Mouse Models of Human GWAS Hits for Obesity and Diabetes in the Post Genomic Era: Time for Reevaluation

Samantha Laber[1,2]* and Roger D. Cox[1]*

[1] *Mammalian Genetics Unit, Medical Research Council Harwell Institute, Oxfordshire, UK*, [2] *Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK*

In recent years, genome-wide association studies (GWAS) have identified hundreds of loci and thousands of single-nucleotide polymorphisms (SNPs) associated with type 2 diabetes mellitus (T2DM) and obesity traits [such as body mass index (BMI) and waist–hip ratio (WHR)] in the human population (1–4). The vast majority of these SNPs are in non-coding regions of the genome and distal to promoters, suggesting they act through gene regulation which makes their functional interpretation challenging (5). Collectively, comparing the epigenetic landscape between mouse and human has established new pathways involved in obesity and diabetes, and in fact, inter-species conservation has successfully been used as criteria in finding functional and disease-relevant elements (6–8). By contrast, genome-wide comparative analysis of the mouse and human epigenome across tissues has highlighted the presence of cis-regulatory divergence (9, 10). New mouse engineering approaches together with bioinformatics dissection of trait-associated regions, for example, epigenetic modifications and genome interactions hold great promise to fully understand the underlying mechanisms of human disease-associated non-coding variants in T2DM and obesity.

THE CONTEXT-SPECIFIC NATURE OF HUMAN GWA SIGNALS IN HUMAN

Over 80% of loci identified by GWAS are in intergenic and intronic regions and many of these genetic risk regions are enriched for histone modifications (5), suggesting they act as regulatory elements which appear to function in a highly cell-selective manner. Due to the tissue specificity as well as the developmental and epigenetic complexity of gene regulation, functional approaches require the study of the relevant tissue and cell type as well as genetic and bioinformatics approaches that reliably assess the regulatory role of non-coding variants (7, 11, 12). Ongoing progress in high-throughput sequencing and the development of new experimental tools are greatly advancing our capacity to study chromatin biology and genome function. In particular, ChIP-seq allows identification of transcription factor binding sites and chromatin states; chromosome conformation capture-based techniques (including 3C, 4C, 5C, CaptureC, and HiC) allow the study of chromatin interactions; and DNase hypersensitivity or ATAC-seq can identify accessible chromatin (13–16). Additionally, tools like HaploReg (17), Enlight (18), RegulomeDB (19), and The Islet Regulome Browser (11) are emerging that allow the integration of GWAS results with genetic and epigenetic annotations that can be used to dissect the gene regulatory networks that underpin genomic association signals.

By integrating the information gained from functional genomics efforts such as the ENCODE (5) and Roadmap Epigenomics projects (20) together with expression quantitative trait loci (eQTL) results and functional studies, it becomes increasingly clear that adipose tissue is one of several key effectors of genetic risk loci for T2DM and obesity trait associations, particularly for WHR signals (2, 7, 21, 22). However, there is currently still a lack of comprehensive maps linking distal elements that harbor disease-associated variants with their target genes in relevant tissues and developmental

stages. Furthermore, extensive fine-mapping of risk associations is crucial in order to narrow down the association signal to the likely causative variants which can then be functionally investigated (23). This has resulted in many studies being performed assuming that the closest gene to a given disease-associated signal is the causative one. Traditionally, target genes based on proximity to a signal were selected to model in the mouse using global or tissue-specific gene knockout or overexpression alleles to characterize gene function (24). However, with this approach, many target genes for GWA signals have potentially been overlooked, for example, in the case of the BMI-associated variants in *Fto* (25, 26). An additional level of complexity comes with the possibility for an association signal—that usually harbors dozens of SNPs—to potentially contain a number of disease-causing variants that might act in different tissues and/or at different times, affecting different genes. For example, there is currently evidence for intronic *FTO* risk variants to alter the expression of nearby genes in both adipose tissue and brain. An eQTL in human cerebellum links re9930509 to altered *IRX3* expression (25), rs1421085 has very convincingly been shown to be located within an enhancer for *IRX3* and *IRX5* in adipocyte precursors (7, 26) and rs1421085 and rs8050136 have been proposed to selectively alter *FTO* and *RPGRIP1L* expression in human-induced pluripotent stem cell-derived neurons (27). Therefore, mouse models which could help pinpoint variants, target genes, and relevant tissues would prove invaluable in the mechanistic dissection of human disease-associated sequence variants. However, whether it is possible to use the mouse for modeling regulatory variants (which is essential to capture the relevant spatiotemporal effects) will depend on the functional conservation of the regulatory circuitry of a given signal in human and mouse.

