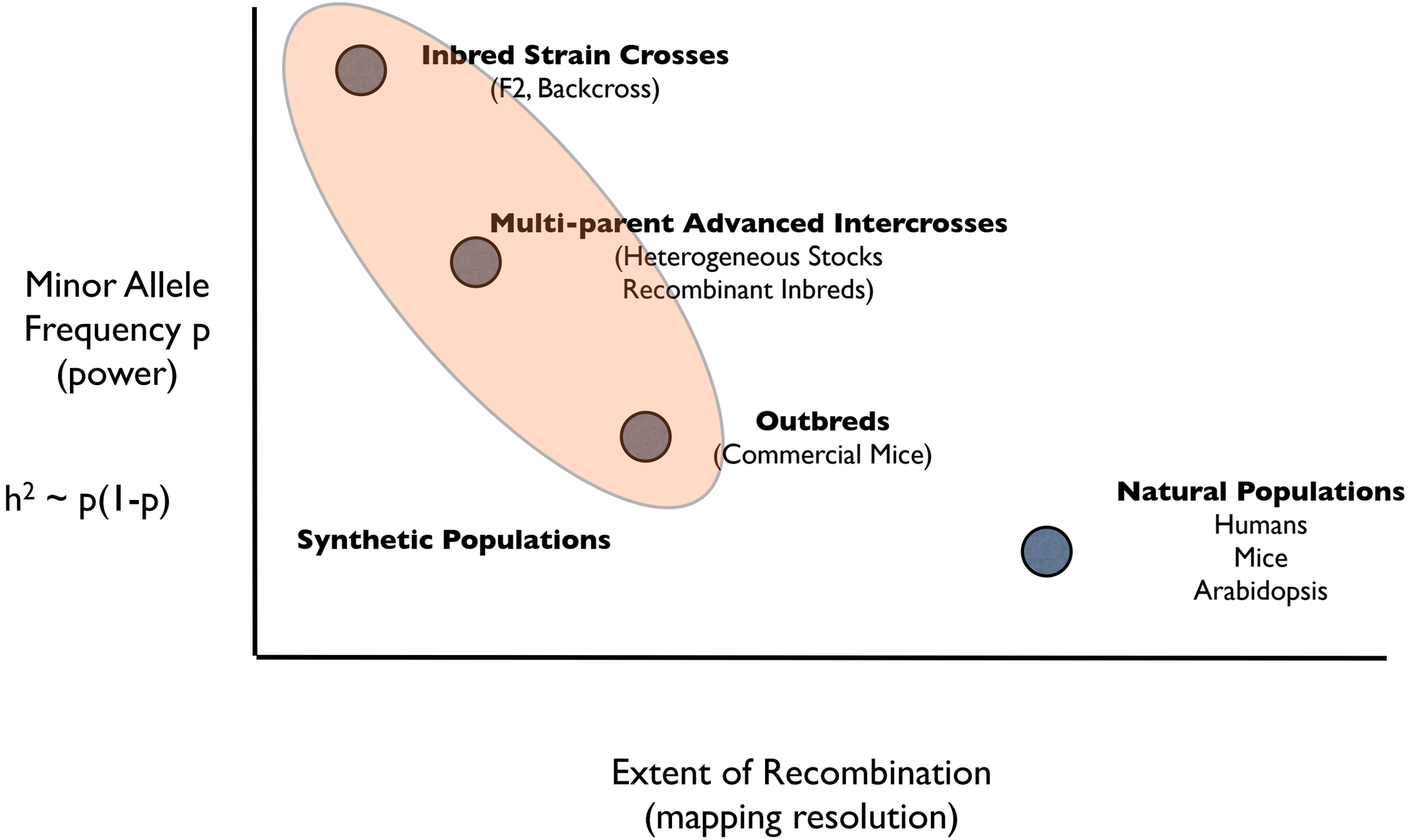
MAGIC Rodents

Richard Mott
Wellcome Trust Centre for Human Genetics
University of Oxford UK

Mice

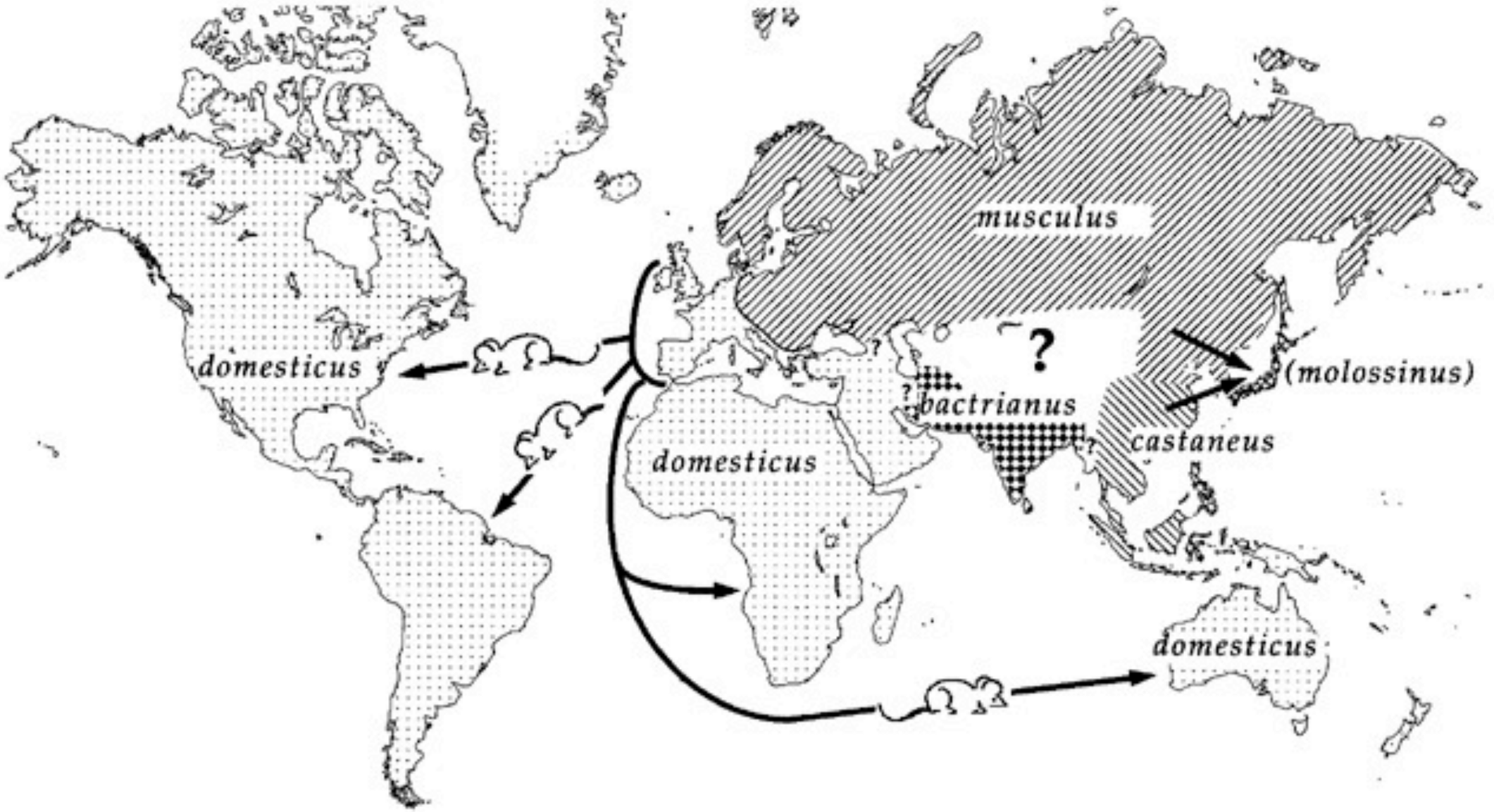
Rats

Synthetic *vs* Natural Populations



Mice

Genome size 2.8 Gb



Lee Silver

ARTICLE

doi:10.1038/nature10413

Mouse genomic variation and its effect on phenotypes and gene regulation

Thomas M. Keane[1]*, Leo Goodstadt[2]*, Petr Danecek[1]*, Michael A. White[3], Kim Wong[1], Binnaz Yalcin[2], Andreas Heger[4], Avigail Agam[2,4], Guy Slater[1], Martin Goodson[2], Nicholas A. Furlotte[5], Eleazar Eskin[5], Christoffer Nellåker[4], Helen Whitley[2], James Cleak[2], Deborah Janowitz[2,6], Polinka Hernandez-Pliego[2], Andrew Edwards[2], T. Grant Belgard[4], Peter L. Oliver[4], Rebecca E. McIntyre[1], Amarjit Bhomra[2], Jérôme Nicod[2], Xiangchao Gan[2], Wei Yuan[2], Louise van der Weyden[1], Charles A. Steward[1], Sendu Bala[1], Jim Stalker[1], Richard Mott[2], Richard Durbin[1], Ian J. Jackson[7], Anne Czechanski[8], José Afonso Guerra-Assunção[9], Leah Rae Donahue[8], Laura G. Reinholdt[8], Bret A. Payseur[3], Chris P. Ponting[4], Ewan Birney[9], Jonathan Flint[2] & David J. Adams[1]

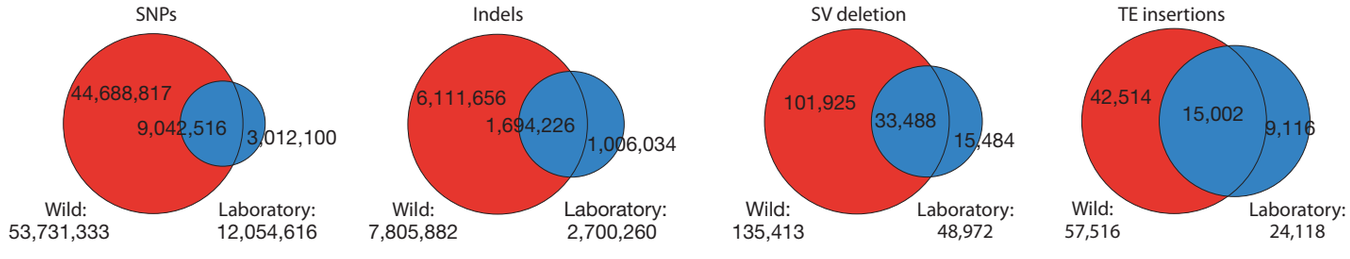
1 SNP per 500 bp (classical)
1 SNP per 130 bp (wild)

Table 1 | An overview of the sequence and variants called from 17 mouse genomes.

