Bacteroidetes (42%). Specifically, the authors identified almost 400 genes of metabolic pathways that were enriched or depleted in the gut microbiome of obese individuals compared with lean controls, with an interesting example being enrichment of phosphotransferase systems responsible for microbial processing of carbohydrates. Despite the apparent absence of a core microbial community in the human gastrointestinal tract, another recent study did find enrichment of particular microbial groups in the gut of obese individuals: the H_2-producing *Prevotellaceae* (ironically, a family belonging to the *Bacteroidetes*) and the H_2-using *Methanobacteriales* (an order of methanogenic archaea)[5]. Methanogens increase the extraction of energy by the host from otherwise indigestible polysaccharides[7].

The finding by Gordon and colleagues[1] of gene-level core microbiomes may have important implications. With regard to the Human Microbiome Project, it raises the question of whether community profiling using 16S rRNA genes should be used to select samples for metagenomic sequencing as unifying functional patterns may be missed in samples with variable community profiles. It is also possible that drug targets or drug candidates for the treatment of obesity could be identified from the obesity-associated microbiome.

The sudden obesity epidemic is likely to be the result of changes upstream of the human gastrointestinal microbiome (**Fig. 1b**). The role of host genetics, for example, is underscored by the fact that monozygotic twins exhibit a higher degree of co-variation in body adiposity compared with dizygotic twins[8], despite the similar microbial community structures of monozygotic and dizygotic twins demonstrated by Gordon and colleagues[1]. Future progress toward understanding the role of the gut microbiome in body-fat regulation should include collaborative approaches aimed at linking microbial genetic studies with established models of molecular bodyweight regulation. Different gut microbiota may well have differential impacts on (i) afferent gastrointestinal peptide hormones that regulate appetite, energy metabolism and body weight; (ii) portal vein–sensed gut glucose production[9]; and (iii) gut-derived lipid signals to the brain[10]. Changes in the microbiome may also affect what has been termed "metabolic endotoxemia"—increases in plasma lipopolysaccharide (and, presumably, bacterial lipopeptide) concentrations in mice and humans exposed to high-fat diets[11]. In this context, it should be noted that mice with genetic deletions in a variety of pro-inflammatory mediators exhibit exacerbated diet-induced obesity.

To test the mechanistic hypotheses arising from the results of Gordon and colleagues[1] and to develop preventive and therapeutic strategies, we will need to go beyond analysis of the gut microbiome to interdisciplinary collaborations among genome researchers, microbial ecologists, immunologists and obesity experts. An impressive example of such teamwork, also out of the Gordon group[12], who identified short-chain fatty acid products of microbial polysaccharide fermentation as ligands of Gpr41, a gut epithelial cell G protein–coupled receptor that regulates peptide hormone secretion, and went on to show that conventionally raised and gnotobiotic (but not germ free) $Gpr41^{-/-}$ mice showed a leaner phenotype compared with similarly raised $Gpr41^{+/+}$ controls. These results highlight a potentially relevant mechanistic connection between gut microbial function and endogenous molecular pathways controlling energy balance.

Interesting as the new molecular data are, they represent only a dip into the lean and obese gut microbiota gene pools. As the cost of sequencing continues to fall, the next decade should be a productive one for gastrointestinal biology in which microbial-community analysis will be expanded to thousands of individuals and along the complete length of the gastrointestinal tract, and fully integrated with other analytical approaches.

1. Turnbaugh, P.J. *et al. Nature* **457**, 480–484 (2009).
2. Bäckhed, F. *et al. Proc. Natl. Acad. Sci. USA* **101**, 15718–15723 (2004).
3. Ley, R.E., Turnbaugh, P.J., Klein, S. & Gordon, J.I. *Nature* **444**, 1022–1023 (2006).
4. Turnbaugh, P.J. *et al. Nature* **444**, 1027–1031 (2006).
5. Zhang, H. *et al. Proc. Natl. Acad. Sci. USA* **106**, 2365–2370 (2009).
6. Ley, R.E. *et al. Proc. Natl. Acad. Sci. USA* **102**, 11070–11075 (2005).
7. Samuel, B.S. & Gordon, J.I. *Proc. Natl. Acad. Sci. USA* **103**, 10011–10016 (2006).
8. Stunkard, A.J., Foch, T.T. & Hrubec, Z. *J. Am. Med. Assoc.* **256**, 51–54 (1986).
9. Troy, S. *et al. Cell Metab.* **8**, 201–211 (2008).
10. Gillum, M.P. *et al. Cell* **135**, 813–824 (2008).
11. Cani, P.D. *et al. Diabetes* **56**, 1761–1772 (2007).
12. Samuel, B.S. *et al. Proc. Natl. Acad. Sci. USA* **105**, 16767–16772 (2008).

