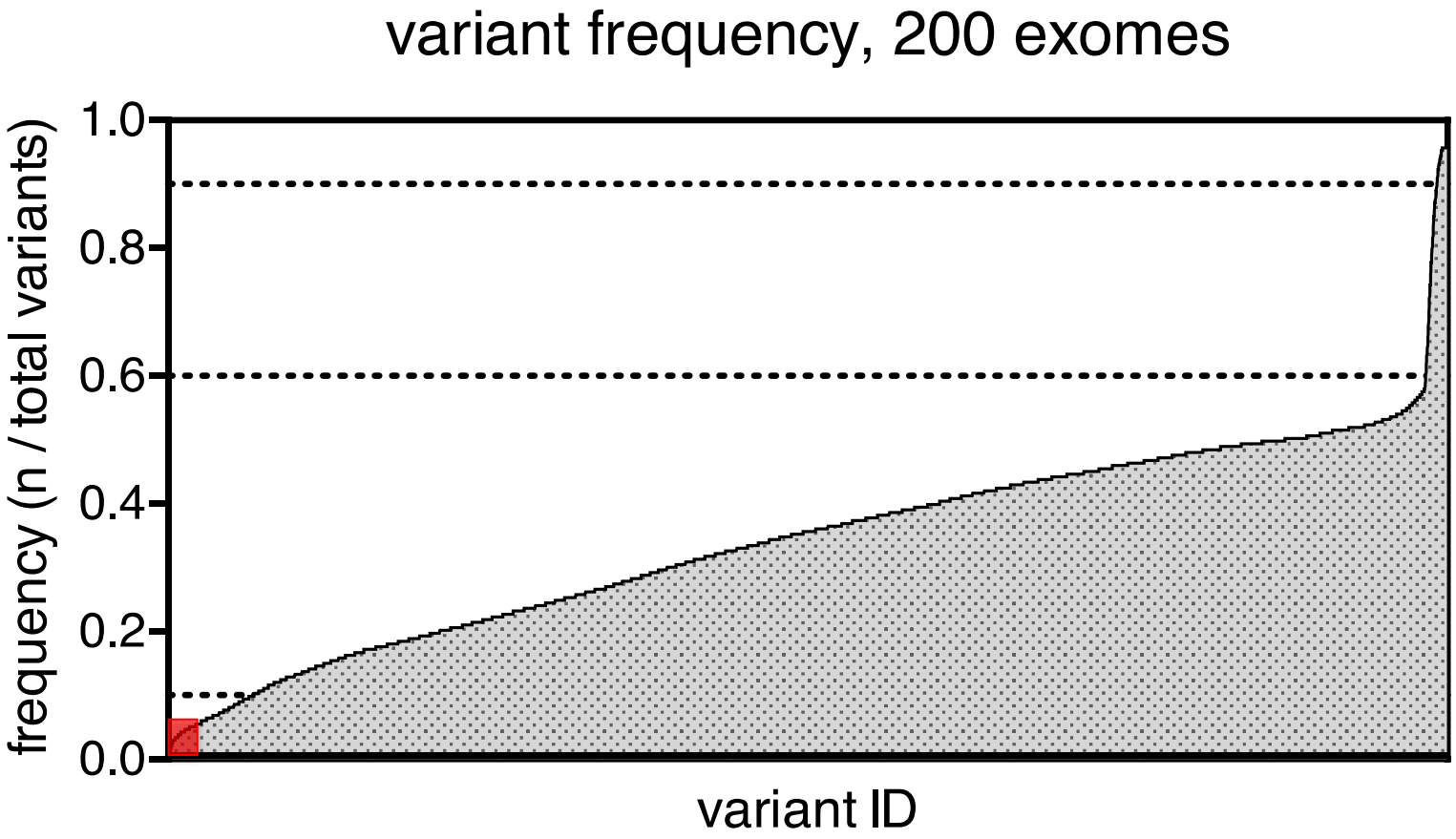
Stepwise reduction of candidate variants, single sample

"*bloodline*", recessive

- reduced bone mass
- abnormal tooth pulp



Exome variant frequency distribution, >200 mouse exomes



Exome variants → putative causative mutations

https://mmrdb.jax.org

1,251,572
total variants

8,360
flagged candidate mutations

Flag if:

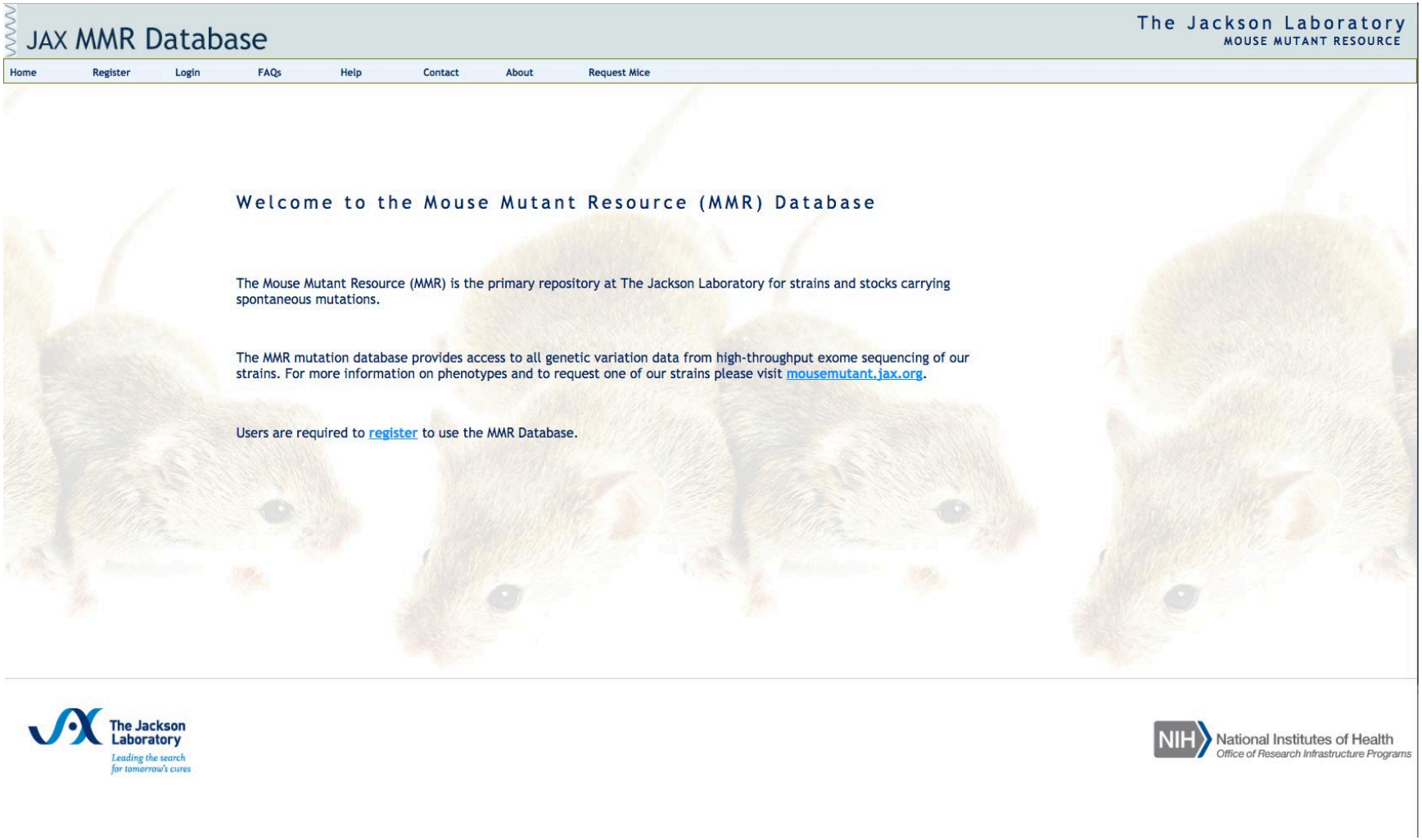
- Variant allele frequency = inheritance
- Chr pos = linkage data
- Novel variant (not in dbSNP)
- absent in WT control
- < 5% frequency in database

