

## Advances in comparative genetics: influence of genetics on obesity

Wendy Foulds Mathes<sup>1</sup>, Scott A. Kelly<sup>1</sup> and Daniel Pomp<sup>1,2,3\*</sup>

<sup>1</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>2</sup>Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>3</sup>Department of Cell and Molecular Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

(Received 2 November 2010 – Revised 1 February 2011 – Accepted 12 February 2011)

### Abstract

Obesity has reached epidemic proportions and is recognised as a significant global health problem. Increased food intake and decreased physical activity are traditionally to blame for the development of obesity; however, many variables such as behaviour, diet, environment, social structures and genetics also contribute to this multifactorial disease. Complex interactions among these variables (for example, gene–environment, gene–diet and gene–gene) contribute not only to individual differences in the development of obesity, but also in treatment response. Mouse models have historically played valuable roles in understanding the genetics of traits related to energy balance and obesity. In the present review, we survey past use and examine new advances in mouse models designed to uncover the genetic architecture of obesity and its component traits. We discuss traditional models such as inbred strains and selectively bred lines and their contributions and shortcomings. We consider the evolution of mouse models into more informative resources such as outbred crosses and the Hybrid Mouse Diversity Panel, as well as novel next-generation approaches such as the Collaborative Cross. Moreover, the genetic architecture of voluntary exercise and the interactive relationship between host genetics and the gut microbiome are presented as novel phenotypes that augment studies using body weight and body fat percentage as endpoints. Understanding the intricate network of phenotypic, genotypic and environmental variables that predispose individuals to obesity will elucidate biological networks involved in the development of obesity. Knowledge obtained from advances in mouse models will inform human health and provide insight into inter-individual variability in the aetiology of obesity-related diseases.

**Key words:** Mice: Genes: Body weight: Body fat: Physical activity: Gut microbiome

The prevalence of overweight and obese individuals is a significant global problem<sup>(1)</sup>. Obesity develops due to an imbalance between energy intake and energy expenditure. Energy balance represents a conglomerate of traits, each influenced by numerous variables such as behaviour, diet, environment, social structures and genetics. Complex interactions among these variables (for example, gene–environment, gene–diet, gene–gene) contribute not only to individual differences in the development of obesity, but also in the response to interventions and treatments (Fig. 1).

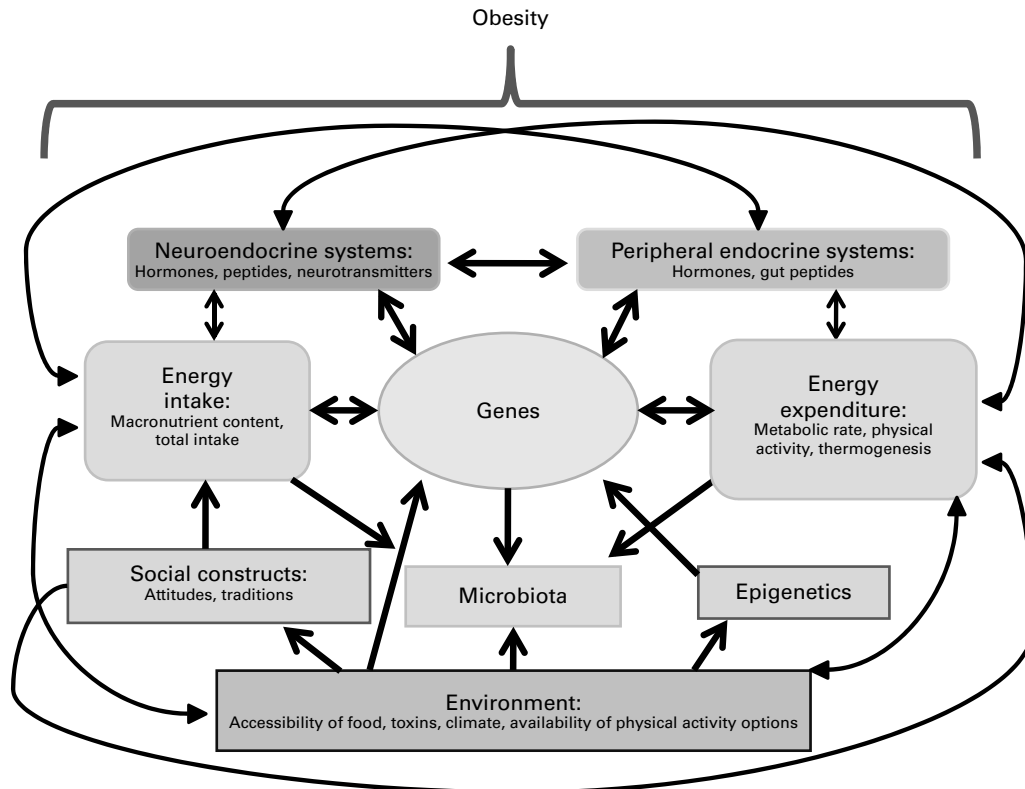
Mouse models have historically played valuable roles in understanding the genetics of traits related to energy balance and obesity in many mammalian species including companion animals, agricultural species and man. Initially, mouse models were successful in identifying individual genes that had large effects and that were inherited in a simple Mendelian pattern<sup>(2)</sup>. Mice with naturally occurring mutations, as well as those with mutations that were genetically engineered or

chemically induced, have played a crucial role in furthering our understanding of how specific genes may contribute to disease aetiology. In many instances (for example, the ob/ob and db/db mice, and the melanocortin-4 receptor (MC4R) knock-out mouse)<sup>(3–5)</sup>, these mutations have identified key proteins involved in obesity and the regulation of energy balance. Yet, investigation into the role of these target proteins has revealed that they do not work in isolation and that the alteration or absence of a single protein cannot explain common obesity-related phenotypes in human populations<sup>(6)</sup>. Thus, it is generally accepted that obesity and energy balance are highly complex and polygenic traits, regulated by numerous genes that have an impact on many physiological systems, and that interact not only with each other but also with many environmental stimuli.

Recent advancements in genetics research have employed the use of genome-wide association studies (GWAS) to determine the genetic underpinnings of energy balance and obesity

**Abbreviations:** ALL, advanced intercross lines; CC, Collaborative Cross; GAWAS, genome-wide association studies; HAM, haplotype association mapping; HMDP, Hybrid Mouse Diversity Panel; HR, high running; NZO, New Zealand obese; QTL, quantitative trait loci.

