Heritability and coefficient of genetic variation analyses of phenotypic traits provide strong basis for high-resolution QTL mapping in the Collaborative Cross mouse genetic reference population

Fuad A. Iraqi · Hanifa Athamni · Alexandra Dorman · Yasser Salymah · Ian Tomlinson · Aysar Nashif · Ariel Shusterman · Ervin Weiss · Yael Houri-Haddad · Richard Mott · Morris Soller

Abstract Most biological traits of human importance are complex in nature; their manifestation controlled by the cumulative effect of many genetic factors interacting with one another and with the individual's life history. Because of this, mouse genetic reference populations (GRPs) consisting of collections of inbred lines or recombinant inbred lines (RIL) derived from crosses between inbred lines are of particular value in analysis of complex traits, since massive amounts of data can be accumulated on the individual lines. However, existing mouse GRPs are derived from inbred lines that share a common history, resulting in limited genetic diversity, and reduced mapping precision due to long-range gametic disequilibrium. To overcome these limitations, the Collaborative Cross (CC) a genetically highly diverse collection of mouse RIL was established. The CC, now in advanced stages of development, will eventually consist of about 500 RIL derived from reciprocal crosses of eight divergent founder strains of mice, including three wild subspecies. Previous studies have shown that the CC indeed contains enormous diversity at the DNA level, that founder haplotypes are inherited in expected frequency, and that long-range gametic disequilibrium is not present. We here present data, primarily from our own laboratory, documenting extensive genetic variation among CC lines as expressed in broad-sense heritability (H^2) and by the well-known "coefficient of genetic variation," demonstrating the ability of the CC resource to provide unprecedented mapping precision leading to identification of strong candidate genes.

Background

Human populations display wide variation in their susceptibility to and manifestations of infectious, metabolic, or psychiatric disease. This variation has been shown by biometrical analyses to have a genetic component of greater or lesser magnitude, depending on the disease. Generally, susceptibility traits belong to the class of "complex" (also termed "quantitative") traits. That is, their specific manifestation in an individual is controlled by the cumulative effect of many genetic factors, interacting with one another and with the life history of the individual. Because of this complex etiology, it has proven exceedingly difficult in humans to identify the individual genetic elements ("QTG," taken in the broadest sense to include functional RNA transcripts, up- and down-stream regulatory sites, enhancers, and so on) or sequence variants within the QTG ("QTN," taken in the broadest sense to include all sequence variants: SNPs, indels, CNVs, and larger chromosomal re-arrangements) contributing to genetic variation in susceptibility to specific diseases. This has prompted the development of mouse genetic resources for genetic analysis of complex traits. Humans and mice

F. A. Iraqi (✉) · H. Athamni · A. Dorman · Y. Salymah
Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, 69978 Tel Aviv, Israel
e-mail: fuadi@post.tau.ac.il

I. Tomlinson · R. Mott
Welcome Trust human Genome center, University of Oxford, Oxford, UK

A. Nashif · A. Shusterman · E. Weiss · Y. Houri-Haddad
Department of Prosthodontics, Dental School at Hadassah, Hebrew University, Jerusalem, Israel

M. Soller
Department of Genetics, Silberman Institute of Life Sciences, Hebrew University, Edmund Safra Campus, Givat Ram, Jerusalem, Israel

share widely in their genome architecture, and in their basic biology and disease susceptibility. Large numbers of inbred mouse lines that vary in their genetic architecture have been developed and are commercially available. These have formed the basis for genetic reference populations (GRPs) based on collections of inbred lines or crosses between them (Roberts et al. 2007; Silver 1995). GRPs are popular for the study of complex traits and biological systems in both medical and life sciences because genotyping is only required once (the "genotype once, phenotype many times" paradigm), replicate individuals can be produced with the same genotype in larger cohorts allowing for optimal case/control and gene-by-environment designs (Broman 2005); data from numerous experiments can be accumulated for the population allowing deep bioinformatic data mining. GRPs can be studied under defined environmental conditions, providing model systems through which the genetic elements responsible for the genetic variation of complex traits, including disease susceptibility can be mapped to defined chromosomal regions [termed "quantitative trait loci, (QTL)"] (see e.g., Groden et al. 1991; Hernandez-Valladares et al. 2004a; b; Houle 1992; Iraqi et al. 2000, 2003). Yet, with some exceptions, even in such model systems, it has not been possible to identify the actual functional QTG and QTN, in part due to the complex genetic architecture of these lines which precludes high-resolution mapping. This has led the mouse genetics community to propose and design a new "next generation" GRP, the Collaborative Cross, "CC" (Threadgill et al. 2002). The CC is a mouse GRP specifically tailored for high-resolution QTL mapping of complex traits, with special emphasis on traits relevant to human health in its broadest aspects. However, the CC goes beyond this to include expanded possibilities for analysis of epistatic interactions among QTG, identification of the biological systems and networks within which the QTG are embedded and through which their QTN exert their effect at the whole organism phenotypic level, and the interaction of these biological systems with the environment as mediated through epigenetic markings (Churchill et al. 2004; Silver 1995).

The CC resource

This unique genetic reference population will eventually comprise a set of approximately 500 recombinant inbred lines (RIL) created from almost full reciprocal matings of eight divergent strains of mice (some of the F1 crosses were not viable). These include five classical inbred lines (A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, and NZO/HiLtJ) and three wild-derived strains: CAST/Ei, derived from *M. m. castaneum* mice trapped in Thailand in 1971, and PWK/

PhJ and WSB/EiJ, derived, respectively, from wild *M. m. musculus* and wild *M. m. domesticus* mice trapped near Prague in 1974 and the Eastern Shore of Maryland in 1976, respectively (Beck et al. 2000).

Currently, over 350 pre-CC lines are in more or less advanced stages of development at three locations: Tel Aviv University, Israel (TAU) (Iraqi et al. 2008); University of North Carolina, USA (UNC) (Chesler et al. 2008); and Geniad Ltd, Western Australia (GND) (Morahan et al. 2008). In addition to TAU, UNC, and GND, The Jackson Laboratory and Oxford University also participated in the initial development of the CC resource lines. To facilitate community access to the CC, a material transfer agreement was executed among all five parties and can be obtained from any of them (Welsh et al. 2012). Genotypes are available at a dedicated Web site: http://csbio.unc.edu/CCstatus with a browser to facilitate visualization and interaction with the genomes of the individual CC lines: http://csbio.unc.edu/CCstatus/?run==CCV/.

Controlled randomization was performed during the breeding process to break up large linkage disequilibrium (LD) blocks and to recombine the natural genetic variation present in these inbred strains with the aim to create a unique resource of RI strains exhibiting a large phenotypic and genetic diversity (Roberts et al. 2007).

