Thematic review series: Systems Biology Approaches to Metabolic and Cardiovascular Disorders

Approaches to lipid metabolism gene identification and characterization in the postgenomic era

Karen Reue[1] and Laurent Vergnes

Department of Human Genetics and Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, and Department of Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA 90073

Abstract **The availability of genomic resources has already had a tremendous impact on biomedical research. In this review, we describe how whole genome sequence and high-throughput functional genomics projects have facilitated the identification and characterization of important genes in lipid metabolism and disease. We review key approaches and lipid genes identified in the first years of this century and discuss how genomic resources are likely to streamline gene identification and functional characterization in the future.**—Reue, K., and L. Vergnes. **Approaches to lipid metabolism gene identification and characterization in the postgenomic era.** *J. Lipid Res.* 2006. **47:** 1891–1907.

Supplementary key words comparative genomics • positional cloning • functional genomics • gene-trap mutagenesis • ethylnitrosourea mutagenesis

In 2001, the human genome sequence was published (1). The availability of this information has accelerated the pace of discovery and characterization of genes with important roles in lipid metabolism and disease. Our review focuses primarily on gene discovery and characterization in the mouse, because it has been the experimental model of choice for lipid research and biomedical research in general. The identification of important lipid metabolism genes has been facilitated in the mouse by the existence of a vast number of genetically distinct mouse strains that harbor mutations leading to differences in susceptibility to dyslipidemias, atherosclerosis, hypertension, obesity, and diabetes. These mouse strains have been the starting point for the identification of many genes influencing lipid metabolism, and **Fig. 1** illustrates some current strategies that have been aided by the availability of genomic resources. These include the identification of specific mutations underlying altered lipid metabolism via

positional cloning on the one hand and high-throughput gene-trap mutagenesis on the other. The current genomic resources also facilitate the translation of findings in mouse or other experimental model organisms to humans. Thus, the human counterparts of genes identified in animal models may be immediately obtained and used as candidate genes for human diseases, in genetic association studies, or to evaluate their role in disease by examination of gene expression levels (Fig. 1, bottom). In addition, in silico studies comparing human and mouse genome sequences allow the identification of new genes and regulatory elements. Here, we describe current technologies and recent findings in the identification and characterization of lipid metabolism genes.

IDENTIFICATION OF LIPID METABOLISM GENES BY GENOME SEQUENCE ANALYSIS

The current "finished" version of the human genome sequence released in 2004 covers >99% of the euchromatic regions with an error rate of <1 in 100,000 nucleotides and contains only ∼300 remaining sequence gaps (reduced from 150,000 gaps in the draft sequence released in 2001) (1, 2). Inherent in the nucleotide sequence is information about the chromosomal landscape, including the presence of local and long-range duplications, disperse and tandem repeat sequence families, and, of greatest interest, an estimated 20,000–25,000 protein-coding genes. However, the genes constitute the vast minority of the genome, accounting for <2% of the sequence, whereas >45% of the genome is composed of interspersed repeat sequences, many of which are relics of retroviruses and DNA transposons. The high ratio of repetitive sequence to recognizable functional sequence in the genome has made it challenging to identify the complete complement of protein-coding genes.

---

[1] To whom correspondence should be addressed.
  e-mail: reuek@ucla.edu
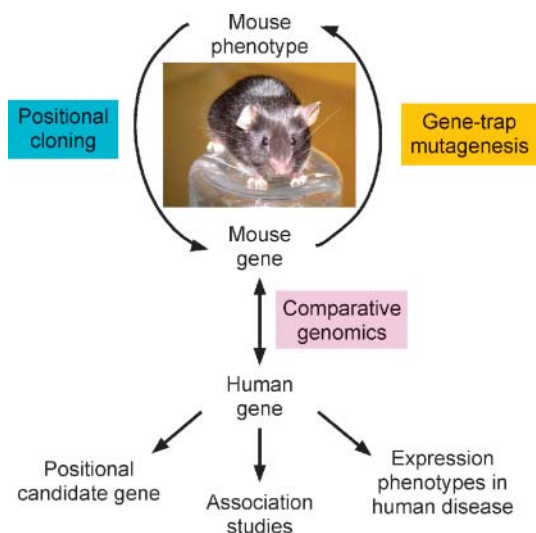
Fig. 1. Strategies for gene identification and characterization in the mouse. The availability of human and mouse genome resources facilitates gene identification on several levels. First, mutations and genetic variations underlying mouse phenotypes can be more rapidly identified by positional cloning, often revealing novel genes or previously unknown gene functions. Functional genomic approaches, such as gene-trap mutagenesis, can be used to generate live mouse models carrying mutations of interest to study physiological function. Comparative genomics allows the translation of findings in the mouse to human disease and can also be used to identify genes and regulatory sequences.