\* **Corresponding author:** Dr Daniel Pomp, fax +1 919 843 4682, email dpomp@unc.edu



**Fig. 1.** Weight maintenance is complex, involving both central and peripheral inputs that are simultaneously affected by highly interactive genetic architecture (involving 100s or even 1000s of genes) and multiple environmental stimuli. Here, we have attempted to depict only a fraction of the components that contribute to obesity, specifically focusing on the novel phenotypes described in the body of the text (exercise, diet and the gut microbiome) and the potential interactions among them.

in human populations. Several SNP have been implicated in obesity, yet many have failed to replicate across studies<sup>(7–9)</sup>, and confirmed findings account for only a small percentage of the phenotypic variance across the population tested<sup>(10)</sup>. To date, GWAS have consistently identified two SNP significantly associated with obesity, one in an intron of the *FTO* gene and one downstream from the *MC4R* gene. These two SNP have relatively small effects, accounting for less than 2% of the variance in adult BMI<sup>(11,12)</sup>. In conjunction with the relatively low level of explanatory power with regard to percentage phenotypic variance, GWAS have additional limitations. For example, hundreds of thousands to million of SNP may be evaluated in a single sample, reducing the statistical power to identify significant associations. Moreover, heterogeneity requires extremely large sample sizes to reveal moderate to small genetic effects, limiting the discovery of contributions of rare alleles<sup>(13–15)</sup>.

In mice, early studies on the genetic architecture of complex traits employed linkage analysis based on simple F2 crosses of phenotypically divergent inbred strains to identify obesity loci, but the mapping resolution of such crosses proved prohibitive for identifying target genes<sup>(2)</sup>. Currently, in light of the ability to evaluate human populations directly using GWAS, mouse models have evolved to better address the complexity of the genetic architecture of obesity and related traits. Novel approaches include more complex population structures of

advanced intercross lines (AIL) derived from inbred lines and heterogeneous outbred stocks, as well as association analysis based on haplotype–phenotype relationships across large numbers of inbred strains. Moreover, a unique and novel mouse population, the Collaborative Cross (CC), was designed specifically to model the genetic diversity of the human population<sup>(16,17)</sup>. Modern mouse models offer specific advantages over GWAS in that it is possible to completely control environmental variables, population structure and trait variants, lending more power to detect statistically significant associations<sup>(15)</sup>. Furthermore, studies using mouse models enable the interrogation of obesity-related phenotypes that are not easily amenable to large-scale analysis in humans, including, for example, levels of voluntary exercise.

In the present review, we summarise recent advances in the development and use of mouse models designed to understand the genetic architecture of obesity and its component traits. We highlight innovative and novel approaches to understanding the complexity of obesity, such as the CC, which will provide unprecedented resolution for mapping complex traits. In addition, we describe recent advancements in applying mouse models to begin to dissect genetic predisposition to previously understudied traits such as voluntary exercise, and the interactive relationship between host genetics and composition of the gut microbiome.

## Mouse models for understanding the genetic architecture of obesity and related traits

### Inbred mouse strains

Phenotypic differences in body weight and adiposity across multiple inbred mouse strains reveal how genetic background influences phenotypic expression<sup>(18,19)</sup>. Exploiting this phenotypic variability, interbreeding strains with high and low values for a specific phenotype, creates segregating populations allowing for the detection of genome regions associated with the phenotype of interest, or quantitative trait loci (QTL). To date, hundreds of QTL for body weight and adiposity have been described<sup>(20)</sup>, yet little is known about the actual genes and the responsible genetic variation that may contribute to the expression of obesity-related phenotypes<sup>(20)</sup>. Many of these QTL were identified using F2 intercross or backcross populations. QTL mapping of F2 intercross and backcross populations is generally low in power and generalisability since identified QTL typically account for such small proportions of the phenotypic variation and are background strain dependent<sup>(21)</sup>. Furthermore, few F2 intercross QTL studies have used obese or obesity-prone mouse strains such as New Zealand obese (NZO) or KK, biasing the results toward more common inbred strains<sup>(20)</sup>. Recombinant inbred lines offer increased mapping resolution due to increased recombination resulting from generations of inbreeding<sup>(22–25)</sup>. However, many traits may not be mapped to any region of the genome because the two founder strains of the recombinant inbred lines share a large proportion of genetic variation. For example, 48.6% of genomic regions in BXD strains will not be mapped due to shared variation<sup>(26)</sup>.

Recent advances in computational analysis such as haplotype association mapping (HAM) enable QTL mapping by simultaneously evaluating dozens of inbred strains, providing more statistical power for identifying significant QTL and incorporating more of the variation within the mouse genome. HAM overcomes the resolution issue of traditional F2 and backcrosses<sup>(26)</sup> by enabling mapping of QTL to much smaller confidence intervals, potentially harbouring fewer candidate genes. HAM is not without limitations, however. For example, the power to detect genetic contributions of small effect is low, making it more difficult to identify QTL associated with complex traits such as obesity<sup>(26–28)</sup>. In addition, population structure must be taken into account when performing the analysis to eliminate false-positive associations<sup>(26–28)</sup>.

The Hybrid Mouse Diversity Panel (HMDP) has been proposed as a solution to challenges with population structure in association studies<sup>(2)</sup>. Comprised of 100 inbred strains and seventy-one recombinant inbred lines, the HMDP is specifically designed to map traits with small variance contributions and to increase power to detect loci with small effects<sup>(2)</sup>. The combination of recombinant inbred and classical inbred strains will disrupt long spans of linkage disequilibrium within each of the populations, thereby increasing mapping resolution. Preliminary mapping studies using the HMDP have replicated existing QTL identified in classical inbred crosses, yet with greater mapping resolution<sup>(2)</sup>.

### Selectively bred mouse lines

Selective breeding experiments, which have been used for decades to gain an understanding of the genetic underpinnings of growth, fatness and other traits related to obesity and energy balance, can capture extreme phenotypic variation beyond the range of the founding population<sup>(29)</sup>. For example, high growth selection lines have been established with mean body weights of up to 77 g, far exceeding the body weights within the base populations from which these lines were selected, as well as body weight in at least 400 inbred mouse strains<sup>(29,30)</sup>.

Experiments evaluating the genetic underpinnings of obesity-related traits using selection lines have identified several QTL for early growth, obesity, fatness and exercise behaviour (wheel running)<sup>(31–33)</sup>. A large QTL analysis of obesity-related traits in 993 mice from an F2 population derived from a cross between M16i, a mouse strain selected for high growth and body weight at 3–6 weeks of age and inbred, and L6, selected for low growth at 6 weeks of age, revealed many significant QTL for growth and fatness<sup>(31,32)</sup>. In addition, several epistatic effects (gene × gene interactions) were identified<sup>(34)</sup>. Evaluating epistasis in QTL analyses aids in the discovery of loci that may have small main effects but large effects when combined with interacting loci<sup>(34)</sup>. Similarly, extensive research examining adiposity and body composition traits has been performed using Large (LG/J) and Small (SM/J) mice, inbred lines selected for large and small body size at age 60 d<sup>(35–37)</sup>. Several QTL for adiposity and body weight were identified, each of only small effect. Furthermore, pleiotropic QTL for adiposity and tail length demonstrated interaction in the genetic control of obesity and body size<sup>(35,36)</sup>.

