



# Dark matter: are mice the solution to missing heritability?

Clarissa C. Parker<sup>1</sup> and Abraham A. Palmer<sup>1,2\*</sup>

<sup>1</sup> Department of Human Genetics, The University of Chicago, Chicago, IL, USA

<sup>2</sup> Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago IL, USA

## Edited by:

Theodore Garland, University of California Riverside, USA

## Reviewed by:

Bret Payseur, University of Wisconsin – Madison, USA  
Scott A. Kelly, University of North Carolina Chapel Hill, USA

## \*Correspondence:

Abraham A. Palmer, Department of Human Genetics, University of Chicago, 920 E 58th Street, CLSC-507D, Chicago, IL 60637, USA.  
e-mail: apalmer@bsd.uchicago.edu

Genome-wide association studies (GWAS) in humans have identified hundreds of single nucleotide polymorphisms associated with complex traits, yet for most traits studied, the sum total of all these identified variants fail to explain a significant portion of the heritable variation. Reasons for this “missing heritability” are thought to include the existence of rare causative variants not captured by current genotyping arrays, structural variants that go undetected by existing technology, insufficient power to identify multi-gene interactions, small sample sizes, and the influence of environmental and epigenetic effects. As genotyping technologies have evolved it has become inexpensive and relatively straightforward to perform GWAS in mice. Mice offer a powerful tool for elucidating the genetic architecture of behavioral and physiological traits, and are complementary to human studies. Unlike F<sub>2</sub> crosses of inbred strains, advanced intercross lines, heterogeneous stocks, outbred, and wild-caught mice have more rapid breakdown of linkage disequilibrium which allow for increasingly high resolution mapping. Because some of these populations are created using a small number of founder chromosomes they are not expected to harbor rare alleles. We discuss the differences between these mouse populations and examine their potential to overcome some of the pitfalls that have plagued human GWAS studies.

**Keywords:** GWAS, quantitative trait loci, complex traits, forward genetics, advanced intercross lines, heterogeneous stock, outbred mice, wild mice

## INTRODUCTION

Genome-wide association studies (GWAS) revolutionized the field of human genetics and helped to implicate hundreds of polymorphisms underlying diseases and other traits (Altshuler et al., 2008; Hindorff et al., 2009). This success can be attributed to four major factors: (1) the large number of accumulated recombinations present among unrelated human subjects, (2) the catalog of human variants created by the HapMap project, (3) commercial high-throughput single nucleotide polymorphism (SNP) arrays that capture a large proportion of common variation in the genome and, (4) the development of statistical tools and methods for association studies. Nevertheless, there has also been great frustration in the small effect sizes of the alleles identified (Manolio et al., 2009). Even where GWAS was able to find many significant associations, the fraction of the heritable variance that can be explained was distressingly low (Yang et al., 2010). The discrepancy between the variation that is known to be heritable and the fraction that can be explained by GWAS results has been referred to as the “dark matter”. Several explanations have been offered for the apparent existence of this “dark matter,” including additional common as well as rare variants not included in the current genotyping arrays, structural variants that also go undetected by existing array technology, insufficient power to identify gene-by-gene interactions, small sample sizes, and inadequate accounting for environmental and epigenetic effects (Manolio et al., 2009). Model organisms have been used extensively to gain insights about the underlying architecture of complex traits. In this review, we will explore how mice can be used to explore possible explanations for the dark matter.

## BENEFITS OF MOUSE MODELS

Mice offer a powerful tool for elucidating the genetic architecture of behavioral and physiological traits, and are complementary to human studies. By using mouse models, researchers can cross strains with measurable phenotypic differences and quickly and cheaply generate large litters of offspring from a limited number of founder genotypes (Lawson and Cheverud, 2010). In addition, environmental factors can be held constant or systematically varied in order to explore interactions between genotype and environment. The mouse genome has been fully sequenced and annotated (Waterston et al., 2002); and most mouse genes have a human homolog, thus allowing rapid translation of results from mice back to humans. Because mice are an experimental model organism, researchers can perform potentially invasive or stressful procedures not possible in humans. Finally, multiple tools are available for altering the mouse genome; permitting the efficient and precise manipulation of genes and allowing for direct experimental testing of observed statistical associations.

## THE PAST

Forward genetic strategies seek to identify the relationship between genetic polymorphisms and traits. Traditional approaches to the analysis of quantitative traits in mice evolved in an environment where genotyping was the most difficult and expensive step. As a result, populations with low levels of recombination [such as backcrosses (BC), F<sub>2</sub> intercrosses (F<sub>2</sub>), recombinant inbred (RI), congenic, and consomic strains] were chosen in order to minimize the number of markers needed to provide adequate coverage of the entire genome. The confidence intervals for the QTLs identified in

these populations were generally quite large, often encompassing 20–40 cM and contained hundreds of genes (Flint et al., 2005). While such regions can be refined using congenic and consomic mice, many generations of genotyping and phenotyping were required in order to achieve a resolution of ~1 cM. Furthermore, efforts at subsequent dissection were frequently derailed by the discovery that an apparently single QTL of large effect identified in an F<sub>2</sub> study was in fact caused by multiple loci of small effect positioned in the same chromosomal region (Legare et al., 2000; Mott et al., 2000; Cheng et al., 2010; Shao et al., 2010). Thus, previous QTL studies in mice, with several notable exceptions (Fehr et al., 2002; Phelan et al., 2002); have generally failed to identify specific genes.