1,918
high / moderate impact

108
subchromosomal map position

80
confirmed

Demo

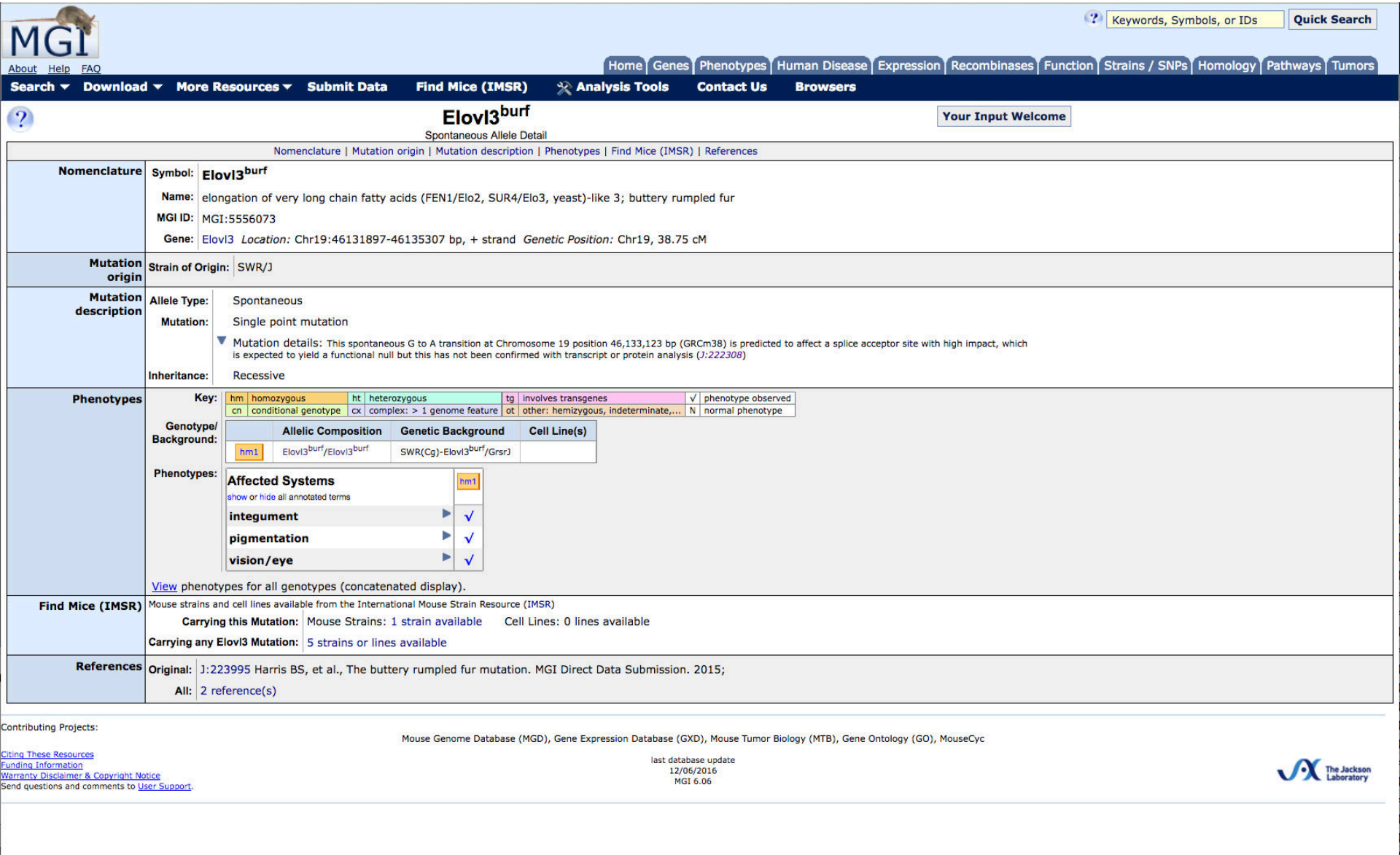


https://mmrdb.jax.org

Candidate gene analysis

- MGI batch query – alleles, phenotypes
- MouseMine – gene expression
- Validation – PCR genotyping of samples from the colony (segregation analysis)
- Allele or complementation tests (limited)
- Data release…

MGI Direct Data Submission



MGI

About Help FAQ

Keywords, Symbols, or IDs | Quick Search

Home | Genes | Phenotypes | Human Disease | Expression | Recombinases | Function | Strains / SNPs | Homology | Pathways | Tumors

Search | Download | More Resources | Submit Data | Find Mice (IMSR) | Analysis Tools | Contact Us | Browsers

Elovl3^{burf}

Spontaneous Allele Detail

Your Input Welcome

Nomenclature | Mutation origin | Mutation description | Phenotypes | Find Mice (IMSR) | References

Nomenclature

Symbol: **Elovl3^{burf}**

Name: elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 3; buttery rumpled fur