Although evidence from mice, other model species such as *Caenorhabditis elegans*<sup>(38)</sup> and humans point to the contribution of literally hundreds of genes to the genetic architecture of body weight, a recent study in dogs paints a distinctly different picture<sup>(39)</sup>. Boyko *et al.*<sup>(39)</sup> found that only a few genes of large effect collectively explain most variability in canine body size. This unusual genetic architecture was attributed to several unique aspects of selection in dogs. For example, the authors speculated that many of the modern breeds of dogs were created during the Victorian era where novelty was a focus of selection and breeders favoured the preservation of discrete mutations<sup>(39)</sup>. Repeated selection for rare or novel traits in dogs contributed to genetic variation, thus increasing the power to identify causal genes.

### Advanced intercross lines (AIL)

AIL have been used to refine the mapping of genetic loci associated with obesity- and energy balance-related traits, as well as to identify new genetic variation. AIL offer increased mapping resolution of QTL through increased recombination resulting from random intercrosses over multiple generations of breeding<sup>(40)</sup>. Initial AIL experiments using the LG/J by SM/J cross with continued interbreeding through generations F9 and F10 resulted in five times the mapping resolution of the original F2 population<sup>(41)</sup>. This F10 AIL identified ninety

significant single trait QTL for phenotypes related to obesity such as increased size, adipose deposition and body weight, many of which were multiple narrow QTL with small effect size resolved from original large single peak QTL in F2/F3 populations. Such studies only further highlight the complexity of the genetic architecture of obesity, and the potentially enormous number of genes involved.

Additional AIL resources for fine-mapping complex traits such as obesity include heterogeneous outbred stocks of mice. Outbred stocks are derived from multiple founder inbred strains and numerous random matings, resulting in a population with closely spaced recombinants (<2cM) and fine mapping resolution (<1cM) for QTL of small effect<sup>(42,43)</sup>. GWAS performed in outbred stocks identified hundreds of QTL with effect sizes less than 2%. Many of these QTL replicated existing QTL identified in crosses between inbred strains that had been included in the creation of the heterogeneous stock<sup>(42)</sup>. In some instances, large-effect QTL were identified with mapping intervals encompassing three or fewer genes, uncovering significant targets for causal genes<sup>(42)</sup>.

The mapping approaches discussed above have provided valuable information as to the general genetic architecture of obesity and related traits. Numerous genetic loci have been identified, emphasising the complex and polygenic nature of the genetic architecture of obesity, but specific target genes have yet to be determined. Although these approaches have been informative, several shortcomings have been recognised. For example, the lack of genetic diversity among traditional inbred laboratory mouse strains has left large regions of the genome unmapped, and has afforded little power to detect small effects and identifying QTL that are largely strain dependent<sup>(21)</sup>. AIL and outbred stocks have been established to increase mapping resolution and increase power to identify loci of small effect. However, the lack of genetic diversity still exists, and mapping power remains too low for identification of the underlying causative genetic variation<sup>(42,43)</sup>. To address these shortcomings, a unique and innovative mouse model designed to mimic human genetic diversity, the CC, has been created.

### *The Collaborative Cross, a next-generation mouse model*

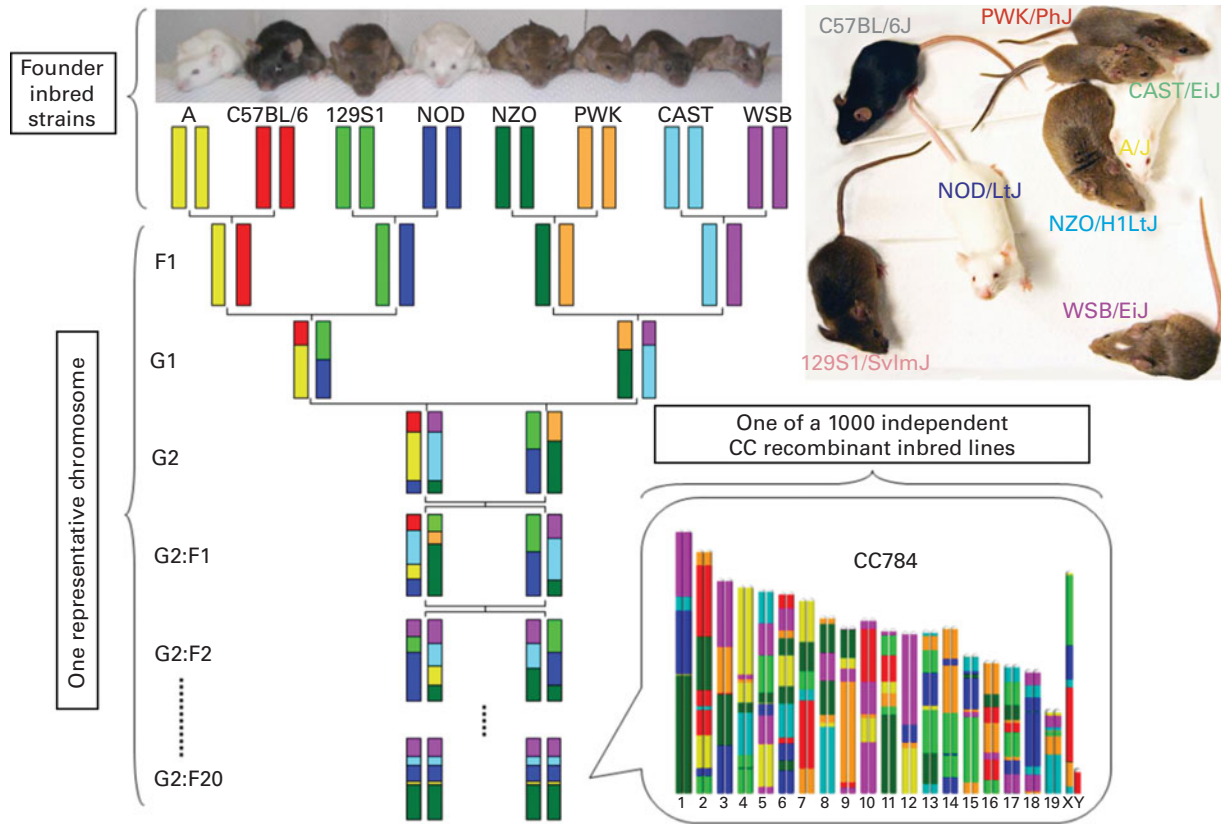
The CC is a large panel of recombinant inbred mouse lines derived from a genetically diverse set of eight founder strains created to advance discovery of the genetic underpinnings of complex traits, such as energy balance and body weight regulation (Fig. 2)<sup>(16,17)</sup>. The eight founder strains consist of five standard laboratory strains (C57BL A/J, C57BL/6 J, 129S1/SvImJ, NOD/LtJ and NZO/H1LtJ) and three wild-derived strains (WSB/Eij, CAST/Eij and PWK/PhJ), representing each of the subspecies of *Mus musculus* and capturing an unprecedented high degree of genetic diversity<sup>(44,45)</sup>. Not only were the eight founder strains chosen specifically for their genetic diversity, but also to include genetic variants specifically linked to known complex disorders such as type 1 and 2 diabetes, obesity and insulin insensitivity (NOD/LtJ and NZO/H1LtJ)<sup>(46)</sup>. This genetic diversity affords the CC

