Computational Mapping Tools

Daniel M. Gatti Ph.D.

Mar. 19, 2015

NIEHS Population Based Rodent Resources

The Jackson Laboratory

Leading the search for tomorrow's cures

Steps in QTL Mapping

Haplotype Reconstruction

Linkage Mapping

- using a kinship correction

Association Mapping

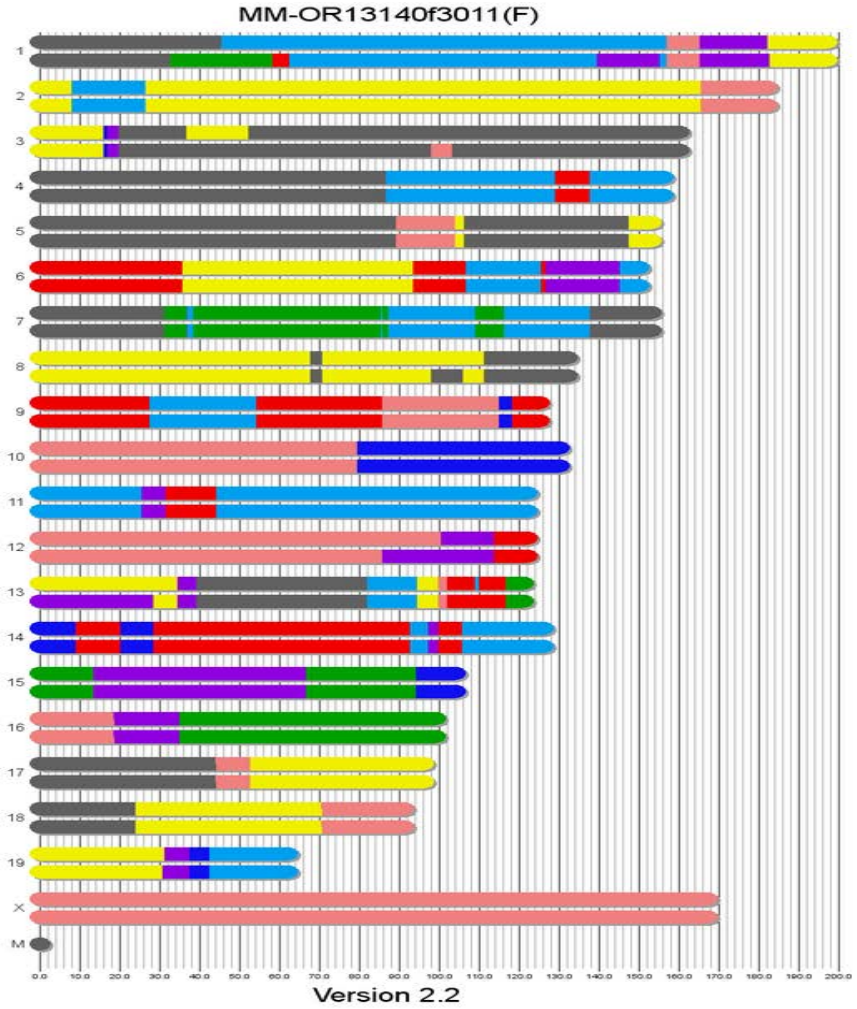
Use founder sequences to search for causal variants and candidate genes.

CC/DO Founders

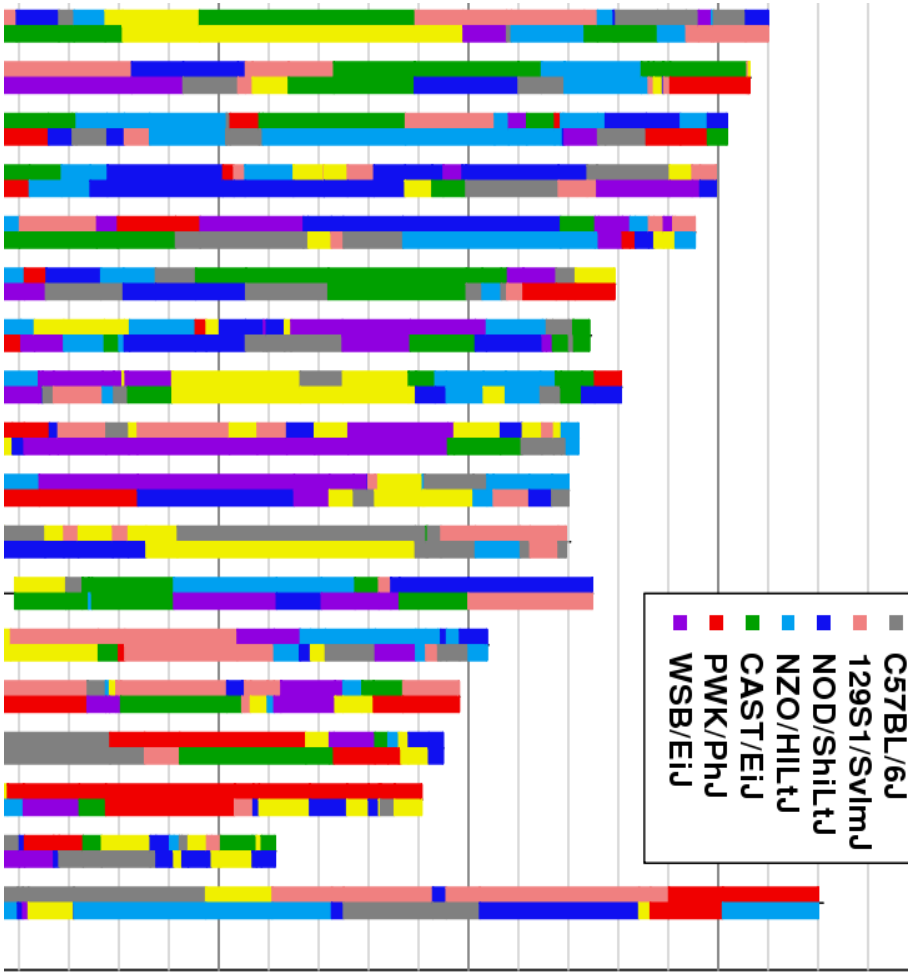
A/J	A
C57BL/6J	B
129S1/SvImJ	C
NOD/ShiLtJ	D
NZO/HlLtJ	E
CAST/EiJ	F
PWK/PhJ	G
WSB/EiJ	H

CC/DO genomes are mosaics of the 8 founders

Collaborative Cross



Diversity Outbred



Genotype the mice by microarray

- MegaMUGA: 77,808 markers
- Mean 33 Kb spacing.

- GigaMUGA: 143,446 markers
- Mean 18 Kb spacing.

MegaMUGA

High Density Mouse Universal Genotyping Array

Content and Design Characteristics

The Mega Mouse Universal Genotyping Array (MegaMUGA) provides up to 77,800 SNP markers built on the Illumina® Infinium platform. The SNP markers are distributed throughout the mouse genome (average spacing of 33 Kb) and with a slight excess of probes in the telomeric regions of each autosome to facilitate detection of recombination events throughout the chromosome. The SNPs were selected to be maximally informative — making the array a highly cost-effective tool for many genotyping applications.

Up to 77,800 SNPS
$90/$100
Per DNA Sample

The MegaMUGA design criteria make it optimal for detecting heterozygous regions and discriminating between haplotypes in homozygous regions. SNPs were selected to be informative in most mouse populations, including wild mice and multiple Mus species, but with a special emphasis for markers that are informative in the Collaborative Cross and Diversity Outbred population resources.

Performance

- The MegaMUGA array provides robust calls for over 74,000 SNP markers in classical strains with greater than 99.9% concordance (based on over 300 controls).
- Initial analysis in classical inbred strains shows that on average the number of informative SNPs in pairwise combinations is ~20,000. MegaMUGA allows to discriminate between closely related sister strains (e.g. C57BL/6J and C57BL/6CR) and between related wild derived strains of similar origin (e.g. PWK/PhJ and PWD/PhJ; ZALENDE/EIJ and TIRANO/EIJ).
- Genotypes for many inbred strains can be viewed at the UNC Systems Genetics Core Facility website (http://www.csbio.unc.edu/CCstatus/index.py?run).

Additional Features

- MegaMUGA has probes on the Mitochondria and Y Chromosome that support detailed analysis of matrilineal and patrilineal inheritance.
- Some probes were designed to discriminate between known structural variants (insertions, deletion and duplications).
- Some probes were designed to detect sequences present only in genetically engineered mice (Cre, Luciferase, etc).

illumina CSPro CERTIFIED SERVICE PROVIDER

MegaMUGA content and design were done at the University of North Carolina at Chapel Hill by Catherine Welsh, Chen-Ping Fu, John Didion, John Calaway, Leonard McMillan and Fernando Pardo-Manuel de Villena. This work was partially supported by a contract from the Neogen Corporation. A complete description of MegaMUGA will appear in an upcoming publication.

