Background
The past three decades of studies have unveiled some of the genetic underpinnings of human disease. For complex diseases, those with obscure genetic roots, discoveries have accelerated recently owing to a bloom of genome-wide association studies (GWASs) [1]. Nevertheless, even for the most successful cases (such as inflammatory and ulcerative bowel disease [2,3]), discoveries account for only a fraction, often small, of the disease’s heritability. These yet to be discovered genetic variants comprise the ‘missing heritability’ or the genetic ‘dark matter’ for disease.

State of dark matter
Heritability, the proportion of trait variability explained by genetic factors, has two somewhat different meanings. Narrow-sense heritability involves only the additive effects of genes. Broad-sense heritability involves both additive and non-additive effects. The difference between the two makes a difference when hunting for dark matter. If genetic variation were all to act additively, the best predictor of an offspring’s trait value would be the average of his/her parents’ values. Human height is an excellent example, after adjusting for gender. Hunting for dark matter for a trait such as human height will be more straightforward than for a disease such as schizophrenia, for which the evidence for substantial gene-gene interaction is compelling [4]. Yet when researchers refer to heritability of human height, they implicitly mean narrow-sense heritability; for schizophrenia, it is heritability in a much broader sense.

Why should we care about the genetic basis of disease? Greater understanding of the genetics equals greater understanding of molecular etiology and, with it, eventually more cogent treatments. However, the origins of some human diseases, especially those of the mind, can be mysterious. For diseases of the mind, few environmental or genetic risk factors are understood; instead the hope is that identified genetic factors will lead to a subtler understanding of why diseases such as schizophrenia arise and how they can be treated effectively. Even for cardiovascular disease, for which environmental risk factors are well characterized, new insights into its genetics could produce more targeted treatment. This leads to the other expectation - that greater genetic knowledge will pave the way for ‘personalized’ medicine. The rapid technological advances in genomics will soon make it feasible to sequence whole genomes at relatively low cost. The idea that each individual will have meaningful sequence variation in their medical records and will have interventions tailored to their risk profile and likely treatment response is quite appealing. The goal of personalized medicine, however, is hindered because so much molecular etiology remains in the dark.

One way to explain more of the dark matter is to develop more efficient ways to use existing data. Naukkarinen et al. [5] develop an innovative approach that integrates gene expression and genotype data. They apply these ideas to a GWAS of obesity, as measured by body mass index (BMI). Studies estimate BMI’s heritability at 45 to 85%, but identified genetic variants...
explain about 1% of the total variance [6]. To discover more variants, the authors [5] examined gene expression of adipose tissue in a sample of monozygotic (MZ) twins discordant for BMI and in a sample of unrelated individuals. Because MZ twins are genetically identical, or nearly so, the authors reasoned that genes showing expression differences between twins are ‘reactive’ genes with differences that are due to regulatory or epigenetic changes in response to environmental factors. By contrast, genes uncovered in unrelated individuals are a combination of reactive and genetically ‘causal’ genes. By contrasting results from the unrelated sample and discordant MZ twins, the authors identified 27 causal genes that were differentially regulated. They then tested 197 single nucleotide polymorphisms (SNPs) falling in and around these genes in a sample of 21,000 subjects. They discovered a significant excess of small P-values in this set of SNPs. Neither the set of SNPs defined by reactive genes nor the individual SNPs in the reactive set were associated with BMI. Notably, this work identifies a new gene, F13A1, which encodes the coagulation factor XIII A chain, with variation that affects BMI. This gene has also been identified by meta-analysis of 12 studies of venous thromboembolism [7]. Obesity is well known to predispose to vein thromboses; however, the study of Naukkarinen et al. [5] reveals a potential biological pathway for the relationship between obesity, thrombosis and cardiovascular risk.

The methods advanced by Naukkarinen and colleagues [5] require discordant MZ twins, which were available for BMI. This experimental design could prove highly informative for similar quantitative traits, for which extremes are easily identified and by which the pathology or phenotype of interest is defined. For some diseases, especially diseases of the brain, quantitative traits that map precisely onto risk are not yet available. In addition, because reactive genes are environment-dependent, successful implementation of this design might require a sample exposed to a homogeneous environment, limiting its generality. Regardless, this study shows how innovative research can cast more light on dark matter. Moreover, the study design could also inform us about pathways of correlated gene expression and how much these correlations are influenced by genetic and environmental variation.

Many other methods and designs are available to illuminate dark matter [8-15]. One appealing approach teams gene-expression results with genome-wide association data to produce targeted hypothesis tests [8]. One possibility is to organize tests by expression quantitative trait loci affecting genes in pathways meaningful for the disease. Statistical methods for targeted testing are available, whether on the basis of prior information of the likelihood of an association between a SNP and the phenotype or on the basis of plausible disease pathways [9,10]. Genetic variants with parental origin effects, or whose effects depend on the parent from whom they were inherited, could be part of the dark matter; methods are now available to determine the parental origin of alleles and haplotypes even in the absence of genotyped parents [13]. Studies of copy number variants and their inheritance in families could also reveal insight into plausible biological pathways for disease [14,15]. It is also safe to say that rare variants account for some of the dark matter [16], possibly the majority of it in some cases. Next-generation sequencing promises to fill some of our void in knowledge by identifying more penetrant but rarer variants.