increased mapping precision for complex traits, a distinct advantage over AIL and existing recombinant inbred lines<sup>(16,17)</sup>. It is expected that the mapping resolution of the fully inbred lines will be approximately 1 Mb, more precise than the finest resolution presented thus far<sup>(23,42,47)</sup>. Furthermore, the CC lines will be completely sequenced, and sequences for each line will be archived and available for shared use. Contrarily, each new individual in a specific cross or AIL is genetically unique and thus must be individually genotyped or sequenced, making experiments in these populations costly and labour intensive<sup>(16,17,46)</sup>. The genetic variability and eternal reproducibility of the CC makes it a superior resource for systems genetics research. This large reference panel of recombinant inbred strains will enable the evaluation of complex gene–environment and gene–gene interactions, providing a platform for the investigation into the genetic architecture underlying complex interactions among diverse phenotypes and molecular networks<sup>(16)</sup>.

Initial analysis of genetic architecture confirmed that generations F5–F12 of the incipient CC lines (i.e. during the inbreeding process) encompassed the total genetic diversity of the eight founder strains with alleles from all eight strains dispersed among the genomes of each incipient CC line<sup>(44)</sup>. Furthermore, phenotypic analysis of the eight founder strains and 176 incipient CC lines revealed extreme diversity in obesity- and metabolism-related traits. For example, body fat percentages for the founder strains varied from 2 to 32% for the A/J and NZO/H1LtJ strains, respectively<sup>(48)</sup>. Likewise, body weights for the founder strains ranged from 13 g for WSB/Eij to 48 g for NZO/H1LtJ. Variation in adiposity and body weight was mirrored in the incipient CC lines. Body fat percentages ranged from 1 to 32%<sup>(48)</sup>. Body weights for the incipient CC lines were slightly less varied, spanning from 15 to 35 g. However, several of the traits measured in the incipient CC lines surpassed the phenotypic variation observed in the founder strains. For instance, average wheel-running distances on days 11 and 12 of wheel access ranged from 0 to 13 km in the founders and from 0 to 19 km in the incipient CC lines<sup>(48)</sup>. Moreover, body weight change in response to wheel running varied from a 6 g loss to a 3 g gain in the founder strains and from a 7 g loss to an 8 g gain in the incipient CC lines<sup>(48)</sup>. These findings demonstrate that the genetic diversity highlighted in the eight founder strains was captured in the incipient CC lines resulting in equivalent or increased phenotypic diversity for complex obesity-related traits<sup>(48)</sup>.

Mapping experiments in the incipient CC lines as a proxy for the power of the CC resource to uncover the genetic architecture of complex traits revealed significant novel QTL for wheel-running traits and replicated previously described QTL for body weight-, body fat- and metabolism-related traits (Aylor *et al.*<sup>(44)</sup>; WF Mathes, DL Aylor, DR Miller, GA Churchill, EJ Chesler, F Pardo-Manuel de Villena, DW Threadgill and D Pomp, in press). A significant body weight QTL was identified on proximal chromosome 4, *Bwq14*, which replicated existing QTL for body weight identified using the KK strain, a strain phenotypically similar to NZO<sup>(49,50)</sup>. Expression QTL analysis for *Bwq14* revealed





**Fig. 2.** The Collaborative Cross (CC) is a large panel of recombinant inbred mouse lines designed to model human genetic diversity. Each individual line of the CC (one theoretical example depicted here) will represent a genetic mosaic of the eight founder strains. Derived from the crossbreeding of five classic inbred lines (C57BL A/J, C57BL/6 J, 129S1/SvImJ, NOD/LtJ and NZO/H1LtJ) and three wild-derived mouse lines (WSB/EiJ, CAST/EiJ and PWK/PhJ), the CC captures more than 90% of the genetic diversity across the mouse genome. The genetic and phenotypic diversity of the eight founder strains captures genetic variants specifically linked to body size and known complex disorders such as type 1 and 2 diabetes, obesity and insulin insensitivity. Some images in this figure are courtesy of Fernando Pardo Manuel de Villena.

twenty-one genes with significant local expression QTL. Using allele effect estimates and existing SNP data sources for the eight founder strains, the genes underlying *Bwq14* could be narrowed to a single positional candidate, aspartate- $\beta$ -hydroxylase, *Asph*<sup>(44)</sup>.

The genetic and phenotypic diversity that has manifested in the incipient CC lines provides strong evidence that the CC, once fully inbred, will be a powerful tool for complex trait analysis. Early success in mapping QTL for body weight in the incipient CC lines confirms that the CC will aid in the discovery of the genetic variables that predispose individuals to obesity and related traits. The advantages of the CC over existing resources are clear. The CC will offer increased genetic and phenotypic diversity that will translate into the discovery of QTL within regions that were previously 'blind' to investigation. The elaborate mating scheme used to create the CC will eliminate strain-specific bias and reduce linkage disequilibrium, allowing for mapping resolution precise enough to identify specific genes that underlie complex traits. Moreover, the CC will be an invaluable tool for the examination of gene–environment interactions that contribute to energy balance and obesity-related phenotypes. Taken together, the CC will inform human health by elucidating biological networks involved in the development of obesity as well as providing

predictive models from which one can determine the variables that may contribute to the aetiology of obesity-related diseases on an individual basis.

### Novel phenotypes: exercise

As described above, genetic architecture unquestionably plays a major role in weight dysregulation and traits related to obesity. However, behaviours such as food consumption and exercise are also crucial in maintaining a balance between energy intake and expenditure, and ultimately constancy of body weight. In addition to weight loss and reductions in adiposity, physical activity has also been associated with reductions in TAG and LDL levels, increased HDL and enhanced insulin sensitivity<sup>(51,52)</sup>. Despite the growing body of evidence of positive health benefits, especially with regard to weight reduction<sup>(53)</sup>, there remains considerable variation in the amount that children and adults voluntarily exercise<sup>(51,54)</sup>. While a voluntary behaviour, the predisposition to engage in exercise is heritable<sup>(55)</sup>, and thus, the genes that underpin physical activity traits are integral to our understanding of the genetic architecture of obesity. As such, multiple mapping studies have been directed at understanding the genetic architecture of physical activity-related traits and

how they may have an impact on body weight and obesity. Moreover, as changes resulting from exercise remain considerably variable within populations, especially with regard to weight and adiposity, studies are now aiming to uncover the genetic architecture regulating the change in weight and adiposity in response to exercise (i.e. Mitchell *et al.*<sup>(56)</sup>).

