The Center for Genome Dynamics Research Overview

Randy Von Smith, Hyuna Yang, Petko M. Petkov, Timothy A. Bell,, Fernando Pardo-Manuel de Villena, Y. Zhang, R. Kirova, M. A. Langston, E. J. Chesler, Renhua Li, Keith Shockley, David Witmer, Beverly Paigen, Gudrun A. Brockmann, Christina Neuschl, Gary A. Churchill

The Jackson Laboratory , Bar Harbor, Maine 04609 USA

Abstract

The Center for Genome Dynamics mission is to evaluate the role of genome-wide organization in mammalian biology by developing detailed maps of interactions that encompass allelic diversity, functional categories, gene expression, recombination hotspots, and phenotype associations. We believe that understanding these principles will provide new genetic approaches to understanding human health and disease at the systems level. This presentation will highlight recent progress at the Center in the areas of gene expression response to high fat diet, structural model analysis for adiposity in mice, analysis of the genetic structure of mouse populations by the combinatorial analysis of long-range linkage disequilibrium, the use of linear model genome scans for expression QTL analysis, distribution and control of recombination activities in the mouse, and the subspecific origin of the laboratory mouse.

Center for Genome Dynamics http://cgd.jax.org/ NIGMS: GM076468 National Centers for Systems Biology http://www.nigms.nih.gov/Initiatives/SysBio/

On the subspecific origin of the laboratory mouse

Hyuna Yang, Timothy Bell, Gary A. Churchill, Fernando Pardo-Manuel de Villena

The genome of the laboratory mouse is thought to be a mosaic of regions with distinct subspecific origins. We have developed a high-resolution map of the origin of the laboratory mouse by generating 25,400 phylogenetic trees in 100 kb intervals spanning the genome. On average 92% of the genome is of *M. m. domesticus* origin and the distribution of diversity is strikingly non random among the chromosomes. There are large regions of extremely low diversity, representing blind spots for studies of natural variation and complex traits, as well as hot spots of diversity. In contrast with the mosaic model we found that the majority of the genome has intermediate levels of variation of intrasubspecific origin. Finally, the wild-derived mouse strains that are supposed to represent different mouse subspecies show substantial intersubspecific introgression. This has serious implications for evolutionary studies that assume these are pure representatives of a given subspecies.

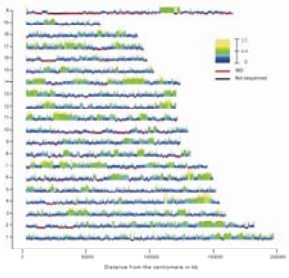


Figure 1. Frequency and spatial distributions of the mean normalized genetic variation observed among 11 resequenced strains. Spatial distribution of the mean normalized variation in the 11 resequenced classical strains is shown as vertical bars of different color and height for each 100 kb interval.

Genetics of recombination

Petko M. Petkov

In mammals, recombination occurs in highly localized regions, usually 1-2 kb in size, termed hotspots. Hotspots can be positioned relatively close to each other or separated by long distances, up to several hundred kilobases.
Our lab investigates the distribution, location, sex specificity of hotspots on entire chromosomes and the influence of genetic background on their activity. We have set up four genetic crosses for large-scale mapping of recombinational activity, with 3000 offspring of female and 3000 offspring of male meioses. The maps show substantial differences in crossover rates along the chromosome, with alternating regions of high and low recombination rates. Recombination is distributed very unevenly between hotspots of different activity. Highly active hotspots, located 1.5 Mb apart on average, account for more than 50% of total recombination. All these features make recombination maps very granular. Generally, female recombination rates are higher than in males near centromeres and in the middle section of the chromosome, whereas males show elevated recombination near telomeres. This sex specificity is mostly due to different activity of the same hotspots rather than to presence of different hotspot sets (Fig.1).
Particularly exciting is our work on crossover interference, and its relation to the sex differences in recombination rate. Crossover interference appears to act on a physical level; when the difference in the compaction of chromosomes is taken into account, the distances between crossovers in males and females is nearly identical (Fig.2). We hypothesize that crossover interference, and the difference in the compaction of chromosomes in males and females, is the primary cause of the sex difference in recombination rate. (Females have longer chromosomes at meiosis, and so may exhibit more crossovers.)
We have found evidence for the role of trans-acting genes on the activity of individual hotspots. Experiments for mapping these genes are ongoing, and several exciting avenues for investigation of mechanisms regulating recombination are pursued.