The CC population exhibits about fourfold map expansion compared with a single generation cross, increasing accuracy of QTL map location in proportion. Because of their inbred nature, all genetic traits involve homozygotes; thus, increasing genetic variation associated with each QTL (Falconer and Mackay 1996); there may, of course, be exceptions to this rule. In addition, multiple individuals can be phenotyped in each line reducing environmental sources of variation. In this way, the effective mapping power of the set of RILs is increased many-fold relative to standard F2 mapping populations (Valdar et al. 2006). Initially, all CC mice were genotyped with the mouse diversity array (MDA), which contains 620,000 SNP markers (Yang et al. 2009), and their genome reconstruction was presented (Durrant et al. 2011). Recently, all mice were regenotyped at advanced generations with the new 7500 custom-designed mouse universal genotype array (MUGA), which provided the genome architecture of the CC lines (CCC 2012). After six inbreeding generations, the entire CC mouse population was re-genotyped by the 77,000 SNP array of MegaMUGA, and all their genotypes will be available. Figure 1 shows the genomic reconstruction of three CC lines after genotyping with MDA (Durrant et al. 2011) by using Happy software (Mott et al. 2000).

With some exceptions, the founder line haplotypes are distributed more or less equally across the population of lines as a whole, although the distribution of founder genome within individual chromosomes or lines can differ

Fig. 1 Reconstructions of the genomes of representative CC lines. Genomes of CC lines IL-127 (*upper panel*), IL-134 (*middle panel*), and IL-135 (*lower panel*) were reconstructed using a hidden Markov model (HMM) implemented by HAPPY program (Mott et al. 2000). The *X*-axis shows the 19 autosomes. The *Y*-axis shows the eight CC founders, with probability of descent from each founder. Regions attributed with high probability to a single founder appear as dark horizontal bands in the lane corresponding to the founder. Regions where two or more putative founders cannot be distinguished are gray. Regions where a founder is not represented at all are *white*
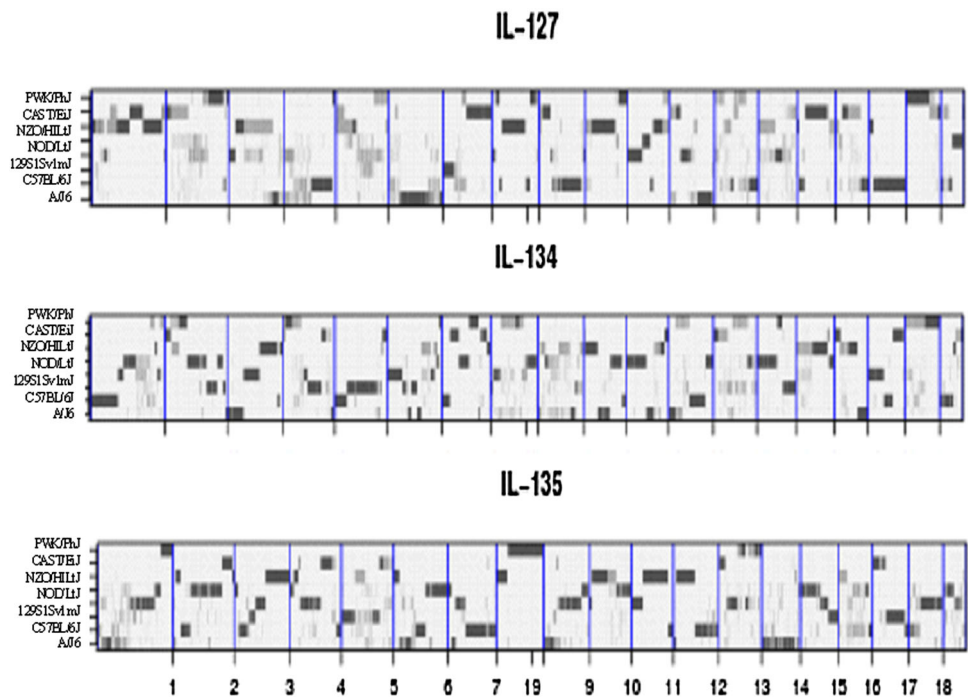
widely from equality. Linkage disequilibrium decayed rapidly with distance as expected for a collection of independent inbred lines, and there were no indications of gametic disequilibrium (i.e., of LD among unlinked markers), which agrees with the report by Broman et al. (2012). Consequently, except for type I error due to sampling, marker–QTG association tests will be significant only between markers and QTN that are closely linked. This is in contrast to the situation found in panels of classical mouse GRP involving collections of inbred strains, where extensive long-range gametic disequilibrium is present due to historical relationships among the lines, and results in a much higher effective type I error than predicted by sampling considerations alone. Thus, the CC resource is well along the way to fulfill the most sanguine hopes of its community.

A recent study characterizing the genome architecture of 350 pre-CC lines (CCC 2012) showed that the wild mice indeed contributed enormous stores of genetic variation that were not found in the standard mouse laboratory strains. Whereas most classical strains differ from the reference C57BL/6J at about 4 million SNPs, PWK/PhJ, and CAST/Ei each differ at about 17 million SNPs, and WSB/EsJ at 6 million (Keane et al. 2011). Among them, the eight founder populations present 36.155 million SNPs! This is many-fold greater than the genetic diversity captured in existing GRPs derived from crosses or panels of standard laboratory strains. Consequently QTL mapping using the CC is likely to uncover novel QTLs involving contrasts between the wild-derived strains. This is exemplified in a pilot experiment in our laboratory in which we fine-mapped eight QTLs associated with post-challenge survival after infection by *Aspergillus fumigatus*. Of these, five QTL involved contrasts with wild-derived strains and would not have been present in a cross between classical strains (Durrant et al. 2011). That study, and others by our collaborators (Aylor et al. 2011; Durrant et al. 2011; Bottomley et al. 2012; Kelada et al. 2012; Kovacs et al. 2011; Mathes et al. 2011; Philip et al. 2011), further showed that by incorporating variation data from the genome sequences of the CC founders—available from the Sanger Mouse Genomes Project (Keane et al. 2011)—and restricting attention to variants whose differences across the founders are consistent with the pattern of action of the QTL (Yalcin et al. 2005), confidence interval of QTL location can be significantly narrowed, and the list of potential candidate genes markedly refined.

In the present report, we present extensive data primarily from our own laboratory, documenting extensive genetic variation among CC lines as expressed by the well-known heritability statistic, representing the proportion of total phenotypic variation that can be attributed to genetic factors, and by the well-known "coefficient of genetic variation (CV_G)," which provides a unit-free measure of the relative magnitude of the genetic variation. In addition, we discuss previous studies from our laboratory and others demonstrating the ability of the CC resource to provide high-resolution mapping of QTL affecting a wide variety of traits, including susceptibility to a spectrum of infectious diseases. Finally, we consider the potential role of the CC as a uniquely powerful resource for systems biology.