Strain	Gb of mapped data	Coverage	% of genome inaccessible	SNPs	(Private)	Indels	(Private)	Structural variants	(Private)
C57BL/6NJ	77.29	29.29	13.21	9,844	(1,488)	22,228	(4,259)	431	(75)
129S1/SvImJ	71.91	27.25	15.30	4,458,004	(1,489)	886,136	(16,140)	29,153	(786)
129S5SvEv^{Brd}	50.27	19.05	15.17	4,383,799	(1,991)	810,310	(21,214)	25,340	(691)
129P2/Ola	115.52	43.78	14.47	4,694,529	(23,677)	1,028,629	(58,173)	32,227	(3,430)
A/J	70.39	26.68	15.90	4,198,324	(44,837)	823,688	(24,502)	28,691	(1,474)
AKR/J	107.16	40.61	14.86	4,331,384	(87,527)	966,002	(64,422)	30,742	(3,576)
BALB/cJ	65.72	24.90	15.09	3,920,925	(29,973)	831,193	(30,998)	25,702	(1,056)
C3H/HeJ	92.81	35.17	15.09	4403599	(16,804)	949,206	(34,834)	28,532	(1,779)
CBA/J	77.43	29.34	14.79	4,511,278	(34,203)	929,860	(35,976)	28,183	(1,178)
DBA/2J	65.11	24.67	15.09	4,468,071	(72,214)	868,611	(37,085)	28,346	(1,469)
LP/J	73.03	27.67	15.29	4,701,445	(53,509)	947,614	(33,817)	30,024	(1,194)
NOD/ShiLtJ	75.88	28.75	17.30	4,323,530	(143,489)	797,086	(41,113)	30,605	(2,479)
NZO/HILtJ	45.68	17.31	16.06	4,492,372	(210,256)	806,511	(60,231)	25,125	(1,938)
PWK/PhJ	66.99	25.38	19.26	17,202,436	(4,461,772)	2,635,885	(833,794)	90,125	(25,383)
CAST/EiJ	64.84	24.57	19.18	17,673,726	(5,368,019)	2,727,089	(956,828)	86,322	(25,232)
WSB/EiJ	48.19	18.26	16.23	6,045,573	(894,875)	1,197,006	(211,348)	35,066	(5,957)
SPRET/EiJ	70.41	26.68	23.26	35,441,735	(23,455,525)	4,456,243	(2,936,998)	157,306	(91,721)
Total	1,238.63	469.36		129,260,574		21,683,297		711,920	

Private variants are strain-specific variants.

b



Combined sequence-based and genetic mapping analysis of complex traits in outbred rats

Rat Genome Sequencing and Mapping Consortium*

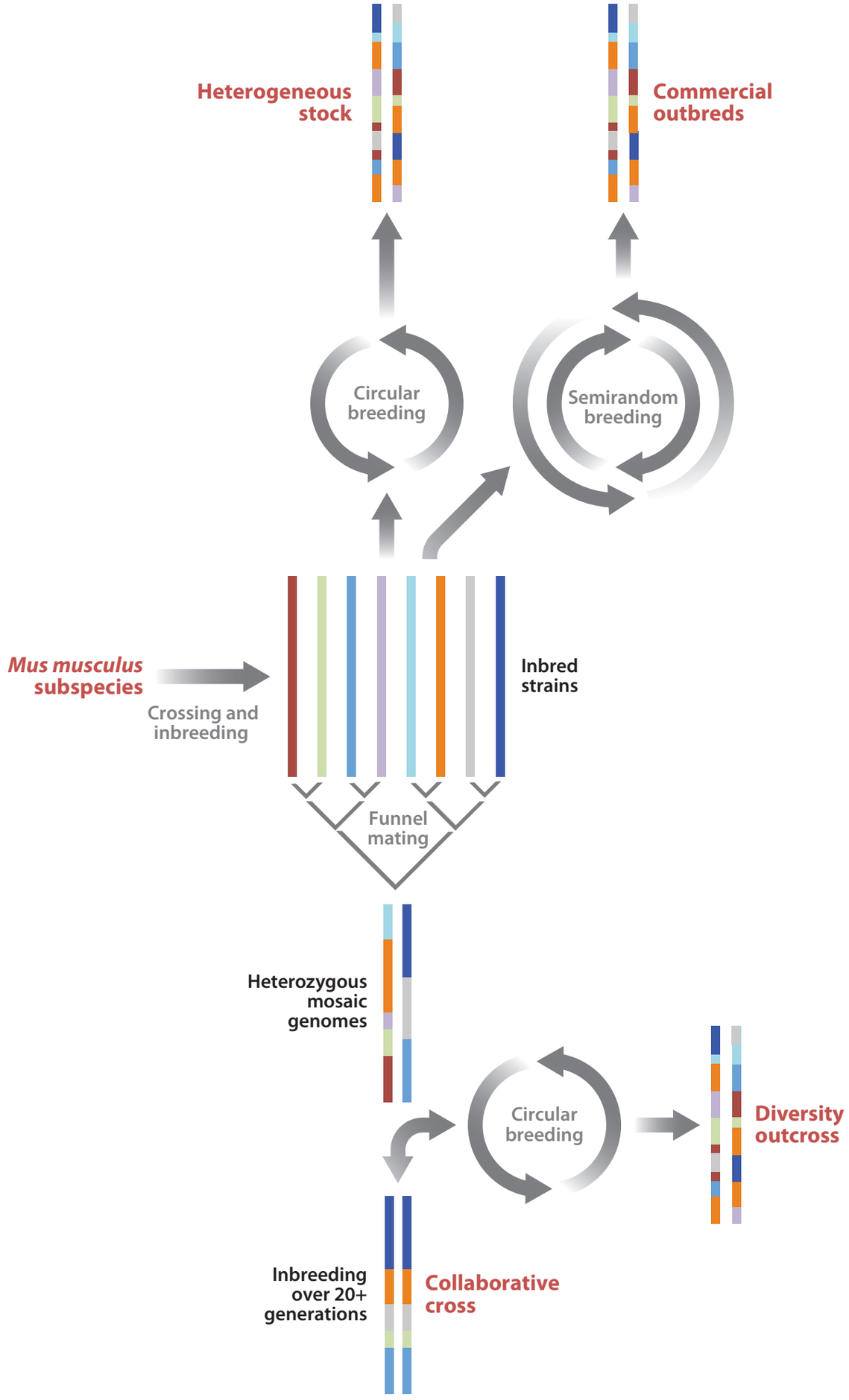
1 SNP per 800 bp

Table 1 Sequence variation in the eight progenitor strains of NIH-HS rats

Strain	Mapped data (Gb)	Coverage	Inaccessible genome (%)	SNPs	Private SNPs	Indels	Private indels	Structural variants	Private structural variants
ACI/N	65.9	26.3	12.6	2,883,405	228,468	166,425	12,646	19,499	756
BN/SsN	54.4	21.7	9.4	71,038	563,308	0	14,839	27	4,203
BUF/N	62.3	24.9	12.7	2,748,633	125,202	172,934	7,195	22,176	1,002
F344/N	77.9	31.1	11.8	2,831,144	97,951	157,522	5,007	25,257	1,003
M520/N	72.5	28.9	12.3	2,836,898	89,277	170,031	5,008	24,090	915
MR/N	62.4	24.9	12.3	2,664,124	223,514	151,099	12,005	18,306	1,004
WKY/N	63.4	25.3	12.1	3,088,953	496,327	164,634	23,979	28,270	3,357
WN/N	62.3	24.9	12.2	2,698,493	249,563	154,769	13,541	18,563	700

Shown for each strain is the amount of sequence mapped to the reference, the coverage, the percent of the genome deemed inaccessible and the counts of the three classes of variants compared to the reference strain. Private variants are variants that distinguish a specified strain from all others; most of the alleles private to BN/SsN are reference alleles.

