Published in final edited form as:

Curr Opin Clin Nutr Metab Care. 2008 July ; 11(4): 366–370. doi:10.1097/MCO.0b013e3283034990.

The fat tail of obesity as told by the genome

Alan Herbert*

Abstract

Purpose of review—Many genes affect pathways that predispose to and protect against obesity. We ask how many different variants affect human obesity and how common are they?

Recent findings—The current generation of genome-wide association scans is moderately powered to detect and replicate associations between single nucleotide polymorphisms, or common copy number variations and common diseases. They are not designed either to find rare germline variants or those somatic changes, unique to an individual, that arise with high frequency in adult stem cells. They do not directly assay the epigenetic reprogramming of outcomes related to maternal or environmental exposures.

Summary—There are more gene variants, more gene-gene and gene-environmental interactions leading to obesity than current GWAS studies can validate. Those genetic associations that can be verified provide valuable insight into the pathways contributing to human obesity.

Keywords

obesity; genome wide association scan; epigenetic; stem cell; homologous recombination

Introduction

A large numbers of genes regulating body mass and fat storage have been found in many organisms. The most recent edition of the Human Obesity Gene Map lists 11 single gene mutations, 244 knock-out or transgenic animal models and 135 different candidate genes that have been associated with obesity-related phenotypes. Included are variants in 22 different human genes in which association with obesity-related phenotypes has been replicated in five or more independent studies [1]. In *C. elegans*, approximately 650 genes affect fat storage when inactivated by RNA interference (RNAi) screen [2]. In mice, a survey of 1,977 knockout strains revealed that 31% of viable knockout mouse strains weighed less and an additional 3% weighed more than did controls, suggesting as many as 6000 genes influence mouse size [3]. How do we evaluate this information and how many genetic variants do we really expect to impact human obesity?

The genetic architecture of Complex traits

On both theoretical and experimental grounds we expect that quantitative traits are determined by the interaction of many genes [4,5]. The number may be surprisingly large [6]. For example, genes determining abdominal bristle number in drosophila were originally linked to 7–11 quantitative trait loci (QTLs) [7]. Fine mapping revealed that each of these QTLs actually consist of multiple QTLs. Currently 42 QTLs have been identified that influence abdominal bristle number, although the number may be as high as 100 [7]. In *C. elegans*, there are approximately 300 gene inactivations that cause fat reduction and

*Corresponding Author: Department of Genetics and Genomics, Boston University School of Medicine, 715 Albany Street, E633, Boston MA 02118, 617.414.1254, aherbert@bu.edu.

approximately 100 gene inactivations that cause fat accumulation without altering growth rate or viability [2]. Another 250 genes affecting fat storage are pleiotrophic in their action; loss of function causes sterility, growth arrest or lethality [2]. Similarly, around a third of viable mice knock-outs alter body weight besides their other effects on phenotype [3].

Even such large numbers of genes fail to explain the observed variation in quantitative phenotypes. Environment and gene-environment effects are also important sources of trait variation. Some environmental effects are so strong as to mask the underlying genetic variation. Indeed, it is possible to discuss the "runaway weight gain train" leading to human obesity without considering any role for genetics [8]. Infectious agents are also part of this spectrum: certain strains of adenovirus have been proposed as a cause of obesity in humans, chickens and mice [9]. In other cases, gene-environment interactions affect trait values in locus-specific ways, for example temperature affects expression of each abdominal bristle number QTL differently in *Drosophila* in a sex-specific manner [7].

Gene-gene interactions (genetic epistasis) also influence the population estimates of effect sizes produced by variants at each of these loci (statistical epistasis): the estimated effect size is significantly different from that expected if the effect of each variant acted additively [10]. This nonlinear mode of interaction can take many forms. For example in a multiplicative model the relative risk associated with each variant is multiplied together to produce an overall estimate of the combined risk.

Lack of statistical epistasis does not imply an absence of biological interaction. An analysis of 12 of the possible motifs based on three interacting gene products reveals that statistical epistasis is much more likely with topologies containing positive feedback loops [11]. Further, only certain combinations of genetic variants produce statistically measurable effects [11]. For a particular variant, penetrance and expressivity depends upon the genetic background and may vary from population to population. For example, in *Drosophila melanogaster*, eye development proceeds on some genetic backgrounds even when the *ey* mutation inactivates the master gene for eye formation [12].