Contact GeneSeek for More Details and Special Offers at: geneseekinfo@neogen.com

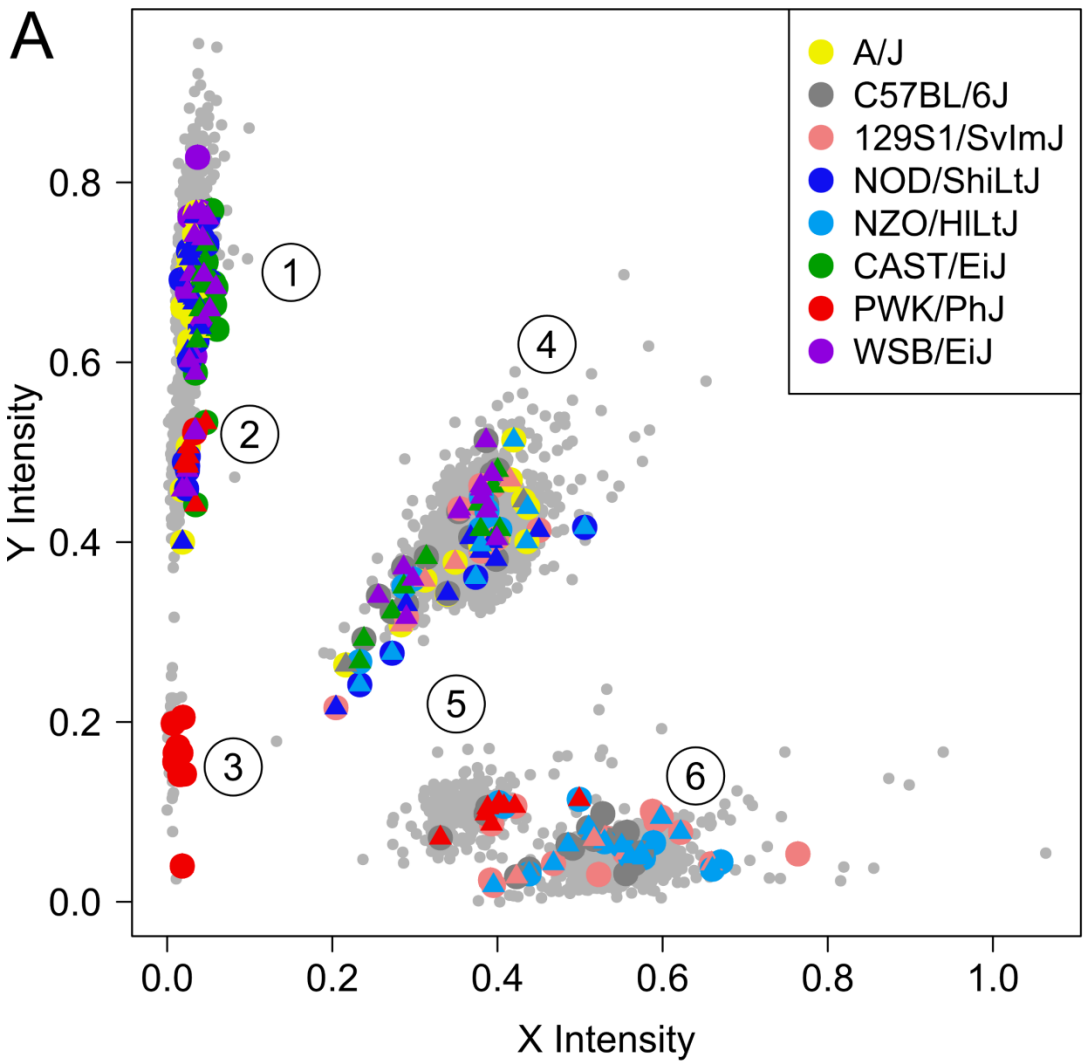
Geneseek can also assist with custom SNP genotpying panels.

GENESEEK a Neogen® Company

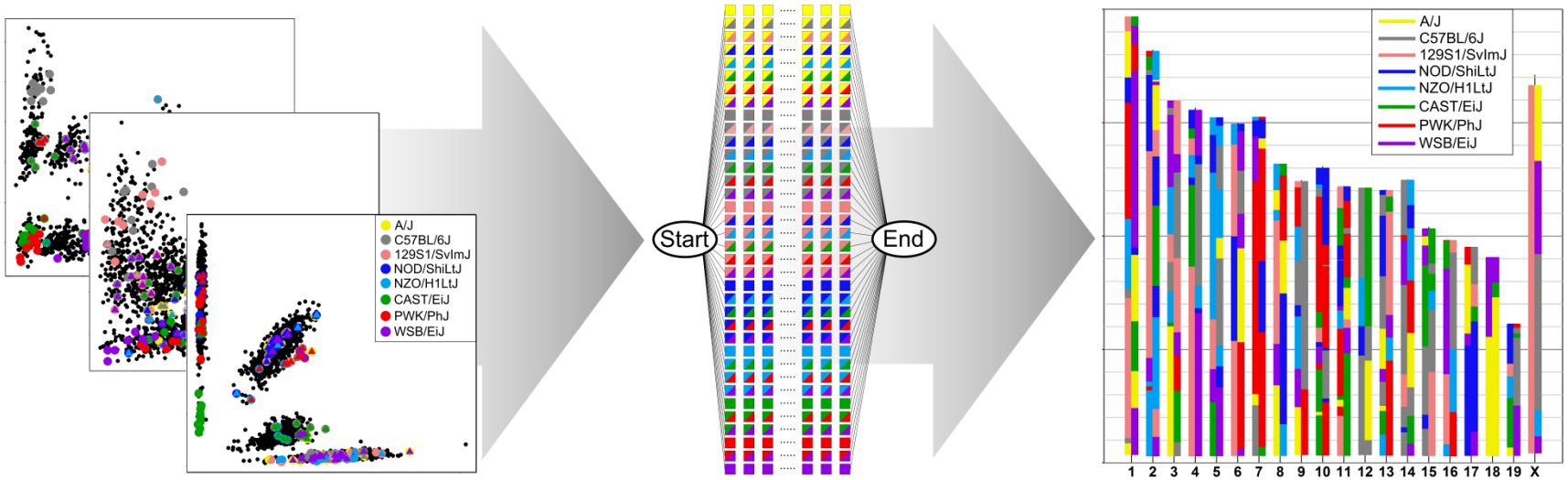
4665 Innovation Drive, Suite 120 • Lincoln, NE 68521
402.435.0665 • Fax: 402.435.0664
geneseekinfo@neogen.com • www.neogen.com

©2012 Neogen Corporation. Neogen and GeneSeek are registered trademarks of Neogen Corporation, Lansing, Mich. All other trademarks are property of their respective companies.

MUGA Markers are highly informative



Haplotype Reconstruction



Array Intensities

Hidden Markov Model

DO Genome

Haplotype Reconstruction

Genotype Probabilities

M	AA	AB	AC	AD	AE	AF	AG	AH	BB	BC	BD	BE	BF	BG	BH	CC	CD	CE	CF	CG	CH	DD	DE	DF	DG	DH	EE	EF	EG	EH	FF	FG	FH	GG	GH	HH
1	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	**½**	**½**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Founder Allele Proportions

Marker	A	B	C	D	E	F	G	H
1	**0.5**	0	0	**0.5**	0	0	0	0
2	**0.5**	0	0	**0.5**	0	0	0	0
3	**0.5**	0	0	**0.25**	**0.25**	0	0	0
4	**0.5**	0	0	0	**0.5**	0	0	0
5	**0.5**	0	0	0	**0.5**	0	0	0

Linkage Mapping Model

Regress phenotype on haplotypes (allele proportions).