Populations of Mice Descended from Inbred Strains



Population	Number of Segregating SNPs
Mouse Heterogeneous Stock	11 million
Outbred Mice	7 million
Collaborative Cross / Diversity Outcross	36 million
Mouse Inbred Strains (incl M Spretus)	57 million
Rat Heterogeneous Stock	7 million

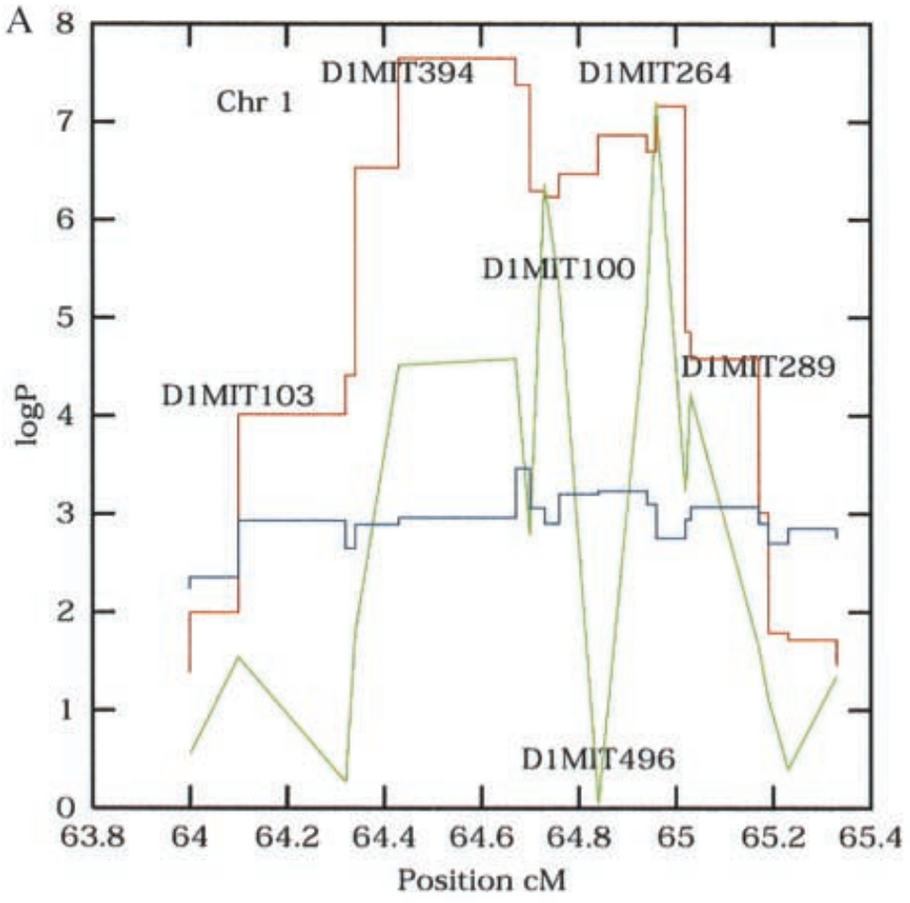
Outline of Genetic Association Analysis

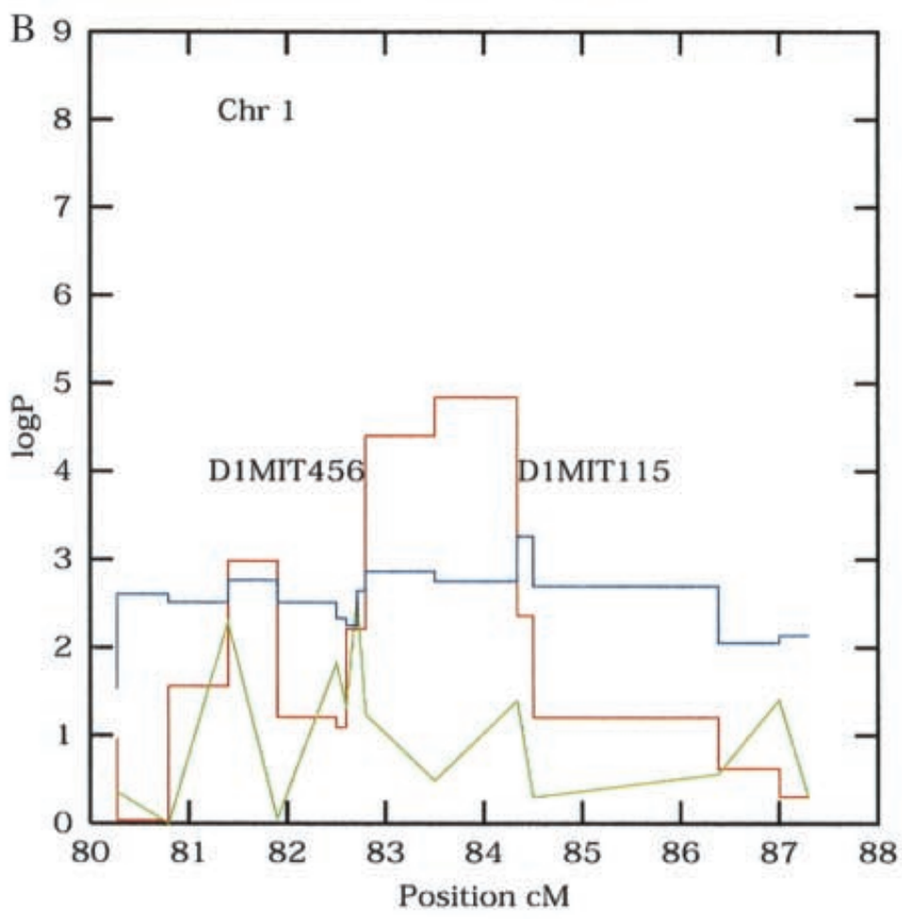
- Collect phenotypes
- Collect genotypes
  - Array
  - Sequencing
- Construct mosaics in terms of founders
  - Hidden Markov Model
- Associate founder haplotypes with phenotypes
- Control for population structure
  - Mixed Models or Resample Model Averaging
- Sequence founders
  - Create catalogue of segregating variation
- Impute genomes of mapping population
- Associate individual sequence variants with phenotypes

Mott *et al PNAS* 2000

A method for fine mapping quantitative trait loci in outbred animal stocks

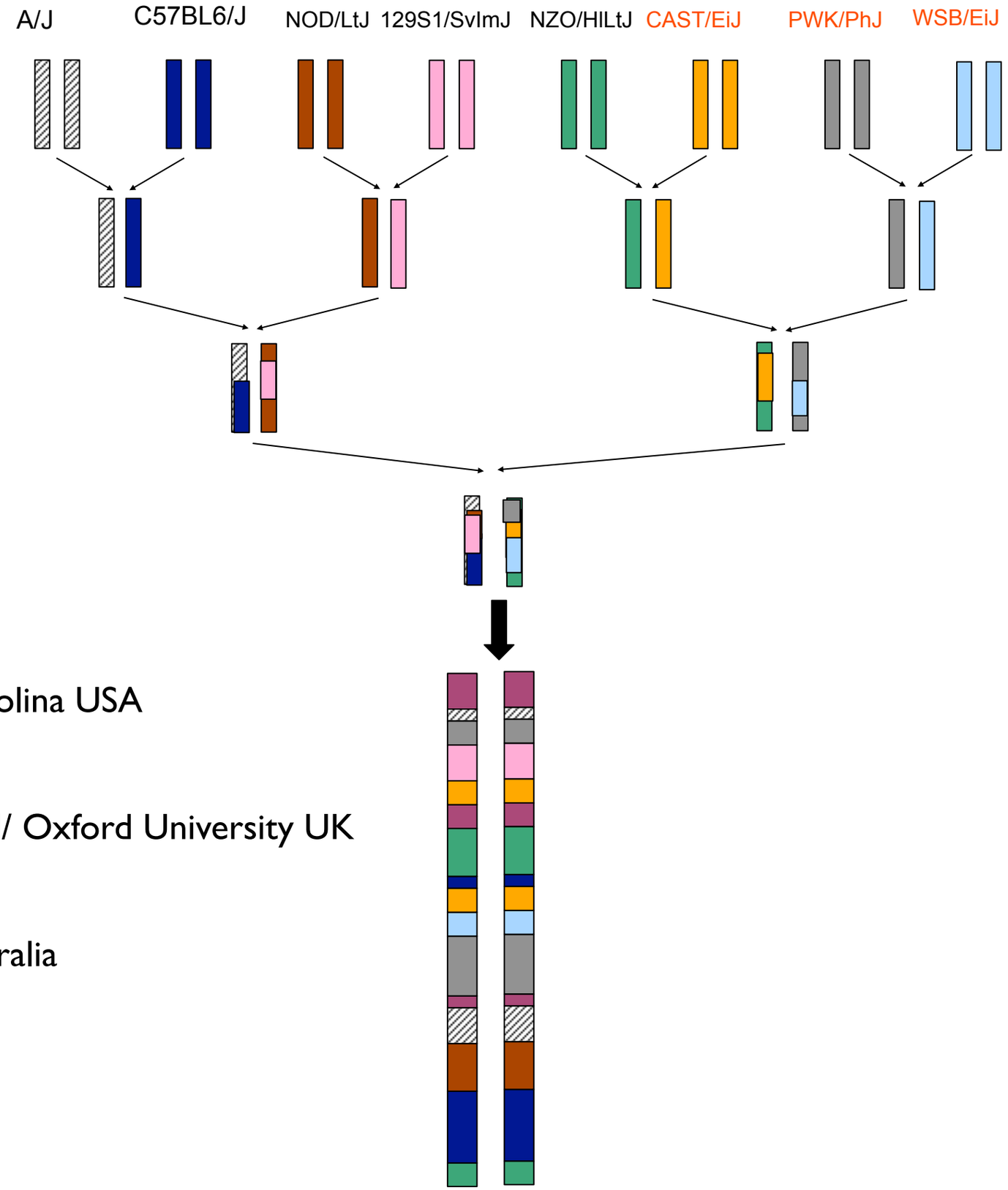
Richard Mott*, Christopher J. Talbot*, Maria G. Turri*, Allan C. Collins†, and Jonathan Flint*‡