Missing lincs in the transcriptome

Thomas Gingeras

Are long, intervening noncoding (linc) RNAs a new class of functional transcripts?

Ask scientists at an RNA meeting whether most of the recently observed non-protein-coding RNAs are functional, and the group is apt to be divided. Many would acknowledge that these RNAs form the bulk of transcriptomes and that they are characterized by remarkable complexity, but skeptics would cite their lack of sequence conservation and the meager results from functional studies using forward and reverse genetics[1]. A recent report by Guttman *et al.*[2] helps to shed new light on this debate by identifying a large number of non-protein-coding RNAs that are enriched in evolutionarily conserved sequences and map to intergenic regions.

The complexity of non-protein-coding RNAs is evidenced by their broad incorporation of genome sequences, their interleaved organization and the variety of types detected, including long and short polyadenylated and nonpolyadenylated RNAs[3–5]. Until recently, however, the extent of this complexity in cells from yeast[6] to humans[7] has been underappreciated.

Interestingly, Guttman *et al.*[2] began not with experimental discovery of novel RNAs but with a computational approach in which they searched outside of protein-coding sequences for chromatin signatures of actively transcribed genes (**Fig. 1a**). This chromatin signature was defined as tracts of trimethylated lysine 4 of histone H3 (indicative of transcriptional initiation at promoters) adjacent to tracts of trimethylated lysine 36 of histone H3 (indicative of elongation of transcribed regions). From an analysis of four mouse cell types, the authors identified 1,250 unannotated intergenic regions at least 5 kb in size.

Subsequent screening of these regions for evolutionarily conserved sequences provided a prioritized list of candidate non-protein-coding RNAs. Expression levels for 350 of these were measured using custom tiling arrays, northern hybridizations and RT-PCR (**Fig. 1b**). Similar to what was previously reported for intergenic non-protein-coding

Thomas Gingeras is at Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. e-mail: gingeras@cshl.edu

transcripts of unknown function[8] in the pilot phase of the ENCODE project[9], these experiments yielded a verification rate of ~70%, supporting the presence of spliced RNAs mapping to these regions. The authors named these transcripts large intervening noncoding (linc) RNAs.

Evidence that lincRNAs are biologically important was inferred from an impressive set of computational and empirical analyses (**Fig. 1c**). These included (i) correlation of lincRNA expression patterns with those of protein-coding genes using RNA profiles obtained from 16 mouse samples; (ii) detailed analysis of lincRNAs' expression changes in $p53^{+/+}$ and $p53^{-/-}$ mouse embryonic fibroblast cells exposed to DNA damaging agents; (iii) inferred association of two core transcription factors (Oct4 and Nanog) to lincRNA promoter sites; (iv) detected binding of Oct4 and Sox2 together at a cloned promoter region for one of the lincRNAs; and (v) identification by means of a literature search of a lincRNA whose perturbed expression (in a short hairpin RNA screen) resulted in an altered proliferation phenotype[10]. These data implicate the identified long noncoding RNAs in a variety of biological processes and provide a method for locating other potentially important noncoding RNAs. Thus, lincRNAs join a long list of functional long noncoding RNAs[11,12].

Guttman *et al.*[2] emphasize two defining characteristics of lincRNAs: (i) they are situated entirely in the intervening regions between genes, and (ii) as a group, they are enriched in evolutionarily conserved sequences. Although the authors note that transcripts lacking sequence conservation could be functional, they seem to favor the view that most noncoding transcripts—generated by the pervasive transcription observed in cells—represent biological noise.

This perspective may very well be correct, but two points in their data related to these defining characteristics are worth review.