Other approaches are less illuminating. Let’s reconsider human height. We know numerous rare variants and about 50 common variants that have an impact on height. Thus far, known genetics account for roughly 5% of the variance. Using many SNPs from GWAS analysis that are not significantly associated with height, Yang et al. [17] estimated the proportion of variance in height explained by SNPs as 0.45 and even got close to the heritability estimate of 0.84 after correcting for incomplete linkage disequilibrium between SNPs genotyped and causal variants. In spirit, this approach [17] is similar to the allele score method [18], which seeks a predictive model for disease status on the basis of thousands of SNPs with modest evidence for association. If their results are correct, both studies [17,18] suggest that the effects of SNPs are small and will be difficult or impossible to detect from simple analyses of GWASs, at least for current sample sizes [19]. These intriguing approaches have some drawbacks: they shed no new light on the molecular etiology of phenotype; and inherent in the calculations are assumptions that could prove difficult to validate.

We all recognize the hidden biases that inflate estimates of heritability. There are other complex pathways for the transmission of a phenotype across generations without the transmission of a specific common or rare variant, namely through epigenetic factors that can result in the inheritance of gene expression patterns without an alteration of the DNA sequence [20]. Gene-environment interactions could also affect the estimates of heritability and when they are in play, they can explain as much of the variance in the phenotype as genetic factors [21].

Conclusions
Concerted effort will almost surely be required to understand the genetic architecture of most complex diseases. Naukkarinen et al.’s [5] novel study design illustrates the impact that concerted effort can have in advancing our knowledge of the genetic etiology of such diseases. There remains ample room for novel analytic methods and
study designs to shed light on the genetic dark matter of disease. It is entirely possible, 10 years hence, that we will realize that much of the missing heritability was hiding in plain sight in common variants.

Abbreviations
GWAS, genome-wide association study, MZ, monozygotic, SNP, single nucleotide polymorphism.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
The authors contributed equally to the writing and preparation of this commentary.

Authors' information
BD is Associate Professor of Psychiatry and Human Genetics, University of Pittsburgh School of Medicine, Pittsburgh. His background and area of expertise is psychiatric epidemiology and statistical genetics. NM is Assistant Professor of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh. Her background and areas of expertise are psychiatric epidemiology and statistical genetics.

Acknowledgements
This work was supported by R01 grants (MH057881, BD) and a K01 grant (MH077930, NM) from the National Institute of Mental Health. The National Institute of Mental Health did not participate in the preparation, review or approval of the manuscript.

Published: 21 October 2010

References
1. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al.: Finding the missing heritability of complex diseases. Nature 2009, 461:747-753.
2. McGovern DP, Gardet A, Torkvist L, Goyette P, Essers J, Taylor KD, Neale BM, Ong RT, Lagace C, Li C, Green T, Stevens CR, Beauchamp C, Flesher PR, Carlson M, D'Amato M, Hallvarsson J, Hibberd ML, Lordard M, Padyukov L, Andriulli A, Colombi E, Laitano A, Palmiiero O, Bernard EJ, Delabidiere C, Hommes DW, de Jong DJ, Stockers PC, Weerom R, et al.: Genome-wide association identifies multiple loci associating with common susceptibility loci. Nat Genet 2010, 42:332-337.
3. Imielinski M, Baldwin SN, Griffiths A, Russell RK, Annese V, Dubinsky M, Griffiths A, Baldassano RN, Griffige M, Leal SM, Moore JH, Nadeau JH: Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 2010, 11:446-450.
4. Undersanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 2010, 464:768-772.
5. Wells PS, Anderson JL, Scavellis DK, Doucette SP, Gagnon F: Factor XIII Val34Leu variant is protective against venous thromboembolism: a HuGE review and meta-analysis. Am J Epidemiol 2006, 164:101-109.
6. Weil MI, Surakka I, Penailken KH, Rissanen A, Salomaa V, Ripatti S, Yki-Järvinen H, van Duijn CM, Wichmann HE, Kaprio J, Taskinen MR, Peltomaki P, ENGAGE Consortium: Use of genome-wide expression data to mine the “Gray Zone” of GWA studies leading to novel candidate obesity genes. PLoS Genet 2010, 6:e1000976.
7. Reed LK, Williams S, Springer M, Brown J, Freeman K, DesRoches CE, Sokolowski MB, Gibson G: Genotype-by-diet interactions drive metabolic phenotype variation in Drosophila melanogaster. Genetcs 2010, 185:1009-1019.