At each marker: $$y_i = s_i\alpha + \sum_{j=1}^{8} h_{ij}\beta_j + \varepsilon_i$$

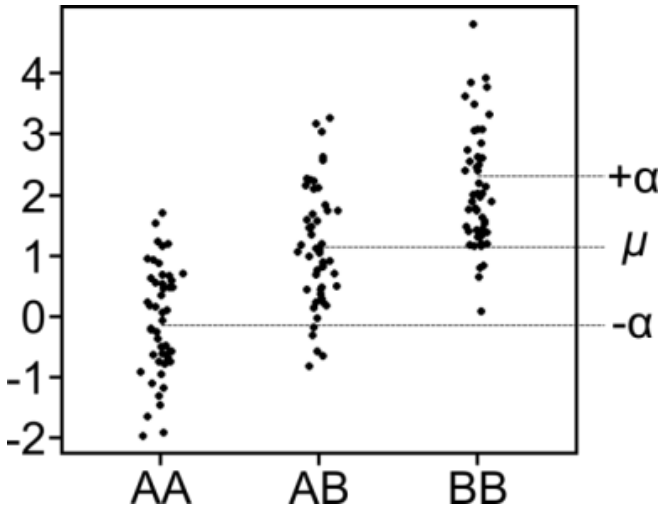
y_i = phenotype of animal *i*
s_i = sex of animal *i*
a = sex effect
h_{ij} = allele proportion of founder *j* in animal *i*
β_j = effect of founder allele *j*
ε_i = residual error for animal *i*

Model Output

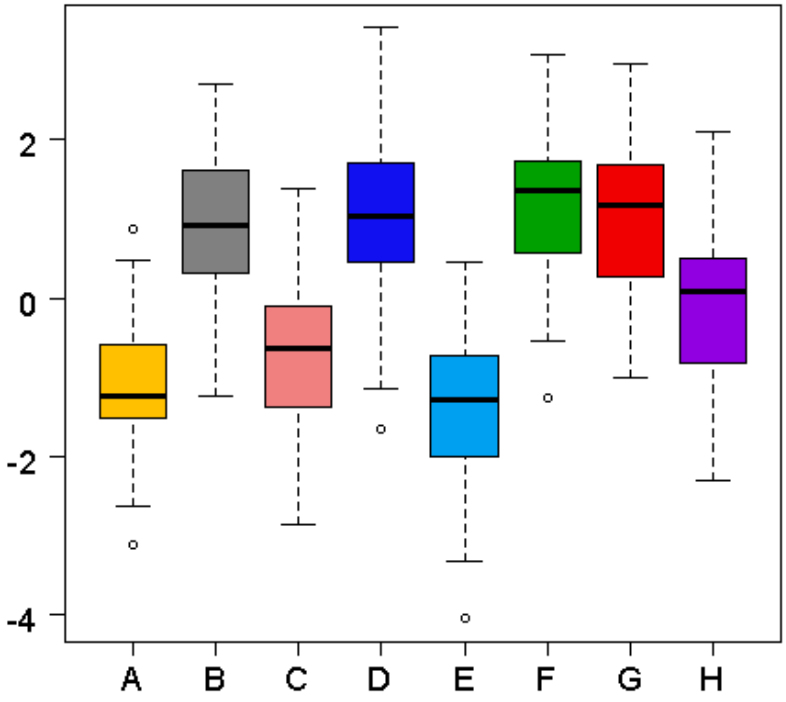
At each marker:

LOD Score

Founder allele effects



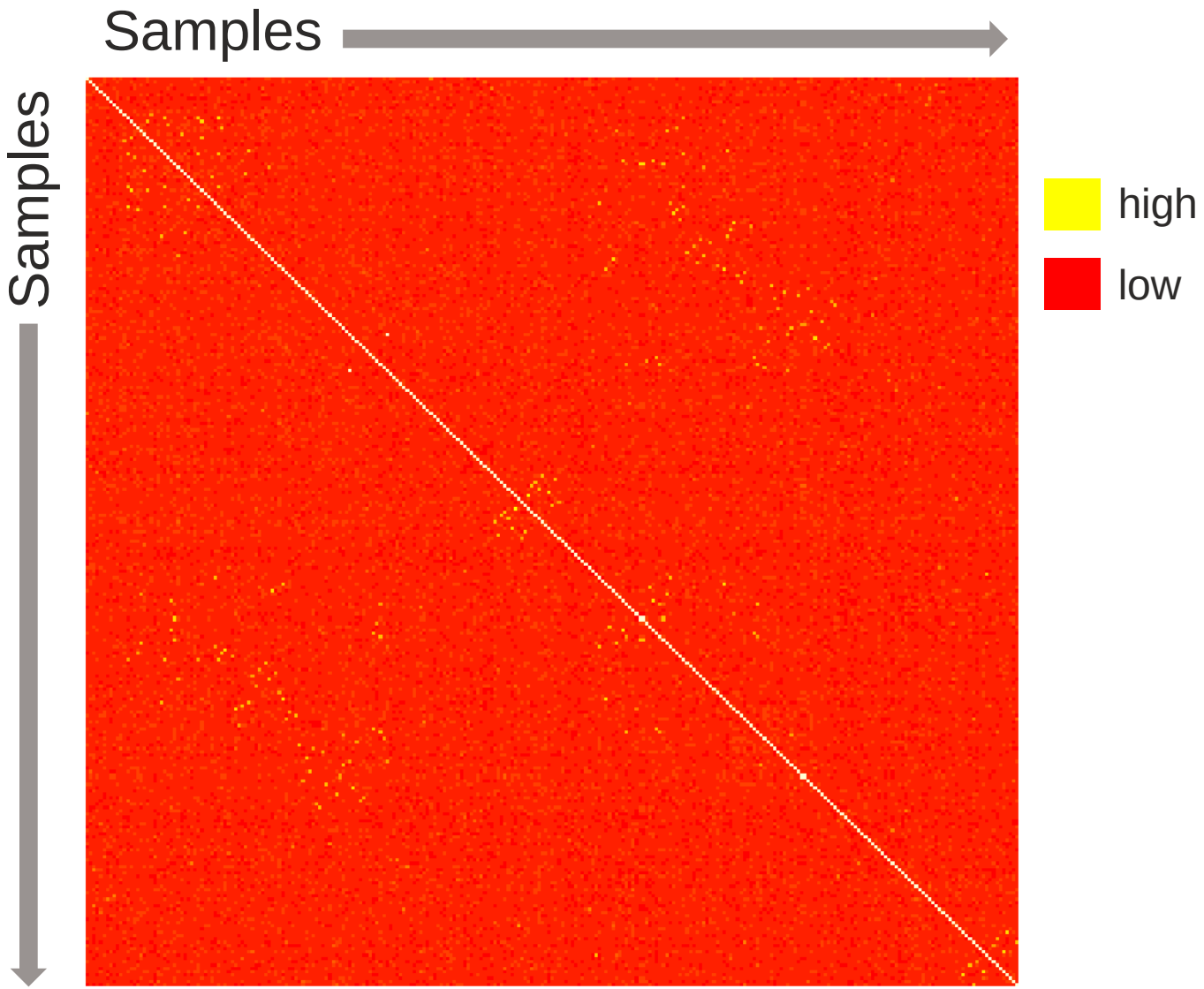
Marker 1



Kinship

- Animals may have different degrees of relatedness.
- This creates correlation structure in the data that may lead to false positives.
- Kinship estimation:
  - From pedigree data
  - From genotype data (allele or haplotype sharing)

Kinship Matrix



Linkage Mapping Model

Regress phenotype on <u>haplotypes</u> (allele proportions).

$$y_i = s_i\alpha + \sum_{j=1}^{8} h_{ij}\beta_j + \varepsilon_i$$

Regress phenotype on <u>haplotypes</u> with kinship correction.

$$y_i = s_i\alpha + \sum_{j=1}^{8} h_{ij}\beta_j + \gamma_i + \varepsilon_i$$

How should we use the kinship matrix?

Use a single kinship matrix constructed from all markers.

or

Leave One Chromosome Out (LOCO) kinship matrices.

Chr 1: Create kinship matrix from markers on Chr 2,… X.

Chr 2: Create kinship matrix from markers on Chr 1,3,… X.

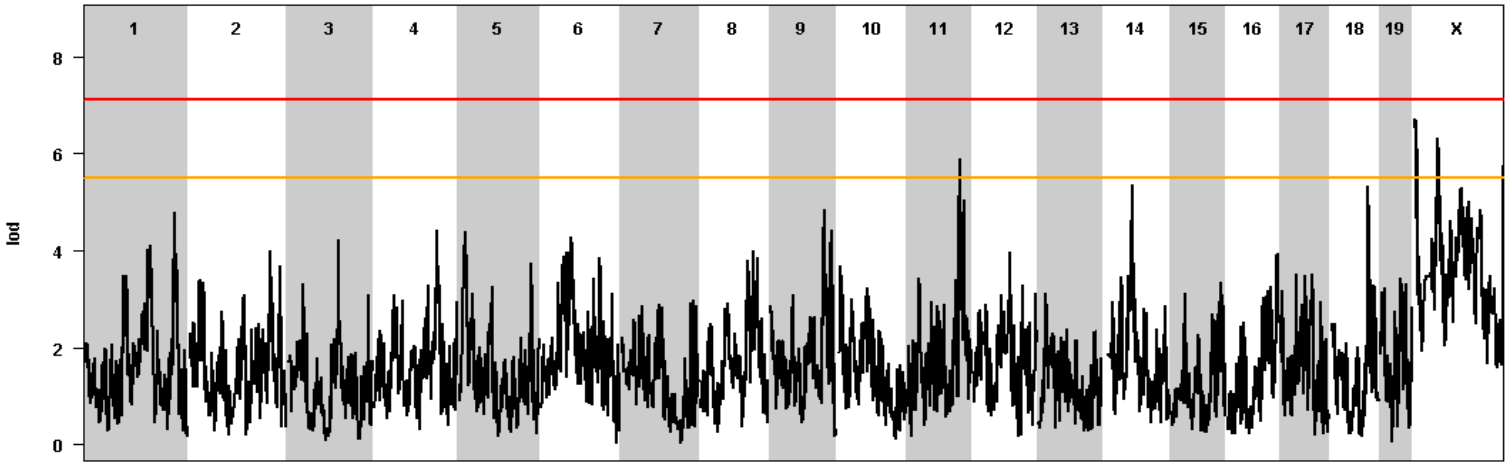
Chr 3: Create kinship matrix from markers on Chr 1,2,4,… X.