### THE FUTURE

These challenges have led some to claim that mice are poorly suited for forward genetics studies. In our view, the problem is not that mice are unsuitable models for forward genetic studies, but rather that we and others have been using the wrong sorts of mice. Traditionally, the biggest advantage of QTL mapping with F<sub>2</sub>, BC, RI, congenic, and consomic strains is that they were specifically designed to minimize the number of recombinations and therefore limited the amount of genotyping that was required. However, human GWAS studies have been successful precisely because they take advantage of the large number of accumulated recombinations observed among unrelated human subjects. Recombination degrades the non-random associations between adjacent polymorphisms; these associations between nearby markers are known as linkage disequilibrium (LD). Populations that have been intercrossed for multiple generations accumulate many recombinations, which causes a rapid breakdown of LD between adjacent markers. Thus, only markers that are very close and thus in LD with a functional polymorphism will show a significant association with the trait of interest. Populations with more degraded LD allow for more accurate mapping of QTLs, provided that enough markers are genotyped.

Now, as technologies for genotyping have evolved rapidly over the last decade, it is no longer expensive or difficult to perform GWAS in mice. The same improvements in genotyping technology that have been widely used in human genetics also possess enormous but largely unrealized potential to revolutionize mouse genetics. The advent of the Affymetrix mouse

diversity array (MDA), which is capable of genotyping hundreds of thousands of SNPs, provides the technical ability to perform GWAS in mice (Yang et al., 2009). Next-generation (NextGen) sequencing technology will also be extremely useful for providing genotypes and should reduce the biases surrounding SNP ascertainment. NextGen sequencing will also allow for identification of novel SNPs and structural variants such as insertions, deletions, duplications, and translocations (Edenberg and Liu, 2009; Perez-Enciso and Ferretti, 2010; Ozsolak and Milos, 2011). These technologies would not be useful for standard F<sub>2</sub> crosses, but are helpful when populations with more degraded LD are used.

## BUILDING A BETTER MOUSE: ADVANCED INTERCROSS LINES, HETEROGENEOUS STOCKS, OUTBRED, AND WILD MICE

### ADVANCED INTERCROSS LINES

Darvasi and Soller (1995) first introduced advanced intercross lines (AILs) as a tool for fine mapping small chromosomal regions, rather than as a tool for genome-wide scans. AILs are created by successive generations of pseudo-random mating after the F<sub>2</sub> generation to produce additional recombinations. Each additional generation leads to the accumulation of new recombinations, which further degrades LD between adjacent markers and thus allows for more precise mapping (Table 1). Thus, the mapping utility of AILs improves with the number of generations that have accrued since the F<sub>2</sub> cross. Because AILs are created by crossing two inbred strains; all polymorphic markers are perfectly informative in terms of identifying which inbred strain the region is inherited from. In addition, because only two inbred strains are crossed, the starting allele frequency is 0.5 for all alleles in AILs, and will be 0.5 on average in each subsequent generation because the two inbred strains have made equal contributions to the populations (drift and selection for fitness and fecundity will distort this in any multigenerational population; the first can be mitigated by maintaining a large effective population size). When using AILs, it is essential to incorporate information about relatedness among pedigree members into QTL mapping methods in order to prevent false-positive findings and maximize the phenotype and genotype information (Abney et al., 2000; Cheng et al., 2010). Accounting for relatedness poses some statistical and computational challenges, but they can be overcome (e.g.,

**Table 1 | Comparison chart scoring advanced intercross lines (AIL), heterogeneous stock (HS), outbred (OB), and wild-caught (WC) mice on seven variables researchers may consider before choosing a population to study.**

	AIL	HS	OB	WC
Breakdown of linkage disequilibrium	+	++	+++	+++
Derived from inbred progenitors	+	+	–	–
Presence of rare alleles within the population	–	+	++	+++
Requirement of large breeding colony	+	+	–	–
Prevalence of genetic polymorphisms	+	++	++	+++
Controlled environment	+	+	+	–
Need to account for relatedness	++	++	+	+

Scores lie on a four-point scale ranging from lowest (–) to highest (+++).

Cheng et al., 2010). Several software packages are available that can be used for these analyses (e.g., Kang et al., 2008; Peirce et al., 2008)<sup>1</sup>.