MGI ID: MGI:5556073

Gene: Elovl3 *Location:* Chr19:46131897-46135307 bp, + strand *Genetic Position:* Chr19, 38.75 cM

Mutation origin

Strain of Origin: SWR/J

Mutation description

Allele Type: Spontaneous

Mutation: Single point mutation

Mutation details: This spontaneous G to A transition at Chromosome 19 position 46,133,123 bp (GRCm38) is predicted to affect a splice acceptor site with high impact, which is expected to yield a functional null but this has not been confirmed with transcript or protein analysis (*J:222308*)

Inheritance: Recessive

Phenotypes

Key:

hm	homozygous	ht	heterozygous	tg	involves transgenes	√	phenotype observed
cn	conditional genotype	cx	complex: > 1 genome feature	ot	other: hemizygous, indeterminate,...	N	normal phenotype

Genotype/ Background:

	Allelic Composition	Genetic Background	Cell Line(s)
hm1	Elovl3^{burf}/Elovl3^{burf}	SWR(Cg)-Elovl3^{burf}/GrsrJ	

Phenotypes:

Affected Systems (show or hide all annotated terms)	hm1
integument	√
pigmentation	√
vision/eye	√

View phenotypes for all genotypes (concatenated display).

Find Mice (IMSR)

Mouse strains and cell lines available from the International Mouse Strain Resource (IMSR)

Carrying this Mutation: Mouse Strains: 1 strain available Cell Lines: 0 lines available

Carrying any Elovl3 Mutation: 5 strains or lines available

References

Original: J:223995 Harris BS, et al., The buttery rumpled fur mutation. MGI Direct Data Submission. 2015;

All: 2 reference(s)

Contributing Projects:

Mouse Genome Database (MGD), Gene Expression Database (GXD), Mouse Tumor Biology (MTB), Gene Ontology (GO), MouseCyc

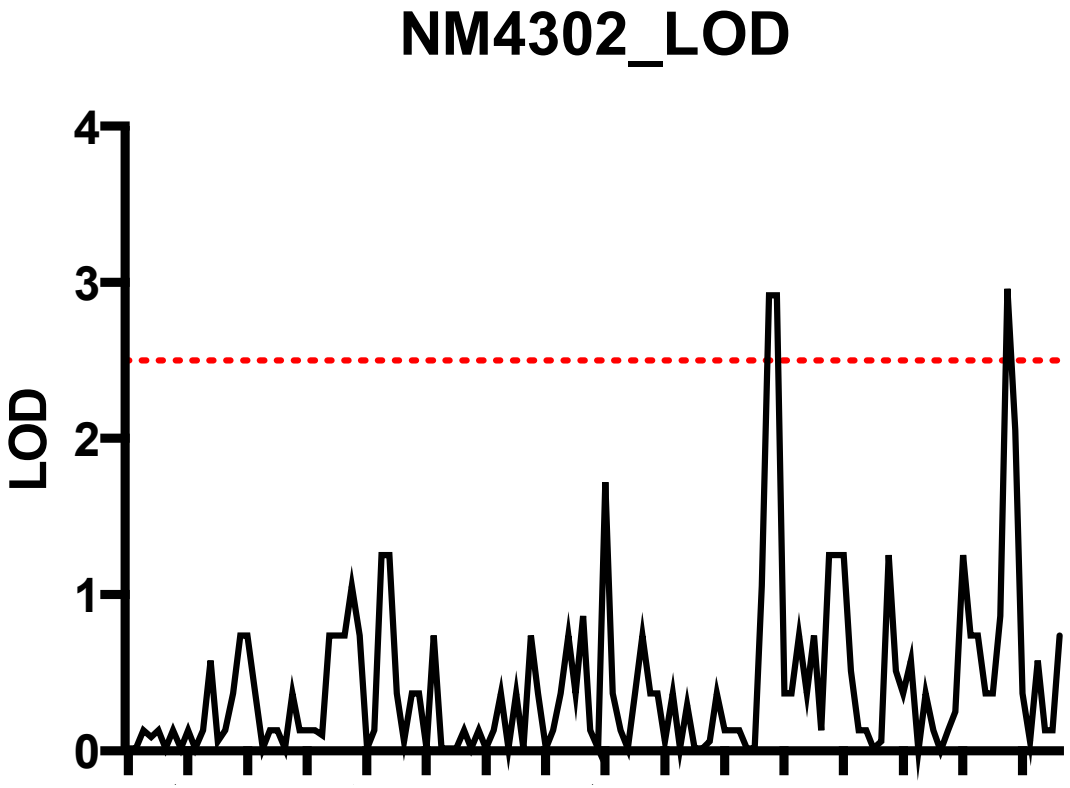
Citing These Resources
Funding Information
Warranty Disclaimer & Copyright Notice
Send questions and comments to User Support.

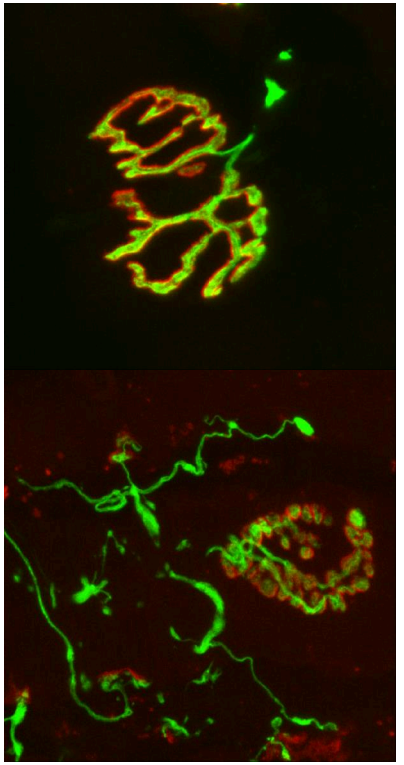
last database update
12/06/2016
MGI 6.06

