Making Biological Sense of GWAS Data: Lessons from the FTO Locus

Mark A. Herman[1,2] and Evan D. Rosen[1,2,*]
[1]Division of Endocrinology, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA
[2]Broad Institute, Cambridge, MA 02142, USA
*Correspondence: erosen@bidmc.harvard.edu
http://dx.doi.org/10.1016/j.cmet.2015.09.018

GWAS have yielded many candidate loci for complex diseases like obesity, but interpreting the biological context of these findings has been difficult. Claussnitzer et al. (2015) use a sophisticated combination of bioinformatic and experimental approaches to address this bottleneck for variants in the *FTO* locus that associate with obesity.

Genome-wide association studies (GWAS) have revolutionized the search for the genetic underpinnings of human disease. Despite the identification of thousands of single-nucleotide polymorphisms (SNPs) associated with complex diseases, GWAS has been the subject of two major criticisms. First, for most conditions, GWAS has failed to account for the lion's share of estimated heritable risk. Second, even the best-documented "hits" have defied biological explanation, and thus true insight into disease mechanisms has lagged. GWAS hits have been hard to interpret because of uncertainty over which exact SNP is causal, given the large number of variants in linkage disequilibrium. This is compounded by the fact that the vast majority of potentially causal SNPs are located in non-coding regulatory regions of the genome (Abecasis et al., 2012). The practice of "mapping" such SNPs to the nearest gene is misleading, as SNPs may regulate gene expression over large genomic distances (Ragvin et al., 2010). Finally, GWAS hits give few clues about which cell types may be most relevant for a SNP, further limiting confidence about the identity and function of the causative SNP. Altogether, it is no surprise that we understand little about how specific SNPs contribute to the pathophysiology of complex disease. Into this morass wade the authors of a study in the *New England Journal of Medicine*, utilizing a battery of computational and experimental strategies to point the way for others trying to make sense of GWAS data (Claussnitzer et al., 2015).

The paper focuses on the FTO locus, which contains intronic SNPs that strongly associate with obesity in diverse populations (reviewed in Tung et al., 2014). FTO is expressed in hypothalamic neurons that control appetite and energy expenditure, and transgenic and knockout studies have demonstrated an effect of FTO copy number on adiposity (Church et al., 2010; Fischer et al., 2009; Gao et al., 2010), suggesting that FTO itself could be the relevant target gene. However, in 2010, Ragvin and colleagues recognized highly conserved synteny in the FTO locus and demonstrated in zebrafish that this locus regulated expression of the distal *Irx3* homeobox gene in multiple tissues during development (Ragvin et al., 2010). Smemo et al., and now Claussnitzer et al., used state-of-the-art chromatin-capture sequencing technologies to confirm direct interactions between regions in the FTO locus and the distant *Irx3* and *Irx5* genes in mouse and zebrafish embryos and human fibroblastic cell lines (Claussnitzer et al., 2015; Smemo et al., 2014). Both groups provide compelling eQTL data that SNPs in this region associate with *IRX3* and/or *IRX5* expression, but *not* with FTO expression, in human cells (see Figure 1, top).

Claussnitzer et al. also identify the causal SNP within the FTO locus using a sophisticated bioinformatics approach. A T-to-C transition (rs1421085) in perfect linkage disequilibrium with the lead SNP (rs1558902) identified by GWAS is predicted to disrupt a conserved motif that binds members of the ARID family of transcription factors. Consistent with this, overexpression or knockdown of ARID5b in pre-adipocytes suppressed or increased IRX3/5 expression, respectively, in cells containing the intact motif, but not the mutated motif. Using gene editing, conversion of T to C at rs1421085 increased IRX3/5 expression to levels observed in carriers of the risk allele, providing strong evidence that this is in fact the causal SNP. And because this site is conserved in mice, gene editing approaches can be used to produce mouse lines replicating the human T-to-C mutation, allowing exploration of the mechanisms by which the FTO locus causes obesity in vivo.

Which tissue mediates the actions of this SNP? Most genetic variants that affect body weight point to hypothalamic regions that regulate food intake and energy expenditure (Speliotes et al., 2010). Consistent with this, Smemo et al. showed that expression of a dominant-negative IRX3 selectively in hypothalamus in mice affects body weight (Smemo et al., 2014). However, Claussnitzer et al. observed that the SNPs in the FTO intron are contained within a "super-enhancer" most active in cells of mesenchymal origin, including adipose tissue-derived mesenchymal cells. This suggests a role for IRX3/5 in adipocyte development and function. They noted increased IRX3 expression in human adipose stromal vascular fractions that bear the risk allele and proposed that one mechanism by which this SNP promotes obesity is by favoring differentiation of progenitor cells toward fat-storing white adipocytes rather than fat-burning beige adipocytes. Consistent with this, expression of IRX3 in pre-adipocytes promoted expression of a gene program implicated in white adipose tissue development and lipid storage, whereas suppression of IRX3 favored a developmental program reminiscent of

beige adipocytes. Additionally, transgenic expression of a dominant-negative IRX3 in mature adipocytes markedly suppressed weight gain, and this was associated with increased energy expenditure.

