Haplotype reconstruction using low-pass whole-genome sequencing in genetically diverse mouse populations

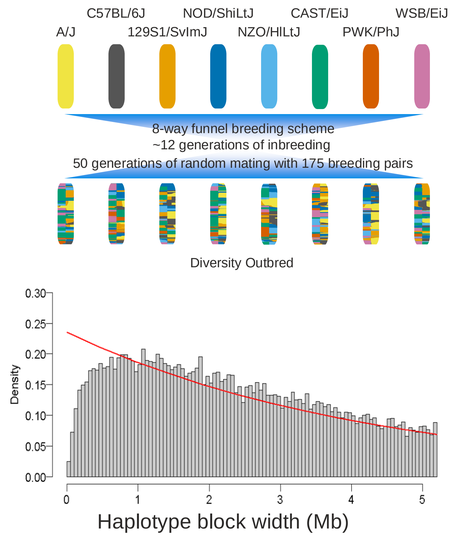
Samuel J. Widmayer[1], Sabina Gude[2], Lydia K. Wooldridge[1], Emily Swanzey[1], Michael C. Saul[1], Mary Barter[1], Qingchang Meng[1], Laura Reinholdt[1], Beth L. Dumont[1], Daniel M. Gatti[1]

[1]The Jackson Laboratory, Bar Harbor, Maine 04609 USA
[2]seqWell, Beverly, Massachusetts 01915 USA

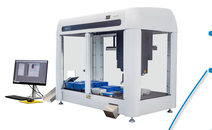
The Jackson Laboratory
Leading the search for tomorrow's cures

MOTIVATION

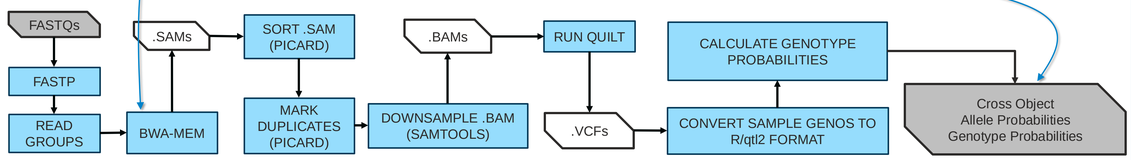
- The Diversity Outbred (DO) mouse population is a premier resource for systems genetics and quantitative trait locus (QTL) mapping.
- Eight founder strains were crossed in a funnel breeding scheme to produce recombinant inbred strains, followed by random mating, resulting in highly recombined outbred mice.
- Each DO mouse is unique and must be genotyped independently, usually with the Giga Mouse Universal Genotyping Array (GigaMUGA).
- Genotyping hundreds of samples on the GigaMUGA can be cost prohibitive; the cost per sample is roughly $100**.**
- With increasing random mating, recombination reduces the size of founder haplotype blocks, increasing mapping resolution.
- Compared to the fit of an exponential distribution (red), there is a depletion of blocks less than 1 Mb in generation 36 of the DO using the GigaMUGA**.**
- **To address these limitations, we have developed a low-coverage genotyping-by-sequencing (lcGBS) approach to genotyping DO individuals that is accurate and cost-effective.**



METHODOLOGY



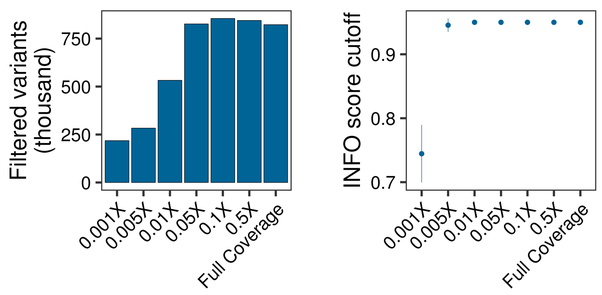
- We prepared Illumina sequencing libraries from 48 DO mice using the seqWell purePlex DNA Library Prep Kit (**Booth 684**) automated on a Revvity Sciclone G3 NGSx (*left*).
- We aligned Illumina short reads to the mouse reference genome build GRCm39.
- Reference haplotypes were constructed using 48 million biallelic SNPs segregating in the DO founders curated by the Sanger Mouse Genomes Project.
- The lcGBS pipeline uses QUILT to impute genotypes using the founder reference haplotypes.
- Imputed sample genotypes were used to reconstruct individual haplotypes with respect to founder alleles (“haplotype reconstructions”) using R/qtl2.