Initial explorations into the genetic architecture of exercise-related traits have been conducted in second-generation intercross (F2), backcross, and advanced intercross (G4) mouse populations. Using F2 crosses between low active (wheel running) C3H/HeJ and high active C57L/J have revealed multiple QTL for running duration, distance and speed<sup>(57,58)</sup>. Moreover, analysis of body weight changes in response to exercise in this population identified additional QTL for energy balance<sup>(59)</sup> and loci that were pleiotropic for both exercise- and body weight-related traits<sup>(60)</sup>. QTL for weight change in response to exercise did not co-localise with QTL for running traits or previously described QTL for body weight.

Nehrenberg *et al.*<sup>(61)</sup> utilised a backcross between mice selectively bred for high voluntary wheel running (high running (HR) line) and the inbred strain C57BL6/J to identify novel exercise QTL as well as loci that were sex-specific. The HR line originated from a long-term replicated artificial selection experiment for high voluntary wheel-running behaviour on days 5 and 6 of a 6d wheel exposure (for a review, see Swallow *et al.*<sup>(62)</sup>). By generation 16, and continuing through generation 50 and beyond, the HR lines (four replicates) had diverged from the control lines (C lines; four replicates) with an approximate 2.5- to 3.0-fold increase in total revolutions per d. As demonstrated by the use of the HR line in Nehrenberg *et al.*<sup>(61)</sup>, selective breeding has not only facilitated the uncovering of the genetic architecture of weight-related phenotypes, but has aided in the search for genetic loci associated with exercise phenotypes as well.

Using a different HR replicate (out of the possible four), Kelly *et al.*<sup>(63)</sup> created an advanced intercross line population (G4) and used it to identify forty-one significant and twenty suggestive QTL for running traits (distance, duration and speed) and body mass. Although several of these loci are close in proximity to physical activity QTL identified in F2 and backcross populations, they do not overlap, suggesting that they may be independent loci. However, the strains used to create each of these populations differ in their phenotypic expression for physical activity- and body weight-related traits, making a direct comparison between the two populations difficult. Each of the QTL identified in the AIL explained only a small percentage of the phenotypic variations observed in this population, with the largest effect accounting for only 6.6%<sup>(64)</sup>. In addition to daily running traits, Kelly *et al.*<sup>(63)</sup> identified seven significant and five suggestive QTL representing the slope and intercept of a linear regression across all 6d of running, some representing a combination of the daily traits. The QTL associated with the trajectory of running across multiple days of wheel exposure were the first of their kind and may provide valuable information with regard to the continuation of exercise programmes.

An additional strategy recently employed by Lightfoot *et al.*<sup>(65)</sup> utilises HAM in thirty-eight inbred mouse strains. In addition to discovering new and refining previously mapped QTL for wheel running, the use of HAM across a diverse set of genetic backgrounds allowed the expansion of genomic coverage and the ability to more confidently generalise the results. In the study by Lightfoot *et al.*<sup>(65)</sup>, 448 mice from thirty-eight inbred strains were phenotyped and genotyped for physical activity-related traits such as running distance, speed and time. Results identified twelve significant QTL for running distance and speed, some with sex-specific effects. QTL identified in this experiment did not replicate existing QTL; however, existing QTL for physical activity traits are based solely on crosses using C3H/HeJ, C57L/J and HR lines<sup>(65)</sup>. The three strongest QTL identified in this HAM study had relatively narrow confidence intervals, spanning genomic regions with few genes. In fact, only eight predicted genes and one annotated gene fell within the confidence intervals of these peaks<sup>(65)</sup>.

Collectively, the findings discussed above have just begun to uncover the genetic architecture of physical activity and how it may relate to obesity and its related traits. Further investigation is warranted to understand the complex relationship among these traits and the genetic underpinnings of the said relationship.

### Host genetic control over composition of the gut microbiome

The primary cause of obesity is an imbalance between energy intake and expenditure, with traits related to energy balance being exceptionally complex, simultaneously affected by genetic architecture and interacting environmental factors. Nonetheless, as we have described, mouse models have been, and continue to be, fruitful in the elucidation of the underlying genetic architecture of obesity. A vast majority of studies, only a fraction of which are discussed above, attempting to identify genetic susceptibility of obesity have focused on the penultimate outcome trait (i.e. weight and/or adiposity), with fewer studies examining traits associated with the behavioural and physiological mechanisms contributing to obesity (for example, food choice, intake and assimilation efficiency; RMR; exercise propensity). As described above, literature is now emerging that attempts to uncover the genetic architecture of behaviours that negatively affect energy balance, specifically exercise. Here we present an additional new paradigm for the study of obesity, highlighting a recent example<sup>(66)</sup>, namely the analysis of host genes that influence composition of the gut microbiome, a climax population of microbial species that enter into intimate metabolic and immune interactions with host gastrointestinal tissues and potentially affect many nutritionally relevant traits and diseases.

Recent literature examining the role of the gut microbial ecology may now enlighten the view that obesity simply results from an excess of energy intake *v.* energy expenditure (for a review, see Turnbaugh & Gordon<sup>(67)</sup>). This emerging hypothesis predicts that lean and obese individuals possess unique gut microbial populations that may in turn affect

energy extraction and subsequent deposition of fat stores from food sources. Thus, the impact of the gut microbial community is most relevant for the proportion of intake that can be metabolised. In other words, even when energy intake and expenditure are held constant, variability in energy metabolism and fat deposition will vary depending on the composition of the gut microbes.

In mammals, the colonisation and establishment of micro-organisms in the gastrointestinal tract begin at birth and stabilise around the time of weaning<sup>(68)</sup>. This dense microbial population, up to 100 trillion organisms, is collectively known as the gut microbiota<sup>(67)</sup>. The gut microbiota is populated by thousands of bacterial species belonging to a small number of phyla, and varies dramatically among communities<sup>(69)</sup> and individuals within a single population<sup>(70)</sup>. Contributing to the individuality of the gut microbiota are differences in the relative ratios of dominant phyla and variation in genera and species found in an individual host<sup>(70)</sup>.

The composition of the gut microbiota is undoubtedly shaped by environmental factors, highly resistant to change once established, and relevant to our understanding of human diseases such as obesity, CHD<sup>(71)</sup>, diabetes<sup>(72)</sup> and digestive maladies<sup>(73)</sup>. However, few studies have directly tested, in the absence of confounding genetic diversity (in human populations) and strong environmental effects, the hypothesis that host genetic architecture may also play a role in the gut microbiota composition<sup>(67,74,75)</sup>.

In an attempt to control the strong environmental effects exerted over the gut microbiota, several studies have tested the impact of the host genotype on the gut microbiota utilising comparisons of genetically distinct mouse inbred strains. Friswell *et al.*<sup>(76)</sup> performed uterine transplantations in order to eliminate any non-genetic maternal effects, and found that mice with different genotypes that were born together had similar microbiota. In another comparison, Loh *et al.*<sup>(77)</sup> used germ-free inbred strains (C57BL/10, C3H, BALB/c), inoculated individuals from a single mouse donor, and compared the strains over a 13-week period. Between 4 and 8 weeks post-inoculation, cluster analyses revealed that C3H and BALB/c mice were similar and distinct from C57BL/10. But at 8 weeks post-inoculation, environmental effects (housing in different cages) appeared to play a greater role relative to host genetic control on intestinal microbiota composition.