### HETEROGENEOUS STOCKS

Heterogeneous stocks (HS) mice are created by interbreeding more than two (often eight) inbred strains followed by many generations of pseudo-random mating (McClern and Meredith, 1970; Hitzemann et al., 1994). The number of generations of pseudo-random mating determines the degree to which LD from the original founder chromosomes is degraded (Mott and Flint, 2002). To the extent that the polymorphisms in the founder strains are known, haplotype mapping can be performed even at markers that are not directly genotyped (Flint et al., 2005). However, the frequencies of minor alleles may be less favorable in an HS compared to an AIL, and there will tend to be a larger number of QTLs for any given trait, making them individually more difficult to identify. Additionally, random fluctuations in allele frequencies and unrecognized selective pressures during creation and subsequent breeding of HS mice may reduce heterozygosity and lower mapping resolution (Valdar et al., 2006a). Furthermore, an HS, like an AIL, requires maintaining a large breeding colony and accounting for relatedness is also essential when performing GWAS.

### LABORATORY OUTBRED MICE

While many laboratory mice have been deliberately inbred, a number of populations have also been maintained using outbred breeding schemes that maintain large numbers of individuals in each generation and that avoid crosses between closely related individuals (Aldinger et al., 2009; Yalcin et al., 2010). Outbred mice (OB) mice exhibit hybrid vigor with longer life spans, high disease resistance, early fertility, large and frequent litters, low neonatal mortality, rapid growth, and large size (Silver, 1995). The primary advantage for using outbred lines for genetic studies is the ability to achieve mapping resolution at the single gene level (Yalcin et al., 2004, 2010; Aldinger et al., 2009) due to much greater breakdown of LD (Chia et al., 2005; Flint et al., 2005). Most outbred stocks are derived from small and relatively homogeneous populations, suggesting a low proportion of rare alleles. In addition, outbred stocks can be purchased from commercial vendors, which avoids the expense associated with maintaining a AIL or HS colony. Most extant outbred stocks are descendants of the same domesticated/laboratory mice that gave rise to inbred laboratory strains and are therefore easy to handle and are likely to segregate many of the same alleles. Thus, SNPs already identified using inbred populations will likely be informative in outbred stocks as well (Yalcin et al., 2010). However, it is possible that many relatively rare variants will exist in these populations as a result of uneven sampling for the foundation population, genetic drift, and new mutations; thus some of the problems familiar to human genetics will also exist when performing GWAS using outbred stocks. It is also important to note that the somewhat limited genetic diversity of outbred colonies means that they (like the AIL and HS mice) cannot be used to analyze the effects of all variants that potentially

exist across *Mus musculus* (Yalcin et al., 2010). Because some parts of the genome may possess little functional variation it might be necessary to study multiple populations to capture the full spectrum of genetic variation (Roberts et al., 2007). Yalcin et al. (2010) reported the finest mapping resolution achieved in an OB colony was approximately twice that obtainable in human populations; the decay of LD ranges from 0.53 to 4.06 Mb.

### WILD-CAUGHT AND WILD-DERIVED MICE

An alternative to using laboratory mice is to capture wild mice. These mice might be directly phenotyped and genotyped, or else they might be bred for a few generations in captivity, in an effort to normalize the uncontrolled environmental variance associated with wild-caught (WC) animals. One advantage to using WC or recent progeny of WC mice (hereafter referred to as “wild-derived” or WD mice) for GWAS is that their genetic makeup has evolved in response to natural conditions (Harper, 2008). Furthermore, WC mice show LD at rates similar to human populations, making them favorable for high resolution association studies (Laurie et al., 2007). In addition, WC and WD mice may have qualitatively different genetic diversity than is available in laboratory mice. Due to their shared ancestry, the genomes of most laboratory mice (including AIL, HS, and OB) are 92% *M. m. domesticus* in origin, and display non-random distributions of diversity across the chromosomes (Yang et al., 2007). WC and WD mice from subspecies *M. m. musculus*, *M. m. castaneus*, or *M. m. domesticus* would be expected to have more uniform patterns of genetic variation (Roberts et al., 2007).

Wild populations of mice are likely to contain large numbers of rare variants that contribute to phenotypic variation; thus, even large studies will not have enough power to identify these rare variants. This is a serious problem that is being addressed in the field of human genetics, but that mouse geneticists might hope to avoid by using laboratory strains. Studies utilizing wild mice face unique problems of their own: (1) trapping and transporting large numbers of wild mice for the purposes of genetic mapping studies is likely to be expensive and difficult, (2) accounting for relatedness in WC populations, (3) environmental variation (nutrition, age, life experience, infections, etc.), (4) WC mice are likely to harbor pathogens and therefore cannot be imported into a standard specific pathogen free mouse vivariums, (5) many standard phenotyping procedures (including most behavioral assays) are extremely difficult to perform in WD and WC mice (Wahlsten et al., 2003; Chesler et al., 2008), and (6) many WC mice do not reproduce well in captivity, requiring a larger than normal founder population to preserve genetic heterogeneity in the offspring (Harper, 2008).