…

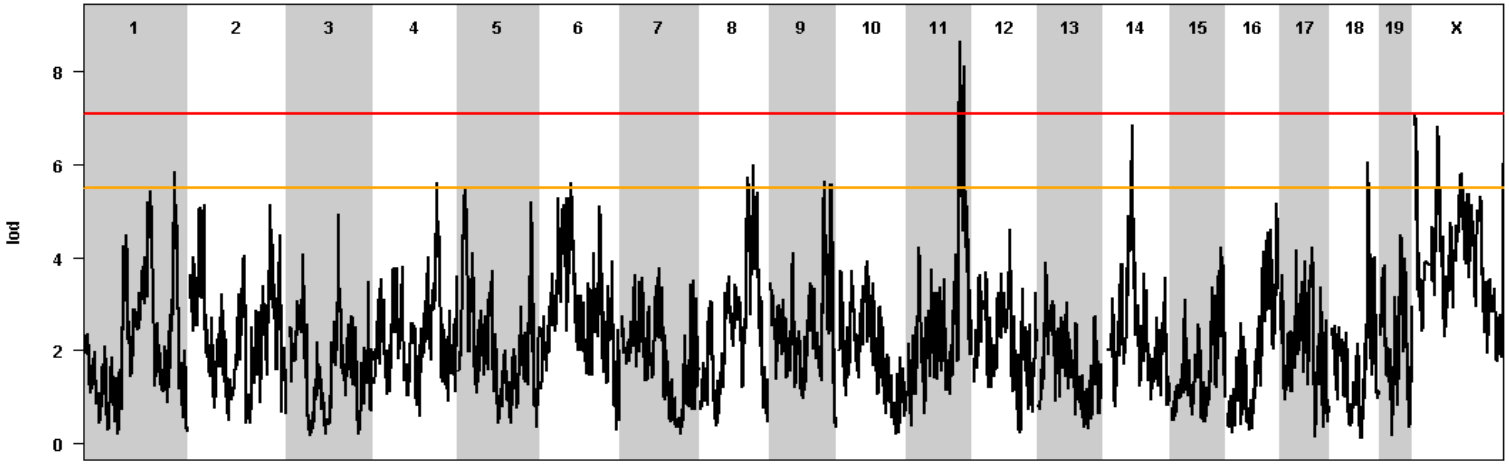
...