SEQUENCING METRICS

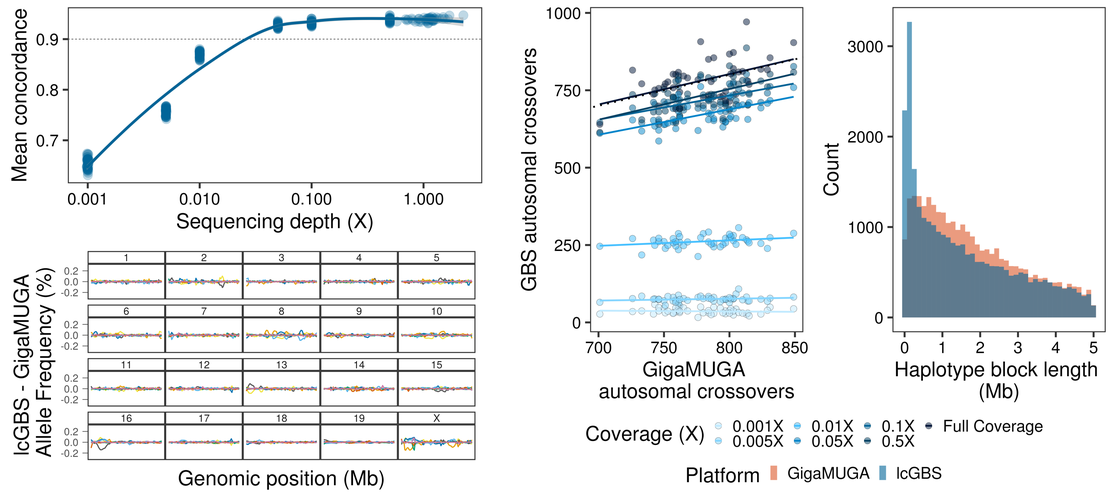
Pre-imputation → Filter to 1M Grid, HWE, INFO > 0.95 → Post-imputation

Mapped HQ reads (million)	19.0 ± 4.77
Sequencing depth	1.14X ± 0.29X
Genome coverage (>Q20)	0.79X ± 0.21X
Coverage at captured sites	1.91X ± 0.17X
Captured variants (million)	1.57 ± 0.38

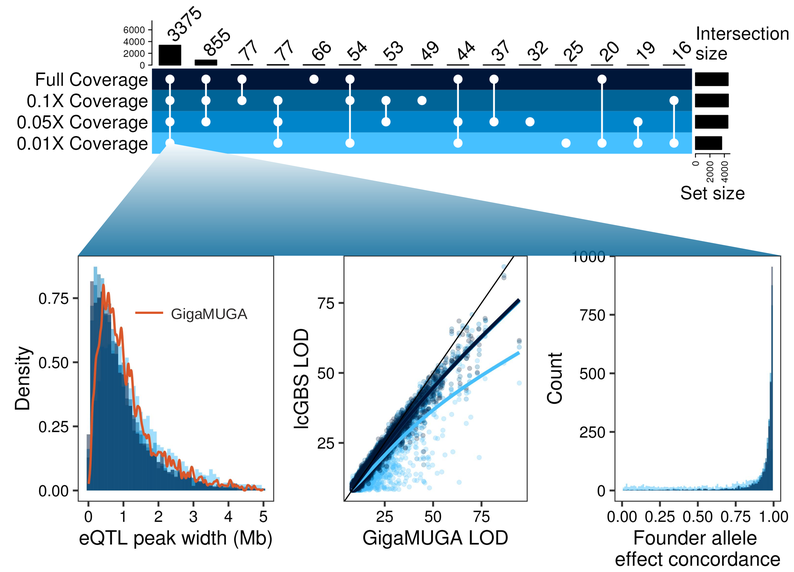


ACCURATE HAPLOTYPE RECONSTRUCTION AT A COMPETITIVE PRICE POINT

- GigaMUGA and lcGBS and haplotype reconstructions were 94% concordant at full coverage, and 93% concordant at 0.1X coverage.
- Crossover counts did not differ significantly between samples genotyped with full coverage lcGBS or GigaMUGA.
- Haplotype blocks shorter than 1 Mb were significantly more abundant in lcGBS samples than in GigaMUGA.
- Founder strain allele frequencies were faithfully preserved over most of the genome.



POWERFUL AND PRECISE SYSTEMS GENETICS



- We genotyped 183 embryoid bodies derived from DO embryonic stem cell lines using lcGBS and GigaMUGA.
- We mapped expression QTL using haplotype reconstructions from both methodologies.
- 3,375 local eQTL closest to their regulated gene were detected in GigaMUGA and lcGBS at all tested coverage levels.
- Low-coverage GBS produced mappings with greater resolution, similar power, and unbiased founder allele effects.
- **We recommend lcGBS as a viable replacement for genotyping DO mice for scaled systems genetics.**

ACKNOWLEDGEMENTS

We kindly acknowledge the JAX Genome Technologies Laboratory for their services and constant support. We also thank the JAX Center for Systems Neurogenetics and Addiction for sourcing the crossover data shown in the Motivation section.

Funding: JAX Scientific Services Innovation Fund

www.jax.org