## PITFALLS OF GWAS IN HUMANS THAT CAN BE ADDRESSED BY MOUSE MODELS

### RARE VARIANTS

The underlying rationale for GWAS is the common disease, common variant hypothesis, which asserts that common diseases are most likely attributable to allelic variants present in >~5% of the population (Collins et al., 1997; Pritchard, 2001). Most common diseases and quantitative traits have heritability estimates between 30 and 90% (Visscher, 2008), yet the majority of loci identified by GWAS account for less than 1% of population variation. This

<sup>1</sup><http://cran.r-project.org/web/packages/QTLRel/index.html>

has led to researchers to focus on the possible contribution of rare variants (minor allele frequency, or MAF <0.5%). While these variants are individually rare, they are collectively common among human populations (Sebat et al., 2004; Conrad et al., 2010) and contribute to disease risk (Cohen et al., 2004; Ji et al., 2008; Meford and Eichler, 2009; Eichler et al., 2010). Unfortunately power is very low for detecting rare alleles, unless the effect size is very large (Manolio et al., 2009). The main advantage of mice in this regard is the ability to select a population where recombination is high, but the presence of rare alleles is low to non-existent (e.g., AIL, HS), thus sidestepping the problems caused by rare alleles. Some alleles segregating in these populations might have been common in the ancestral wild populations, while others might be rare, but in both cases such alleles will be common in an AIL or HS due to the bottle neck imposed during the creation of these populations.

### EPISTASIS

The interaction between alleles at two or more loci is known as epistasis. Many have argued that epistasis is likely to be a common factor in determining phenotypes, yet almost all mouse and human studies instead focus on the detection of main effects (McCarthy and Hirschhorn, 2008; Moore and Williams, 2009). One reason for this is the heavy computational burden imposed by searching for interactions. In order to detect epistatic interactions, a large number of pair-wise tests for marker–marker interactions must be performed. More important than the computational burden is the more stringent significance threshold that is required when so many tests are performed. The power to detect epistasis is also dependent on the frequency of the rarest genotype in the putative epistatic interaction. In humans, the frequency of the rare double homozygous group may be extremely low due to rare alleles, in which case the power of the test for epistasis is extremely poor even with a very large sample size. By selecting a mouse population where no alleles are especially rare, the power to detect epistatic interactions will be improved. As a result, investigating epistasis should be much more powerful in populations like AILs. For example, using pedigree information in an AIL, numerous epistatic interactions have been successfully identified in 1500 F<sub>9/10</sub> LG/J × SM/J mice contributing to variation in obesity-related traits (Cheverud et al., 2001; Fawcett et al., 2010). Two-locus interaction models indicated the presence of epistasis, and further methodological improvements have led to the identification of even more epistatic QTLs in the same AIL cross (Yi et al., 2007). However, it has been suggested that fear-related behavior in HS mice was mostly influenced by alleles with additive (not epistatic) effects (Flint et al., 2004; Valdar et al., 2006a). AILs preserve the power to examine epigenetic effects due to a higher frequency of the minor allele (usually ~50%); thus, they may be more suitable for exploring epistatic interactions than HS. Differences between AIL and HS mice may also be due to differences in the analysis used as well as possible differences in the underlying genetic architecture of the traits examined, effects of selection, population history, and mutation rates (Eichler et al., 2010).

### GENE-BY-ENVIRONMENT INTERACTIONS

It is well accepted that the size and even direction of genetic effects on a phenotype can vary across different environments, a

phenomenon known as gene-by-environment interaction (GEI). GEIs have been extensively documented for most phenotypes for which they were examined in mice (Crabbe et al., 1999; Wahlsten et al., 2006), and have also been shown to be highly important for complex traits in humans (Caspi et al., 2002, 2003). While context-dependent genetic effects may be pervasive, a recent simulation study suggests that astronomically large sample sizes would be needed to robustly and reliably detect GEIs and the extent to which they influence complex traits in human populations (Munafò et al., 2009). Thus, one of the major reasons why GEIs are not discovered in human GWAS is a lack of statistical power, especially if only a small portion of subjects experience the adverse exposure (Eichler et al., 2010). Furthermore, it is often neither possible nor ethical to manipulate environmental influences in humans, and identifying what the relevant environments are can be quite difficult.

One of the attractions of mice is the ability to create a uniform environment such that alleles can be studied in the absence of environmental perturbations. Alternatively, analysis of gene-by-environment interactions can be explicitly performed in mice. For example, in a population of 2448 HS mice, Valdar et al. (2006b) estimated the heritability of a multitude of behavioral and physiological phenotypes, assessed the impact of numerous environmental factors, and measured the size of GEIs. More recently, Lawson et al. (2011) examined the effects of sex and diet on diabetes-related traits in 1002 F<sub>16</sub> LG/J × SM/J AIL mice. They demonstrated that the effects of genetic variation on diabetes-related traits were highly context-dependent and that the same alleles influenced traits differently depending on the sex and/or dietary environment of the individual. The results of these studies clearly demonstrate that gene detection studies suffer when environmental effects are not adequately accounted for in the experimental design and analysis. While these issues can be studied in more traditional populations like F<sub>2</sub> crosses, RI lines, and panels of inbred strains, highly recombinant populations such as AIL, HS, outbred, and WD mice provide sufficient breakdown in LD to efficiently identify the specific genes that interact with environmental influences.