www.well.ox.ac.uk/happy

Collaborative Cross Mice

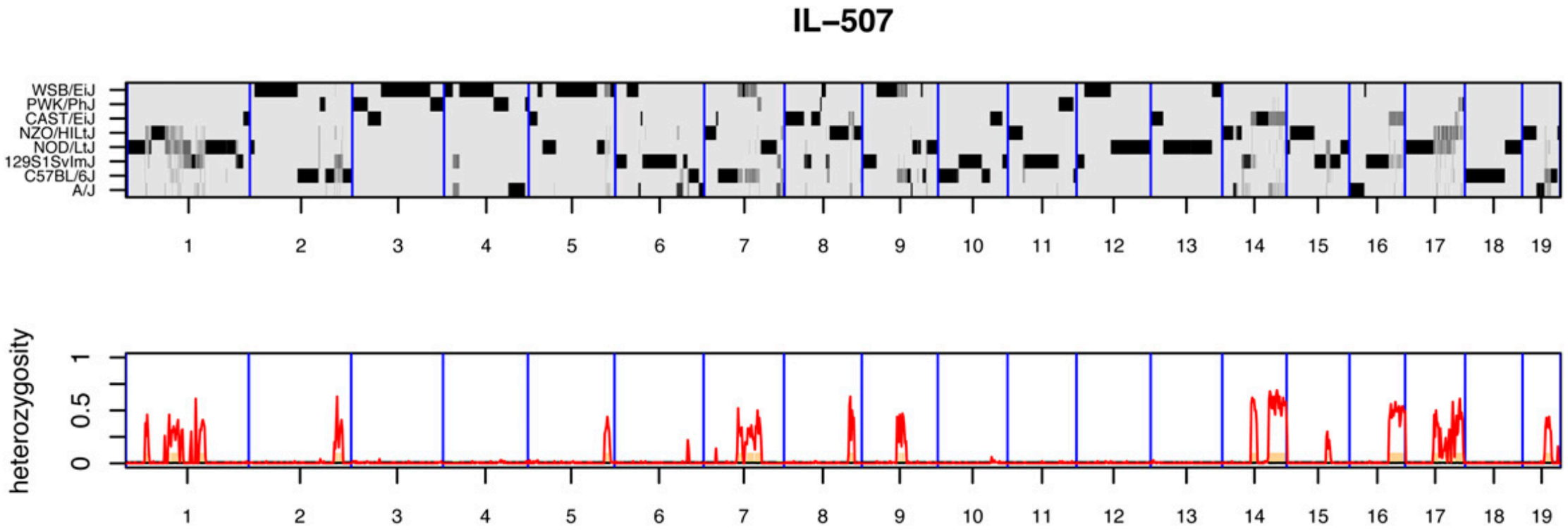


University of North Carolina USA

Tel-Aviv University Israel/ Oxford University UK

University of Perth, Australia

Collaborative Cross Genomes

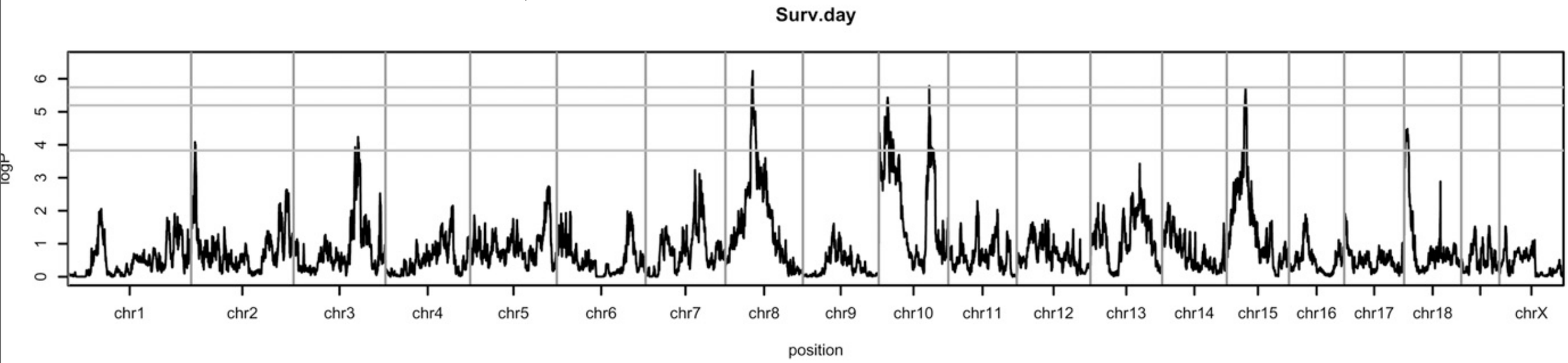


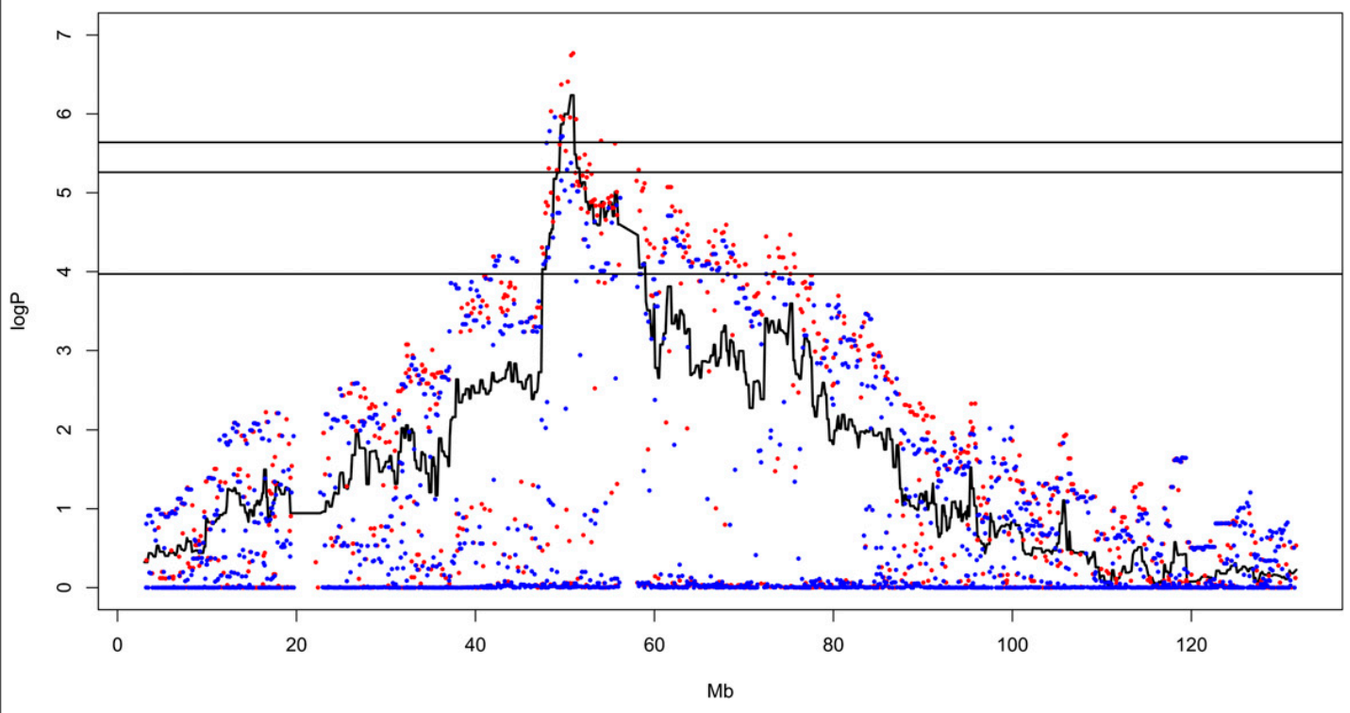
HMM reconstruction based on 170k SNPs

Durrant et al 2011 Genome Res

Susceptibility to Aspergillosis

66 CC lines, 371 mice





Coding variants are enriched for associations

12% of coding variants are associated

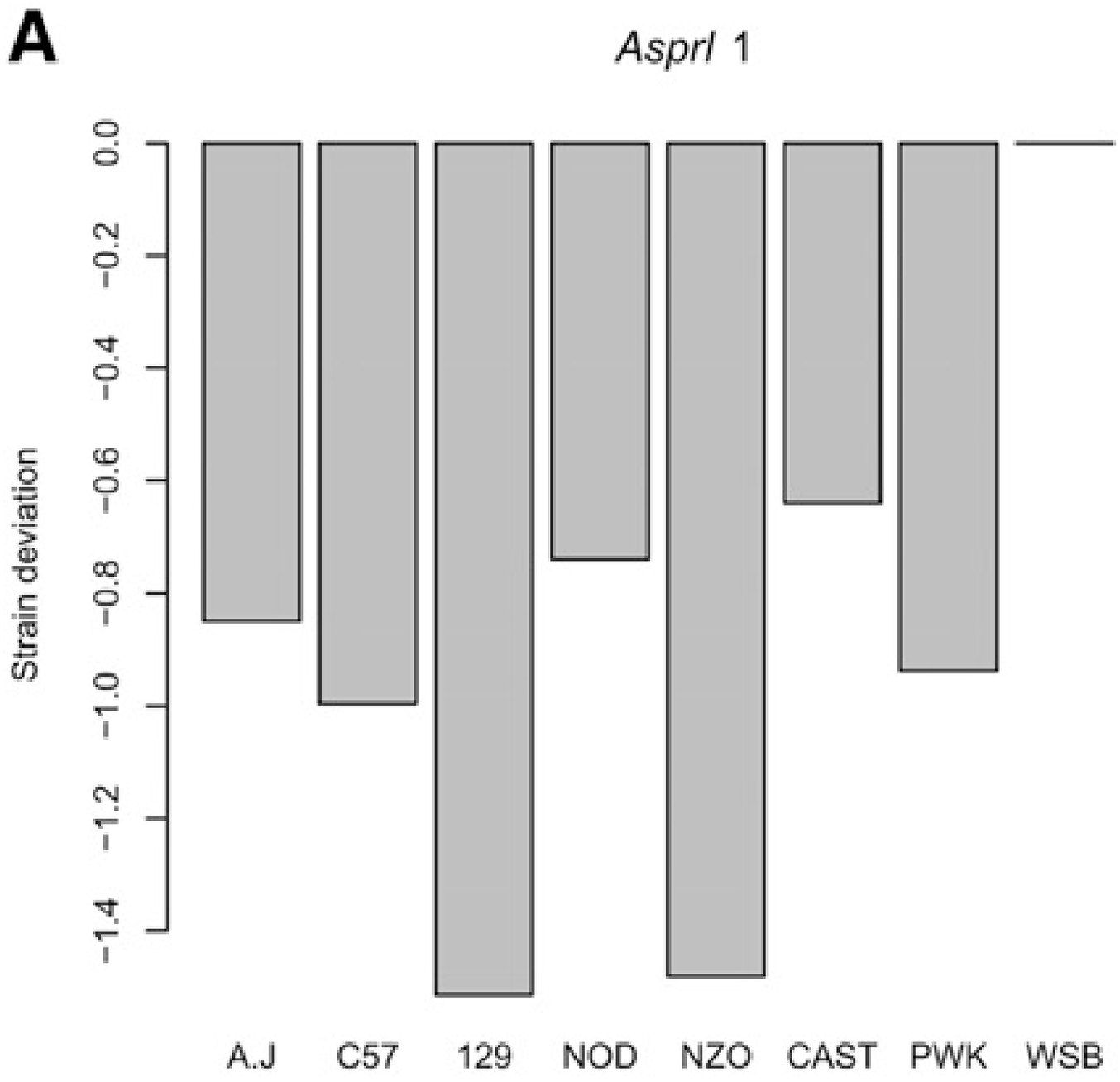
4% of intergenic variants

(Fisher Exact test $P < 3 \times 10^{-14}$)