CONSERVATION BETWEEN HUMAN AND MOUSE

It is estimated that our last common ancestor with the mouse was about 90 million years ago (28). At this point, many of the core physiological regulatory mechanisms had evolved, for example, mouse and human share the same basic mechanisms for controlling food intake *via* leptin and hypothalamic anorexigenic and orexigeneic neurons, and similarly insulin and glucagon are core effectors in glucose homeostasis. However, there have clearly been many evolutionary changes over this long period of time. At the level of the genome, chromosome number and organization have changed, although it is striking how large tracts of DNA have conserved their order of genes and show high coding sequence conservation (29). Thus, if we wish to use the mouse as model of human metabolic disease we can rely on much of the core conservation of ancient metabolic pathways and their regulation but cannot ignore the fact of their continued evolution that adapts and changes these mechanisms for the survival of two very different organisms. The mouse ENCODE Consortium reported that comparative gene expression data from human and mouse reveals that some sets of genes tend to cluster more by species than by tissue and *vice versa* (29). More recently, it has been suggested that gene clustering by tissue rather than species

is much stronger than originally thought (30). Interestingly, single-cell sequencing of human and mouse pancreatic alpha and beta cells showed good cross-species correlation of transcriptomes although with some important species differences (31). Finally, Breschi et al. (32) describe how transcriptomes show a continuum of variation from species dominated clustering to organ dominated clustering. Importantly, for modeling GWA signals in other species, genes that varied little between species (and are more organ-specific) are more likely to overlap with human risk variants (32).

EPIGENOMIC CONSERVATION BETWEEN HUMAN AND MOUSE—INSIGHTS FROM THE MOUSE ENCODE CONSORTIUM

Some of the other key findings of ENCODE in the mouse genome were that human and mouse trans-regulatory networks (transcription factor networks) are considerably more conserved than the cis-regulatory landscape, which in fact accounts for the majority of regulatory plasticity between human and mouse (28, 29). At the same time, the degree of divergence of regulatory elements varies widely between different types of elements that are active in different tissue contexts (9, 28). The Mouse ENCODE Consortium (29) demonstrated that 79.3% of mouse candidate enhancers (predicted by patterns of histone modifications) and 66.7% of transcription factor binding sites have sequence orthologs in humans. Further, 61.5% of tested candidate mouse-specific enhancers also show enhancer activity in human embryonic stem cells in a reporter assay (29), suggesting a degree of functional conservation between human and mouse gene regulation. Based on this level of conservation, it is intriguing to ask the question whether mouse chromatin states could be used to identify potential sites for functional characterization in mouse for human GWAS hits. Mapping 4,265 SNPs from human GWAS studies onto the mouse genome using 15 mouse samples revealed that human GWAS hits are associated with specific chromatin states in relevant mouse tissues (29). For example, in mouse kidney, H3K4me1 is enriched in specific GWAS hits associated with urate levels and metabolites. For mouse liver-specific H3K36me3, GWAS hits related to HDL cholesterol and triglyceride levels are enriched. Together, 55% of mapped SNPs overlapped with at least one histone mark in mouse (29). These results suggest that histone modification marks can be used to inform about human risk variants and for the identification of candidate functional sequences for characterization of human GWAS hits in mouse.