LOCO Method Increases Power

Bone Mineral Density: Single Kinship Matrix



Bone Mineral Density: LOCO Kinship Matrices



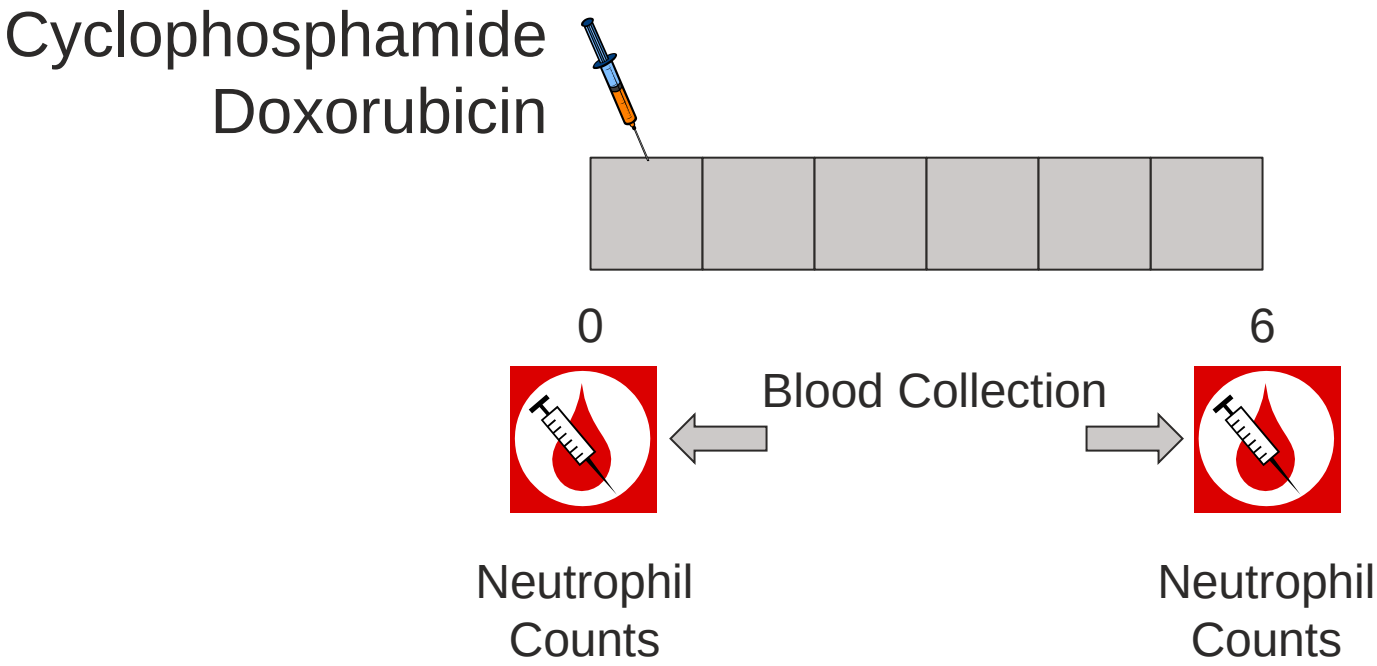
Linkage Mapping Software

Bagpipe: http://valdarlab.unc.edu/software/bagpipe/_build/html/bagpipe.html

QTLRel: http://www.biomedcentral.com/1471-2156/12/66

DOQTL: http://cgd.jax.org/apps/doqtl/DOQTL.shtml

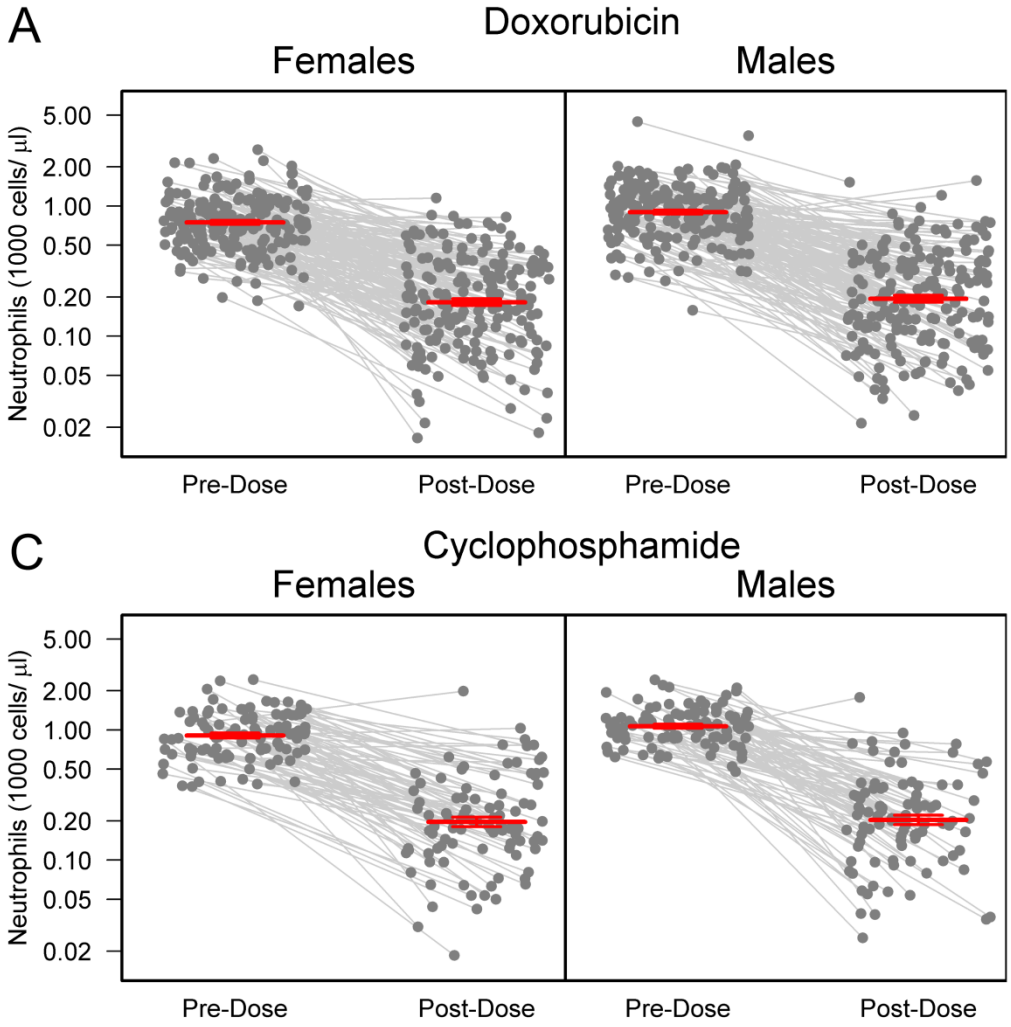
DO Chemotherapy Study Design



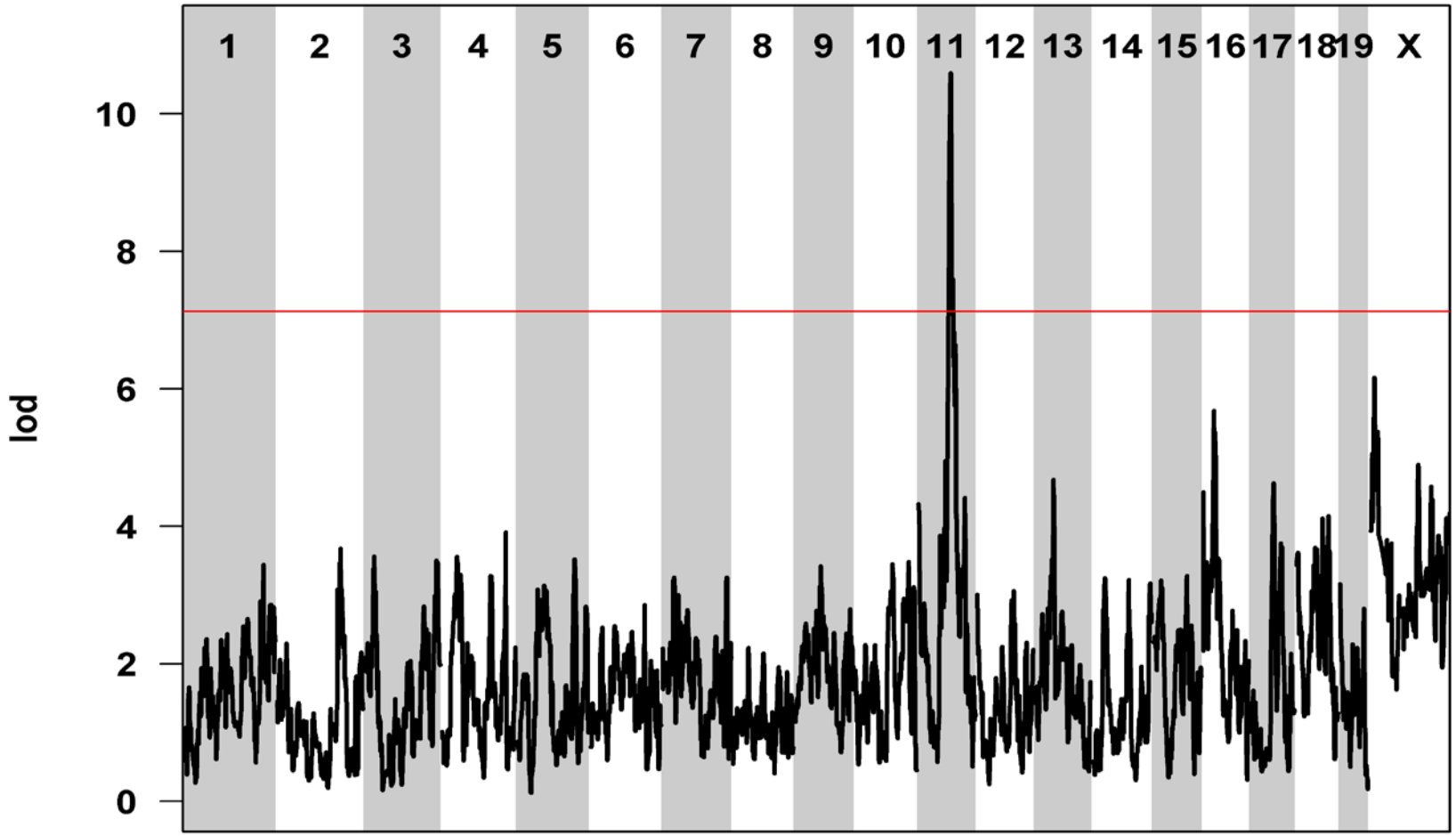
200 to 400 mice per drug

Phenotype of interest: Neutropenia

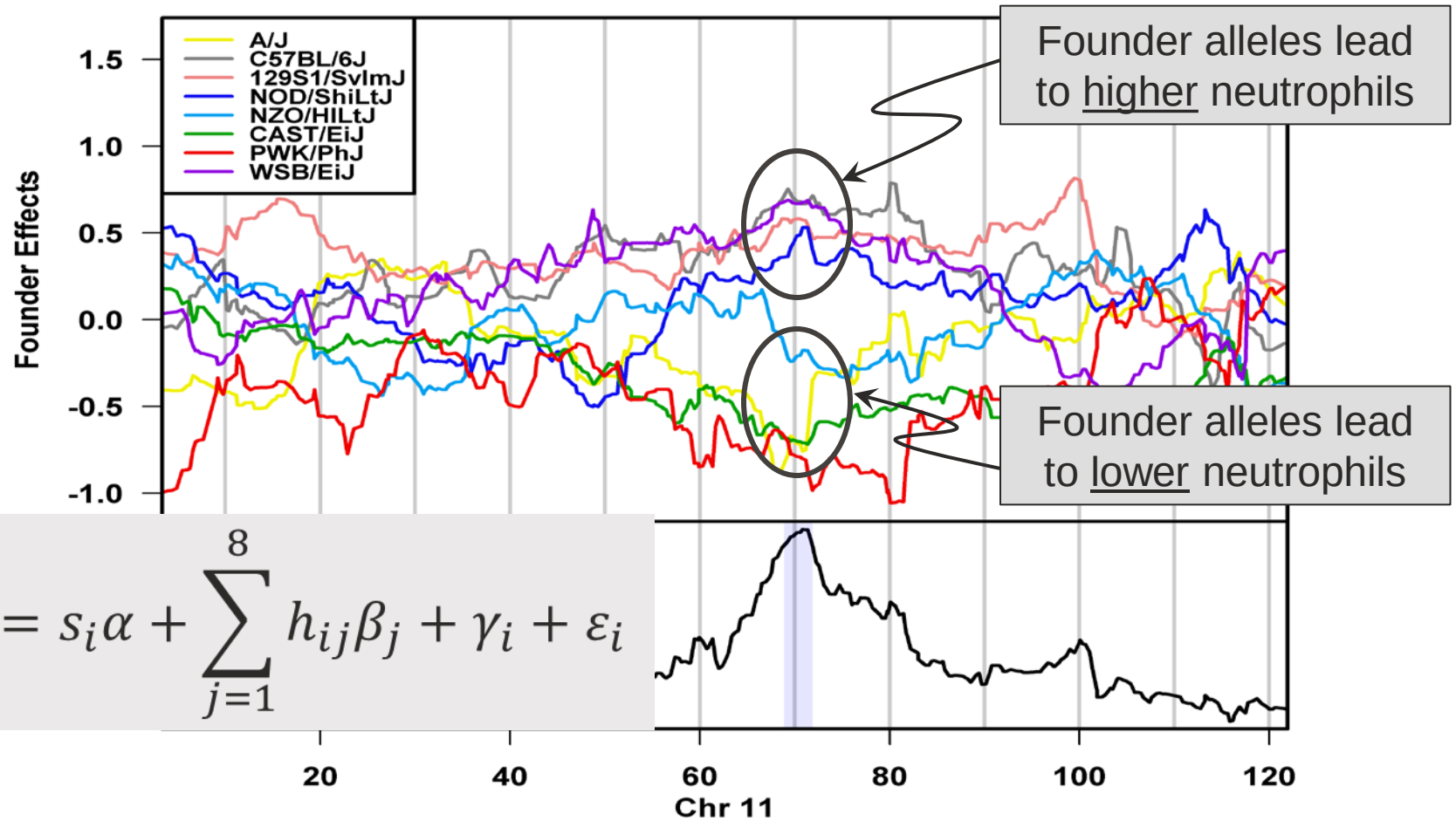
Neutrophil counts drop in DO mice after chemotherapy



Cyclophosphamide: Gene(s) on Chr 11 influence neutropenia.



We can dissect the founder allele effects in the DO.



$$y_i = s_i\alpha + \sum_{j=1}^{8} h_{ij}\beta_j + \gamma_i + \varepsilon_i$$

Association Mapping Model

Regress phenotype on genotypes with kinship correction.