Durrant *et al* 2011 *Genome Res*

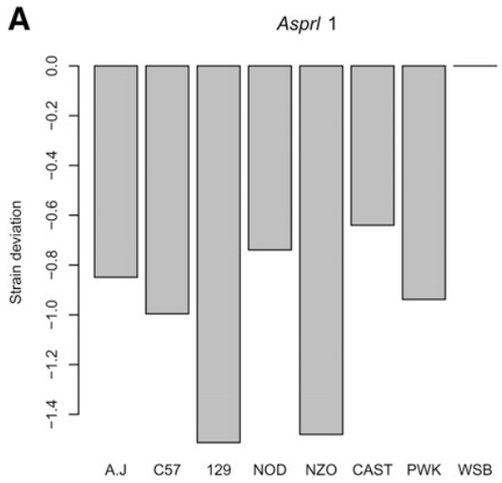
Wild-derived Alleles

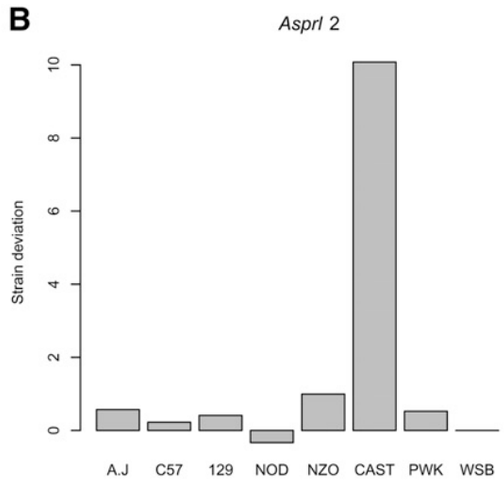
susceptible

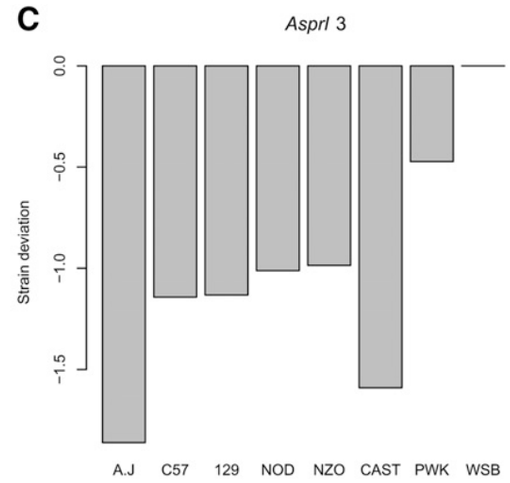


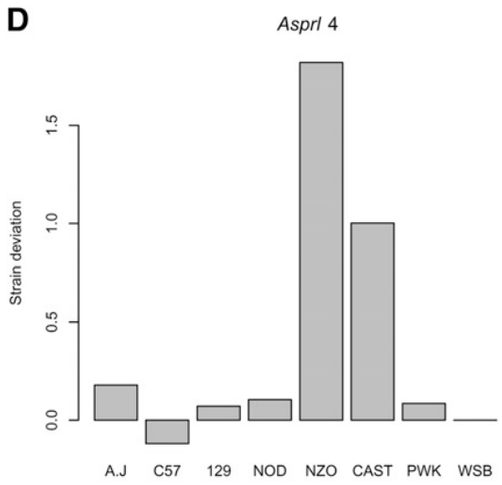
resistant

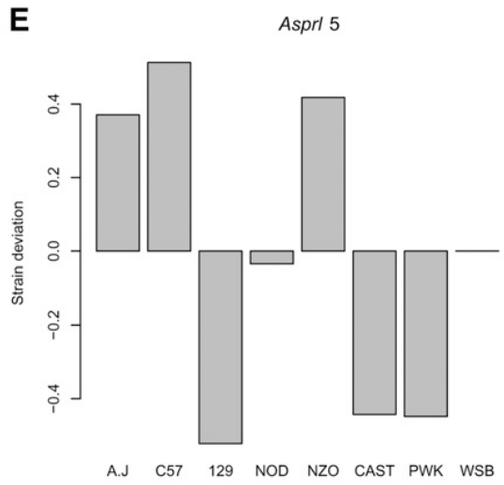
Wild-derived alleles

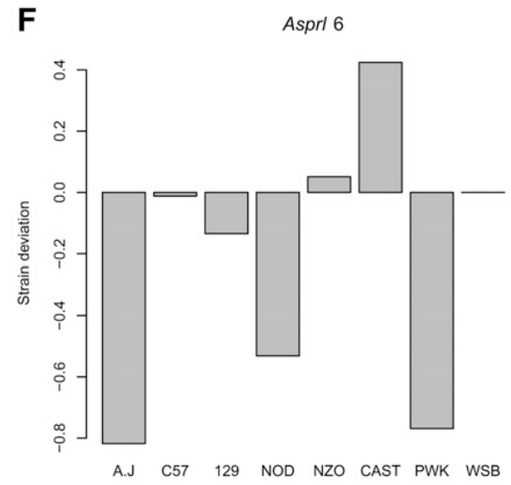


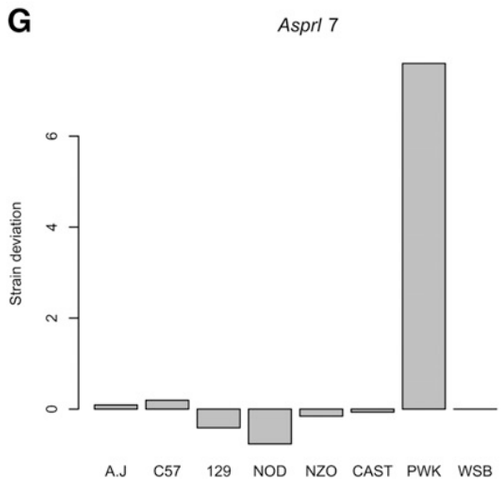












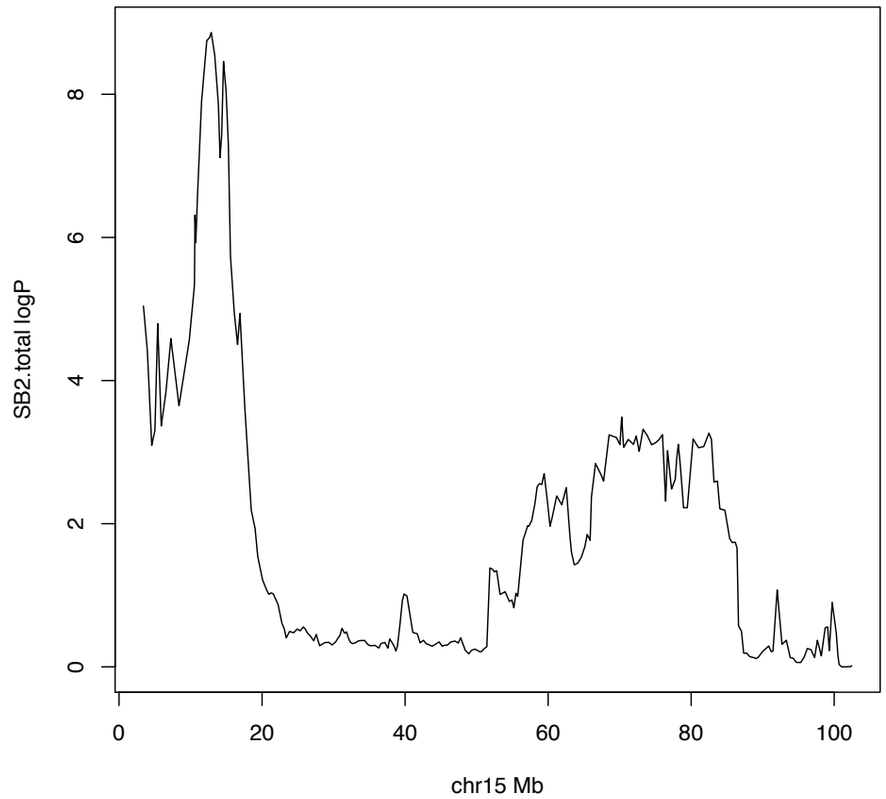
Genetic Modifier QTLs

- Mutations in APC cause colorectal cancer in humans and mice

- Cross APC$^{-/-}$ knockouts onto CC lines to generate F1 mice (no need to genotype)