### EPIGENETIC EFFECTS

In addition to epistasis and GEIs, epigenetic effects may also explain some of the missing inherited variance. Epigenetic effects are transmitted trans-generationally without changes to the DNA sequence; mechanisms may include methylation, genomic imprinting, or histone modifications. These factors cause heritable changes in gene expression in the absence of any DNA mutation, and may be important for human disease (Petronis et al., 2000; Bjornsson et al., 2004; Weaver et al., 2004). Since GWAS studies focus on DNA sequence, epigenetic factors are potentially confounding. AILs and HS mice are very useful for examining trans-generational and molecular epigenetic effects by studying multiple generations of mice with known pedigrees, and by easy access to relevant tissue in which to analyze DNA methylation. For example, Morgan et al. (1999) demonstrated in mice that that changes in DNA methylation could be generated via alterations in diet and then passed on to subsequent offspring even after the dietary modification had been eliminated. Cheverud et al. (2011) mapped genome-wide significant genomic imprinting

effects (which some would argue represent an epigenetic process) in 1002 LG/J  $\times$  SM/J F<sub>16</sub> AIL mice fed both high and low-fat diets; and reported that inconsistent imprinting patterns exist across all genotypes and environments.

### eQTLs

One of the attractions of performing genetic screens in animals is the ability to obtain tissue for gene expression analysis. In the context of GWAS studies, this allows for the identification of eQTLs, which may be molecular intermediates between DNA-level polymorphisms and more complex traits. Nicolae et al. (2010) argue that connecting SNPs with information on gene expression provides a biologically relevant means to identify those associations likely to be replicated while also imparting a better understanding of the biological mechanisms driving the associations. Steps to obtain relevant human tissues for the purpose of eQTL mapping are making progress<sup>2</sup>; unfortunately, such efforts must often utilize sub-optimal tissues that have become available either post-mortem or as a byproduct of surgery. Obtaining brain tissue is especially challenging because people with psychiatric conditions have typically received psychotropic drugs for decades, which is very likely to confound the analysis of gene expression. Mouse models allow the experimenter to obtain tissue samples from mice that have been raised in homogenous environmental conditions, while employing potentially dangerous or invasive procedures that would not be permissible in human subjects (Huang et al., 2009).

### SAMPLE SIZE

A notable difference between GWAS in mice and humans is sample size. In humans, it is now widely recognized that the genetic effect size attributable to a single genetic variant underlying a complex trait or disease is likely to be very small. As a result, studies in human populations now routinely require tens of thousands of participants. Failure to obtain sufficient sample sizes may result in low statistical power, which will increase the ratio of false-positives to true positives (Sterne and Davey Smith, 2001). However, the sample size needed for GWAS in mice is far lower for several reasons. First, in many cases there is more LD among mice than in unrelated humans; while this is bad for mapping precision, it decreases the effective number of tests and thus thresholds for significance are less stringent. Second, the fraction of the total phenotypic variance explained by genetic variation depends on both allele and genotype frequency. Use of simpler systems (such as AIL, HS, and OB mice) will permit smaller sample sizes because the genotypic value of each strain can be measured with higher accuracy. AILs require the fewest number of subjects, due to the smaller number of causal alleles within the population and resulting larger

effect sizes and the more balanced allele frequencies. Because HS and OB mice are derived from more than two inbred strains, they have greater genetic variance contributing to a given trait. As a result, the effect sizes in these populations are smaller, and studies using HS or OB mice need greater numbers of subjects in order to detect them (Flint and Mackay, 2009). In the case of WC and WD mice, factors familiar to human genetics such as highly degraded LD, rare alleles, and uncontrolled environmental variance suggest that sample sizes may need to be similar to those used in human GWAS.

### CONCLUSION

The use of highly recombinant mouse populations offers the ability to discover the DNA sequence variation that gives rise to quantitative phenotypic variation. Mouse populations that do not have rare variants may be useful in understanding the reasons for the “missing heritability” observed in human GWAS. In addition, environmental variation that might lead to GEI can be controlled or systematically varied. Populations without rare alleles may also be useful for investigating the importance of epistasis. Thus mice may be a useful complement to human genetic studies.