Furthermore, SNPs with high regulatory potential are enriched in conserved transcription factor binding sites (19). Cheng et al. (33) show that conserved sequences occupied by orthologous transcription factors in human and mouse are enriched for GWAS variants. When investigating whether this is true for individual phenotypes, they found that SNPs associated with type I diabetes and several other traits are significantly enriched in conserved transcription factor binding sites, with 13 out of 20 type 1 diabetes SNPs being in conserved binding sites. By contrast, all of the SNPs associated with pulmonary function

were found to be human-specific, suggesting that besides GWAS SNPs generally being enriched in conserved regulatory elements, that this enrichment is dependent on the trait (33). Whether this is the case for T2DM and obesity traits association is yet to be investigated. With continuous efforts and the increase in available mouse genome data sets, it will become possible to draw conclusions about the human–mouse conservation of transcription factor occupancy and enrichment of GWAS SNPs in adipose tissue. Indeed, on a cellular level, a systematic comparison between the human and mouse epigenome during adipocyte development and in different fat depots is largely missing, and Mouse ENCODE has currently only limited adipose tissue datasets that could be matched to human. Though, Mikkelsen et al. (34) generated a comparative analysis of chromatin state maps together with gene expression profiles from human adipose tissue and mouse 3T3-L1 at four time points during differentiation. They showed that although a significant amount of open chromatin in orthologous regions were shared between the two models (15–30%), most of them were species-specific. While we are not proposing that this affected the key findings of this study, it is worth pointing out that

comparing a mouse cell line and primary human tissue-derived pre-adipocytes with their accompanying ontogenetic differences can potentially hinder the interpretation when using these data sets for dissecting specific GWAS loci with the aim to establish relevant functional sites.

Taken together, although the cis-regulatory landscape has substantially diverged between human and mouse on a global level, human trait-associated SNPs are enriched in sites that are conserved between the two species for the majority of traits investigated.

THE POTENTIAL FOR NEW STRATEGIES IN MOUSE MODEL ENGINEERING

With our current knowledge of the context-specificity of gene regulation and consequently the many layers of complexity of most GWAS signals, it becomes increasingly clear that it is necessary to study and understand the underlying regulatory network in the relevant human tissue (**Figure 1**). In the past,

FIGURE 1 | Functional validation of variants in human and mouse. An integrative approach for the generation of meaningful and informative mouse models of human Genome-Wide Association Study (GWAS) signals. Deciphering the underpinnings of an association signal in the human context is essential. For mechanistic studies *in vivo*, a human-to-mouse epigenomic comparison can guide the choice of a relevant mouse model, e.g., in the case of low or insufficient functional conservation of a regulatory site (or the lack of data sets that can determine the latter) a classical tissue-specific target gene manipulation can prove valuable; in the case of a high functional conservation (based on genomics and bioinformatics dissection of the loci), a model that selectively manipulates the regulatory region can in principle be useful. Comparing the human–mouse epigenome can be expected to become increasingly powerful with the improvement of quality as well as the comprehensiveness of genomic data sets and tools. The translational utility will depend on the mouse model chosen and the information gained can feedback and help interpret human GWAS signals.

a successful approach to studying the function of individual candidate genes *in vivo* has been achieved by generating global knockout and overexpression models (24). However, these models do not resemble the tissue-specific nature of alterations in regulatory elements. Tissue-specific target gene manipulation using CRE drivers can be a powerful tool to overcome this problem. However, another challenge comes with the current lack of reliable pre-adipocyte-specific CRE lines that can be used to assess the tissue-specific effect of identified target genes in cases of pre-adipocyte-specific signals. Recent advances in genome engineering, namely CRISPR/Cas9, opened the opportunity to conveniently alter any regulatory sequence of interest (35). In other words, it is now possible to genome edit transcription factor binding sites and enhancer elements in the mouse which in principle has the potential to create mouse models of human risk variants that (i) are cell type-specific; (ii) alter all target genes; (iii) alter target genes at the relevant level and direction; and (iv) alter target genes at the relevant time of development (**Figure 1**).

CONCLUSION AND FUTURE DIRECTIONS

The majority of human genetic variants associated with common metabolic disease traits are located within distal regulatory elements. With our current knowledge of gene regulation and the context-specificity of the signal, it is necessary to understand the signal in human. Identifying targets and context is crucial in engineering a *relevant* mouse model. A comprehensive human-to-mouse epigenomic comparison can be informative about human risk variants. Although intriguingly, whether manipulation of regulatory elements will become a tool to dissect human obesity/T2DM risk variants in the mouse will depend on the functional conservation of a given signal. This is yet to be established and offers an exciting avenue to explore.