Materials and methods

Heritability and genetic coefficient of variation analysis in the CC lines

Heritability estimates were obtained from unpublished data for a wide variety of traits presently being studied at TAU and from the published results of Philip et al. (2011) for selected morphological and behavioral traits recorded for the lines under development at Oak Ridge National Laboratory [(ORNL), later transferred to North Carolina State University]. Traits at ORNL were monitored from the initial crosses of the funnel breeding scheme (G1 generation), through the seventh inbreeding generation (G2:7 generation).

Brief descriptions of the study traits are given in Box 1. For TAU data, broad-sense heritability estimates (H^2) including epistatic but not dominance effects were obtained from a one-way ANOVA with CC lines as the main effect as follows:

$$H^2 = V_g / (V_g + V_e)$$

where

V_e is the environmental component of variance within lines = MS_{within}

V_g is the genetic component of variance among CC lines = $(MS_{between} - V_e)/n$

n is the average number of mice per line

For the ORNL data, narrow-sense heritability (h^2) estimates (not including epistatic effects) were obtained by Philip et al. 2011 from parent–offspring regressions.

The genetic coefficient of variation

The heritability statistic estimates the proportion of observed phenotypic variation that is due to genetic factors. However, it does not tell us whether the absolute amount of genetic variation generated by these genetic factors (the "genetic component of variation") is great or small. A

Box 1 Data sets from TAU and ORNL for calculation of heritability and coefficient of genetic variation

Susceptibility to *Klebsiella pneumoniae* infection (KLEB). *Klebsiella pneumoniae* (Kp) is a common pulmonary pathogen causing severe pneumonia in immunocompromised hosts and often associated with sepsis (Keynan and Rubinstein 2007). Here, we initiated a study aimed at mapping and subsequently identifying the host susceptibility genes to Kp infection following intraperitoneal challenge. Survival and body weight were monitored for 15 days post-challenge.

Susceptibility to periodontitis (PER). Periodontal infection (Periodontitis) is the most common chronic inflammatory disease in humans, which results in destruction of tooth-supporting tissues and eventually leading to tooth loss. This process is characterized by destruction of the periodontal ligament, formation of periodontal pockets, and alveolar bone resorption (Baker et al. 2000; Baker and Roopenian 2002; Wilensky et al. 2005). We use the oral mixed infection system of the two anaerobic gram-negative bacteria *Porphyromonas gingivalis* and *Fusobacterium Nucleatum* (Polak et al. 2009; Shusterman et al. 2013a, b) (40, 43, 44) to examine variation in susceptibility to periodontitis. The phenotype is measured as initial alveolar bone volume prior to infection, residual bone value after infection, and bone volume loss, in mm³ (49).

Modifiers of *Apc* gene (APC). *Familial adenomatous polyposis* (FAP) is a rare dominant genetic disease, which is characterized by multiple intestinal polyps (Fearnhead et al. 2001; Fostira et al. 2010). It is caused by a functional mutation in the *adenomatous polyposis coli* (*Apc*) gene. In order to investigate the syndrome in vivo, the $Apc^{Min+/-}$ mouse was developed, modeling the human polyp phenomenon (Moser et al. 1992, 1993; Cormier and Dove 2000; Cormier et al. 2000) (13, 14, 34, 35). We generated an (Apc^{Min-} x CC) F1 mouse population, by crossing male mice of C57Bl/6J carrying $Apc^{Min+/-}$ with females of advanced generations CC lines. After 5 months, mice were terminated, and the number and sizes of polyps in the small intestines and colon were recorded.

Susceptibility to type 2 diabetes (DEXA T2D). The worldwide prevalence of type 2 diabetes (T2D), in the past decades, has reached epidemic proportions, (150 million in the year 2000) expected to double by 2025 (Buse et al. 2007). Diabetes can be considered as a disease, where the challenge is the Western diet. In this study, 8 week-old mice per line were maintained on high-fat diet (42 % fat) for a period of 3 months, and subsequently assessed for glucose level by intraperitoneal glucose tolerance test (IPGTT), and for body composition and body mass index (BMI) index by dual-energy X-ray absorptiometry (DEXA) scan (Broman 2005).

Normal body composition by DEXA scan analysis (DEXA normal). This study aims to determine the genetic factors underlying body composition as obtained through the DEXA scan technique (Brommage 2003). The following phenotypes (among others) were recorded: body length (BL) and weight (BW), bone mineral density (BMD), and body and liver fat percentages (%). The body mass index (BMI = $BW/(BL^2)$) was also obtained.

Immune response cell lineages in peripheral blood (Immuno). Mapping quantitative trait loci (QTLs) effecting the immune response cells lineages in the peripheral blood and subsequently identifying the genes underlying these QTL can lead to better understanding of the host response to these different infectious diseases (Petkova et al. 2008). Here, we quantified the relative proportions of peripheral blood T, B, and macrophage cells of 10 week-old mature mice (3–5 mice per line at 10 weeks), by immunophenotyping using flow cytometry analysis and fluorescence-activated cell staining (FACS) technology.

Studied at ORNL (39) Table 2 of Philip et al. 2011 present estimates of h^2 in generation G2:7 with accompanying data from which CV_A can be estimated for selected morphological and behavioral traits, among them: body weight, heart weight, tail length, open-field test (average distance from the center of the field, and total distance traveled, in first 3 min of exposure to the open field), and hot-plate latency test (time to respond to being placed on a hot metal surface).