At each SNP: $$y_i = s_i\alpha + g_i\beta + \gamma_i + \varepsilon_i$$

y_i = phenotype of animal *i*
s_i = sex of animal *i*
a = sex effect
g_i = genotype of animal *i* (coded 0, 1 or 2)
β = effect of SNP allele
γ_i = kinship adjustment for animal *i*
ε_i = residual error for animal *i*

Association Mapping Software

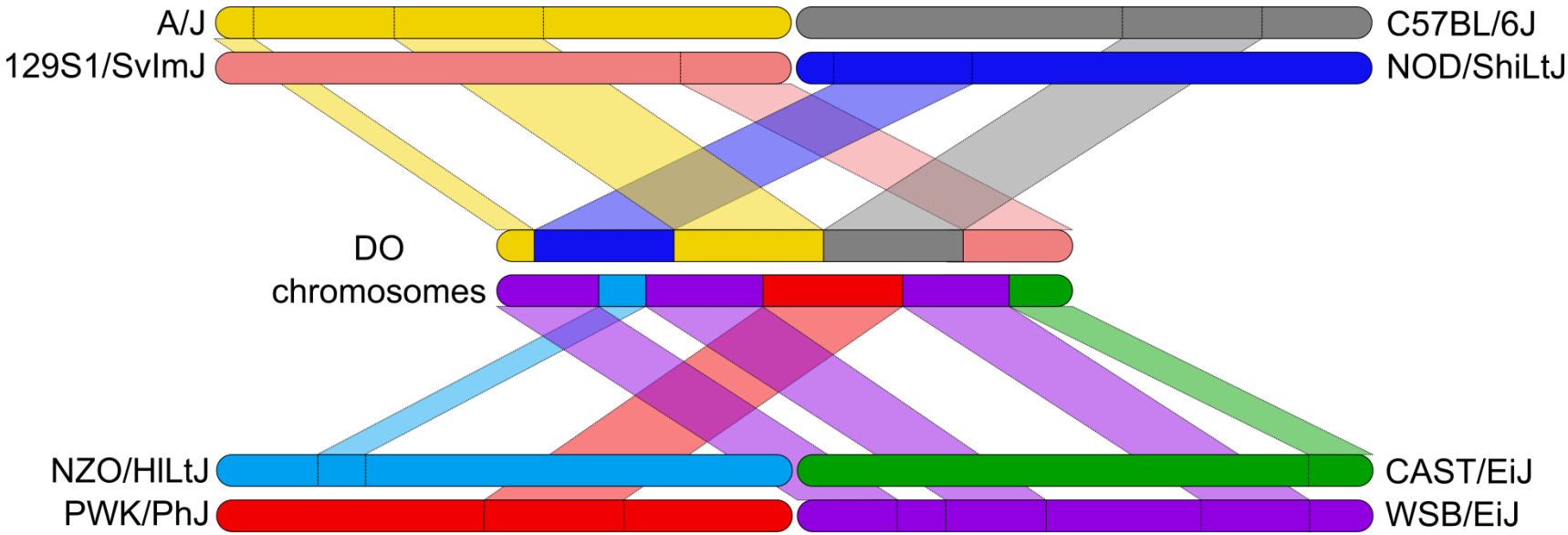
EMMA(X): http://genetics.cs.ucla.edu/emmax/

FastLMM: http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/Fastlmm/

GEMMA: http://www.xzlab.org/software.html

DOQTL: http://cgd.jax.org/apps/doqtl/DOQTL.shtml

Association Mapping



Impute the founder genotypes onto DO mice.
Perform GWA mapping using SNPs.

Baud, A. et.al., *Nat. Gen.*, 2013
Gatti, D. et.al., *G3*, 2014

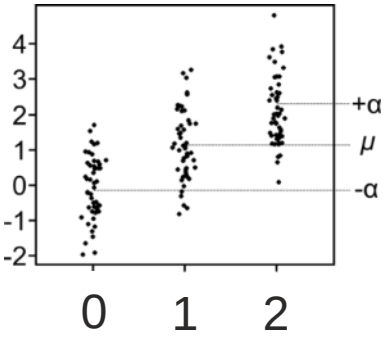
Association Mapping

- Founders have been sequenced.
- We have founder haplotypes for the DO.
- Map founder sequences onto DO.
- QTL map using smaller, three genotype model.

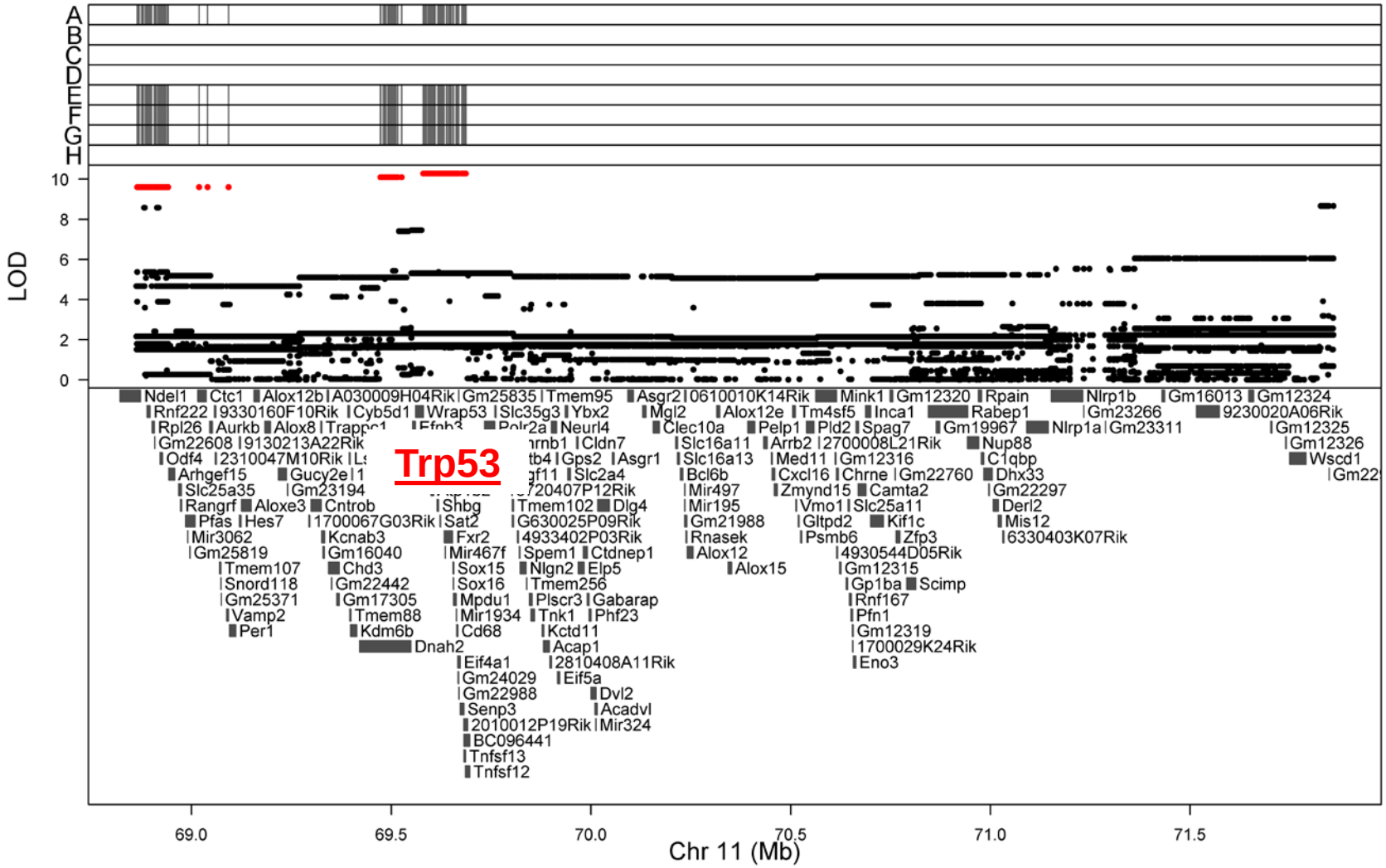
One Sanger SNP

A	B	C	D	E	F	G	H
0	1	0	0	1	1	0	0

DO	#	Geno
AB	0+1	1
BB	1+1	2
AC	0+0	0
BE	1+1	2
DE	0+1	1



P53 is a major DNA repair gene on Chr 11.



Sanger Mouse Genome Project

http://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1410

wellcome trust sanger institute

Search for Enter search here...