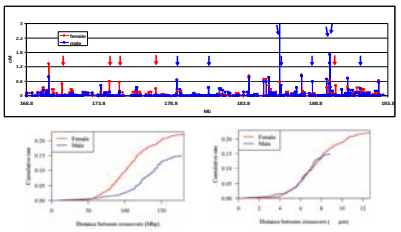


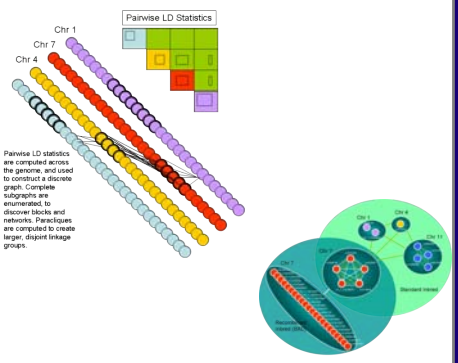
Fig. 1. Interference as a function of intercrossover distances. A. Cumulative rates of double crossovers in female and male meioses, intercrossover distances in megabases of DNA length. B. Cumulative rates of double crossovers in female and male meioses, intercrossover distances expressed in physical length of pachytene bivalents (µm).

A comparison between panels A and B shows that interference in both sexes is the same in terms of bivalent length (µm) rather than genomic length (Mb).

Comparing Mouse Population Architecture Through Combinatorial Analysis of Linkage Disequilibrium Networks.

Y. Zhang, R. Kirova, M. A. Langston, E. J. Chesler

Linkage disequilibrium has been observed lengthwise and pairwise across regions of the mouse genome, and evidence suggests that higher order LD structures exist. The nature and extent of these LD blocks and LD networks is determined both by the relations among individuals (breeding history) and phenomena such as co-adaptive allele selection. Combinatorial algorithms allow the complete enumeration of these networks, their stringency (magnitude of LD) and their localization. The LD network or group is the theoretical limit of precision of QTL mapping, and enumeration and comparison of conserved networks will facilitate development and reduction of complex QTL models, and application of strain sets with high local or distributed LD.



Modeling of systems genetics reveals genetic buffering mechanisms involved in HDL homeostasis

Renhua Li, Gary Churchill

Most common human diseases are genetically complex and may involve gene-gene and gene-environment interactions. Mouse models provide a powerful tool to identify genes involved in complex interactions. Stressful environmental factors such as a high-fat, high-cholesterol diet can induce specific regulatory responses that vary in different strain backgrounds. Modeling the effects of the epistatic interactions on genome wide mRNA expression profiles and a clinic trait (HDL) in a mouse intercross population (C57BL6/J x DBA/2J), we are able to identify candidate regulatory and target genes involved in the genetic buffering mechanisms that play an important role in the homeostasis of HDL. The genetic system is characterized by multiple epistatic interactions involving gene expression traits, and balanced allelic effects in different physiological pathways. Epistatic interactions affect HDL phenotype through specific transcripts that help to bridge the gap between genotypes and phenotypes. Understanding of the genes involved in this buffering system may provide new therapeutic targets for drug discovery.

Figure 1. System-level modeling

We fit a model (left) with genetic variation at QTLs, transcription abundances of corresponding candidate genes identified, and HDL cholesterol, using SEM (Li et al. 2006). This model points to a third mechanism in the homeostasis of plasma HDL response to a high-fat diet; that is, balance of multiple pathways with opposite directions. The other two mechanisms include epistatic buffering and allele-specific coupling of transcripts (data not shown here). This model provides a plausible mechanism for HDL regulation that can be tested experimentally. Interestingly, four out of the selected candidate genes are documented in the gene network from Ingenuity Database (right). The four genes are involved in lipid metabolism that is connected to cell cycle and apoptosis pathways.