The Jackson Laboratory

Connecting with clinicians

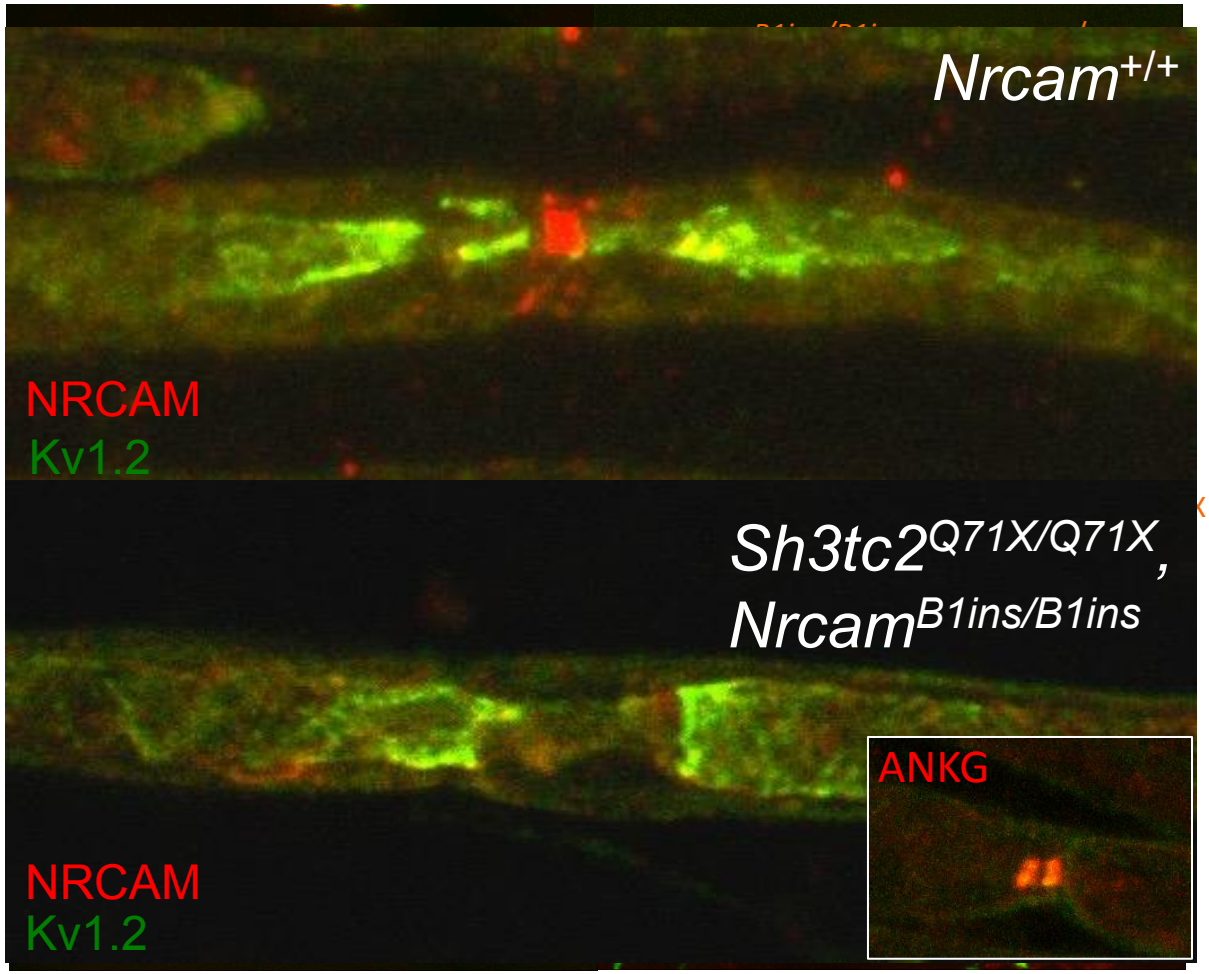
NM4302, progressive paralysis and death at 5 months





Robert Burgess, Greg Cox, Dave Schroeder

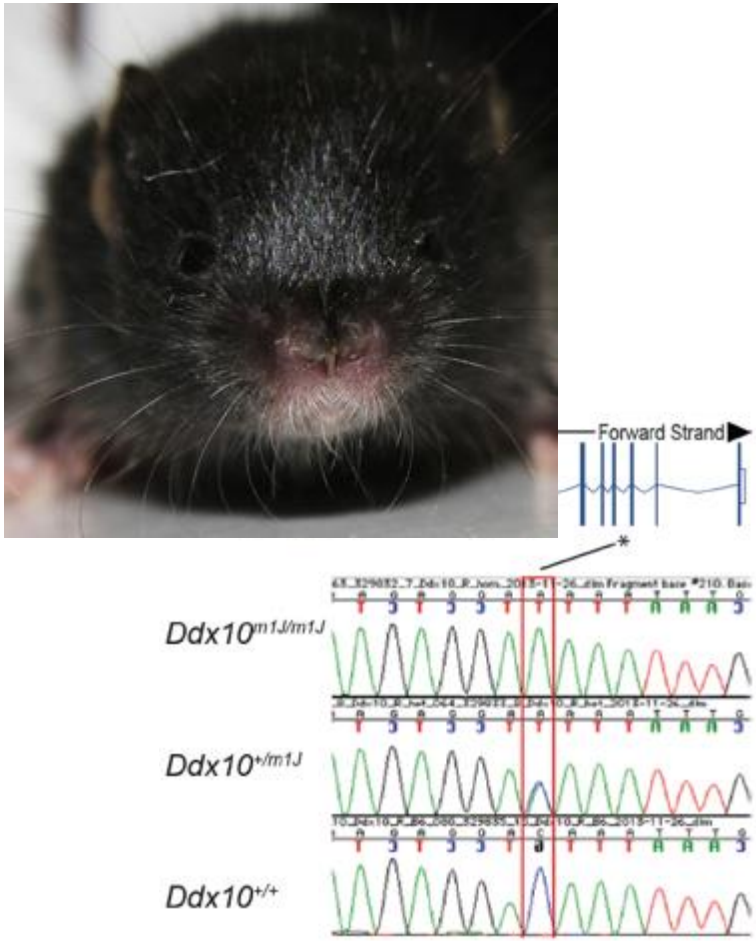
NM4302, progressive paralysis and death at 5 months



Nrcam (Chr12): B2 element insertion in exon 26 (g.44576714_44576715insB2)
Sh3tc2 (Chr18): C to T mutation (c.211C>T), results in p.Q71X