Fig. 2 Variation among CC lines: modifiers of familial adenomatous polyposis (*Apc*) gene. Mean polyp count in F1 cross of CC line x Apc^{Min-}. The *X*-axis represents the 15 tested CC lines (2–4 mice per line) while the *Y*-axis represents the mean polyps count (with standard error) in small intestines and colon at age 5 months. Mean polyp count in control C57BL/6J ($APC^{min+/-}$) mouse is also presented
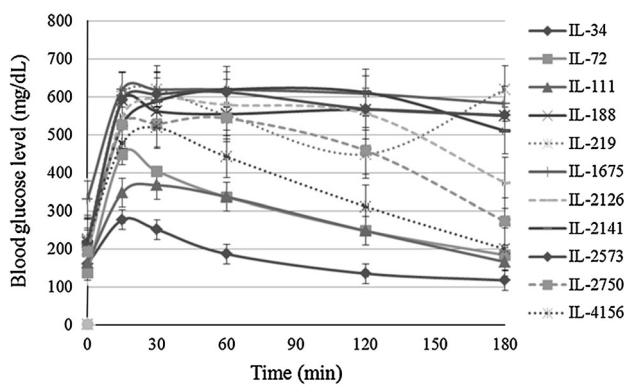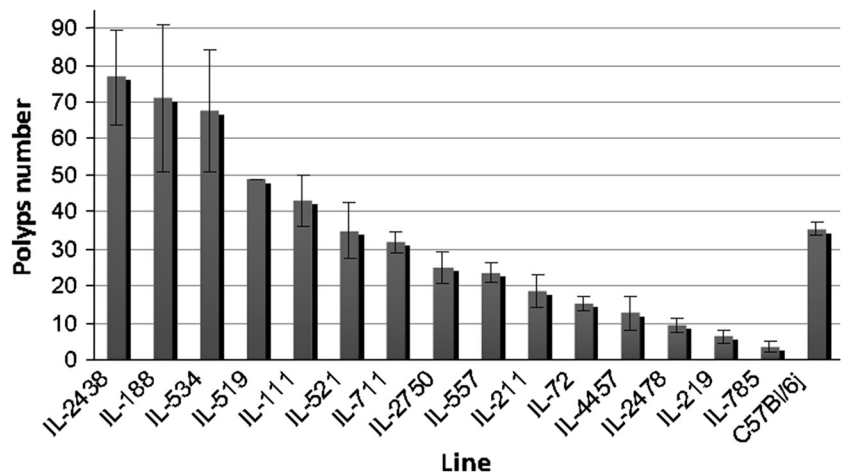

Fig. 3 Variation among CC lines: intraperitoneal glucose tolerance test (IPGTT). Blood glucose levels during an IPGTT, performed on 11 CC mouse lines (4–6 mice per line), maintained on high-fat diet for a period of 12 weeks. The *X*-axis represents the time points (min) used for testing blood glucose levels (mg/dL). Blood glucose levels (with standard error) are represented on the *Y*-axis

high-heritability value is compatible with very little absolute genetic variation (if total phenotypic variation is also very low and mostly due to genetic factors), while a low heritability value is compatible with a large genetic component of variation (if phenotypic variation is very large). The absolute value of the genetic variation is readily obtained as the genetic standard deviation ($Vg^{0.5}$). This value, however, depends on the measurement unit of the trait and is not meaningful for comparison among traits. The coefficient of variation (ratio of standard deviation to the mean) is a commonly accepted unit-free measure of dispersion. We used the well-known evolvability parameter, the ratio of the genetic standard deviation to the mean, (also termed the genetic coefficient of variation, CV_G), as the comparable measure for unit-free evaluation of genetic dispersion (Garcia-Gonzalez et al. 2012; Houle 1992).

It is of interest to give a benchmark for judging whether CV_G values are large or small. As a rough estimate, within any given outcrossing population, quantitative traits generally have a phenotypic coefficient of variation equal to about 10 % of the mean, i.e., SD/mean = 0.1. If we consider a trait with heritability 0.50, the genetic variance will be 0.5 of the phenotypic variance, and the genetic standard deviation will be the square root of this, or 0.71 SD. Thus, the CV_G will be about 0. 071. Thus, comparison of CV_G among the CC lines with this benchmark will tell us to how the genetic variation among the CC lines compares to that found for many quantitative traits, within a typical outcrossing population.

For TAU data, CV_G was estimated as

$$SD_G/Mean$$

where

SD_G = the broad-sense genetic standard deviation among CC lines = $V_G^{0.5}$
Mean = mean trait value across all CC lines

For ORNL data that provided information on narrow-sense heritability, SD_A was calculated from values for mean, h^2, and SD presented in Table 2 of Philip et al. 2011.

Results and discussion

Here, we present selected phenotypic profiles of traits that were characterized on TAU CC lines. Figure 2 shows total polyps variation among different (Apc^{Min-} x CC) F1 mouse lines. Figure 3 shows variation of blood glucose level during intraperitoneal glucose tolerance test (IPGTT) assessments of different CC lines. Figures 4 and 5 show distribution of body lengths and body mass index (BMI) of 20 week-old mice of different CC lines, respectively.

Fig. 4 Body length of
20 week-old mice of CC lines.
Body length was defined the
distance between nose and anus.
The x-axis represents the 15
tested CC lines while y-axis
represents the mean body length
(cM) with standard error
included. On average 3–6 mice
per line were assessed. Mean
body lengths of the different
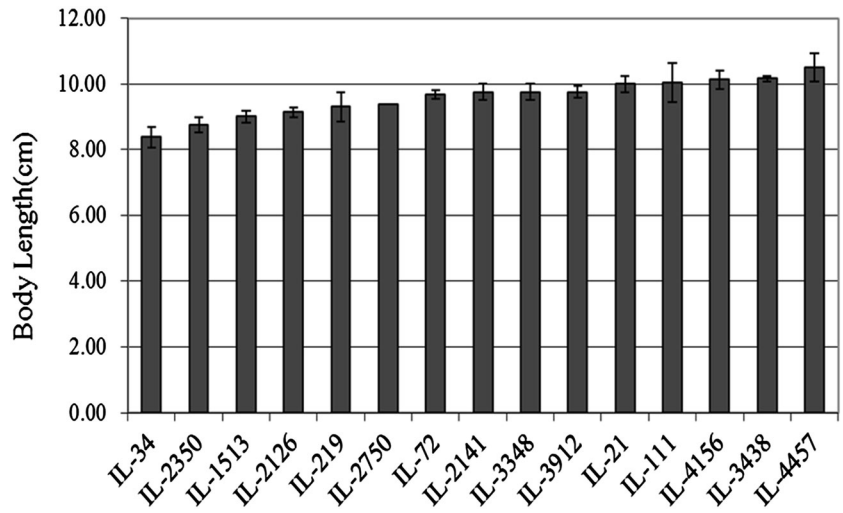tested CC lines were ranged
between 8.2 and 10.3 cM