One approach to better decipher the information contained in the human genome has been to compare it with genome sequences from other organisms (3) (see **Table 1** for genome browsers and other pertinent resources). Comparative genomics is based on the premise that functional sequences have been positively selected over evolutionary time, whereas random mutations have been tolerated in sequences without specific function (for reviews, see 4–6). The choice of which genomes to compare depends on the type of information being sought. A comparison of genome sequences from human and chimpanzee, which diverged ∼8 million years ago and exhibit 95% genome-wide sequence similarity, will not distinguish between functional and nonfunctional sequences because the species diverged so recently. On the other hand, the divergence of mouse and human 75–80 million years ago provides a suitable evolutionary distance for the identification of functional sequences without the high background of similarity across the entire genome that occurs between primates. Importantly, functional similarities between human and mouse are easily established, as >99% of human genes have a detectable homolog in the mouse, with high conservation of gene order and exon-intron organization. Although only ∼2% of the genome consists of protein-coding sequences, there is ∼5% sequence conservation between human and mouse. This suggests that in addition to protein-coding sequences, regulatory elements and non-protein-coding genes have been conserved as a result of functional constraints (7).

The comparison of human and mouse genome sequences has proven to be a valuable tool for the identifi-

cation of genes, including new genes in lipid metabolism. A notable example of the power of this approach was the identification of the apolipoprotein A-V (apoA-V) gene (*APOA5*), a new member of the apolipoprotein gene family. By comparing human and mouse genomic sequences in the vicinity of the *APOA1/C3/A4* gene cluster using an algorithm know as VISTA, Pennacchio et al. (6, 8) detected a region of high interspecies conservation corresponding to the exons of a previously unidentified gene located 35 kb downstream of *APOA4*. The existence of expressed sequence tags in sequence databases provided evidence that the gene was indeed expressed as mRNA, and it was subsequently shown that *APOA5* is expressed primarily in the liver. The generation of transgenic and gene-knockout mouse models revealed a role for apoA-V in triglyceride metabolism (8). Expression of an *APOA5* transgene reduced circulating triglyceride levels to one-third of normal, whereas apoA-V deficiency led to a 4-fold increase over normal triglyceride levels. Furthermore, polymorphisms in the human *APOA5* gene were shown to be associated with triglyceride levels in multiple populations, indicating that this is an important gene in triglyceride homeostasis (8, 9).

Although the example of *APOA5* is striking, it raises the question of whether this was a unique case or does comparative genomics have further applicability for the identification of lipid metabolism genes? In fact, despite the availability of nearly all of the genome sequence, many genes have not been annotated accurately, or at all, and comparative genomics can be extremely useful in this regard. We previously used the available mouse genome sequence to characterize the mouse aldo-keto reductase (AKR) gene cluster and to identify new AKR family members based on similarities in protein-coding sequences (10). AKRs represent a superfamily of oxidoreductases that catalyze the NADP(H)-dependent reduction of a wide variety of aldehydes and ketones, including steroid hormones, prostaglandins, and many xenobiotic compounds (11). The AKR1C subfamily members function in steroid hormone homeostasis and also act as bile acid binding proteins. To better characterize the members of the AKR1C subfamily, we analyzed mouse genome sequence using sequences of four known *Akr1c* genes and identified four additional members of the AKR gene superfamily residing in a cluster on chromosome 13 (10). The high degree of similarity in predicted amino acid sequence among the eight AKR1C proteins suggested that they have arisen through repeated gene duplication events but have diverged in substrate specificity and expression patterns. Because of the high sequence similarity between these AKR gene family members, it had been impossible to unambiguously identify the corresponding human gene homologs simply by standard sequence comparisons, such as BLAST (Basic Local Alignment Search Tool). However, this is possible through analysis with the VISTA alignment tool (Table 1) that was used to identify the *APOA5* gene, because it integrates both DNA sequence and position within the genome. Applying the VISTA tool to a region of human chromosome 10, we can now identify not only the six previously described human AKR genes but also three