The impact of a particular genetic interaction on trait values is determined by the allele frequency of the variants at each locus. When the actions of a variant are averaged over many different genetic backgrounds, there may be no statistically detectable main effect, whereas in more homogeneous populations such effects may be readily observable [13]. Additionally, epistasis may be measurable under one set of environmental conditions, but not others [14]. Those variants that are most easily found in Genome Wide Association Studies (GWAS) in humans may result from positive selection of the variant on different genetic backgrounds in different environments. Given the robustness of the association, epistasis may therefore be difficult to demonstrate. Indeed, complement factor H (CFH) polymorphisms and variants at other loci independently increase risk for age-related macular degeneration, but epistasis has been reported in one [15] but not in another study [16].

Detecting statistical epistasis computationally by randomly sampling variations from the genome becomes difficult (but not impossible) when the number of pair-wise comparisons is large and is limited by the frequency of each variant [17]. For example a pair-wise comparison of 1 million SNP would require $\sim 10^{12}$ calculations. Another approach, similar to those applied in model organisms, may be to identify homogeneous genetic backgrounds in isolated populations in which penetrance of a particular variant is high. Even then, the effect size is unlikely to approach that seen in Mendelian diseases [18].

What is the nature of Genetic variants causing common disease?

Common diseases are likely to result from the interaction of many different loci and reflect gene-gene and gene-environment interactions. Two models have been proposed: the common variants-common disease model and the rare variants-common disease models [19,20]. With the common variant model, the occurrence of frequent risk alleles of modest effect has been explained in terms of recent human population growth, in which a bottleneck resulting in a small number of breeding individuals was followed by rapid population expansion [20]. In the absence of strong stabilizing selective pressures, such as an infectious disease, and assuming a nucleotide mutation rate of 10^{-8} corresponding to 120 new point mutations per diploid genome per generation, it is considered unlikely that new allelic variants would have arisen with sufficient frequency to cause a major impact on the population risk of disease, even given a large risk associated with a particular deleterious mutation [19,20]. Under this scenario, quantitative trait values reflect the action or interaction of a large number of genes of modest effect and by implication, any disease associated with an extreme quantitative value will also have a large number of genetic determinants. While each child is likely to possess many risk alleles for a common disease due to their high frequency in the population, the combination is different from that of their parents, as each risk allele will be transmitted independently of all others. As a consequence, the familial risk of disease for a sibling of a proband or to a child of an affected parent (as compared to that of the general population) will be low and will not approximate that seen with Mendelian diseases where transmission of even a single allele can create a high-risk genome [21].

The view of a few common variants causing common disease has been challenged on the grounds that the nucleotide mutation rate may be much higher than 10^{-8} at some loci, meaning that there is a high frequency of rare variants in the population for some genes [19]. One class of frequent mutation results in duplication or deletion of nucleotide blocks producing copy number variation (CNV) [22,23]. Often CNVs contain several genes. They also can affect expression of neighboring genes [24]. CNVs may comprise as much as 12% of the human genome, although the precise figure is still uncertain [25]. They arise at high frequency in germinal tissue. One estimate is that in humans there may be a novel germline segmental deletion in between 1 in 25 and 1 in 4 live births [26]. Their formation is favored in hotspots of genomic instability. For example, a recent study in of B6 inbred mice reported the frequency of germline CNV generation at some loci to be as high as 1 per 46–139 live births [27].

CNV arise commonly in somatic tissue and are unique to an individual. They increase with age as a result of homologous recombination in stem cells to repair double-stranded breaks. Studies of transgenic mice containing reporter genes reveal that cells from small intestine, liver and heart in mice 32 months of age contain ~60–120 large-scale genomic rearrangements per cell. [28]. Loss in homozygosity due to homologous recombination causes loss of endogenous APRT gene activity in aged humans and mice at a frequency of 10^{-4} [29,30]. Up to one in 10^3 cells in pancreas is recombinant in older mice scored using homologous recombination between two defective tandem gene copies of a yellow-fluorescent reporter protein [31]. In hematopoietic and neural stem cells, homologous recombination is necessary for normal development [32] and offsets age-related decline in stem cell function [33]. Hematopoietic stem cell maintenance also requires de novo methylation of DNA [34]. which in some systems marks sites repaired by homologous recombination [35]. Aneuploidy in human brain is also surprisingly common: up to 5.2% of human hippocampal cells have chromosome 21 monosomy or trisomy [36]. Aneuploidy is a frequent outcome when homologous recombination is impaired in dividing cells [37]. By extrapolation, these frequent chromosomal alterations have the potential to increase disease