- Phenotype F1s (count intestinal polyps)

- Map modifier QTL



Preliminary data based on ~35 CC lines

Heterogeneous Stocks

Baud *et al Nature Genetics* 2013

Combined sequence-based and genetic mapping analysis of complex traits in outbred rats

Rat Genome Sequencing and Mapping Consortium*

Valdar *et al Nature Genetics* 2006

Genome-wide genetic association of complex traits in heterogeneous stock mice

William Valdar[1], Leah C Solberg[1,4], Dominique Gauguier[1], Stephanie Burnett[1], Paul Klenerman[2], William O Cookson[1], Martin S Taylor[1], J Nicholas P Rawlins[3], Richard Mott[1] & Jonathan Flint[1]

Mouse *vs* Rat

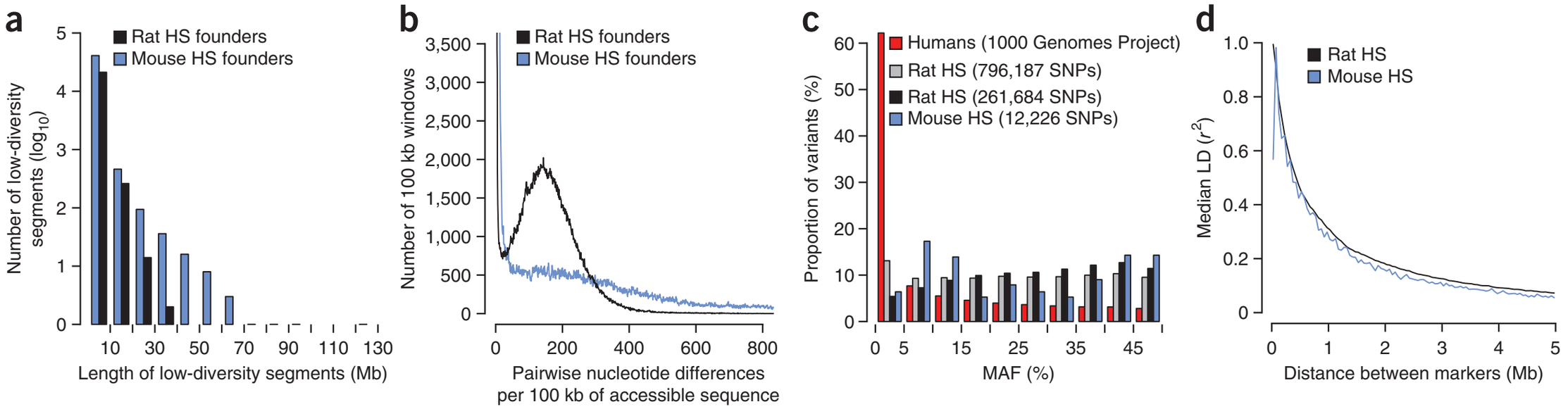
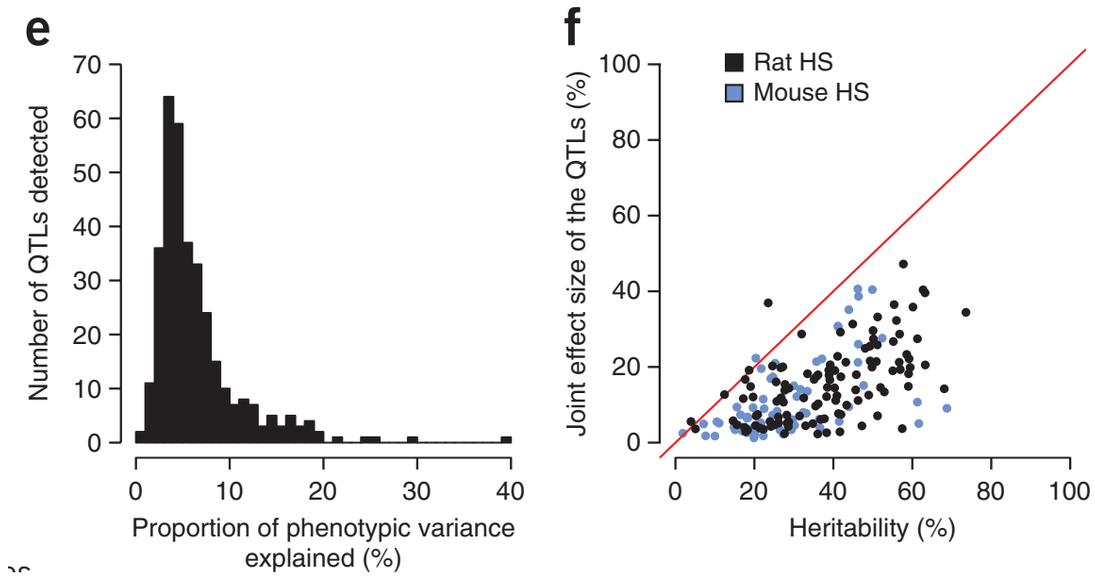


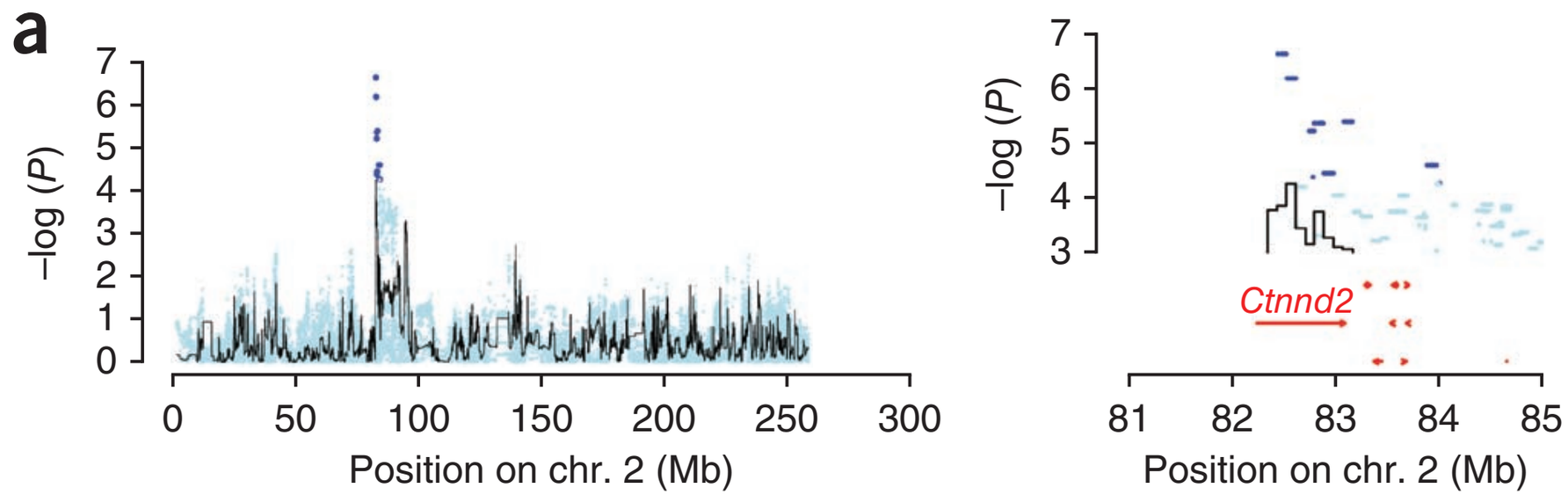
Table 2 Summary of phenotypes collected

Phenotype	Disease model	Number of measures	Age (weeks)
Coat color		4	7
Wound healing		1	7, 17
Fear-related behaviors	Anxiety	10	8–10
Glucose tolerance	Type 2 diabetes	6	11
Cardiovascular function	Hypertension	2	12
Body weight	Obesity	1	13
Basal hematology		26	13
Basal immunology		34	13
Induced neuroinflammation	Multiple sclerosis	11	13–17
Bone mass and strength	Osteoporosis	43	17
Arterial elastic lamina ruptures		6	17
Serum biochemistry		15	17
Renal agenesis		1	17