First, although the 5′-start and 3′-stop positions were determined for a few of the lincRNAs, the boundaries of most of them were loosely defined using custom tiling arrays, which interrogate only limited genomic regions. Thus, the 5′ and 3′ ends of many as-yet-uncharacterized lincRNAs may overlap the transcribed regions of neighboring protein-coding genes, as has been observed in the pilot ENCODE project[9]. If lincRNAs, upon more detailed sequence characterization, are found to overlap known or unknown genes, they would have to be reclassified as among the many previously described overlapping RNAs[5,7]. Consequently, the creation of a new class of noncoding RNAs with a key characteristic of being intervening may not be as helpful as intended.

A second question concerns the way in which the functionality of lincRNAs is assessed. In Guttmann *et al.*[2], the larger the length of evolutionarily conserved sequences in a lincRNA, the more likely it is to be considered functional. However, it has been well documented that long primary non-protein-coding transcripts are precursors for short functional RNAs—such as micro, short interfering, nucleolar, small nuclear, transfer and ribosomal RNAs[10]—and that, in regions outside the embedded shorter functional RNAs, these transcripts show much less evolutionary conservation. Thus, it would seem reasonable to consider that primary transcripts with only short regions of conservation may have a functional role, possibly involving the processing of short functional RNAs. Such observations have been made previously[7], and it would seem judicious to consider whether these less well-conserved transcripts should be included in the lincRNA collection.

The report by Guttmann *et al.*[2] reminds us that there are many more interesting and potentially functional RNAs to be identified. New empirical approaches to discover, categorize and prioritize these RNAs—such as those described here—are very much needed to focus future genetic experiments. Ultimately, however, it may be necessary to embark upon a new era of large-scale systematic full-length cDNA cloning and sequencing to completely catalog all the functional long and short RNAs made by cells. Such future large-scale experiments would be differentiated from the shotgun experiments performed decades ago if they were guided by the results of recent RNA mapping experiments, as exemplified by the work of Guttman *et al.*[2].

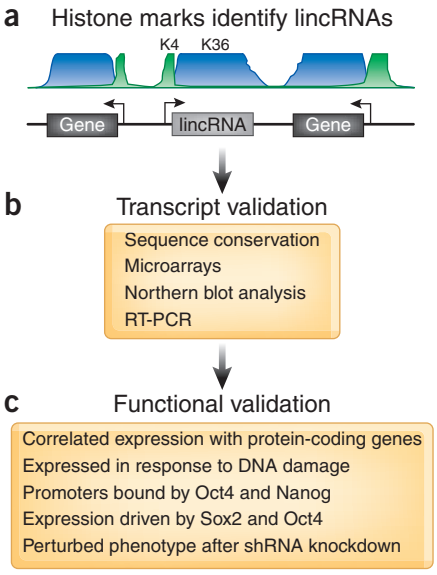

Figure 1 Identification and validation of long intervening noncoding RNAs. (**a**) Patterns of H3K4me3 (trimethylated lysine 4 of histone H3, green) adjacent to H3K36me3 (trimethylated lysine 36 of histone H3, blue) mark transcription units and identify unannotated transcripts positioned between protein-coding genes. (**b**) Large multi-exonic non-protein-coding RNA transcripts are detected by a variety of methods. (**c**) Experimental and computational validation suggests that lincRNAs are biologically functional. shRNA, short hairpin RNA.

1. Struhl, K. *Nat. Struct. Mol. Biol.* **14**, 103–105 (2007).
2. Guttman, M., *et al. Nature* **458**, 223–227 (2009)
3. Kapranov, P. *et al. Nat. Rev. Genet.* **8**, 413–423 (2007).
4. Ponjavic, J. & Ponting, C.P. *Bioessays* **29**, 1077–1080 (2007).
5. Carninci, P. *et al. Curr. Opin. Cell Biol.* **20**, 274–280 (2008).
6. Nagalakshmi, U. *et al. Science* **320**, 1344–1349 (2008).
7. Kapranov, P. *et al. Science* **316**, 1484–1488 (2007).
8. Cheng, J. *et al. Science* **308**, 1149–1154 (2005).
9. ENCODE Project Consortium. *Nature* **447**, 799–816 (2007).
10. Ivanova, N. *et al. Nature* **442**, 533–538 (2006).
11. Mercer, T.R. *et al. Nat. Rev. Genet.* **10**, 155–159 (2009).
12. Ponting, C.P. *et al. Cell* **136**, 629–641 (2009).