GLOSSARY

ATAC-seq—assay for transposase-accessible chromatin followed by high-throughput sequencing. This technique allows the identification of open chromatin.

BMI—body mass index. A measure of body weight that takes account of an individual's size and calculated by dividing body weight by height squared.

ChIP-seq—chromatin immunoprecipitation followed by high-throughput sequencing. This technique allows the identification of DNA fragments that are bound by a specific antibody.

Cis-regulatory—non-coding DNA sequences in or near a gene required for its spatiotemporal expression that characteristically contain transcription factor binding sites.

CRE—Cre recombinase recognizes DNA sequences known as LoxP sites and when a pair of sites is provided in the same orientation this leads to deletion of the intervening sequence. In this way, a segment of DNA such as a key exon (said to be floxed) can be deleted resulting in, for example, a null mutation. This can be done *in vivo* by gene editing to place LoxP sites in the required location and then crossing animals that carry this modification to Cre recombinase strains, which then results in recombination. The expression of Cre recombinase can be driven by a promoter of choice either as a transgene or knocked into an endogenous gene promoter. Thus, the recombinase can be expressed in specific tissues as required allowing cell- or tissue-specific recombination, i.e., for the generation of a conditional knockout.

Epigenome—a network of chemical compounds (for example, DNA methylation or histone modifications) surrounding DNA that modify the genome without altering the DNA sequence itself. These modifying elements play a role in determining which genes are active in a particular cell at a particular time.

eQTL—expression quantitative trait loci are genomic loci that contribute to variation in the expression levels of mRNAs. For example, in individuals in a population inheriting SNP allele A, the expression of gene Y is found to be quantitatively increased or decreased on average relative to the other SNP alleles inherited across the population assayed. This is a correlated trait rather than a direct functional link between a SNP and the expression of a gene. Further, any particular SNP marks a haplotype (a linked co-inherited group) of SNPs and as such represents a locus.

GWAS—Genome-Wide Association Study.

iPSC—induced pluripotent stem cell.

SNP—single-nucleotide polymorphism.

T2DM—type 2 diabetes mellitus.

Transcriptome—the entire mRNA expressed from the genes of a cell.

Trans-regulatory—in the context of transcriptional regulation, a trans-acting element is usually a DNA sequence that contains a gene. This gene encodes for a protein (or other molecules such as microRNA) that will regulate another target gene.

WHR—waist-to-hip ratio.

AUTHOR CONTRIBUTIONS

SL and RC wrote, edited, and approved the manuscript.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

1. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* (2015) 518:197–206. doi:10.1038/nature14177

2. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* (2015) 518:187–96. doi:10.1038/nature14132

3. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic

architecture and pathophysiology of type 2 diabetes. *Nat Genet* (2012) 44:981–90. doi:10.1038/ng.2383

4. DIAbetes Genetics Replication And Meta-Analysis (DIAGRAM) Consortium, Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, South Asian Type 2 Diabetes (SAT2D) Consortium, Mexican American Type 2 Diabetes (MAT2D) Consortium, Type 2 Diabetes Genetic Exploration by Nex-Generation Sequencing in Multi-Ethnic Samples (T2D-GENES) Consortium, Mahajan A, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* (2014) 46:234–44. doi:10.1038/ng.2897

5. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489:57–74. doi:10.1038/nature11247

6. Dickel DE, Barozzi I, Zhu Y, Fukuda-Yuzawa Y, Osterwalder M, Mannion BJ, et al. Genome-wide compendium and functional assessment of in vivo heart enhancers. *Nat Commun* (2016) 7:12923. doi:10.1038/ncomms12923

7. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, et al. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* (2014) 156:343–58. doi:10.1016/j.cell.2013.10.058

8. Multhaup ML, Seldin MM, Jaffe AE, Lei X, Kirchner H, Mondal P, et al. Mouse-human experimental epigenetic analysis unmasks dietary targets and genetic liability for diabetic phenotypes. *Cell Metab* (2015) 21:138–49. doi:10.1016/j.cmet.2014.12.014

9. Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS, et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* (2014) 346:1007–12. doi:10.1126/science.1246426