Fig. 5 Body mass index (BMI)
of CC lines. BMI was defined as
body weight divided by the
square of body length (BW/
BL^2), at 20 weeks of age. The
X-axis represents the 15 tested
CC lines (3–6 mice per line),
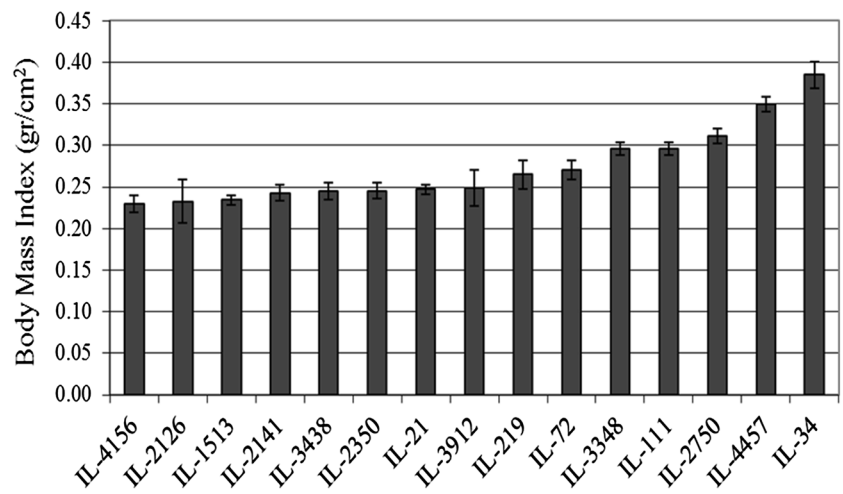while the Y-axis represents the
BMI (with standard error)

Fig. 6 Variation among CC
lines: percentage of T, B, and
macrophage immune cells. The
X-axis represents 15 naïve CC
mouse lines (5–7 mice per line),
while the Y-axis represents the
percentage (with standard
errors) of the three tested
immune cells in peripheral
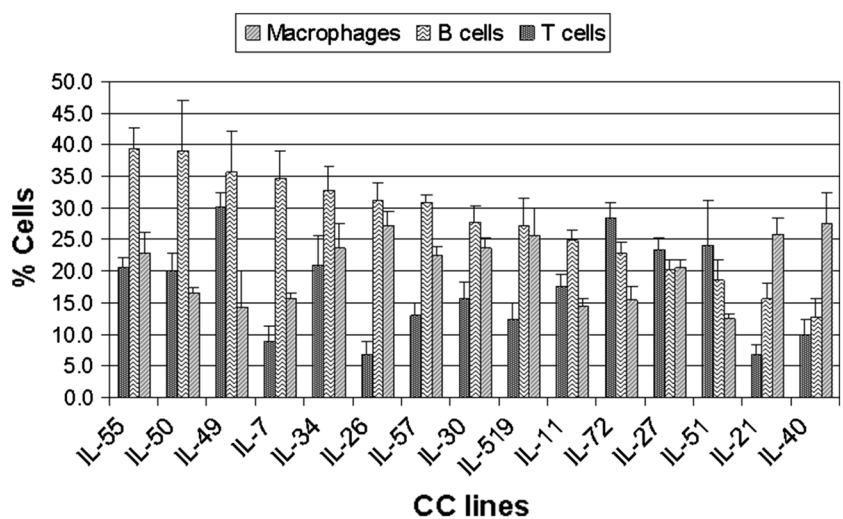blood at age of 8–9 weeks

Table 1 Collaborative Cross, genetic variation among lines

Experiment	Trait	H^2, h^2	Mean	CV_G
PER				
(23, 139)	Initial bone volume (mm^3)	0.40	3.09	0.26
(23, 133)	Residual bone volume (mm^3)	0.40	1.86	0.24
(23,117)	Bone loss (mm^3)	0.20	NA	NA
APC				
(31, 234)	Total polyp lengths (mm)	0.34	28.45	0.53
KLEB				
(60, 345)	Survival time (d)	0.17	4.25	0.33
Immuno				
(41, 355)	T cells	0.34	17.41	0.35
(41, 355)	B cells	0.30	26.53	0.26
(41, 355)	Macrophages	0.16	19.57	0.21
DEXA naive				
(38, 394)	Body weight (g)	0.46	26.49	0.15
(34, 274)	Liver weight (g)	0.40	1.75	0.24
(34, 269)	Spleen weight (g)	0.47	1.59	0.13
(34, 274)	% fat, liver	0.10	1.30	0.22
(38, 394)	% fat, total body	0.44	20.99	0.22
(42, 562)	Body length	0.27	10.25	0.05
(42, 562)	Body mass index	0.30	0.26	0.11
(20, 147)	IPGTT: blood glucose, T0	0.37	170.88	0.15
(20, 147)	BG 30	0.28	476.10	0.14
(10, 147)	BG180	0.25	220.68	0.27
(38, 394)	BMD (g/cm2)	0.00	0.061	0.00
DEXA T2D				
(44, 125)	% fat, liver	0.24	56.90	0.14
(44, 125)	% fat, total body	0.51	32.85	0.24
(44, 125)	Body weight (g)	0.47	29.62	0.18
(43, 111)	Liver (g)	0.43	1.66	0.28
(44, 122)	Spleen (g)	0.27	2.13	0.30
(53, 397)	Body mass index	0.42	0.28	0.17
(53, 397)	BL	0.39	10.64	0.05
(44, 125)	BMD	0.00	0.072	0.00
(40, 81)	BG0	0.16	195.3	0.16
(37, 77)	BG30	0.33	523.6	0.15
(38, 76)	BG180	0.35	362.1	0.30
ORNL				
(NA/181)	Body weight (g)	0.59	24.33	0.21
(NA/141)	Tail length (cm)	0.72	5.28	0.12
(NA/141)	Heart weight (g)	0.66	0.16	0.30
(NA/70)	Hot plate (log s)	0.37	1.62	0.13
(NA/56)	Open field, distance from center (cm)	0.70	5.28	0.67
(NA/56)	Open field, distance traveled 3 min	0.89	1,399.12	0.54

Heritability[1] and genetic coefficient of variation (CV_G) according to trait and experiment. In parentheses under column "Experiment," number of lines, number of mice

[1] TAU studies: broad-sense heritability (H^2); Philips, narrow-sense heritability (h^2)

Box 2 QTL mapping experiments

AYL: Aylor et al. 2011. They used 184 partially inbred CC lines to map QTL affecting body weight and eQTL affecting gene expression. They used analysis based on inferred haplotypes to improve power, increase QTL map resolution, and identify promising candidate genes. The number of uncovered expression QTLs was greater than in all previous efforts at eQTL mapping in mice, and cis-eQTL were mapped at 1-Mb resolution. They demonstrated that the genetic diversity of the CC, which derives from random mixing of eight founder strains, enhanced the ability to map causative loci underlying complex disease-related traits.

DUR: Durrant et al. 2011. They used 371 mice from 66 pre-CC lines to map QTL affecting survival time after challenge by *Aspergillus fumigatus* infection and identified 7 QTL. Most of the QTL involved contrasts between wild-derived founder strains and therefore would not segregate in classical inbred strains. Use of variation data from the genomes of the CC founder strains further refined these QTL.