Home | Research | Scientific resources | Work & study | About us

Mouse Genomes Project - Gene: Trp53; Chr: 11:69,580,359-69,591,873

Notes:

- Release **REL-1410** - all coordinates shown are relative to the **GRCm38** assembly
- All SNPs, indels and structural variants reported have not been experimentally validated
- SNPs file: **mgp.v4.snps.dbSNP.vcf** in directory ftp://ftp-mouse.sanger.ac.uk/REL-1410-SNPs_Indels
- Insertions/Deletions file: **mgp.v4.indels.dbSNP.vcf** in directory ftp://ftp-mouse.sanger.ac.uk/REL-1410-SNPs_Indels
- Structural variants file: **28strains.REL-1410-SV.sdp.tab** in directory ftp://ftp-mouse.sanger.ac.uk/REL-1410-SV
- Variant consequences are predicted using the Ensembl Variant Effect Predictor using the Sequence Ontology. Ensembl has a brief description of the terms.

SNPs | Insertions/Deletions | Structural variants

Export as TSV | Export as CSV

Rows: 1-122

Filter: 122 entries 1 All

Gene	Chromosome	Position	dbSNP	Reference	129S1/SvImJ	A/J	C57BL/6NJ	CAST/EiJ	NOD/ShiLtJ	NZO/HILtJ	PWK/PhJ	WSB/EiJ
Trp53	11	69,580,480	rs13460498	C	-	-	-	-	-	-	T*	-
Trp53	11	69,580,621	rs211771881	G	C*	-	-	-	-	-	-	-
Trp53	11	69,580,876	rs250363375	C	-	-	-	-	-	-	T*	-
Trp53	11	69,581,523	rs29407720	T	-	-	-	C*	-	-	C*	-
Trp53	11	69,581,731	rs29483021	A	-	-	-	G*	-	-	G*	-
Trp53	11	69,581,881	rs26922529	C	-	T*	-	T*	-	T*	T*	-
Trp53	11	69,581,907	rs50318680	A	-	-	-	G*	-	-	G*	-
Trp53	11	69,581,936	rs29468702	T	-	-	-	C*	-	-	C*	-
Trp53	11	69,581,952	rs29383587	A	-	T*	-	-	-	T*	-	-
Trp53	11	69,581,954	rs29467510	T	-	-	-	A*	-	-	A*	-
Trp53	11	69,581,959	rs46428169	A	-	T*	-	T*	-	T*	T*	-
Trp53	11	69,581,960	rs47769637	G	-	A*	-	A*	-	A*	A*	-
Trp53	11	69,582,055	rs26922527	T	-	-	-	G*	-	-	G*	-
Trp53	11	69,582,262	rs26922526	A	-	-	-	G*	-	-	G*	-
Trp53	11	69,582,400	rs250821274	A	-	-	-	C*	-	-	C*	-
Trp53	11	69,582,401	rs223018749	G	-	-	-	A*	-	-	A*	-

Related links

- Mouse Genomes Project
- Sanger Institute mouse resources
- Mouse FTP site

Releases

- **REL-1410** - GRCm38
- REL-1303 - GRCm38
- REL-1211 - NCBIm37

Search

Legend

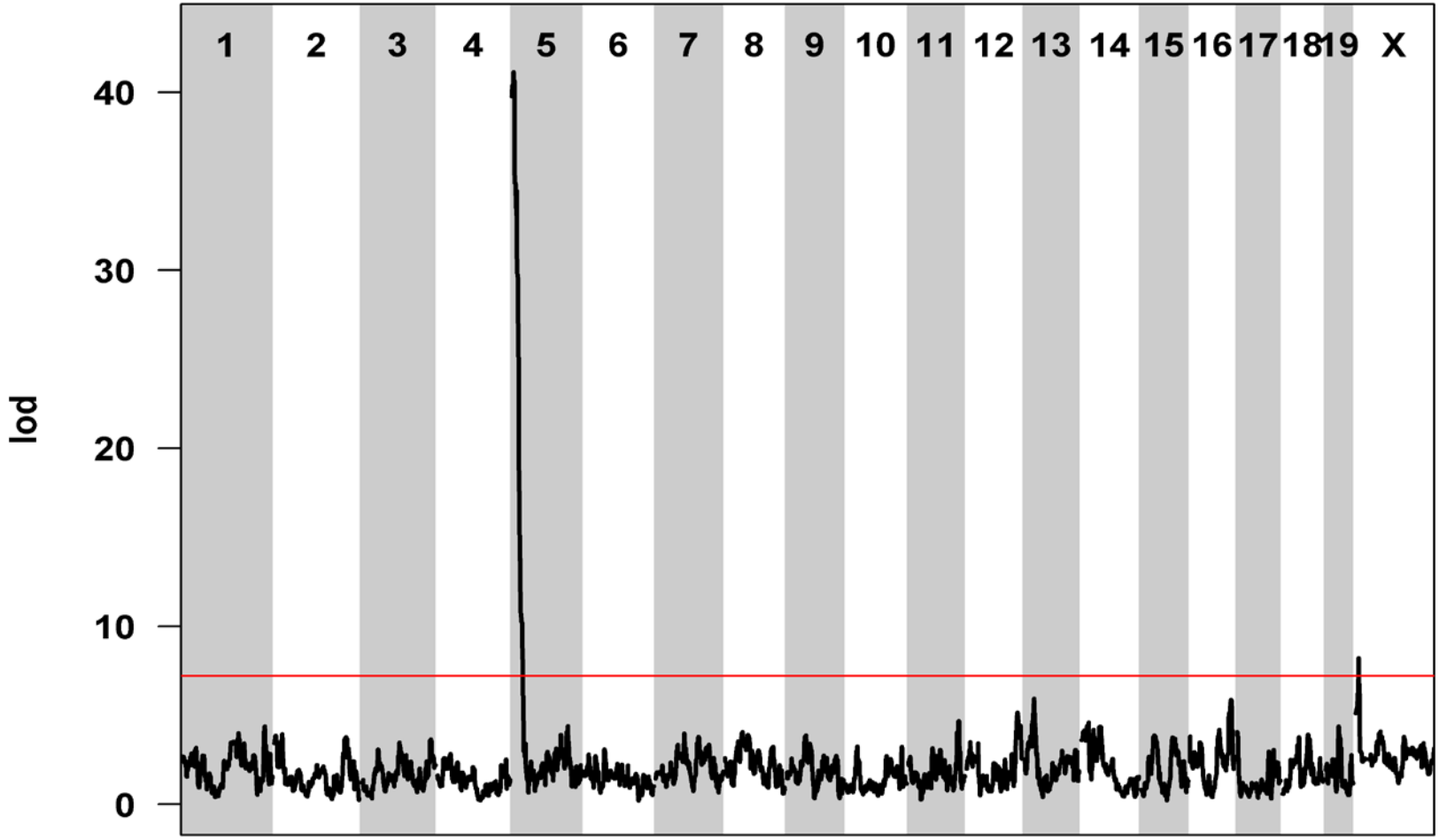
Snp/Indel Consequences:

- 3' UTR variant
- 5' UTR variant
- Coding sequence variant
- Downstream gene variant
- Feature elongation
- Feature truncation
- Frameshift variant
- Incomplete terminal codon variant
- Inframe deletion
- Inframe insertion
- Initiator codon variant
- Mature miRNA variant
- Missense variant
- NMD transcript variant
- Nc transcript variant
- Non coding exon variant
- Regulatory region ablation
- Regulatory region amplification
- Regulatory region variant