Rob Burgess, Greg Cox, Abby Tadenev, The Jackson Laboratory

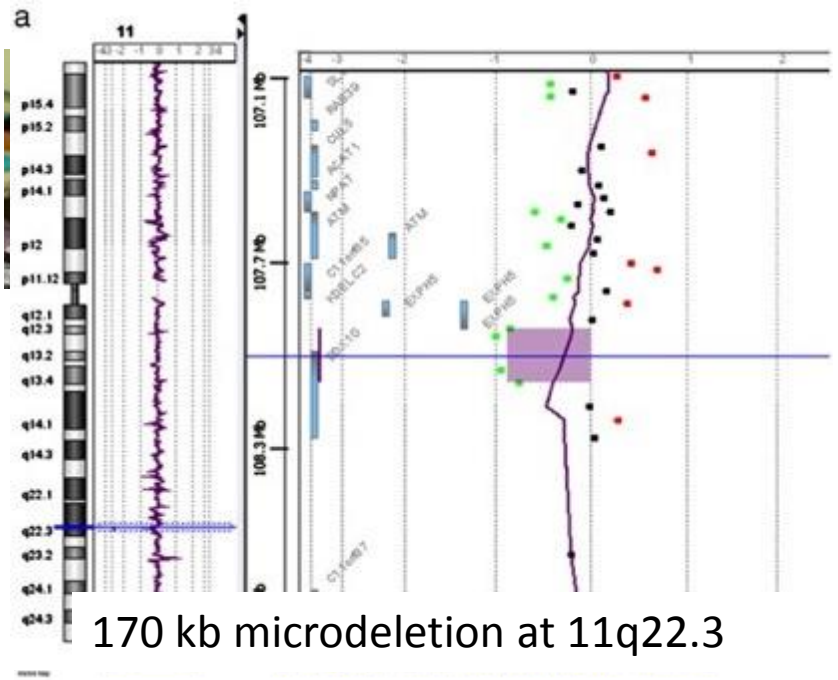
Ddx10 – DEAD box polypeptide10



Contents lists available at ScienceDirect
Gene
journal homepage: www.elsevier.com/locate/gene

Array CGH analysis of a cohort of Russian patients with intellectual disability

Anna A. Kashevarova [a,*], Lyudmila P. Nazarenko [a], Nikolay A. Skryabin [a], Olga A. Salyukova [a], Nataliya N. Chechetkina [a], Ekaterina N. Tolmacheva [a], Elena A. Sazhenova [a], Pamela Magini [b], Claudio Graziano [b], Giovanni Romeo [b], Vaidutis Kučinskas [c], Igor N. Lebedev [a]



170 kb microdeletion at 11q22.3

Fdxr – ferredoxin reductase

nm4877, *Fdxr*^{m1J}, R389Q

Clinical data – Ambry Genetics

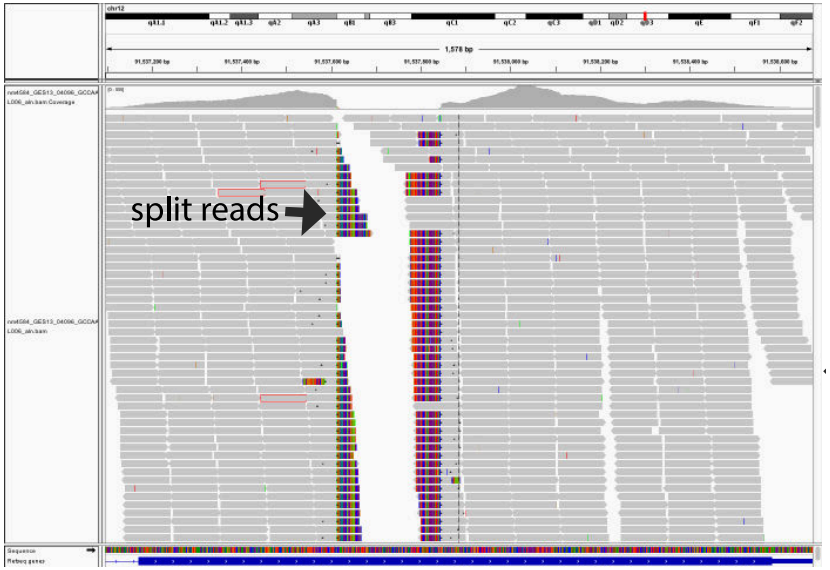
- Recessive missense mutations in three pedigrees
- Suspected mitochondrial disorder with polyneuropathy, gait defects, vision impairment

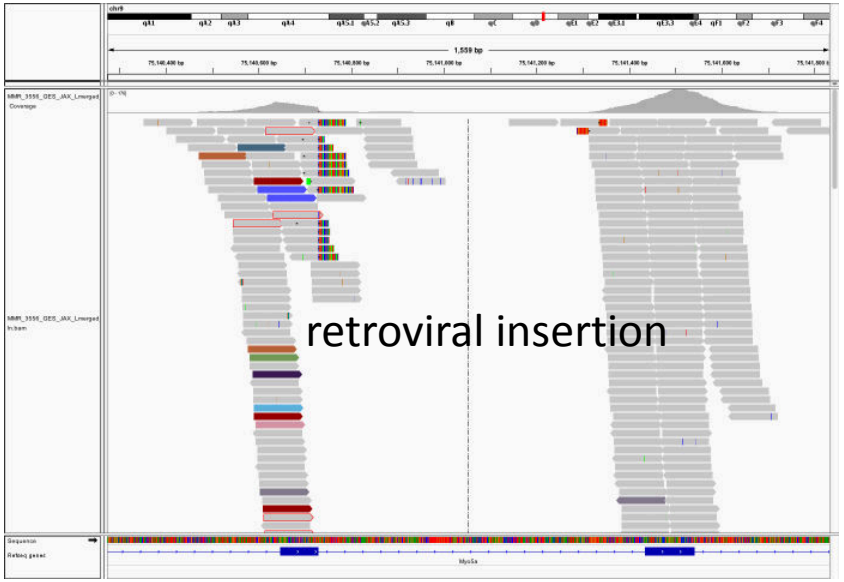
What are we missing?

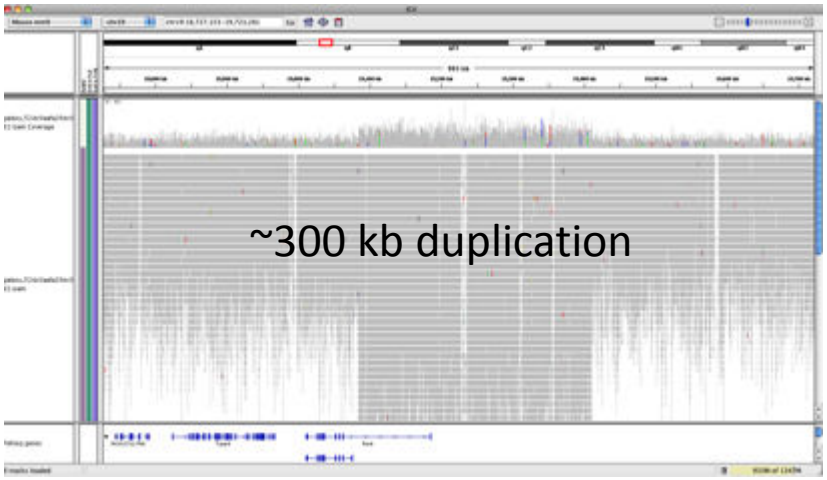
Exome sequencing of mice with proven Mendelian phenotypes provides putative causative mutations with a success rate of ~55%.