An alternative approach to the comparison of inbred strains is a candidate gene approach. This approach, primarily driven by transgenic mouse studies, aims to identify single genes that strongly influence the gut microbiota. Thus far, identified candidate genes have been primarily related to immune system function<sup>(78)</sup> and metabolism<sup>(79)</sup>. For example, the obese phenotype (*ob/ob*) of leptin-deficient mice, compared with lean mice, has been shown to be associated with a 50% reduction in the abundance of Bacteroidetes while proportionally increasing the gut population of Firmicutes<sup>(80)</sup>. Subsequently, Turnbaugh *et al.*<sup>(81)</sup> demonstrated that this shift in relative abundance of Bacteroidetes and Firmicutes, among obese individuals, resulted in greater capacity for energy extraction from the diet.

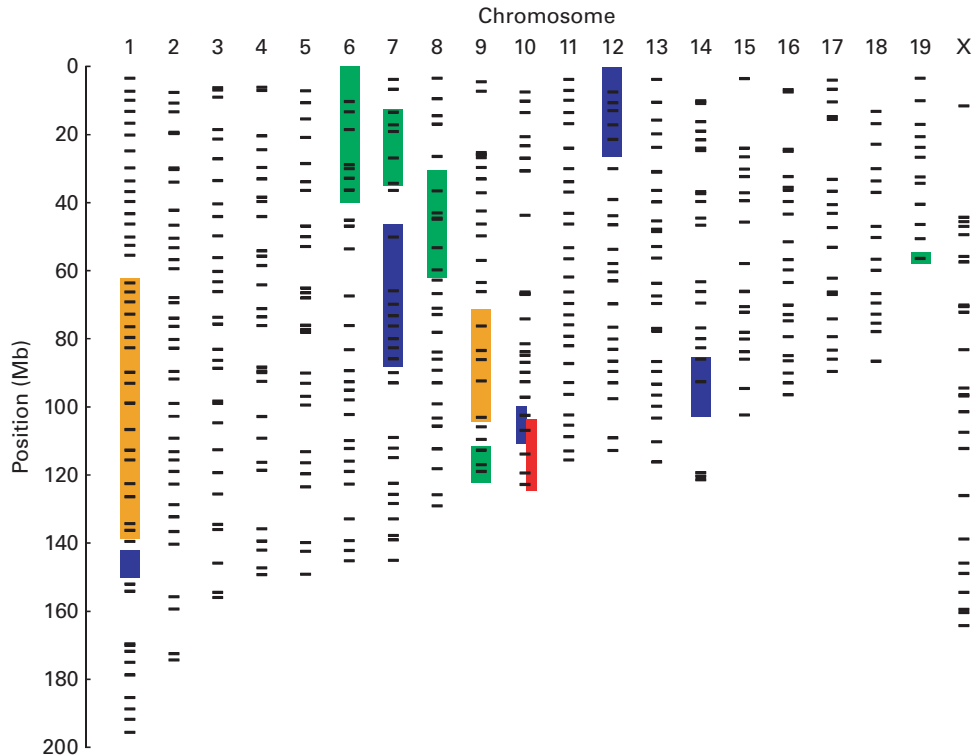
Benson *et al.*<sup>(66)</sup> have most directly tested the assumption of host genetic control of the gut microbiome using a sophisti-

cated murine advanced intercross line model, in which the genetic background was evaluated and environmental factors controlled. This large G4 population ( $n$  645) was originally derived from an intercross between C57BL/6J and an ICR-derived outbred line (HR), and designed to uncover the genetic architecture of voluntary exercise while providing a high level of mapping resolution (as discussed above, see the Novel phenotypes: exercise section). DNA was extracted from faecal samples, and using quantitative pyrosequencing, sixty-four conserved taxonomic groups, which varied across animals in the population, were defined. While a proportion of the variation in the abundance of taxonomic groups could be explained by environmental variables such as litter and cohort, Benson *et al.*<sup>(66)</sup> found a substantial contribution from the host genetic architecture. Using 530 fully informative SNP markers, QTL analyses revealed eighteen loci that show significant or suggestive genome-wide linkage with relative abundances of specific microbial taxa (Fig. 3). Some loci controlled individual microbial species, some controlled groups of related taxa, and some had pleiotropic effects on groups of distantly related organisms. These findings clearly establish host genetics as a factor in determining composition of the gut microbiome.

Benson *et al.*<sup>(66)</sup> have for the first time presented direct evidence for environmental and host genetic loci that control variability in the abundances of different taxa in the mouse gut microbiome, and have therefore hypothesised that the gut microbiota composition as a whole can be understood as a complex, polygenic trait. Additionally, the genetic control identified by Benson *et al.*<sup>(66)</sup> appears to encompass, for example, host genetic factors such as those influencing mucosal immunity. Consequently, host genetic loci that affect composition of the gut microbiome are likely to partially contribute to an individual's overall predisposition to obesity and other nutritionally relevant traits and diseases. How changes in nutrition may influence this host–microbiome relationship, and thus make an impact on weight regulation, remains an interesting yet untested question.

### Summary: the genetic architecture of obesity

The many mapping strategies presented within the present paper have outlined significant progress in uncovering the genetic architecture of obesity and related traits. Initial studies examining obesity and related traits used simple F2 crosses of inbred strains to parse out regions of the genome associated with these traits. The need for finer mapping resolution and greater capacity to detect small effects drove the development of AIL, outbred crosses, and diverse populations such as the HMDP to investigate the genetic architecture of complex traits. From these studies it became clear that obesity is a complex trait with genetic effects originating from literally hundreds of small-effect QTL across the genome. In addition, it was recognised that many environmental factors, and interactions between genes and the environment, contribute to the development of obesity and its related traits. The CC is a next-generation mouse model designed to maximise genetic and phenotypic diversity, refine mapping resolution, and



**Fig. 3.** A summary of host murine quantitative trait loci (QTL) controlling gut microbiome composition as described in Benson *et al.*<sup>(66)</sup>. ■, SNP ( $n$  530) used for QTL mapping, with corresponding chromosomal positions. QTL confidence intervals are shaded in colours corresponding to major phyla: ■, Actinobacteria; ■, Proteobacteria; ■, Firmicutes; ■, Bacteroidetes.

enable the detection of small effects. Early evaluation of the CC demonstrates that it has the potential to be a remarkable tool for complex trait analysis and could potentiate significant discoveries within the fields of genetics and systems biology. The completion of the CC in 2012 may afford unprecedented mapping precision to aid in the search for causal genes for obesity and its related traits. It is important to note that novel phenotypes such as physical activity and wheel running in mice can contribute to the development of obesity. Likewise, genetic sequences of host and gut microbiota may influence energy utilisation contributing to obesity and overweight. Understanding the complex interplay among these traits will provide insight into the intricate network of phenotypic, genotypic and environmental variables that predispose individuals to obesity and its related traits.