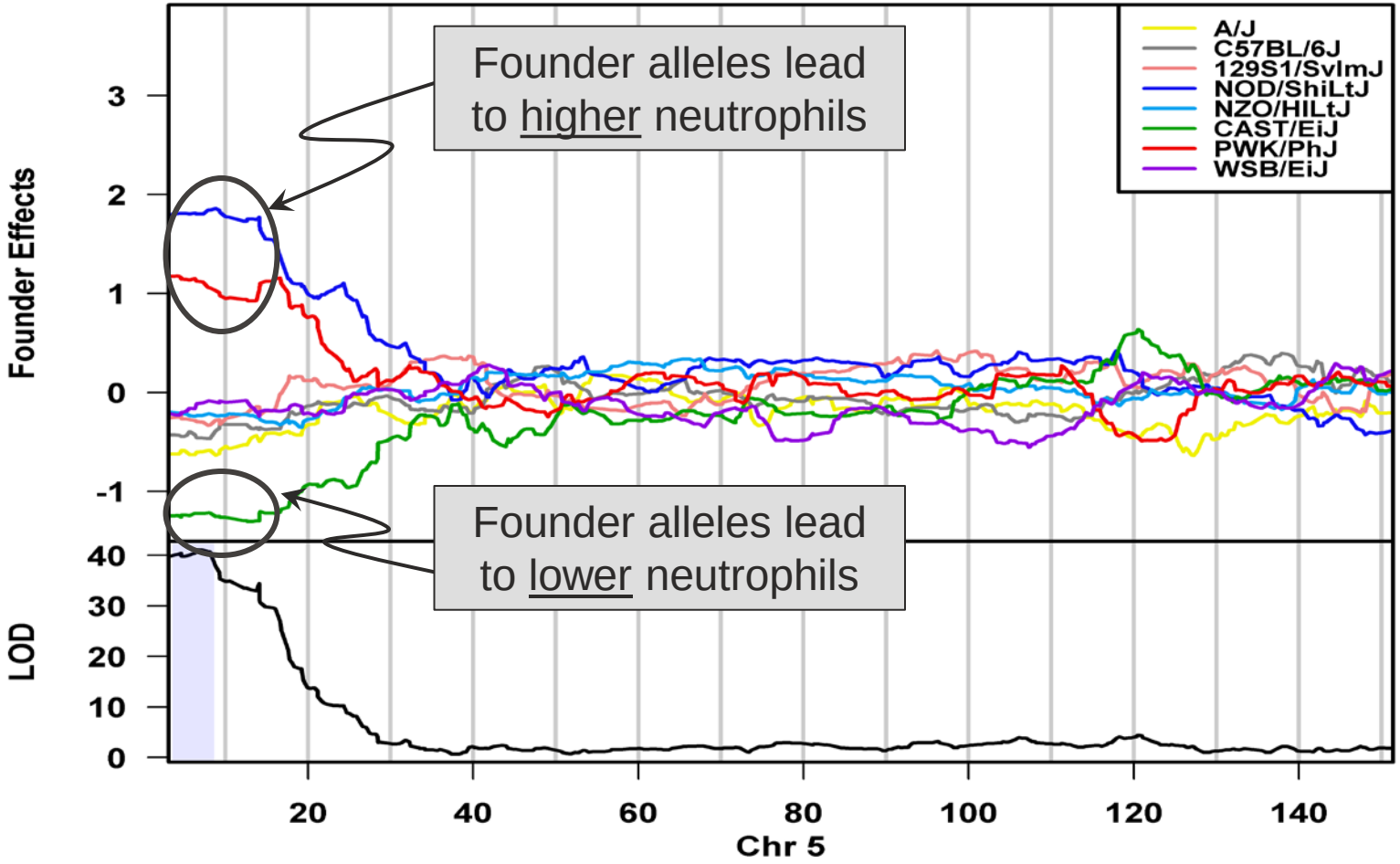
Sanger SNPs that match the founder effect pattern

Chr	Position	Gene	dbSNP	Ref	129S1_SvImJ	A_J	C57BL_6J	CAST_EiJ	NOD_ShiLtJ	NZO_HILtJ	PWK_PhJ	WSB_EiJ
11	69581881	Trp53	rs26922529	C	-	T	-	T	-	T	T	-
11	69581959	Trp53	rs46428169	A	-	T	-	T	-	T	T	-
11	69582531	Trp53	rs49610642	A	-	G	-	G	-	G	G	-
11	69584800	Trp53	rs51392127	A	-	G	-	G	-	G	G	-
11	69584819	Trp53	rs47258361	G	-	A	-	A	-	A	A	-
11	69584824	Trp53	rs47491011	T	-	C	-	C	-	C	C	-
11	69584832	Trp53	rs51564940	C	-	T	-	T	-	T	T	-
11	69584837	Trp53	rs51593866	A	-	G	-	G	-	G	G	-
11	69584954	Trp53	rs26922516	C	-	G	-	G	-	G	G	-
11	69585102	Trp53	rs26922515	C	-	T	-	T	-	T	T	-
11	69585137	Trp53	rs26922514	G	-	A	-	A	-	A	A	-
11	69585351	Trp53	rs26922513	C	-	T	-	T	-	T	T	-
11	69585432	Trp53	rs26922512	T	-	C	-	C	-	C	C	-
11	69585564	Trp53	rs26922511	A	-	G	-	G	-	G	G	-
11	69586165	Trp53	rs29383202	A	-	G	-	G	-	G	G	-
11	69586833	Trp53	rs29422241	A	-	T	-	T	-	T	T	-
11	69587290	Trp53	rs26922507	A	-	G	-	G	-	G	G	-
11	69587985	Trp53	rs29454713	A	-	G	-	G	-	G	G	-
11	69588218	Trp53	rs26922506	G	-	A	-	A	-	A	A	-

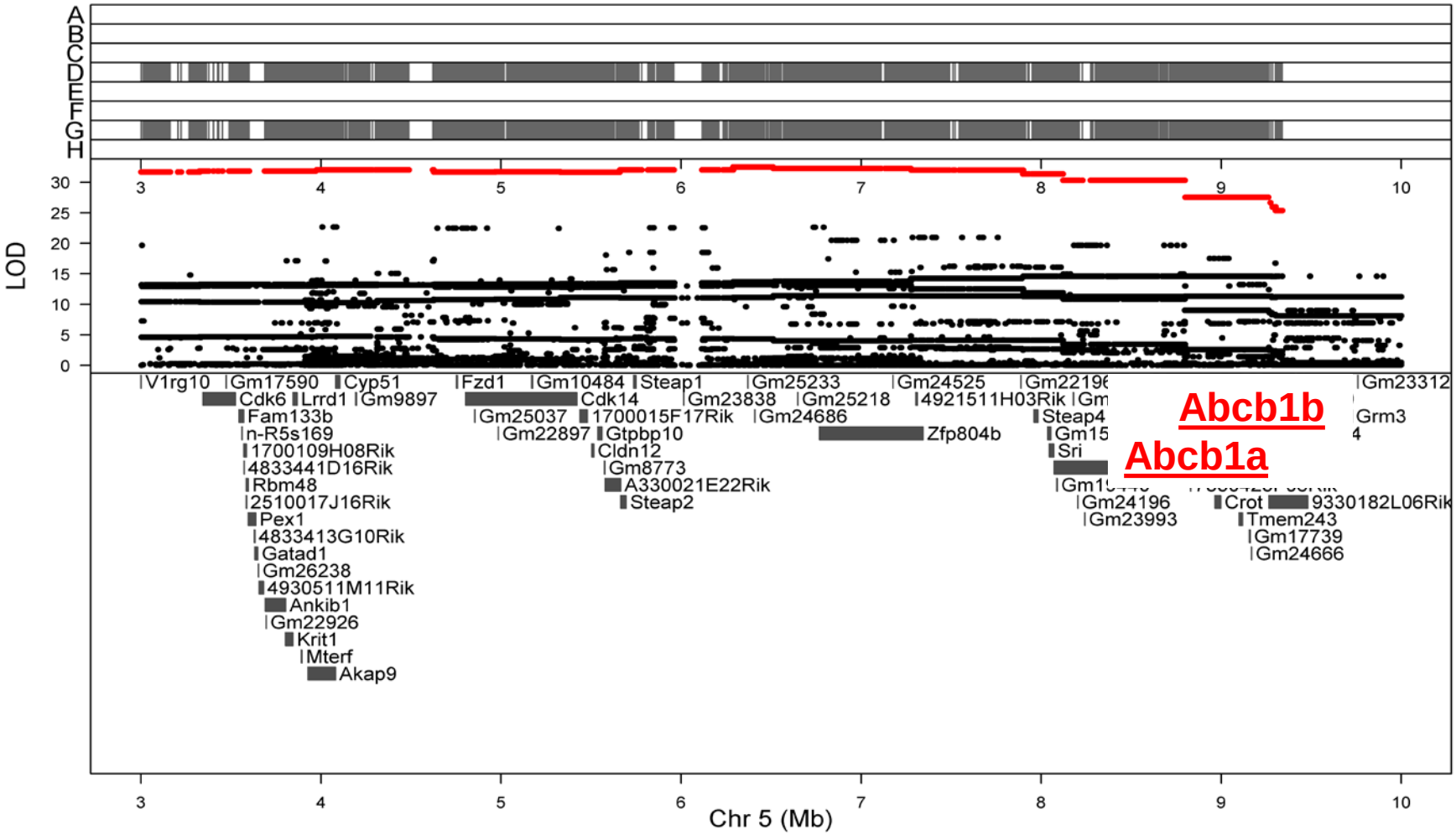
Doxorubicin: Gene(s) on Chr 5 influence neutropenia.



Founder allele effects on Chr 5.



ABCB1 is a multidrug resistance gene on Chr 5.



Future Directions

Website where users can:

Upload genotype data and perform haplotype reconstruction

Upload phenotype data and perform QTL mapping

Deposit data for public release after publication

Summary

Methods for haplotype reconstruction and mapping in the CC and DO are mature and robust.

There are several software packages that implement these methods. (We use DOQTL)

Founder sequences and association mapping facilitate candidate gene discovery.

Acknowledgements

The Jackson Laboratory

Gary Churchill

Elissa Chesler

Karen Svenson

Steven Munger

Neal Goodwin

Univ. of Wisconsin

Karl Broman

Univ. of Chicago

Abraham Palmer

Riyan Cheng

Univ. of NC

Leonard McMillan

Fernando Pardo-Manuel de Villena

Cheng-Ping Fu

Funding:

NCI

NIGMS