Quantitative Trait Loci

Anxiety in Heterogeneous Stock Rats



$$y = X\beta + P_L T_L + u + \varepsilon$$

$$(A^{-1}y) = (A^{-1}X)\,\beta + (A^{-1}P_L)\,T_L + A^{-1}(u + \varepsilon)$$

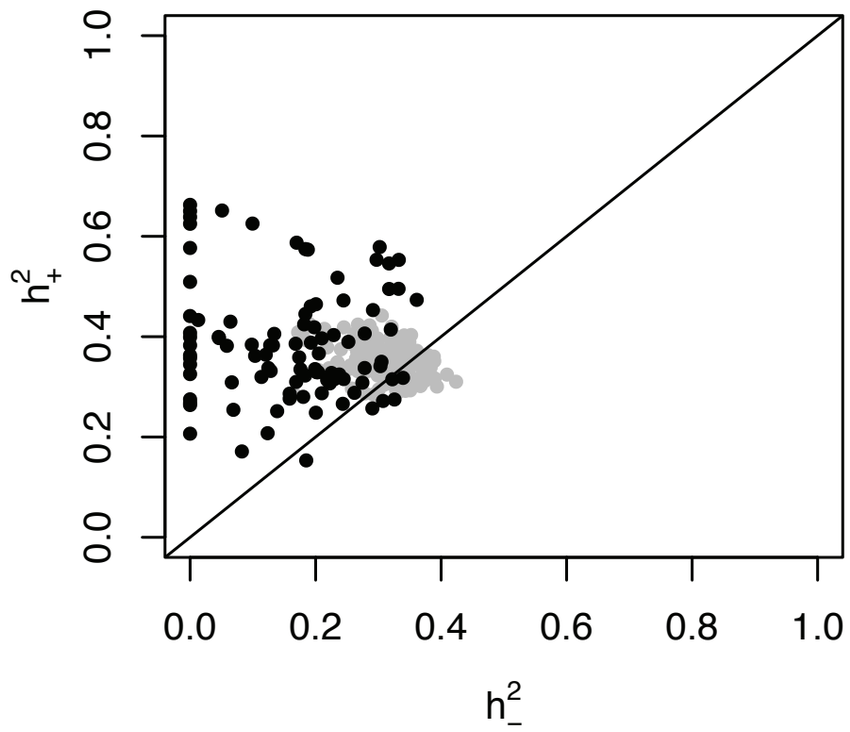
$$V = \sigma_g^2 K + \sigma_e^2 I$$

$$V = A^2$$

EMMA: Kang, H.M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).

Parent of Origin Effects in Heterogeneous Stock Mice

- Parents of 1389 mice were genotyped
- Modified HMM estimates phased haplotype probabilities
- Test for Parent of Origin Effects
- Partition heritability according to Parent of Origin

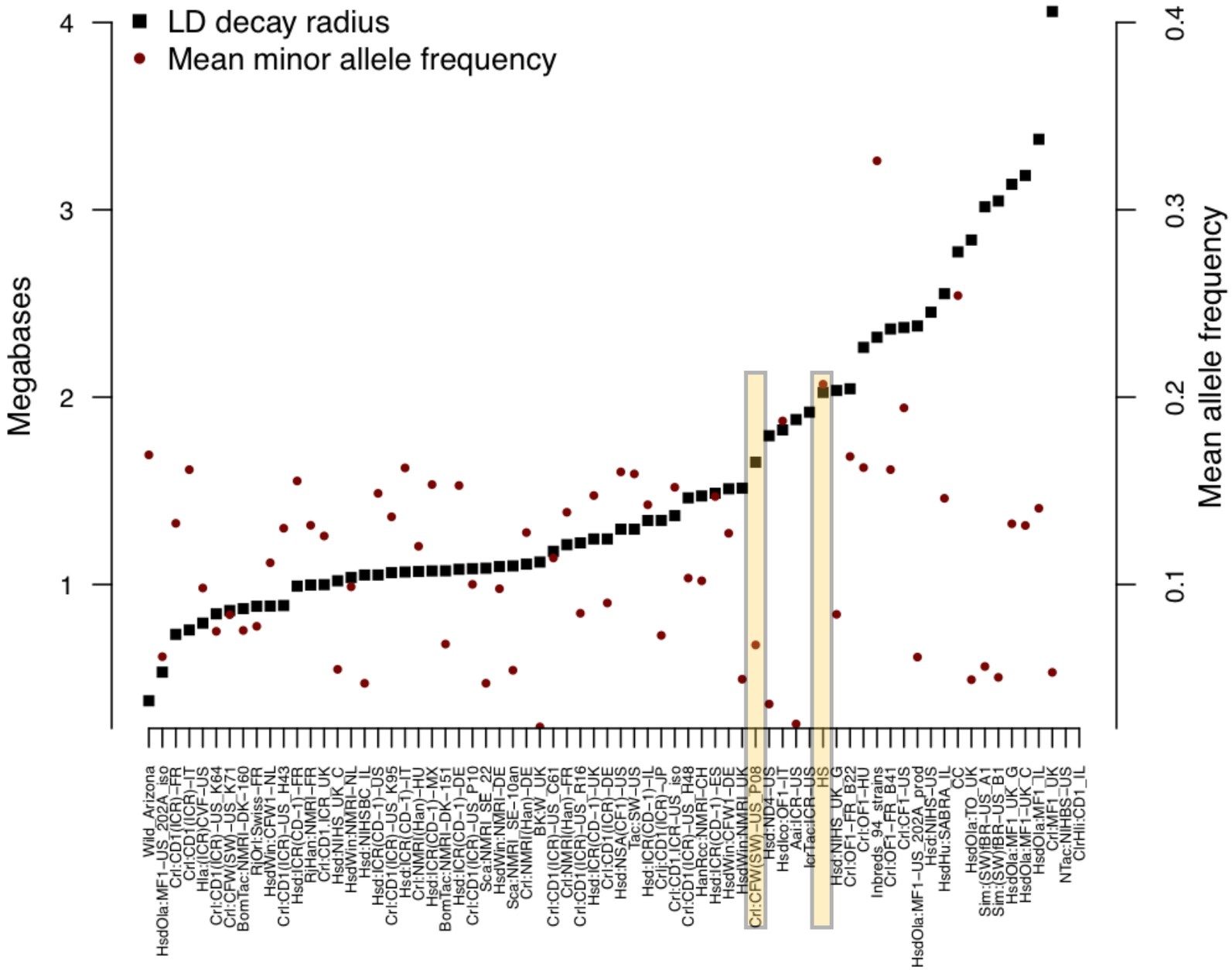


OPEN ACCESS Freely available online

PLoS GENETICS

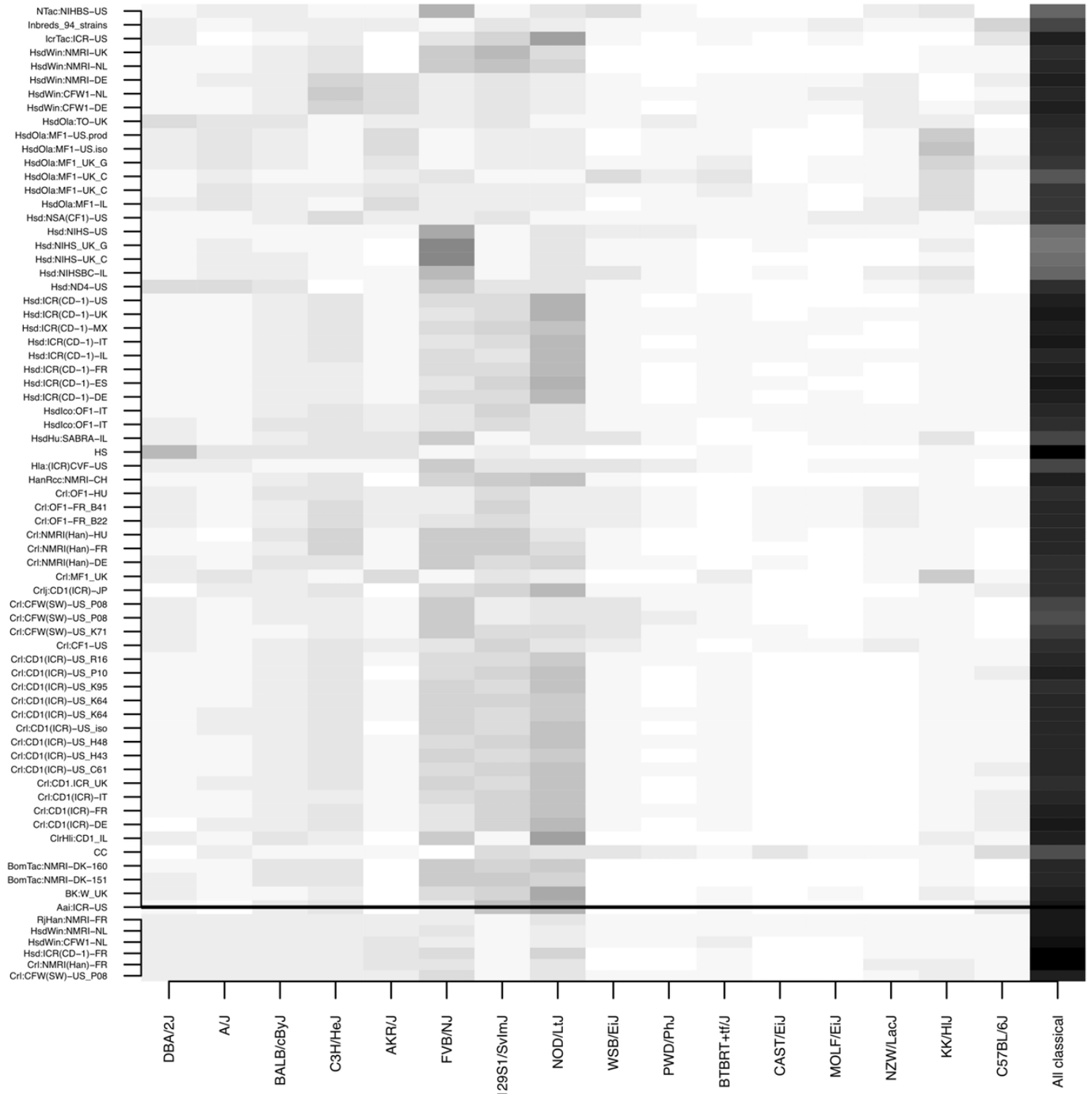
Commercially Available Outbred Mice for Genome-Wide Association Studies