### Acknowledgements

Sections of the present review were supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (grants no. RC1DK087346 and no. DK076050). Some phenotypes were collected using the Animal Metabolism Phenotyping core facility at the University of North Carolina's Nutrition Obesity Research Center (funded by NIDDK grant no. DK056350). W. F. M. was supported by National Institutes of Health grant no. K12 HD01441. S. A. K. was supported through a National Institute of Mental Health-funded (no. 5T32-MH-075854-04) Interdisciplinary Obesity Training (IDOT) programme and from the National Institute

of Diabetes and Digestive and Kidney Diseases (NIDDK) (grants no. RC1DK087346 and no. DK076050).

W. F. M., S. A. K. and D. P. all contributed to the writing of the present review.

There are no conflicts of interest.

### References

1. World Health Organization (2010) *Set of Recommendations on the Marketing of Foods and Non-alcoholic Beverages to Children*. Geneva: WHO.
2. Bennett BJ, Farber CR, Orozco L, *et al.* (2010) A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res* **20**, 281–290.
3. Campfield LA, Smith FJ, Guisez Y, *et al.* (1995) Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science* **269**, 546–549.
4. Butler AA, Marks DL, Fan W, *et al.* (2001) Melanocortin-4 receptor is required for acute homeostatic responses to increased dietary fat. *Nat Neurosci* **4**, 605–611.
5. Ste Marie L, Miura GI, Marsh DJ, *et al.* (2000) A metabolic defect promotes obesity in mice lacking melanocortin-4 receptors. *Proc Natl Acad Sci U S A* **97**, 12339–12344.
6. Considine RV, Considine EL, Williams CJ, *et al.* (1996) The hypothalamic leptin receptor in humans: identification of incidental sequence polymorphisms and absence of the db/db mouse and fa/fa rat mutations. *Diabetes* **45**, 992–994.
7. Hindorff LA, Sethupathy P, Junkins HA, *et al.* (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362–9367.



8. Vogel CI, Greene B, Scherag A, *et al.* (2009) Non-replication of an association of CTNBL1 polymorphisms and obesity in a population of Central European ancestry. *BMC Med Genet* **10**, 14.
9. Andreassen CH, Mogensen MS, Borch-Johnsen K, *et al.* (2008) Non-replication of genome-wide based associations between common variants in INSIG2 and PFKF and obesity in studies of 18,014 Danes. *PLoS One* **3**, e2872.
10. Bogardus C (2009) Missing heritability and GWAS utility. *Obesity (Silver Spring)* **17**, 209–210.
11. Frayling TM, Timpson NJ, Weedon MN, *et al.* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894.
12. Loos RJ, Lindgren CM, Li S, *et al.* (2008) Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* **40**, 768–775.
13. Cohen JC, Kiss RS, Pertsemlidis A, *et al.* (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872.
14. Frikke-Schmidt R, Nordestgaard BG, Jensen GB, *et al.* (2004) Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. *J Clin Invest* **114**, 1343–1353.
15. Cheverud JM, Lawson HA, Fawcett GL, *et al.* (2011) Diet-dependent genetic and genomic imprinting effects on obesity in mice. *Obesity (Silver Spring)* **19**, 160–170.
16. Churchill GA, Airey DC, Allayee H, *et al.* (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* **36**, 1133–1137.
17. Threadgill DW, Hunter KW & Williams RW (2002) Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm Genome* **13**, 175–178.
18. Bachmanov AA, Reed DR, Beauchamp GK, *et al.* (2002) Food intake, water intake, and drinking spout side preference of 28 mouse strains. *Behav Genet* **32**, 435–443.
19. Reed DR, Bachmanov AA & Tordoff MG (2007) Forty mouse strain survey of body composition. *Physiol Behav* **91**, 593–600.
20. Wuschke S, Dahm S, Schmidt C, *et al.* (2007) A meta-analysis of quantitative trait loci associated with body weight and adiposity in mice. *Int J Obes (Lond)* **31**, 829–841.
21. Stylianou IM, Korstanje R, Li R, *et al.* (2006) Quantitative trait locus analysis for obesity reveals multiple networks of interacting loci. *Mamm Genome* **17**, 22–36.
22. Peirce JL, Lu L, Gu J, *et al.* (2004) A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet* **5**, 7.
23. Broman KW (2005) The genomes of recombinant inbred lines. *Genetics* **169**, 1133–1146.
24. Haldane JB & Waddington CH (1931) Inbreeding and linkage. *Genetics* **16**, 357–374.
25. Williams RW, Gu J, Qi S, *et al.* (2001) The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biol* **2**, RESEARCH0046.
26. Kirby A, Kang HM, Wade CM, *et al.* (2010) Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. *Genetics* **185**, 1081–1095.
27. McClurg P, Pletcher MT, Wiltshire T, *et al.* (2006) Comparative analysis of haplotype association mapping algorithms. *BMC Bioinformatics* **7**, 61.
28. Payseur BA & Place M (2007) Prospects for association mapping in classical inbred mouse strains. *Genetics* **175**, 1999–2008.
29. Bunker L, Laidlaw A, Bulfield G, *et al.* (2001) Inbred lines of mice derived from long-term growth selected lines: unique resources for mapping growth genes. *Mamm Genome* **12**, 678–686.
30. Svenson KL, Von Smith R, Magnani PA, *et al.* (2007) Multiple trait measurements in 43 inbred mouse strains capture the phenotypic diversity characteristic of human populations. *J Appl Physiol* **102**, 2369–2378.
31. Rocha JL, Eisen EJ, Van Vleck LD, *et al.* (2004) A large-sample QTL study in mice: II. Body composition. *Mamm Genome* **15**, 100–113.
32. Rocha JL, Eisen EJ, Van Vleck LD, *et al.* (2004) A large-sample QTL study in mice: I. Growth. *Mamm Genome* **15**, 83–99.
33. Allan MF, Eisen EJ & Pomp D (2005) Genomic mapping of direct and correlated responses to long-term selection for rapid growth rate in mice. *Genetics* **170**, 1863–1877.
34. Ankra-Badu GA, Pomp D, Shriner D, *et al.* (2009) Genetic influences on growth and body composition in mice: multi-locus interactions. *Int J Obes (Lond)* **33**, 89–95.
35. Kramer MG, Vaughn TT, Pletscher LS, *et al.* (1998) Genetic variation in body weight gain and composition in the intercross of large (LG/J) and small (SM/J) inbred strains of mice. *Genet Mol Biol* **21**, 211–218.
36. Cheverud JM, Vaughn TT, Pletscher LS, *et al.* (2001) Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. *Mamm Genome* **12**, 3–12.
37. Hrbek T, de Brito RA, Wang B, *et al.* (2006) Genetic characterization of a new set of recombinant inbred lines (LGXSM) formed from the inter-cross of SM/J and LG/J inbred mouse strains. *Mamm Genome* **17**, 417–429.
38. Ashrafi K, Chang FY, Watts JL, *et al.* (2003) Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**, 268–272.
39. Boyko AR, Quignon P, Li L, *et al.* (2010) A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol* **8**, e1000451.
40. Darvasi A & Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199–1207.
41. Fawcett GL, Jarvis JP, Roseman CC, *et al.* (2010) Fine-mapping of obesity-related quantitative trait loci in an F(9/10) advanced intercross line. *Obesity (Silver Spring)* **18**, 1383–1392.
42. Valdar W, Solberg LC, Gauguier D, *et al.* (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* **38**, 879–887.
43. Mott R, Talbot CJ, Turri MG, *et al.* (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci U S A* **97**, 12649–12654.
44. Aylor DL, Valdar W, Mathes WF, *et al.* (2011) Genetic analysis of complex traits in the emerging collaborative cross. *Genome Res* (epublication ahead of print version 15 March 2011).
45. Roberts A, Pardo-Manuel de Villena F, Wang W, *et al.* (2007) The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. *Mamm Genome* **18**, 473–481.
46. Morahan G, Balmer L & Monley D (2008) Establishment of 'The Gene Mine': a resource for rapid identification of complex trait genes. *Mamm Genome* **19**, 390–393.
47. Valdar W, Flint J & Mott R (2006) Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* **172**, 1783–1797.