KEL: Kelada et al. 2012. They used 131 mice from 131 pre-CC lines to identify genetic determinants of hematological parameters, including RBC volume, WBC count, percent neutrophils/lymphocytes, and monocyte number. They used founder line haplotypes to further refine QTL confidence intervals and identified a small number of promising candidate genes

MAT: Mathes et al. 2011. They used 176 mice from 176 pre-CC lines to map QTL for 15 metabolism- and exercise-related phenotypes.

PHI: Philip et al. 2011. (see Box 1)

BOT: Bottomley et al. 2012

They used 44 mice from 44 influenza-infected pre-CC lines determined to have extreme phenotypes with regard to the host response to influenza A virus infection. Global transcriptome profiling identified 2671 transcripts that were differentially expressed between mice that showed a severe ("high") and mild ("low") response to infection, and identified 21 eQTL, which allowed direct examination of genes associated with regulation of host response to Influenza A.

Figure 6 describes percentage of T, B, and macrophage immune cells in different naïve CC mouse lines.

Table 1 presents heritability and genetic coefficient of variation values for a variety of traits studied at TAU and obtained from ORNL data. For the TAU traits, H^2 values are generally in the range of 0.10–0.40; for ORNL, h^2 values were higher, in the range of 0.37–0.89. In both data sets, CV_G or CV_A was in the range 0.15–0.60, much higher than the benchmark of 0.071. Thus, these data show that the absolute magnitude of genetic variation among the CC lines is much greater than found within typical outcrossing populations. This can be attributed to the inclusion of the three wild subspecies among the founder parents. It also suggests that QTL with very large effects may be segregating in the CC lines. This is supported by the initial QTL mapping studies that were conducted using CC lines in various stages of development ("pre-CC lines"). The

Table 2 Collaborative Cross, QTL mapping

Exp	Trait	Des.	Location, CI	Effect	Reduced	CG
AYL 184,184	Body weight	Bwq14	**4** (3.03) *7.3*	0.186	(6.01) 4.61	Asph
	eQTL[a]	7235	0.92	ND	ND	
DUR 66,371	Survival time, *Aspergillus fumigatus* challenge	Asprl1	**8** (42.3) *16.7*[b]	0.162	NS	Irf2
		Asprl2	**10** (93.7) *5.9*	0.154	NS	
		Asprl3	**15** (28.3) *9.2*	0.153	NS	Laptm4b;
		Asprl4	**10** (10.0) *14.6*	0.148	NS	
		Asprl5	**18** (4.4) *10.16*	0.130	NS	
		Asprl6	**3** (108.5) *17.01*	0.125	NS	
		Asprl7	**2** (5.5) *8.8*	0.122	NS	
KEL 131, 131	Red cell volume	Mcvq4	**17** (110.3) *5.9*	ND	0.127	Hbb
	Red cell volume	Mcvq5	**14** (8.9) *14.0*	ND	2.45	Anxa7
	White cell count	Wbcq7	**18** (70.0) *4.1*	ND	NS	Idd212
	% Neutrophils	Ne_pct1	**11** (31.3) *3.8*	ND	NS	oc2
	Monocyte number	Moq1	**1** (87.6) *7.4*	ND	0.167	Lrrfip1
MAT 176, 176	Body weight (BW)	None	**4** (3.3) *7.0*	ND	NS	
	Avg. running distance days 11/12)	None	**16** (32.5) *5.8*	ND	NS	
	Food intake during running phase	None	**12** (82.9) *6.2*	ND	NS	
	BW change after running phase	None	**6** (72.7) *10.7*	ND	NS	
		None	**6** (39.4) *8.7*	ND	NS	
PHI 163–293, 326–586	Body length	None	**2** (152.9) *5.6*	ND	ND	
	Avg. % sleep at night	None	**7** (90.9) *6.0*	ND	ND	
	Hot-plate latency	None	**6** (45.5) *5.5*	ND	ND	
	Red cell width	None	**7** (105.5) *7.0*	ND	ND	
	Peak activity after sleep deprivation	None	**9** (29.7) *0.5*	ND	ND	
	Open field, locomotion,	None	**4** (21.1) *9.2*	ND	ND	
	Open field, distance from center	None	**6** (89.6) *3.6*	ND	ND	
	Periosteal circum.	None	**19** (21.4) *1.1*	ND	ND	
FER 155,155	Host response to influenza: BW Day 4 post-challenge	HrI1	**16** (97.5) *0.7*	0.417	NS	Mx1
		Hrl2	**7** (89.1) *7.6*	0.097	NS	Il16,
	Pulmonary edema	Hrl3	**1** (21.7) *7.3*	0.297	1.01	
	Airway neutrophils	Hrl4	**15** (77.4) *9.2*	0.228	0.091	Grap2
BOT 99,99	eQTL affecting host response to influenza by gene under eQTL control. H, High response; L, Low response. All eQTL were cis.					
	LOC675467 H	None	**14** (20.2) *0.4*	ND	ND	
	Ifi2712a H	None	**12** (107.0) *0.7*	ND	ND	
	Kcmf1 H	None	**6** (68.5) *1.3*	ND	ND	
	Thns12 L	None	**6** (68.0) *2.7*	ND	ND	
	AK153595 L	None	**17** (5.2) *2.1*	ND	ND	
	BC022687 L	None	**12** (112,8) *1.6*	ND	ND	
	Sh3g13 L	None	**7** (87.8) *3.2*	ND	ND	
	Clec16a L	None	**16** (6.2) *4.5*	ND	ND	

Exp. Experiment acronym, (in parentheses no. of lines, no. of mice); Des. QTL designation); Location: in bold, MMU; in parentheses, lower bound of 95 % CI of QTL location; in italics, 95 % CI in Mb. Effect, proportion of observed phenotypic variation explained; Reduced, reduced CI using founder allele effect patterns

CG strong candidate gene, *ND* not done, *NS* not successful

[a] 7235 eQTL were mapped, with median eQTL–mid-gene distance 0.91 Mb

[b] Median distance true QTL location and peak Log P < 1 Mb

experiments are listed with brief descriptions in Box 2. These QTL mapping studies (Table 2) were remarkably successful in mapping QTL for a variety of traits to very narrow confidence intervals, using population sizes much smaller than would normally be required to achieve such results. In a few cases, resolution was sufficiently precise to point to one or a very small number of candidate genes for the QTN.

This unexpected mapping effectiveness, relative to the Valdar et al.'s (2006) simulations can be attributed to a number of factors:

(1) As noted, the CC includes three wild-derived strains, representing three Mus musculus subspecies. Apparently, these introduced genetic variants having much stronger effects than those uncovered in QTL analyses of the standard mouse inbred lines, or in domesticated livestock and poultry.