The potential of beige adipose tissue to increase energy expenditure and promote weight loss is currently a topic of great interest. However, most (but not all) clinical data indicate that the FTO risk alleles are associated with increased food intake and not with reduced energy expenditure (reviewed in Tung et al., 2014), inconsistent with a mechanism that relies on beige fat activity. Measuring subtle changes in energy expenditure that might account for relatively small changes in weight accrued over many years is notoriously difficult. Short of this, it will be important to address whether the FTO locus affects the abundance of beige fat in adult human populations and whether such differences are capable of mediating physiologically significant changes in human weight (Sidossis and Kajimura, 2015). Additionally, both Smemo et al. and Claussnitzer et al. express dominant-negative IRX3 under the control of elements expressed in mature neurons and adipocytes, respectively, which may not be the most relevant system for testing the effects of what are believed to be developmental genes. Altogether, the cell type and mechanism by which the FTO variant affects body weight is still unsettled, and we should be mindful that important activities in brain and adipose tissue are not mutually exclusive (see Figure 1, bottom).

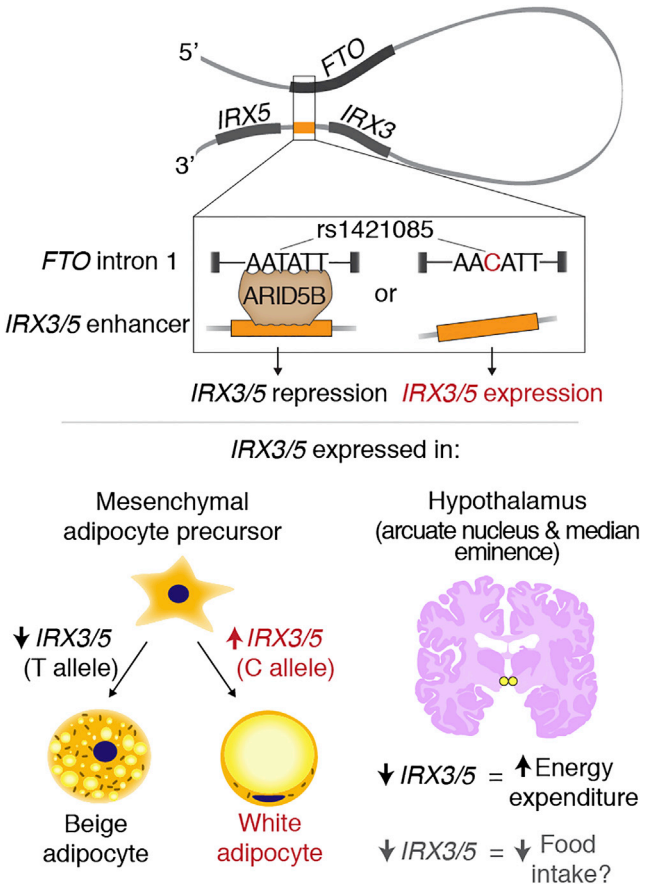

Figure 1. Variants in the FTO Locus Regulate IRX3/5 Expression to Exert Effects on Body Weight

Top: Variants in the FTO locus regulate the expression of IRX3 and IRX5 via binding the transcription factor ARID5B. Bottom: The consequences of altered IRX3/5 expression may be manifested in the adipose lineage, with altered beige fat development, and/or in the hypothalamus, via changes in food intake and energy expenditure. See text for details.

While questions remain, Claussnitzer and colleagues have driven the field forward by identifying the causal SNP and by illuminating one mechanism by which this variant could affect body weight. More importantly, they have generated a roadmap by which the community can begin to elucidate the biology associated with the GWAS data generated over the past decade. These insights will be essential to move us toward the world of precision medicine that we aspire to.

ACKNOWLEDGMENTS

We would like to thank Christina Usher for help with the figure.

REFERENCES

Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). Nature *491*, 56–65.

Church, C., Moir, L., McMurray, F., Girard, C., Banks, G.T., Teboul, L., Wells, S., Brüning, J.C., Nolan, P.M., Ashcroft, F.M., and Cox, R.D. (2010). Nat. Genet. *42*, 1086–1092.

Claussnitzer, M., Dankel, S.N., Kim, K.H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., et al. (2015). N. Engl. J. Med. *373*, 895–907.

Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Brüning, J.C., and Rüther, U. (2009). Nature *458*, 894–898.

Gao, X., Shin, Y.H., Li, M., Wang, F., Tong, Q., and Zhang, P. (2010). PLoS ONE *5*, e14005.

Ragvin, A., Moro, E., Fredman, D., Navratilova, P., Drivenes, Ø., Engström, P.G., Alonso, M.E., de la Calle Mustienes, E., Gómez Skarmeta, J.L., Tavares, M.J., et al. (2010). Proc. Natl. Acad. Sci. USA *107*, 775–780.

Sidossis, L., and Kajimura, S. (2015). J. Clin. Invest. *125*, 478–486.

Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Nature *507*, 371–375.

Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al.; MAGIC; Procardis Consortium (2010). Nat. Genet. *42*, 937–948.

Tung, Y.C., Yeo, G.S., O'Rahilly, S., and Coll, A.P. (2014). Cell Metab. *20*, 710–718.