- Poor gene annotation
- Structural mutations
- Non-coding mutations

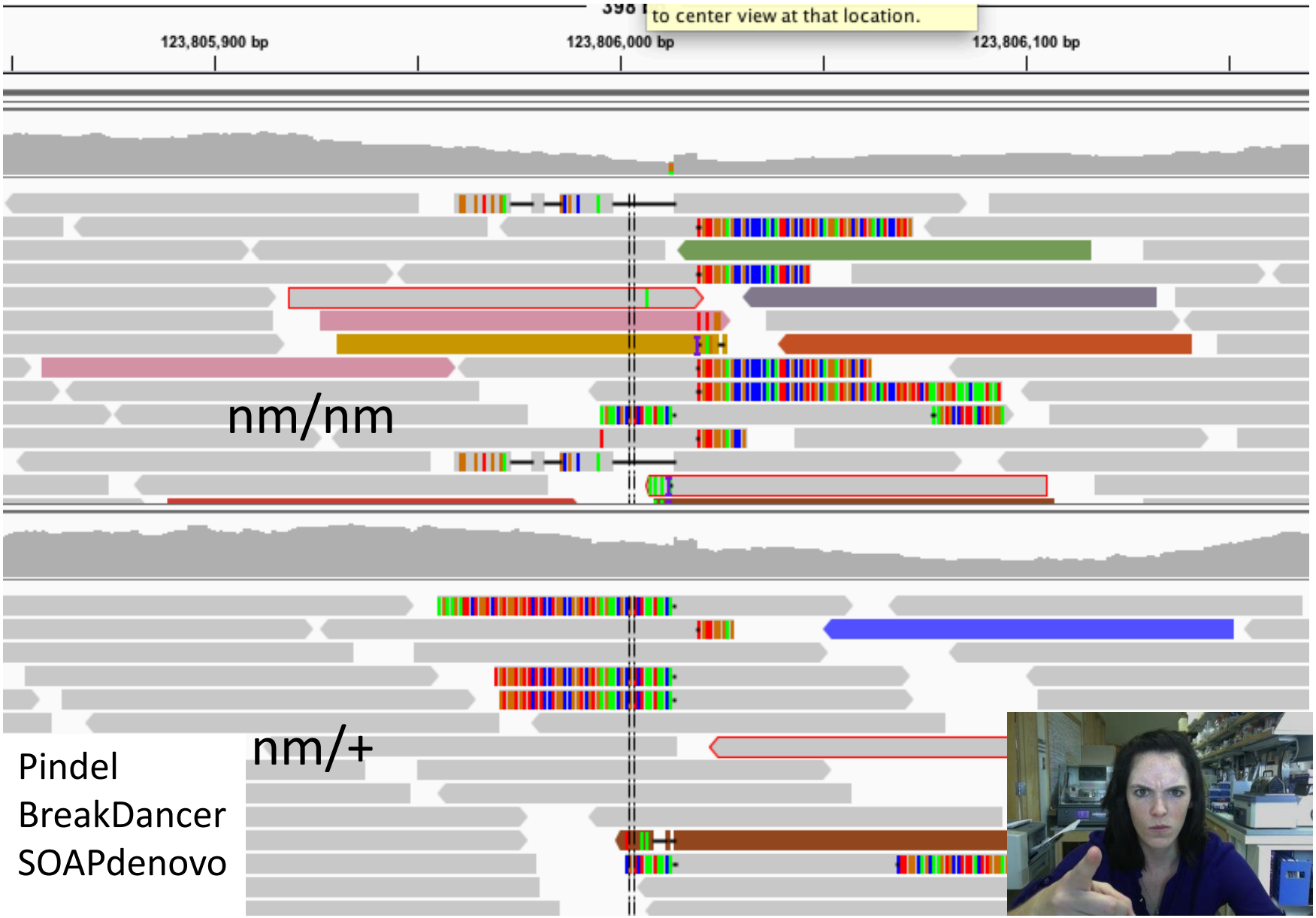
Structural mutations







Non-coding mutation

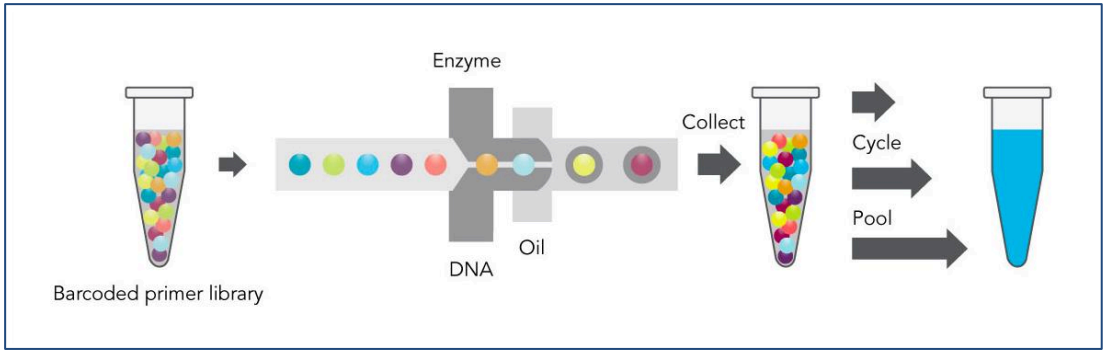


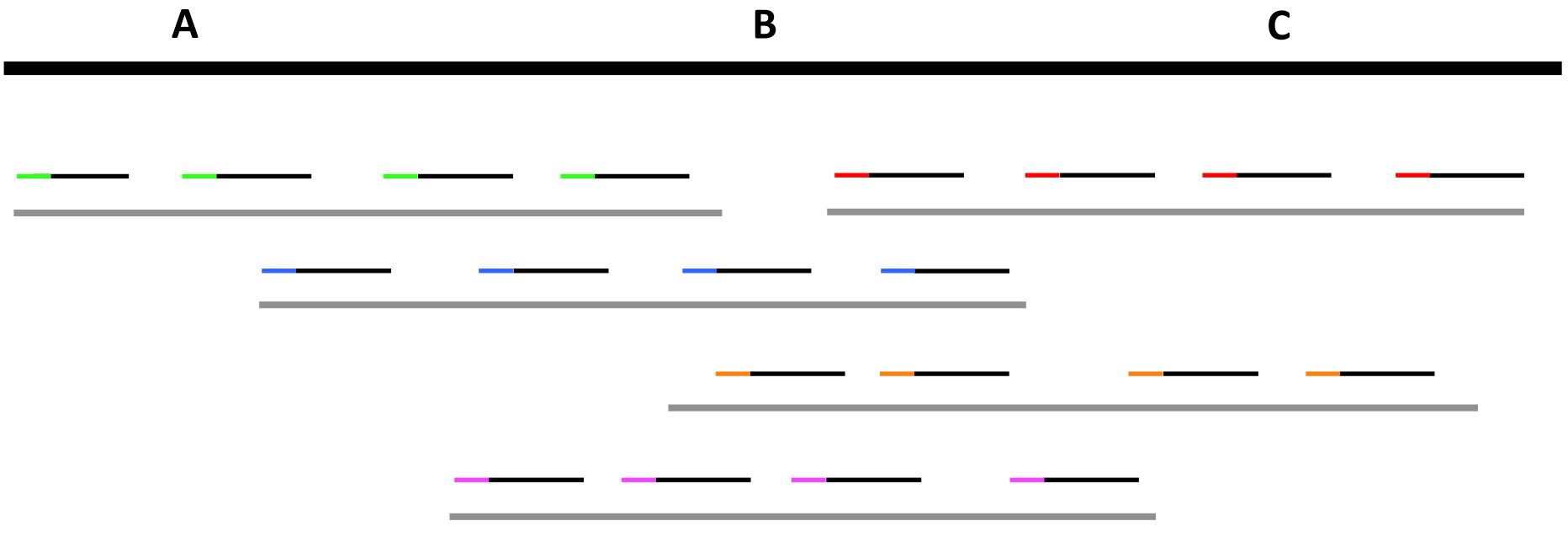
Pindel
BreakDancer
SOAPdenovo

Approaches for solving failed exomes

- RNA Seq
  - correct tissue and relevant time points, live mice
- CGH
  - works well for CNV detection > ~5 Kb
- Whole genome sequencing → 100 pending, stay tuned
- 10X Genomics → synthetic long reads