How exactly will the identification of genes in mice assist in our understanding of human genetics? First, some of the genes that influence a trait in mice will surely also influence the analogous trait in humans. Such genes could then be examined in existing human GWAS datasets. By approaching these datasets with strong prior hypotheses it might be possible to accept associations that do not obtain genome-wide significance. Along the same lines, attention is quickly shifting from an effort to identify associations with common polymorphisms to the search for rare variants (Asimit and Zeggini, 2010; Bacanu et al., 2011; Neale et al., 2011). Because rare variants are prevalent as a class, but individually rare, it is statistically challenging to establish the role of rare variants, even once the hurdle of cost-efficient genome-wide re-sequencing is overcome. Therefore, prior hypotheses about specific genes, as would be developed in mouse models, would be very useful for knowing which genes and pathways to test for rare variants. Finally, genes identified in mice can be used for drug development and might give rise to novel therapeutic approaches – this potential benefit is independent from the goal of facilitating human genetic studies, but may yield the greatest practical benefits.

### ACKNOWLEDGMENTS

This work was supported by NIH grants DA021336, DA027545, DA024845 MH079103 to Abraham A. Palmer and DA007255 to Clarissa C. Parker.

<sup>2</sup><http://commonfund.nih.gov/GTEX/index.aspx>

### REFERENCES

- Abney, M., McPeck, M. S., and Ober, C. (2000). Estimation of variance components of quantitative traits in inbred populations. *Am. J. Hum. Genet.* 66, 629–650.
- Aldinger, K. A., Sokoloff, G., Rosenberg, D. M., Palmer, A. A., and Millen, K. J. (2009). Genetic variation and population substructure in outbred CD-1 mice: implications for genome-wide association studies. *PLoS ONE* 4, e4729. doi: 10.1371/journal.pone.0004729
- Altshuler, D., Daly, M. J., and Lander, E. S. (2008). Genetic mapping in human disease. *Science* 332, 881–888.
- Asimit, J., and Zeggini, E. (2010). Rare variant association analysis methods for complex traits. *Annu. Rev. Genet.* 44, 293–308.
- Bacanu, S. A., Nelson, M. R., and Whittaker, J. C. (2011). Comparison of methods and sampling designs to test for association between rare variants and quantitative traits. *Genet. Epidemiol.* 35, 226–235.
- Bjornsson, H. T., Fallin, M. D., and Feinberg, A. P. (2004). An integrated epigenetic and genetic approach to common human disease. *Trends Genet.* 20, 350–358.
- Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A.,