(2) QTL mapping in the CC resource is based on an eight-allele founder haplotype model, instead of a simple marker-based association test. This avoids confounding of different linkage phase of marker allele and QTL allele in different founder populations.

(3) Sorting parental lines according to allele effect of the QTL identifies regions of similar effect and similar haplotype shared by two or more founders, and contrasting to the effects and haplotypes of the other founders. This limits the location of the QTL to the chromosomal region and marker haplotype that is common to the founders sharing the QTL allele. The reduction in confidence interval of QTL location by this means can be dramatic.

These studies confirm that by phenotyping a relatively modest number of CC lines (around 100 lines), with sufficient replication, it is possible to map QTLs to a resolution of about 1 Mb (Aylor et al. 2011; Durrant et al. 2011; Philip et al. 2011; Bottomley et al. 2012; Kelada et al. 2012), subsequently leading to identification of strong candidate genes. Knockouts of candidates within QTL can then be used to confirm function. Deep RNA sequencing (Ansorge 2009; Ng et al. 2010; Voelkerding et al. 2009) can also be employed to identify the pathways activated by these genes. We believe that these achievements cannot be obtained with any other currently available mouse resource populations.

References

Ansorge WJ (2009) Next-generation DNA sequencing techniques. New Biotechnol 25(4):195–203

Aylor DL, Valdar W, Foulds-Mathes, Buus RJ, Verdugo RA, Baric RS, Ferris MT, Frelinger JA, Heise M, Frieman MB, Gralinski LE, Bell TA, Didion JD, Hua K, Nehrenberg DL, Powell CL, Steigerwalt J, Xie Y, Kalada S, Collins FS, Yang LV, Schwartz DA, Branstetter LA, Chesler EJ, Miller DR, Spence J, Yi Liu E, McMillan L, Sarkar A, Wang J, Wang W, Zhang Q, Broman KW, Korstanje R, Durrant C, Mott R, Iraqi FA, Pomp D, Threadgill D, de Pardo-Manuel Villena F, Churchill GA (2011) Genetic analysis of complex traits in the emerging collaborative cross. Genome Res 21:1213–1222

Baker PJ, Roopenian DC (2002) Genetic susceptibility to chronic periodontal disease. Microbes Infect 4:1157–1167

Baker PJ, Dixon M, Roopenian DC (2000) Genetic control of susceptibility to Porphyromonas gingivalis-induced alveolar bone loss in mice. Infect Immun 68:5864–5868

Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MFW, Fisher EMC (2000) Genealogies of mouse inbred strains. Nat Genet 24:23–25

Bottomley D, Ferris MT, Aicher LD et al (2012) Expression quantitative trait loci for extreme host response to Influenza A in pre-Collaborative Cross Mice. G3 2:213–221

Broman KW (2005) The genomes of recombinant inbred lines. Genetics 169(2):1133–1146

Broman KW, Kim S, Sen S, Ané C, Payseur BA (2012) Mapping quantitative trait loci onto a phylogenetic tree. Genetics 192(1):267–279

Brommage R (2003) Validation and calibration of DEXA body composition in mice. Am J Physiol Endocrinol Metab 285:454–459

Buse JB, Ginsberg HN, Bakris GL, Clark NG, Costa F, Eckel R, Fonseca V, Gerstein HC, Grundy S, Nesto RW, Pignone MP, Plutzky J, Porte D, Redberg R, Stitzel KF, Stone NJ (2007) Primary prevention of cardiovascular diseases in people with diabetes mellitus: a scientific statement from the American Heart Association and the American Diabetes Association. Circulation 115:114–126

Chesler EJ, Miller DM, Branstetter LR, Galloway LD, Jackson BL et al (2008) The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. Mamm Genome 19:382–389

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ, Geiger H, Gershenfeld H, Gordon T, Gu J, Gu W, de Haan G, Hayes NL, Heller C, Himmelbauer H, Hitzemann R, Hunter K, Hsu HC, Iraqi FA, Ivandic B, Jacob HJ et al (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 36:1133–1137

Collaborative Cross Consortium, Iraqi FA, Mahajne M, Salaymah A, Sandovsky H, Tayem, Vered K, Balmer L, Hall M, Manship G, Morahan G, Petit K, Scholten J, Tweedie K, Weerasekera L, Cleak J, Durrant C, Goodstadt L, Mott R, Yalcin B, Aylor DL, Baric R, Bell TA, Bendt KM, Brooks JD, Buus RJ, Crowley JJ, Calaway JD, Calaway ME, Cholka A, Darr DB, Didion JP, Dorman A, Everett E, Ferris MT, Mathes WF, Fu CP, Gooch TJ, Goodson SG, Garlinski LE, Hansen SD, Heise M, Hoel J, Lee S, Lenarcic AB, Liu EY, McMillan L, Magnuson TR, Manly KF, O'Brien DA, Odet F, Pan W, de Pardo-Manuel Villena F, Perou C, Pomp D, Quackenbush CR, Robinson NN, Sharpless N, Shaw GD, Spence JS, Sullivan PF, Sun W, Tarantino LM, Valdar W, Wang J, Wang W, Welsh CE, Whitmore A, Wiltshire T, Wright F, Xie Y, Yu Zn, Zhabotynsky V, Zhang Z, Zou F, Powell C, Steigerwalt J, Threadgill DW, Chesler EJ, Churchill GA, Gatti

DM, Svenson KL, Yang H, Shusterman A, Nashef A, Weiss EI, Houri-Haddad Y, Soller M, Schughart K, French JE, Collins FS, Crawford N, Hunter K, Kelada SNP, Peck BCE, Reilly K, Tavarez U, Bottomly D, Hitzeman R, McWeeney SK, Frelinger J, Phillippi J, Spritz RA, Benson AK, Kim J, Legge R, Low SJ, Ma F, Martinez I, Walter J, Williams RW, Aicher L, Katze M, Rosenzweig E, Broman KW (2012) The genome architecture of the Collaborative Cross mouse genetic reference population. Genetics 190(2):389–402

Cormier RT, Dove WF (2000) Dnmt1 N/1 reduces the net growth rate and multiplicity of intestinal adenomas in C57BL/6-Multiple intestinal neoplasia (Min)/1 mice independently of p53 but demonstrates strong synergy with the Modifier of Min 1AKR resistance allele. Cancer Res 60:3965–3970

Cormier RT, Bilger AJ, Lillich RB, Halberg KH (2000) The Mom1AKR intestinal tumor resistance region consists of Pla2g2a and a locus distal to D4Mit64. Oncogene 19:3182–3192