10X Genomics – synthetic long read technology

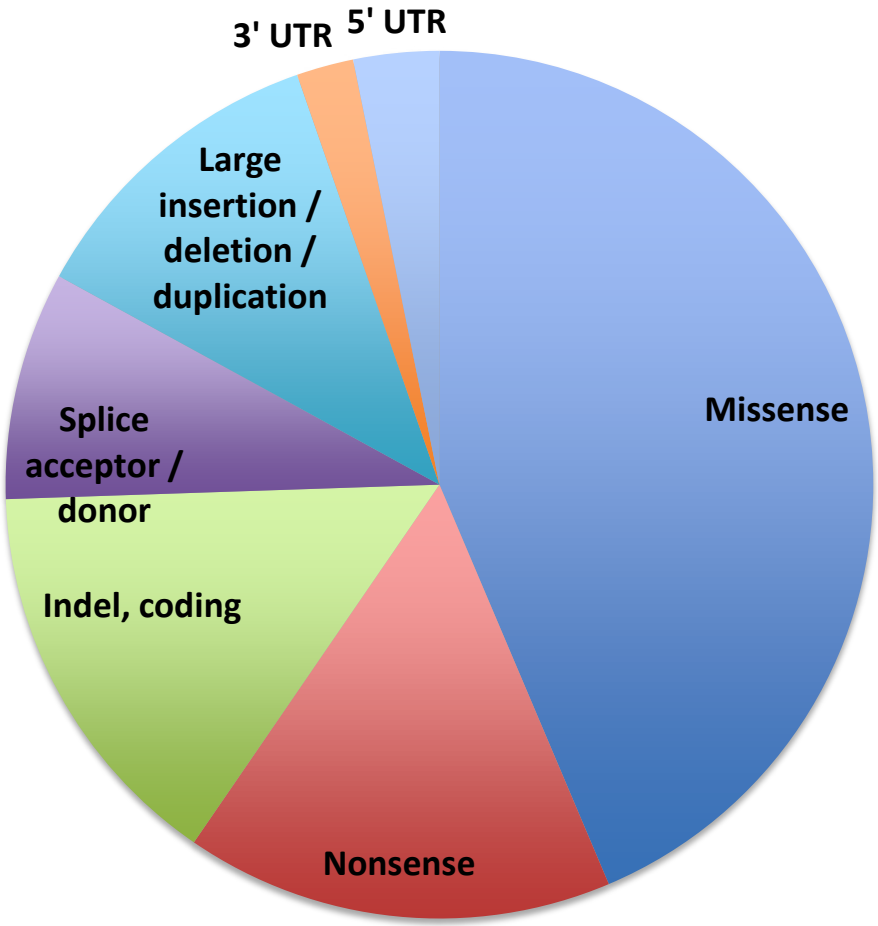




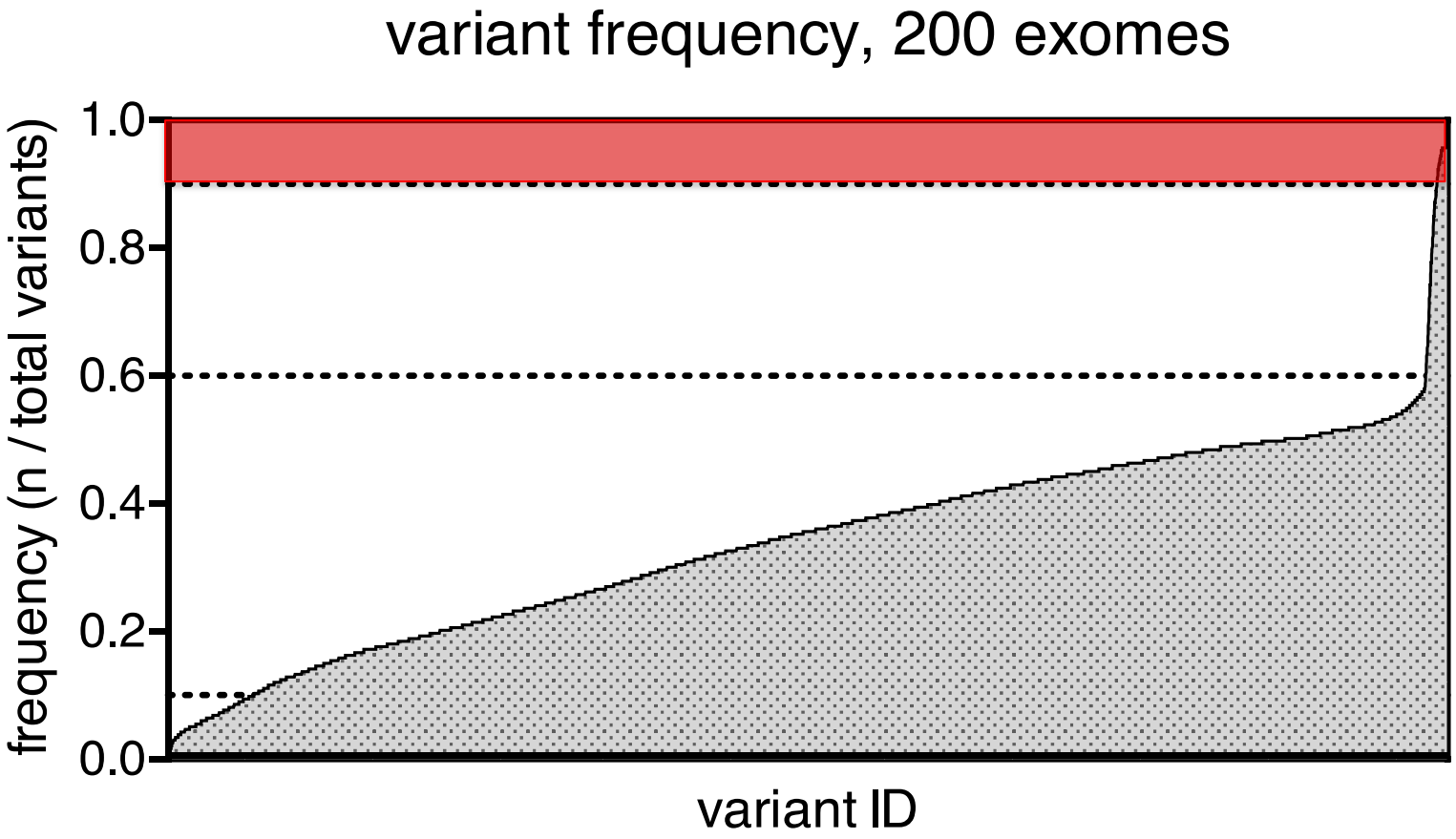
Summary

- Using exome sequencing, our rate of causative gene discovery in spontaneously arising mouse models of Mendelian disease has increased >10X.
- ~10% are the first phenotypic alleles and over 80% are informative for human Mendelian disease modeling.
- Unique resource of exome recalcitrant, Mendelian disease causing mutations.
- Using both forward and reverse genetic approaches we are creating new mouse models with unprecedented speed and precision.

Putative pathogenic mutation spectrum, 91 spontaneous mutations

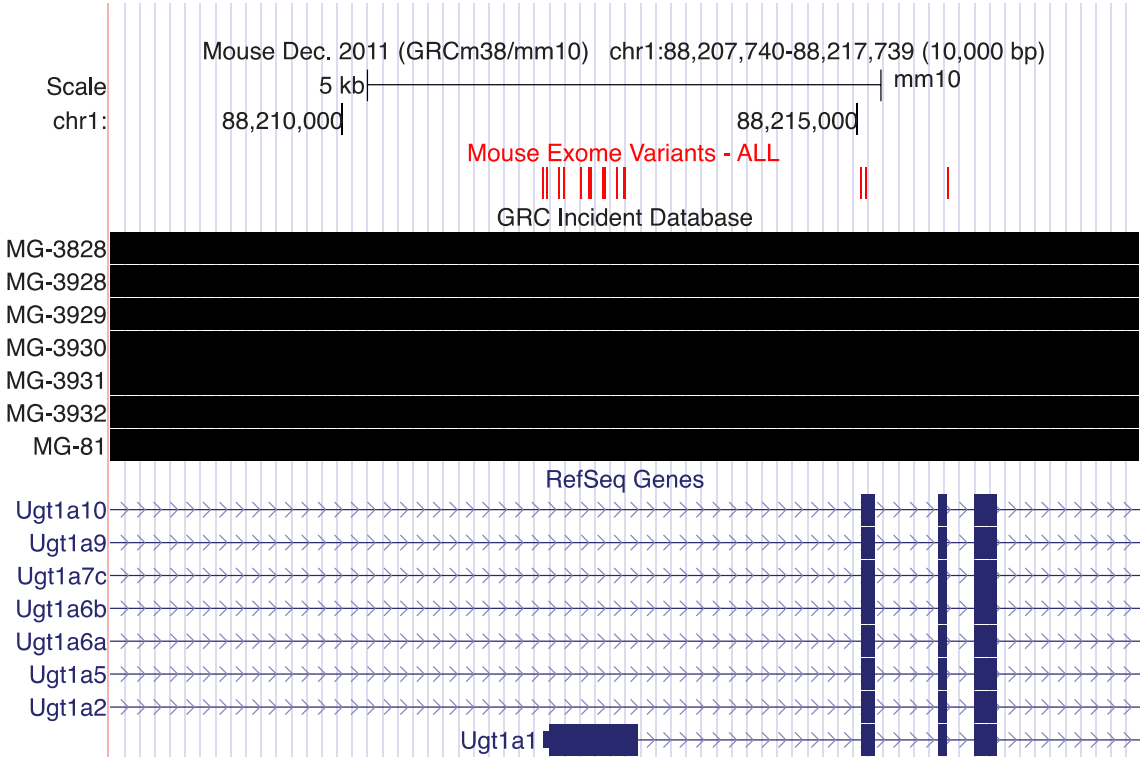


Exome variant frequency distribution



All variants passing soft filter

Variants with allele frequency >95% map to regions with reported assembly problems



151/272, p-value = 9.999e-05; mean=28.74, SD=5.02

Acknowledgements

JAX Genetic Resources
Steve Murray, JAX KOMP
Cat Lutz, JAX Repository

Resource for Comparative Mendelian Genomics, Mouse Mutant Resource
David Bergstrom
Heather Fairfield
Belinda Harris
Son Yong Karst
Melissa Berry

Reinholdt Lab
Candice Byers
Anne Czechanski
Whitney Martin

JAX Mammalian Genetics Research
Rob Burgess
Greg Cox
Kevin Seburn

JAX Computational Sciences
Anuj Srivastava
Guru Ananda
Roger Liu
Anu Lakshminaryana

Centers for Mendelian Genomics Baylor/Hopkins
David Valle
Nara Sobreira

Ambry Genetics
Deepali Shinde
Kelly Hagman
Sha Tang
Zoe Powis

NIH National Institutes of Health

NIH P40 OD010972
NIH R24 OD021325