- and Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science* 297, 851–854.
- Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Brithwaite, A., and Poulton, R. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301, 386–389.
- Cheng, R., Lim, J. E., Samocha, K. E., Sokoloff, G., Abney, M., Skol, A. D., and Palmer, A. A. (2010). Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. *Genetics* 185, 1033–1044.
- Chesler, E. J., Miller, D. R., Branstetter, L. R., Galloway, L. D., Jackson, B. L., Phillip, V. M., Voy, B. H., Culiati, C. T., Threadgill, D. W., Williams, R. W., Churchill, G. A., Johnson, D. K., and Manly, K. F. (2008). The collaborative cross at oak ridge national laboratory: developing a powerful resource for systems genetics. *Mamm. Genome* 19, 382–389.
- Cheverud, J. M., Lawson, H. A., Fawcett, G. L., Wang, B., Pletscher, S., Fox, A. R., Maxwell, T. J., Ehrlich, T. H., Kenney-Hunt, J. P., Wolf, J. B., and Semenovich, C. F. (2011). Diet-dependent genetic and genomic imprinting effects on obesity in mice. *Obesity (Silver Spring)* 19, 160–170.
- Cheverud, J. M., Vaughn, T. T., Pletscher, L. S., Peripato, A. C., Adams, E. S., Erikson, C. F., and King-Ellison, K. J. (2001). Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. *Mamm. Genome* 12, 3–12.
- Chia, R., Achilli, F., Festing, M. F., and Fisher, E. M. (2005). The origins and uses of mouse outbred stocks. *Nat. Genet.* 37, 1181–1186.
- Cohen, J. C., Kiss, R. S., Pertsemlidis, A., Marcel, Y. L., McPherson, R., and Hobbs, H. H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869–872.
- Collins, F. S., Guyer, M. S., and Charkravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation. *Science* 278, 1580–1581.
- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., MacArthur, D. G., Macdonald, J. R., Onyiah, I., Pang, A. W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Wellcome Trust Case Control Consortium, Tyler-Smith, C., Carter, N. P., Lee, C., Scherer, S. W., and Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–712.
- Crabbe, J. C., Wahlsten, D., and Dudek, B. C. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science* 284, 1670–1672.
- Darvasi, A., and Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141, 1199–1207.
- Edenberg, H. J., and Liu, Y. (2009). Laboratory methods for high-throughput genotyping. *Cold Spring Harb. Protoc.* 11, pdb.top62.
- Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., and Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex diseases. *Nat. Rev. Genet.* 11, 446–450.
- Fawcett, G. L., Jarvis, J. P., Roseman, C. C., Wang, B., Wolf, J. B., and Cheverud, J. M. (2010). Fine-mapping of obesity-related quantitative trait loci in an F9/10 advanced intercross line. *Obesity (Silver Spring)* 18, 1383–1392.
- Fehr, C., Shirley, R. L., Belknap, J. K., Crabbe, J. C., and Buck, K. J. (2002). Congenic mapping of alcohol and pentobarbital withdrawal liability loci to a <1 centimorgan interval of murine chromosome 4: identification of Mpdz as a candidate gene. *J. Neurosci.* 22, 3730–3738.
- Flint, J., DeFries, J. C., and Henderson, N. D. (2004). Little epistasis for anxiety-related measures in the DeFries strains of laboratory mice. *Mamm. Genome* 15, 77–82.
- Flint, J., and Mackay, T. F. C. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* 19, 723–733.
- Flint, J., Valdar, W., Shifman, S., and Mott, R. (2005). Strategies for mapping and cloning quantitative trait genes in rodents. *Nat. Rev. Genet.* 6, 271–286.
- Harper, J. M. (2008). Wild-derived mouse stocks: an underappreciated tool for aging research. *Age (Omaha)* 30, 135–145.
- Hindorf, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., and Manolio, T. A. (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.
- Hitzemann, B., Dains, K., and Hitzemann, R. (1994). Further studies on the relationship between dopamine cell density and haloperidol response. *J. Pharmacol. Exp. Ther.* 271, 969–976.
- Huang, G. J., Shifman, S., Valdar, W., Johannesson, M., Yalcin, B., Taylor, M. S., Taylor, J. M., Mott, R., and Flint, J. (2009). High resolution mapping of expression QTLs in heterogeneous stock mice in multiple tissues. *Genome Res.* 19, 1133–1140.
- Ji, W., Foo, J. N., O’Roak B. J., Zhao, H., Larson, M. G., Simon, D. B., Newton-Cheh, C., State, M. W., Levy, D., and Lifton, R. P. (2008). Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat. Genet.* 40, 592–599.
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723.
- Laurie, C. C., Nickerson, D. A., Anderson, A. D., Weir, B. S., Livingston, R. J., Dean, M. D., Smith, K. L., Schadt, E. E., and Nachman, M. W. (2007). *PLoS Genet.* 3, e144. doi: 10.1371/journal.pgen.0030144
- Lawson, H. A., and Cheverud, J. M. (2010). Metabolic syndrome components in murine models. *Endocr. Metab. Immune Disord. Drug Targets* 10, 25–40.
- Lawson, H. A., Lee, A., Fawcett, G. L., Wang, B., Pletscher, L. S., Maxwell, T. J., Ehrlich, T. H., Kenney-Hunt, J. P., Wolf, J. B., Semenovich, C. F., and Cheverud, J. M. (2011). The importance of context to the genetic architecture of diabetes-related traits is revealed in a genome-wide scan of a LG/J x SM/J murine model. *Mamm. Genome* 22, 197–208.
- Legare, M. E., Bartlett, F. S. II, and Frankel, W. N. (2000). A major effect QTL determined by multiple genes in epileptic EL mice. *Genome Res.* 10, 42–48.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., N. S., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarrroll, S. A., and Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753.
- McCarthy, M. I., and Hirschhorn, J. N. (2008). Genome-wide association studies: potential next steps on a genetic journey. *Hum. Mol. Genet.* 17, R156–R165.
- McClern, G. E. W. Jr., and Meredith, W. (1970). “The use of isogenic and heterogenic mouse stocks in behavioral research,” in *Contributions to Behavior-Genetic Analysis: The Mouse as Prototype*, ed. G. T. Lindzey (New York: Appleton Century Crofts), 3–22.
- Mefford, H. C., and Eichler, E. E. (2009). Duplication hotspots, rare genomic disorders, and common disease. *Curr. Opin. Genet. Dev.* 19, 196–204.
- Moore, J. H., and Williams, S. M. (2009). Epistasis and its implications for personal genetics. *Am. J. Hum. Genet.* 85, 309–320.
- Morgan, H. D., Sutherland, H. G., Martin, D. I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318.
- Mott, R., and Flint, J. (2002). Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. *Genetics* 160, 1609–1618.
- Mott, R., Talbot, C. J., Turri, M. C., Collins, A. C., and Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12649–12654.
- Munafo, M. R., Durrant, C., Lewis, G., and Flint, J. (2009). Gene X environment interactions at the serotonin transporter locus. *Biol. Psychiatry* 65, 211–219.
- Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S. M., Roeder, K., and Daly, M. J. (2011). Testing for an unusual distribution of rare variants. *PLoS Genet.* 7, e1001322. doi: 10.1371/journal.pgen.1001322
- Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., and Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 6, e1000888. doi: 10.1371/journal.pgen.1000888
- Ozsolak, E., and Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98.
- Pearce, J. L., Broman, K. W., Lu, L., Chesler, E. J., Zhou, G., Airey, D. C., Birmingham, A. E., and Williams, R. W. (2008). Genome reshuffling for advanced intercross permutation (GRAIP): a permutation method for quantitative advanced intercross population analysis. *PLoS ONE* 3, e1977. doi: 10.1371/journal.pone.0001977