burden in an adult with 10– 100 trillion somatic cells, either by CNV or by altered gene expression associated with de-novo methylation at the site of repair[35]. Those stem cell variants that survive will depend on environmental exposures, a process that can be tracked by methylation changes observed when stem cells age [38]. Selection may initiate a cascade of events leading multiplicatively to disease [39]. Since as many as 10% of cells in human crypts exhibiting stem-cell like properties, the potential impact of these processes is some tissues may be high [38]. Possible outcomes include diminished tissue function or localized pathology such as atherosclerotic plaque.

Programmed epigenetic modifications also contribute to variability in trait values [40,41]. These are processes that modify DNA or histones but not the DNA sequence. They are established during embryogenesis and can restrict expression from either paternally or maternally derived chromosome or act stochastically, for example, to inactivate one copy of the X-chromosome in females. Epigenetic silencing renders the individual haploid at a locus with no readout from the homologous allele. Through such processes, maternal exposures can reprogram fetal development [40], alter DNA methylation patterns [42] and modify the risk of obesity by altering synaptic plasticity within hypothalamic feeding centers [43–46]. These outcomes may epigenetically increase the correlation between maternal and offspring phenotypes and so boost the apparent heritability of a trait [47]. They also have the potential to bias the sites where de novo CNVs arise, increasing the chance that adult stem cells in offspring from the same mother evolve similar CNVs over a lifetime, thereby increasing the correlation between siblings, another measure of heritability. Taken together, these epigenetic processes imply a high risk of disease recurrence in offspring from the same mother.

What do we expect from genome-wide association scans?

Current designs to find genetic variants affecting particular traits focus on individuals from the tails of the population distribution for that phenotype, whether these are quantitative traits, or diseases defined by a threshold. Under the common disease common variant hypothesis, with large samples the tails are assumed to be thin and described by a normal density function in accordance with the standard central limit theorem. Under the alternative hypothesis of rare variants, the distribution is skewed with tails that are mathematically "fat" [48]. The tails are best modeled, depending on the structure of the underlying data, with lognormal or power-law approximations [48–51]. They arise from different sets of multiplicative processes involving either genes or environmental exposures or both affecting a small subset of individuals. For BMI, such outcomes translate into heavier individuals gaining more weight over time than lighter individuals [52].

The current generation of genome-wide scans is moderately powered to find common SNP variants, whether they represent nucleotide or copy number changes, with a moderate relative risk of 1.2 and above [53]. Up to a million sites of nucleotide or copy number variation are mapped. Even a carefully thought-out design like the Wellcome case-control study involving 2000 cases and 3000 common controls [54], is limited in what it can find. For example, of the variants so far found and replicated on the basis of GWAS, only in two instances does the Wellcome design have a power greater than 80% to find a genome-wide significant result in at $\alpha=0.01$, whereas the power for four other variants was less than 1%. Studies of 50,000 or more participants may be needed for discovery to ensure greater than 90% power to discover associations similar to those already found [53].

Even when a true association is discovered, replication is not as easy as commonly proposed [55,56]. If the initial study and the replication attempt have 80% power, then it is often thought that the probability of finding and independently confirming an association is 64%

($=80\% \times 80\%$): this is only true if the when the estimate of the population mean is the same for both studies. However, a 95% confidence level from the initial study will, however, not always include the true population mean. On average, just 83.4% of future replication means will lie within these limits, assuming participants in both studies are randomly drawn from the same population [57]. The discovery and replication probability for a true association thus drops to 53.4% when both stages have 80% power ($=80\% \times 80\% \times 83.4\%$): given the additional sampling variability inherent in large studies due to allelic heterogeneity and differences in environmental exposures arising from sampling of geographically and ethnically distinct populations, the probability of replication in the real world is even lower. Under the current schema, many true positives may be inappropriately dismissed because of failure to replicate in one or two follow-up studies.

It is unlikely that genome scans based on the current generation of SNP assays will be powered to detect the full extent of CNV and other chromosomal changes present in the germline due to the low correlation between these classes of variation [58]. This is especially true of duplications and less so of common deletions [59]. It is even more unlikely that the current generation of scans will detect chromosomal changes arising in adult stem cells that increase risk of disease.