10. Denas O, Sandstrom R, Cheng Y, Beal K, Herrero J, Hardison RC, et al. Genome-wide comparative analysis reveals human-mouse regulatory landscape and evolution. *BMC Genomics* (2015) 16:87. doi:10.1186/s12864-015-1245-6

11. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman I, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* (2014) 46:136–43. doi:10.1038/ng.2870

12. Zhang X, Bailey SD, Lupien M. Laying a solid foundation for Manhattan – 'setting the functional basis for the post-GWAS era'. *Trends Genet* (2014) 30:140–9. doi:10.1016/j.tig.2014.02.006

13. Knight JC. Approaches for establishing the function of regulatory genetic variants involved in disease. *Genome Med* (2014) 6:92. doi:10.1186/s13073-014-0092-4

14. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* (2013) 14:390–403. doi:10.1038/nrg3454

15. Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* (2014) 46:205–12. doi:10.1038/ng.2871

16. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* (2013) 10:1213–8. doi:10.1038/nmeth.2688

17. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* (2012) 40:930–4. doi:10.1093/nar/gkr917

18. Guo Y, Conti DV, Wang K. Enlight: web-based integration of GWAS results with biological annotations. *Bioinformatics* (2015) 31:275–6. doi:10.1093/bioinformatics/btu639

19. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* (2012) 22:1790–7. doi:10.1101/gr.137323.112

20. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, et al. Integrative analysis of 111 reference human epigenomes. *Nature* (2015) 518:317–30. doi:10.1038/nature14248

21. Dahlman I, Rydén M, Brodin D, Grallert H, Strawbridge RJ, Arner P. Numerous genes in loci associated with body fat distribution are linked to adipose function. *Diabetes* (2016) 65:433–7. doi:10.2337/db15-0828

22. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* (2010) 42:579–89. doi:10.1038/ng.609

23. Liu CT, Buchkovich ML, Winkler TW, Heid IM, African Ancestry, Anthropometry Genetics, et al. Multi-ethnic fine-mapping of 14 central adiposity loci. *Hum Mol Genet* (2014) 23:4738–44. doi:10.1093/hmg/ddu183

24. Yazdi FT, Clee SM, Meyre D. Obesity genetics in mouse and human: back and forth, and back again. *Peer J* (2015) 3:e856. doi:10.7717/peerj.856

25. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* (2014) 507:371–5. doi:10.1038/nature13138

26. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med* (2015) 373:895–907. doi:10.1056/NEJMoa1502214

27. Stratigopoulos G, Burnett LC, Rausch R, Gill R, Penn DB, Skowronski AA, et al. Hypomorphism of Fto and Rpgrip1l causes obesity in mice. *J Clin Invest* (2016) 126:1897–910. doi:10.1172/JCI85526

28. Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, Zhang M, et al. Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* (2014) 515:365–70. doi:10.1038/nature13972

29. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* (2014) 515:355–64. doi:10.1038/nature13992

30. Gilad Y, Mizrahi-Man O. A reanalysis of mouse ENCODE comparative gene expression data. *F1000Res* (2015) 4:121. doi:10.12688/f1000research.6536.1

31. Xin Y, Kim J, Okamoto H, Ni M, Wei Y, Adler C, et al. RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab* (2016) 26:608–15. doi:10.1016/j.cmet.2016.08.018

32. Breschi A, Djebali S, Gillis J, Pervouchine DD, Dobin A, Davis CA, et al. Gene-specific patterns of expression variation across organs and species. *Genome Biol* (2016) 17:151. doi:10.1186/s13059-016-1008-y

33. Cheng Y, Ma Z, Kim BH, Wu W, Cayting P, Boyle AP, et al. Principles of regulatory information conservation between mouse and human. *Nature* (2014) 515:371–5. doi:10.1038/nature13985

34. Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, et al. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* (2010) 143:156–69. doi:10.1016/j.cell.2010.09.006

35. Aida T, Imahashi R, Tanaka K. Translating human genetics into mouse: the impact of ultra-rapid in vivo genome editing. *Dev Growth Differ* (2014) 56:34–45. doi:10.1111/dgd.12101

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.