Durrant C, Tayem H, Yalcin B, Cleak J, Goodstadt L, de Pardo-Manuel Villena F, Mott R, Iraqi FA (2011) Collaborative Cross mice and their power to map host susceptibility to *Aspergillus fumigatus* infection. Genome Res 21:1239–1248

Falconer, Mackay, Falconer DS, Mackay FC (1996) Introduction to quantitative genetics. Longman, Essex

Fearnhead NS, Britton MP, Bodmer WF (2001) The ABC of APC. Hum Mol Genet 10:721–733

Fostira F, Thodi G, Sandaltzopoulos R, Fountzilas G, Yannoukakos D (2010) Mutational spectrum of APC and genotype-phenotype correlations in Greek FAP patients. BMC Cancer 10:389–396

Garcia-Gonzalez F, Simmons LW, Tomkins JL, Kotiaho JS, Evans JP (2012) Comparing evolvabilities: common errors surrounding the use and calculation of coefficients of additive genetic variation. Evolution 66:2341–2349

Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, Joslyn G, Stevens J, Spirio L, Robertson M (1991) Identification and characterization of the familial adenomatous polyposis coli gene. Cell 66:589–600

Hernandez-Valladares M, Naessens J, Gibson JP, Musoke AJ, Rihet P, ole-MoiYoi OK, Iraqi FA (2004a) Confirmation and dissection of QTL controlling resistance to malaria in mice. Mamm Genome 15:390–398

Hernandez-Valladares M, Rihet P, ole-MoiYoi OK, Iraqi FA (2004b) Mapping of a new quantitative trait locus for resistance to malaria in mice by a comparative approach of human chromosome 5q31-q33. Immunogenet 56:115–117

Houle D (1992) Evolvability and variability of quantitative traits. Genetics 130:195–204

Iraqi F, Clapcot S, Kuman P, Heley C, Kemp S, Teale A (2000) Fine mapping of trypanosomiasis resistance QTLs in mice using advanced intercross lines. Mamm Genome 11:645–648

Iraqi FA, Behnke JM, Menge DM, Lowe AM, Teale AJ, Gibson JP, Baker LR, Wakelin DR (2003) Mapping chromosomal regions controlling resistance to gastro-intestinal worms in mice. Mamm Genome 14:184–191

Iraqi F, Churchill G, Mott R (2008) The Collaborative Cross, developing a resource for mammalian systems genetics: a status report of the Wellcome Trust cohort. Mamm Genome 19:379–381

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G et al (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 14 477(7364):289–294

Kelada SNP, Aylor DL et al (2012) Genetic analysis of hematological parameters in incipient lines of the Collaborative Cross G3(2):157–166

Keynan Y, Rubinstein E (2007) The changing face of Klebsiella pneumoniae infection in the community. Int J Antimicrob Agent 30:385–389

Kovacs A, Ben-Jacob N, Tayem H, Halperin E, Iraqi FA, Gophna U (2011) Genotype is a stronger determinant than sex of the mammalian gut microbiota. Microb Ecol 61(2):423–428

Mathes WF, Aylor DL, Miller DR, Churchill GA, Chesler EJ, de Villena F, Threadgill DW, Pomp D (2011) Architecture of energy balance traits in emerging lines of the Collaborative Cross. Am J Physiol Endocrinol Metab 300:E1124–E1134

Morahan G, Balmer L, Monley D (2008) Establishment of "The Gene Mine": a resource for rapid identification of complex trait genes. Mamm Genome 19(6):390–393

Moser AR, Dove WF, Roth KA, Gordon JI (1992) The Min (multiple intestinal neoplasia) mutation: its effect on gut epithelial cell differentiation and interaction with a modifier system. J Cell Biol 716:1517–1526

Moser AR, Mattes EM, Dove WF, Lindstrom MJ, Haag JD, Gould MN (1993) Min, a mutation in the murine Apc gene, predisposes to mammary carcinomas and focal alveolar hyperplasias. PNAS 90:8977–8981

Mott R, Talbot CJ, Turri MG, Collins AC, Flint J (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. PNAS 97:12649–12654

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42(1):30–35

Petkova SB, Yuan R, Tsaih SW, Schott W, Roopenian DC, Paigen B (2008) Genetic influence on immune phenotype revealed strain-specific variations in peripheral blood lineages. Physiol Genomics 34:304–314

Philip VM, Sokoloff G, Ackert-Bicknell CL et al (2011) Genetic analysis in the Collaborative Cross breeding population. Genome Res 2011:1223–1238

Polak D, Wilensky A, Shapira L, Halabi A, Goldstein D, Weiss EI et al (2009) Mouse model of experimental periodontitis induced by *Porphyromonas gingivalis/Fusobacterium nucleatum* infection: bone loss and host response. J Clin Periodontol 36(5): 406–410

Roberts A, de Pardo-Manuel Villena F, Wang W, McMillan L, Threadgill DW (2007) The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. Mamm Genome 18(6–7):473–481

Shusterman A, Durrant C, Mott R, Schaefer A, Weiss EI, Iraqi FA, Houri-Haddad Y (2013a) Host susceptibility to periodontitis: mapping murine genomic regions. J Dent Res 92:438–443

Shusterman A, Salyma Y, Nashef A, Soller M, Wilensky A, Mott R, Weiss EI, Houri-Haddad Y, Iraqi FA (2013b) Genotype is an important determinant factor of host susceptibility to periodontitis in the Collaborative Cross and inbred mouse populations. BMC Genet 14:68–79

Silver LM (1995) Mouse genetics: concepts and applications. Oxford University Press, Oxford

Threadgill DW, Hunt KW, Williams RW (2002) Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. Mamm Genome 16:344–355

Valdar W, Flint J, Mott R (2006) Simulating the collaborative cross: power of QTL detection and mapping resolution in large sets of recombinant inbred strains of mice. Genetics 172:1783–1797

Voelkerding KV et al (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55:641–658

Welsh CE, Miller RD, Manly KF, Wang J, McMillan L, Morahan G, Mott R, Iraqi FA, Threadgill DW, Pardo-Manuel de Villena F (2012) Status and access to the Collaborative Cross population. Mamm Genome 23:706–712

Wilensky A, Gabet Y, Yumoto H, Houri Haddad Y, Shapira L (2005) Three-dimensional quantification of alveolar bone loss in

Porphyromonas gingivalis infected mice using Micro-CT. J Periodontol 76(8):1282–1286

Yalcin B, Flint J, Mott R (2005) Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. Genetics 171:673–681

Yang H, Ding Y, Hutchins LN, Szatkiewicz J, Bell TA, Paigen BJ, Graber JH, de Villena FP, Churchill GA (2009) A customized and versatile high-density genotyping array for the mouse. Nat Methods 6(9):663–666