What have we found with genome-wide association scans for genes affecting lipid levels?

An example of the current generation of GWAS is provided by a recent report has reported identifying 18 loci that are associated with blood lipid levels. Eleven of these had been previously identified through candidate gene studies (ABCA1, the APOA5-APOA4-APOC3-APOA1 and APOE-APOC clusters, APOB, CETP, GCKR, LDLR, LPL, LIPC, LIPG and PCSK9) while six were novel (near MVK-MMAB and GALNT2, near SORT1; near TRIB1, MLXIPL and ANGPTL3 and a locus encompassing several genes near NCAN) [60]. In total 8,816 individuals were tested in the discovery phase and 11,569 participants in the replication phase and it is likely more variants will be discovered when samples sizes are increased further. The discovery of variants in well-studied lipid synthesis pathways provides validation for both the candidate gene and GWAS approaches to mapping QTLs in humans. The GWAS result provides useful biological insight but, for all the reasons given above, the goal of deriving a complete description of the heritable component of diseases through this approach is unlikely. Additional analysis of CNV and deep resequencing of both germline and somatic tissue DNA will provide needed information concerning the relative impact of common and rare variants in such pathways on phenotypes and disease outcomes.

Conclusion

There are many pathways to obesity and like other complex diseases, there will be a combination of monogenic, polygenic, epigenetic etiologies [41]. It is likely, on the basis of the results from model organisms that these outcomes will be affected by variants of a large number of different genes. The designs used in the current generation of GWAS address the common variant– common disease model and have great promise for revealing pathways leading to obesity, but not for the discovery and validation of all the variants contributing to obesity. The role of rare variants, de-novo germinal and somatic stem cell mutations in obesity needs to be addressed using other designs, as does the impact of intrauterine and environmental exposures. There is still much to be learned from the fat tail of the population.

Acknowledgments

This work is supported by 1 R01 DK077120-01A2. Dr. Marc Lenburg is thanked for comments. r.

References

1. Rankinen T, Zuberi A, Chagnon YC, et al. The human obesity gene map: the 2005 update. *Obesity* (Silver Spring). 2006; 14:529–644. [PubMed: 16741264]
2. Ashrafi K. Obesity and the regulation of fat metabolism. *WormBook*. 2007:1–20. [PubMed: 18050496] **An excellent overview of shared metabolic pathways between genera and the use of *C. elegans* to dissect them
3. Reed DR, Lawler MP, Tordoff MG. Reduced body weight is a common effect of gene knockout in mice. *BMC Genet*. 2008; 9:4. [PubMed: 18182103] *Stresses the underlying complexity of a phenotype such as body weight
4. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. *Nat Rev Genet*. 2007; 8:437–449. [PubMed: 17510664]
5. Bussey H, Andrews B, Boone C. From worm genetic networks to complex human diseases. *Nat Genet*. 2006; 38:862–863. [PubMed: 16874324]
6. Yang Q, Khoury MJ, Friedman J, et al. How many genes underlie the occurrence of common complex diseases in the population? *Int J Epidemiol*. 2005; 34:1129–1137. [PubMed: 16043441]
7. Mackay TF, Lyman RF. *Drosophila* bristles and the nature of quantitative genetic variation. *Philos Trans R Soc Lond B Biol Sci*. 2005; 360:1513–1527. [PubMed: 16108138]
8. Swinburn B, Egger G. The runaway weight gain train: too many accelerators, not enough brakes. *BMJ*. 2004; 329:736–739. [PubMed: 15388619]
9. Atkinson RL. Viruses as an etiology of obesity. *Mayo Clin Proc*. 2007; 82:1192–1198. [PubMed: 17908526] Currently a view not widely accepted but a hypotheses that can be experimentally addressed
10. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet*. 2002; 11:2463–2468. [PubMed: 12351582]
11. Gjuvlsland AB, Hayes BJ, Omholt SW, Carlborg O. Statistical epistasis is a generic feature of gene regulatory networks. *Genetics*. 2007; 175:411–420. [PubMed: 17028346] An attempt to merge network approaches with statistical measures commonly used in population studies
12. Rutherford SL. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *Bioessays*. 2000; 22:1095–1105. [PubMed: 11084625]
13. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet*. 2004; 36:512–517. [PubMed: 15052271]
14. Harrison R, Papp B, Pal C, et al. Plasticity of genetic interactions in metabolic networks of yeast. *Proc Natl Acad Sci U S A*. 2007; 104:2307–2312. [PubMed: 17284612] *While many approaches to epistasis in yeast stress the robustness of systems to mutation, this approach examines the degree to which robustness is conditional on environmental conditions.
15. Schaumberg DA, Hankinson SE, Guo Q, et al. A prospective study of 2 major age-related macular degeneration susceptibility alleles and interactions with modifiable risk factors. *Arch Ophthalmol*. 2007; 125:55–62. [PubMed: 17210852]
16. Maller J, George S, Purcell S, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet*. 2006; 38:1055–1059. [PubMed: 16936732]
17. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet*. 2005; 37:413–417. [PubMed: 15793588]
18. Holmdahl R. Primer: comparative genetics of animal models of arthritis--a tool to resolve complexity. *Nat Clin Pract Rheumatol*. 2007; 3:104–111. [PubMed: 17299448] A good review on the strengths and limits of model organisms such as mice in the study of a complex disease such as arthritis
19. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant.or not? *Hum Mol Genet*. 2002; 11:2417–2423. [PubMed: 12351577]

20. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet.* 2001; 17:502–510. [PubMed: 11525833]
21. Hemminki K, Bermejo JL. Constraints for genetic association studies imposed by attributable fraction and familial risk. *Carcinogenesis.* 2007; 28:648–656. [PubMed: 17012223] Why common variants do not account for familial recurrence risk
22. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature.* 2006; 444:444–454. [PubMed: 17122850]
23. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. *Science.* 2004; 305:525–528. [PubMed: 15273396]
24. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science.* 2007; 315:848–853. [PubMed: 17289997] A first look at the impact of SNPs and CNVs on gene expression but not the last word as only transformed lymphocytes were examined, a 3' biased expression array was analyzed and the number of variants examined limited.
25. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet.* 2007; 8:639–646. [PubMed: 17637735] *an overview of recent progress in understanding the impact of copy number variants in common disease
26. Lupski JR. Genomic rearrangements and sporadic disease. *Nat Genet.* 2007; 39:S43–S47. [PubMed: 17597781]
27. Egan CM, Sridhar S, Wigler M, Hall IM. Recurrent DNA copy number variation in the laboratory mouse. *Nat Genet.* 2007; 39:1384–1389. [PubMed: 17965714] Indred strains may be lees identical than believed
28. Vijg J, Busuttil RA, Bahar R, Dolle ME. Aging and genome maintenance. *Ann N Y Acad Sci.* 2005; 1055:35–47. [PubMed: 16387716]
29. Bailey KJ, Maslov AY, Pruitt SC. Accumulation of mutations and somatic selection in aging neural stem/progenitor cells. *Aging Cell.* 2004; 3:391–397. [PubMed: 15569356]
30. Stambrook PJ. An ageing question: do embryonic stem cells protect their genomes? *Mech Ageing Dev.* 2007; 128:31–35. [PubMed: 17134741] *Another system showing a high frequency of homologous recombination in stem cells from older animals
31. Wiktor-Brown DM, Hendricks CA, Olipitz W, Engelward BP. Age-dependent accumulation of recombinant cells in the mouse pancreas revealed by in situ fluorescence imaging. *Proc Natl Acad Sci U S A.* 2006; 103:11862–11867. [PubMed: 16882718] A very visual demonstration of recombination events occurring as mice age
32. Rehen SK, Yung YC, McCreight MP, et al. Constitutional aneuploidy in the normal human brain. *J Neurosci.* 2005; 25:2176–2180. [PubMed: 15745943] **A highly unexpected finding
33. Griffin CS. Aneuploidy, centrosome activity and chromosome instability in cells deficient in homologous recombination repair. *Mutat Res.* 2002; 504:149–155. [PubMed: 12106655]
34. Lee Y, McKinnon PJ. Responding to DNA double strand breaks in the nervous system. *Neuroscience.* 2007; 145:1365–1374. [PubMed: 16934412]
35. Brunet A, Rando TA. Ageing: from stem to stern. *Nature.* 2007; 449:288–291. [PubMed: 17882206] *Useful review of the issues in aging of adult stem cells
36. Tadokoro Y, Ema H, Okano M, et al. De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *J Exp Med.* 2007; 204:715–722. [PubMed: 17420264]
37. Cuzzo C, Porcellini A, Angrisano T, et al. DNA Damage, Homology-Directed Repair, and DNA Methylation. *PLoS Genet.* 2007; 3:e110. [PubMed: 17616978]
38. Yatabe Y, Tavare S, Shibata D. Investigating stem cells in human colon by using methylation patterns. *Proc Natl Acad Sci U S A.* 2001; 98:10839–10844. [PubMed: 11517339]
39. Choi MY, Lee HY, Kim D, Park SH. Entropic sampling and natural selection in biological evolution. *Journal of Physics A: Mathematical and General.* 1997; 30:L749–L755.
40. Hall JG. The importance of the fetal origins of adult disease for geneticists. *Clin Genet.* 2007; 72:67–73. [PubMed: 17661809] *A useful summary of many studies and speculations considering maternal effects on the fetus