- Perez-Enciso, M., and Ferretti, L. (2010). Massive parallel sequencing in animal genetics: wherefroms and wheretos. *Anim. Genet.* 41, 561–569.
- Petronis, A., Gottesman, I. I., Crow, T. J., DeLisi, L. E., Klar, A. J., Macciardi, F., McInnis, M. G., McMahon, F. J., Paterson, A. D., Skuse, D., and Sutherland, G. R. (2000). Psychiatric epigenetics: a new focus for the new century. *Mol. Psychiatry* 5, 342–346.
- Phelan, S. A., Beier, D. R., Higgins, D. C., and Paigen, B. (2002). Confirmation and high resolution mapping of an atherosclerosis susceptibility gene in mice on chromosome 1. *Mamm. Genome* 13, 548–553.
- Pritchard, J. K. (2001). Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* 69, 124–137.
- Roberts, A., Pardo-Manuel de Villena, F., Wang, W., McMillan, L., and Threadgill, D. W. (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.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask, B., Patterson, N., Zetterberg, A., and Wigler, M. (2004). Large-scale copy number variants and hotspots of human genetic disease. *Science* 305, 525–528.
- Shao, H., Sinasac, D. S., Burrage, L. C., Hodges, C. A., Supelak, P. J., Palmert, M. R., Moreno, C., Cowley, A. W. Jr., Jacob, H. J., and Nadeau, J. H. (2010). Analyzing complex traits with congenic strains. *Mamm. Genome* 21, 276–286.
- Silver, L. M. (1995). *Mouse Genetics: Concepts and Applications*. New York: Oxford University Press.
- Sterne, J. A., and Davey Smith, G. (2001). Sifting the evidence—what's wrong with significance tests? *Br. Med. J.* 322, 226–231.
- Valdar, W., Solberg, L. C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W. O., Taylor, M. S., Rawlins, J. N., Mott, R., and Flint, J. (2006a). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* 38, 879–887.
- Valdar, W., Solberg, L. C., Gauguier, D., Cookson, W. O., Rawlins, J. N., Mott, R., and Flint, J. (2006b). Genetic and environmental effects on complex traits in mice. *Genetics* 174, 959–984.
- Visscher, P. M. (2008). Sizing up human height variation. *Nat. Genet.* 40, 489–490.
- Wahlsten, D., Bachmanov, A., Finn, D. A., and Crabbe, J. C. (2006). Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16364–16369.
- Wahlsten, D., Metten, P., and Crabbe, J. C. (2003). A rating scale for wildness and ease of handling laboratory mice: results for 21 inbred strains tested in two laboratories. *Genes Brain Behav.* 2, 71–79.
- Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, E., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyas, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigó, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S. P., Zdobnov, E. M., Zody, M. C., and Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562.
- Weaver, I. C., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., Dymov, S., Szyf, M., and Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nat. Neurosci.* 7, 847–854.
- Yalcin, B., Nicod, J., Bhomra, A., Davidson, S., Cleak, J., Farinelli, L., Ostersas, M., Whitley, A., Yuan, W., Gan, X., Goodson, M., Klenerman, P., Satpathy, A., Mathis, D., Benoist, C., Adams, D. J., Mott, R., and Flint, J. (2010). Commercially available outbred mice for genome-wide association studies. *PLoS Genet.* 2, e1001085. doi: 10.1371/journal.pgen.1001085
- Yalcin, B., Willis-Owen, S. A., Fullerton, J., Meesaq, A., Deacon, R. M., Rawlins, J. N., Copley, R. R., Morris, A. P., Flint, J., and Mott, R. (2004). Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. *Nat. Genet.* 36, 1197–1202.
- Yang, H., Bell, T. A., Churchill, G. A., and Pardo-Manuel de Villena, F. (2007). On the subspecific origin of the laboratory mouse. *Nat. Genet.* 39, 1100–1107.
- Yang, H., Ding, Y., Hutchins, L. N., Szatkiewicz, J., Bell, T. A., Paigen, B. J., Graber, J. H., de Villena, F. P., and Churchill, G. A. (2009). A customized and versatile high-density genotyping array for the mouse. *Nat. Methods* 6, 663–666.
- Yang, J., Beben, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., and Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–571.
- Yi, N., Shrinier, D., Banerjee, S., Mehta, T., Pomp, D., and Yandell, B. S. (2007). An efficient Bayesian model selection approach for interacting quantitative trait loci models with many effects. *Genetics* 176, 1865–1877.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 March 2011; paper pending published: 19 April 2011; accepted: 01 June 2011; published online: 13 June 2011.

Citation: Parker CC and Palmer AA (2011) Dark matter: are mice the solution to missing heritability? *Front. Gene.* 2:32. doi: 10.3389/fgene.2011.00032

This article was submitted to *Frontiers in Genetic Architecture*, a specialty of *Frontiers in Genetics*.

Copyright © 2011 Parker and Palmer. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.