Gene Expression Response to High-Fat Diet

David Witmer, Beverly Paigen, Gary Churchill, Keith Shockley

The Center for Genome Dynamics created an outreach program to Maine's elite public science and math boarding school: the Maine School of Science and Mathematics (MSSM: www.mssm.org). As part of this initiative, we studied the gene expression response to dietary fat. Our experiment consisted of a microarray data set containing males and females from 10 inbred mouse strains fed both high- and low-fat diets. An overall F-test for differences between all 40 treatment groups revealed that 63% out of 45,037 transcripts assayed showed at least one expression difference (FDR < 1%). A k-means clustering analysis uncovered four prevailing patterns associated with these changes corresponding to two sex-specific effects and two diet-specific effects. Processes related to the immune system were strongly perturbed by diet, while metabolic processes were associated with sex. A full set of pairwise contrasts comprising 90 independent statistical tests showed that strain-specific responses to high-fat diet differed between sexes. Understanding the metabolic mechanisms by which mice develop diet-induced obesity could lead to a better understanding of obesity in humans.

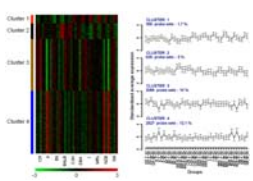


Figure 1. Profiles of each of the 5 clusters generated from by bootstrapped k-means clustering. RMA expression measures were averaged across each group, standardized, and analyzed using a bootstrapping k-means cluster analysis. The heatmap shows standardized average expression in color across the 40 groups along the horizontal axis.

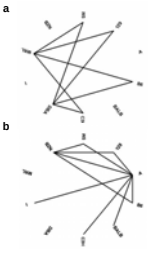


Figure 2. Circle graphs showing sex-specific differences in how strains respond to dietary fat. In (a), connected strains have few differences in females and many differences in males. In (b), connected strains have few differences in males and many differences in females.

Structural model analysis for adiposity in mice

Gudrun A. Brockmann, Renhua Li, Christina Neuschl, Gary A. Churchill

As a polygenic model for obesity, we are using the mouse line NMRI8 which has been selected for high body weight at 8 weeks. Recently, we have mapped QTLs for body weight, adiposity, and muscle weights at 6 weeks in an intercross population between the mouse lines NMRI8 and DBA/2. Using the technique of structural model analysis, we considered body weight as lean mass and fat mass and were able to distinguish genetic loci that affect adiposity from those that affect lean mass. The results show that the NMRI8 alleles of the selection line on chromosomes 7 and 14 have pleiotropic positive effects on both muscle and fat tissue mass, while a locus on chromosome 13 contributed to the selection response only by increased fat deposition. The fat mass was also affected by a complex pattern of interaction between loci on chromosomes 6 and 14. The sex affected the fat mass either directly or indirectly via the muscle as a mediator. The analysis sheds new light on the action of genes controlling body weight as composite trait of fat and muscle tissues.

Structural Equation Model (SEM)

SEM is a descriptive and inferential tool to investigate the simultaneous effects of QTLs on multiple phenotypes and interactions among these phenotypes. The model structure is represented as directed graphs between measured variables. SEM are an extension of multiple regression techniques which emphasizes the correlation structure of continuously distributed variables. Any given variable may be both a response and a predictor.

The development of a Structural Equation Model is performed in five steps:

1. Identification of QTLs for individual phenotypes,
2. Identification of pleiotropic QTLs,
3. Definition of an initial path model,
4. Assessment of the model,
5. Refinement of the model.

The final Structural Equation Model shows that adiposity is influenced by loci that either affect fat deposition alone or have pleiotropic effects on fat tissue and muscle development (see Figure).

Figure. Graphical representation of the final SEM for the direct and indirect actions and interactions of different variables contributing to high adiposity in the NMRI8xDBA/2 intercross population. The QTL positions are given as chromosome number and cM position separated by @. The x indicates epistatic interaction between two genetic loci found by a screen for pair-wise interactions in the F_2 population. The **arrows** indicate paths between variables. The direction of effects is given by the **edges**; the **values** represent the path coefficients. These are standardized values which represent relative effect sizes. **Positive values** indicate increasing and **negative values** reducing effects of the NMRI8 allele compared to the DBA/2 allele effect in the F2 population. **E1** and **E2** denote residual effects.