41. Jiang YH, Bressler J, Beaudet AL. Epigenetics and human disease. *Annu Rev Genomics Hum Genet.* 2004; 5:479–510. [PubMed: 15485357]
42. Thurston A, Lucas ES, Allegrucci C, et al. Region-specific DNA methylation in the preimplantation embryo as a target for genomic plasticity. *Theriogenology.* 2007; 68 Suppl 1:S98–S106. [PubMed: 17482250] *Raises the question of how much plasticity exists in methylation programs during fetal development
43. Horvath TL. The hardship of obesity: a soft-wired hypothalamus. *Nat Neurosci.* 2005; 8:561–565. [PubMed: 15856063]
44. Farooqi S, O'Rahilly S. Genetics of obesity in humans. *Endocr Rev.* 2006; 27:710–718. [PubMed: 17122358]
45. Gao Q, Horvath TL. Neurobiology of feeding and energy expenditure. *Annu Rev Neurosci.* 2007; 30:367–398. [PubMed: 17506645]
46. Ikenasio-Thorpe BA, Breier BH, Vickers MH, Fraser M. Prenatal influences on susceptibility to diet-induced obesity are mediated by altered neuroendocrine gene expression. *J Endocrinol.* 2007; 193:31–37. [PubMed: 17400800] *an examination of maternal effects on fetal gene expression within the hypothalamus in rats
47. Lansdorp PM. Immortal strands? Give me a break. *Cell.* 2007; 129:1244–1247. [PubMed: 17604711] *One look at the issue of how information is partitioned between self-renewing and differentiated cells using asymmetrical epigenetic markers to bias cell fate
48. Mitzenmacher M. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics.* 2004; 1:226–251.
49. Keller EF. Revisiting "scale-free" networks. *Bioessays.* 2005; 27:1060–1068. [PubMed: 16163729]
50. Limpert E, Stahel WA, Abbt M. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience.* 2001; 51:341–352.
51. Reed WJ, Hughes BD. From gene families and genera to incomes and internet file sizes: why power laws are so common in nature. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2002; 66:067103. [PubMed: 12513446]
52. Penman AD, Johnson WD. The changing shape of the body mass index distribution curve in the population: implications for public health policy to reduce the prevalence of adult obesity. *Prev Chronic Dis.* 2006; 3:A74. [PubMed: 16776875] *The fat tail of body mass index distributions highlighted
53. Altshuler D, Daly M. Guilt beyond a reasonable doubt. *Nat Genet.* 2007; 39:813–815. [PubMed: 17597768] *a summary of the power of current GWAS to detect common variants
54. Todd JA. Statistical false positive or true disease pathway? *Nat Genet.* 2006; 38:731–733. [PubMed: 16804532]
55. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature.* 2007; 447:655–660. [PubMed: 17554299]
56. Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered.* 2007; 64:203–213. [PubMed: 17551261] *one view of the difficulties in showing anything with GWAS
57. Cumming G, Maillardet R. Confidence intervals and replication: where will the next mean fall? *Psychol Methods.* 2006; 11:217–227. [PubMed: 16953701] *required reading for population geneticists
58. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet.* 2007; 39:S37–S42. [PubMed: 17597780] A usefuyul introduction into issue surrounding CNV and their detection with SNP arrays.
59. Hinds DA, Kloeck AP, Jen M, et al. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet.* 2006; 38:82–85. [PubMed: 16327809]
60. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet.* 2008; 40:189–197. [PubMed: 18193044] **validates both the candidate gene and GWAS approach to identifying variants that affect quantitative traits such as lipids and their further evaluation as risk factors for coronary artery disease