48. Mathes WF, Aylor DL, Miller DR, *et al.* (2011) Architecture of energy balance traits in the emerging lines of the collaborative cross. *American Journal of Applied Physiology: Endocrinology and Metabolism* (epub ahead of press, March 22, 2011).
49. Suto J (2008) Coincidence of loci for glucosuria and obesity in type 2 diabetes-prone KK-Ay mice. *Med Sci Monit* **14**, CR65–CR74.
50. Suto J, Matsuura S, Imamura K, *et al.* (1998) Genetics of obesity in KK mouse and effects of A(y) allele on quantitative regulation. *Mamm Genome* **9**, 506–510.
51. United States Department of Health and Human Services (1996) *Physical Activity and Health: A Report of the Surgeon General*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.
52. Anonymous (1998) Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults – The Evidence Report. National Institutes of Health. *Obesity Res* **6**, Suppl. 2, 51S–209S.
53. Church TS & Blair SN (2009) When will we treat physical activity as a legitimate medical therapy...even though it does not come in a pill? *Br J Sports Med* **43**, 80–81.
54. Duke J, Huhman M & Heitzler C (2003) Physical activity levels among children aged 9–13 years – United States. *Morb Mortal Wkly Rep* **52**, 785–788.
55. Lightfoot JT, Turner MJ, Daves M, *et al.* (2004) Genetic influence on daily wheel running activity level. *Physiol Genomics* **19**, 270–276.
56. Mitchell JA, Church TS, Rankinen T, *et al.* (2010) FTO genotype and the weight loss benefits of moderate intensity exercise. *Obesity (Silver Spring)* **18**, 641–643.
57. Lightfoot JT, Turner MJ, Pomp D, *et al.* (2008) Quantitative trait loci for physical activity traits in mice. *Physiol Genomics* **32**, 401–408.
58. Leamy LJ, Pomp D & Lightfoot JT (2010) A search for quantitative trait loci controlling within-individual variation of physical activity traits in mice. *BMC Genet* **11**, 83.
59. Leamy LJ, Pomp D & Lightfoot JT (2009) Genetic variation for body weight change in mice in response to physical exercise. *BMC Genet* **10**, 58.
60. Leamy LJ, Pomp D & Lightfoot JT (2009) Genetic variation in the pleiotropic association between physical activity and body weight in mice. *Genet Sel Evol* **41**, 41.
61. Nehrenberg DL, Wang S, Hannon RM, *et al.* (2010) QTL underlying voluntary exercise in mice: interactions with the 'mini muscle' locus and sex. *J Hered* **101**, 42–53.
62. Swallow JG, Hayes JP, Koteja P, *et al.* (2009) Selection experiments and experimental evolution of performance and physiology. In *Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments*, pp. 301–352 [TJ Garland Jr and MR Rose, editors]. Berkeley, CA: University of California Press.
63. Kelly SA, Nehrenberg DL, Pierce JL, *et al.* (2010) Genetic architecture of voluntary exercise in an advanced intercross line of mice. *Physiol Genomics* **42**, 190–200.
64. Kelly SA, Nehrenberg DL, Hua K, *et al.* (2010) Parent-of-origin effects on voluntary exercise levels and body composition in mice. *Physiol Genomics* **40**, 111–120.
65. Lightfoot JT, Leamy L, Pomp D, *et al.* (2010) Strain screen and haplotype association mapping of wheel running in inbred mouse strains. *J Appl Physiol* **109**, 623–634.
66. Benson AK, Kelly SA, Legge R, *et al.* (2010) Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107**, 18933–18938.
67. Turnbaugh PJ & Gordon JI (2009) The core gut microbiome, energy balance and obesity. *J Physiol* **587**, 4153–4158.
68. Tannock GW (2007) What immunologists should know about bacterial communities of the human bowel. *Semin Immunol* **19**, 94–105.
69. Kuczynski J, Costello EK, Nemergut DR, *et al.* (2010) Direct sequencing of the human microbiome readily reveals community differences. *Genome Biol* **11**, 210.
70. Ley RE, Peterson DA & Gordon JI (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837–848.
71. Fava F, Lovegrove JA, Gitau R, *et al.* (2006) The gut microbiota and lipid metabolism: implications for human health and coronary heart disease. *Curr Med Chem* **13**, 3005–3021.
72. Wen L, Ley RE, Volchkov PY, *et al.* (2008) Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature* **455**, 1109–1113.
73. Frank DN, St Amand AL, Feldman RA, *et al.* (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* **104**, 13780–13785.
74. Deloris Alexander A, Orcutt RP, Henry JC, *et al.* (2006) Quantitative PCR assays for mouse enteric flora reveal strain-dependent differences in composition that are influenced by the microenvironment. *Mamm Genome* **17**, 1093–1104.
75. Khachatryan ZA, Ktsoyan ZA, Manukyan GP, *et al.* (2008) Predominant role of host genetics in controlling the composition of gut microbiota. *PLoS One* **3**, e3064.
76. Friswell MK, Gika H, Stratford IJ, *et al.* (2010) Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice. *PLoS One* **5**, e8584.
77. Loh G, Brodziak F & Blaut M (2008) The Toll-like receptors TLR2 and TLR4 do not affect the intestinal microbiota composition in mice. *Environ Microbiol* **10**, 709–715.
78. Vijay-Kumar M, Aitken JD, Carvalho FA, *et al.* (2010) Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231.
79. Jones BV, Begley M, Hill C, *et al.* (2008) Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc Natl Acad Sci U S A* **105**, 13580–13585.
80. Ley RE, Backhed F, Turnbaugh P, *et al.* (2005) Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* **102**, 11070–11075.
81. Turnbaugh PJ, Ley RE, Mahowald MA, *et al.* (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031.