Binnaz Yalcin[1], Jérôme Nicod[1], Amarjit Bhomra[1], Stuart Davidson[1], James Cleak[1], Laurent Farinelli[2], Magne Østerås[2], Adam Whitley[1], Wei Yuan[1], Xiangchao Gan[1], Martin Goodson[1], Paul Klenerman[3], Ansu Satpathy[4], Diane Mathis[4], Christophe Benoist[4], David J. Adams[5], Richard Mott[1], Jonathan Flint[1]*



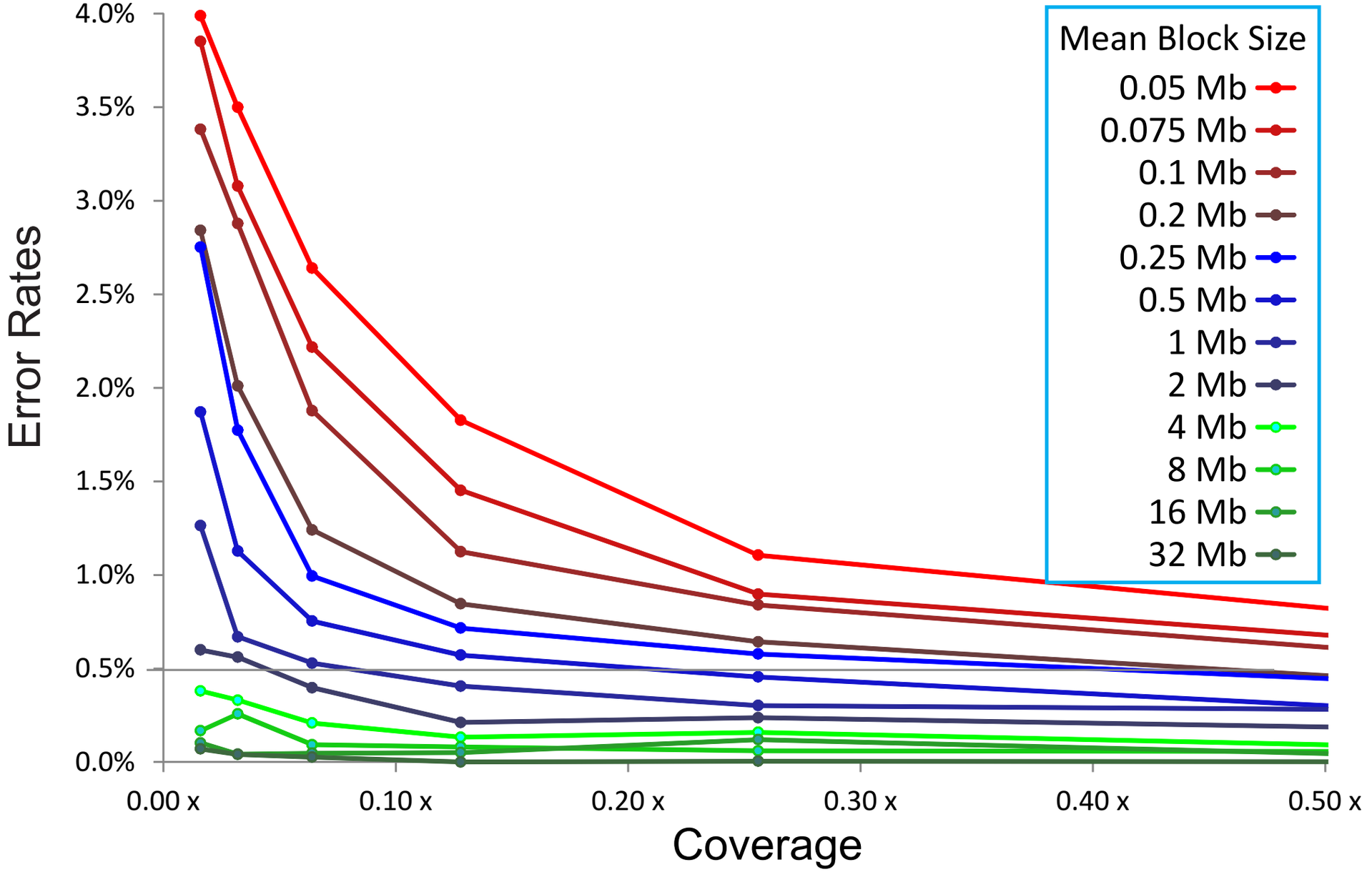
Yalcin *et al PLoS Genetics* 2010

Outbred Mice Resemble Inbred Mosiacs



Genotyping by Low Coverage Sequencing

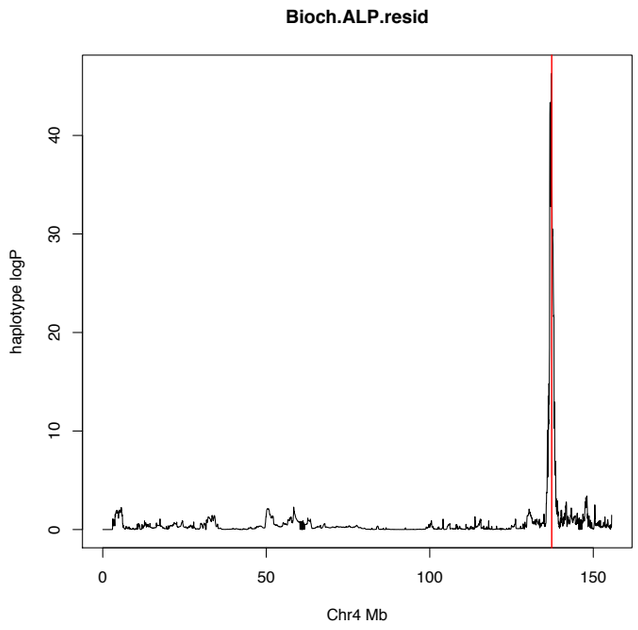
Simulations

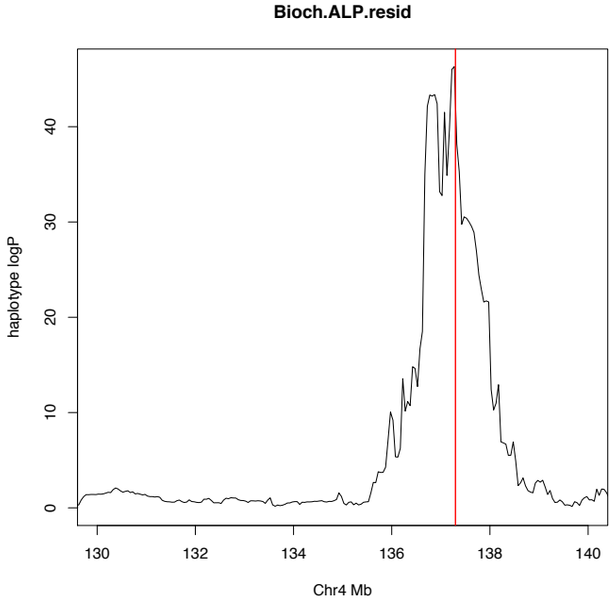


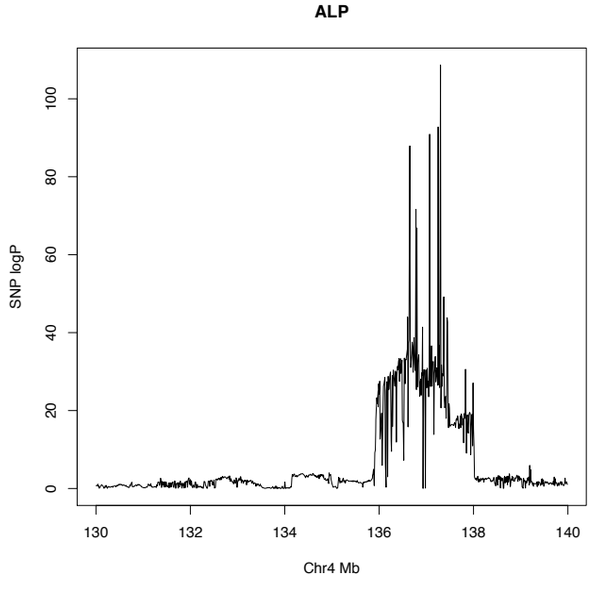
Imputation in Commercial Outbreds

2000 Crl-CFW mice sequenced at 0.1x coverage

- Assume mice are descended from known inbred strains
- Use existing catalogue of variants
- Assume they are descended from two unknown outbred individuals
- Estimate founders and catalogue from all sequence data (FastPhase)







Concluding Thoughts

- Outbred and Inbred populations have similar analysis methodology
- Genetic diversity of founders affects architecture of QTLS
- Genotyping by low coverage sequencing is feasible alternative to SNP arrays

Acknowledgements

- Wellcome Trust
- EU FP7
- CRUK
- Jonathan Flint
- Fuad Iraqi & Collaborative Cross
- Amelie Baud & EURATRANS
- William Valdar
- Caroline Durrant
- Wei Yuan
- Xiangchao Gan
- Robbert Davies
- Jerome Nicod
- Leo Goodstadt
- Ian Tomlinson