TABLE 1. Bioinformatics tools for gene identification and characterization

Bioinformatics Tool	Web Page	Description
Genome browsers and gene information	www.ensembl.org genome.ucsc.edu www.ncbi.nlm.nih.gov/genome	Depositories of many vertebrate and nonvertebrate genomes
	www.genomesonline.org	Status and references for many genome sequencing projects
	www.informatics.jax.org	Mouse genome informatics
	mgc.nci.nih.gov	Database and supplier of full-length cDNA clones for human, mouse, and rat genes
	www.geneontology.org	Gene ontology and functional classifications
Comparative genomics	genome.lbl.gov/vista/index.shtml ecrbrowser.dcode.org	VISTA tool for genome sequence comparison between many organisms
	www.informatics.jax.org/menus/homology_menu.shtml	Mammalian orthology and comparative maps
	eshadow.dcode.org	Tool for comparative genomics between closely related species
Haplotype analysis	mousesnp.roche.com snp.gnf.org/GNF10K www.ncbi.nlm.nih.gov/SNP	Mouse and human single nucleotide polymorphism databases
Gene expression	symatlas.gnf.org	Database of gene expression levels in human and mouse tissues
	www.ncbi.nlm.nih.gov/geo	Depository of microarray database for different organisms under a variety of conditions
	pga.tigr.org	Microarray profiles of rodent models of human disease
Functional genomics	www.genetrap.org	Database of publicly available gene-trap embryonic stem cell clones
	omnibank.lexgen.com www.deltagen.com	Commercially available collections of gene-trap embryonic stem cell clones and mice
	pga.jax.org/resources	Ethylnitrosourea mutagenesis program at the Jackson Laboratory
	www.mgu.har.mrc.ac.uk/mutagenesis	Ethylnitrosourea mutagenesis program at Harwell
	www.wormbase.org	*C. elegans* functional genomics database
	www.celeganskoconsortium.omrf.org	*C. elegans* knockout consortium

previously unannotated members, which likely correspond to mouse *Akr1c21*, *Akr1c19*, and *Akr1c14* (**Fig. 2**, regions A–C). This analysis reveals that the human and mouse AKR gene clusters are conserved in number of genes, exon-intron organization, and orientation of genes within the cluster.

In addition to protein-coding genes, the identification of *cis*-acting regulatory sequences can be facilitated using the comparative genome approach. An example is the identification of a new enhancer for brain expression near the apoE gene cluster. Using a human and mouse genome comparison, it was possible to identify a 600 bp noncoding region with 68% identity between the two mammals (12). The region, 42 kb distal to the apoE gene, was located at a substantial distance from known genes and regulatory regions and was identified solely based on the high level of conservation between human and mouse sequences. The in vivo function of this enhancer element was demon-

strated by the production of three transgenic mouse lines in which the element effectively drove expression in restricted regions of the brain. These results suggest that this element may play a role in the regulation of apoE or other nearby apolipoprotein genes within the cluster.

A related, but distinct, comparative genomic approach has been used to identify regulatory elements upstream of the human apolipoprotein [a] (apo[a]) gene. Apo[a] is unusual in that no homolog exists in the mouse or most other non-primates, so comparison of distant mammalian genomes is not feasible (13). Therefore, Rubin and colleagues (14) instead analyzed apo[a] gene and upstream sequences in several closely related species using an approach known as "phylogenetic shadowing." Phylogenetic shadowing is based on the premise that, even in closely related species, DNA sequences that are not critical will vary, at least some of the time (15). Thus, by analyzing
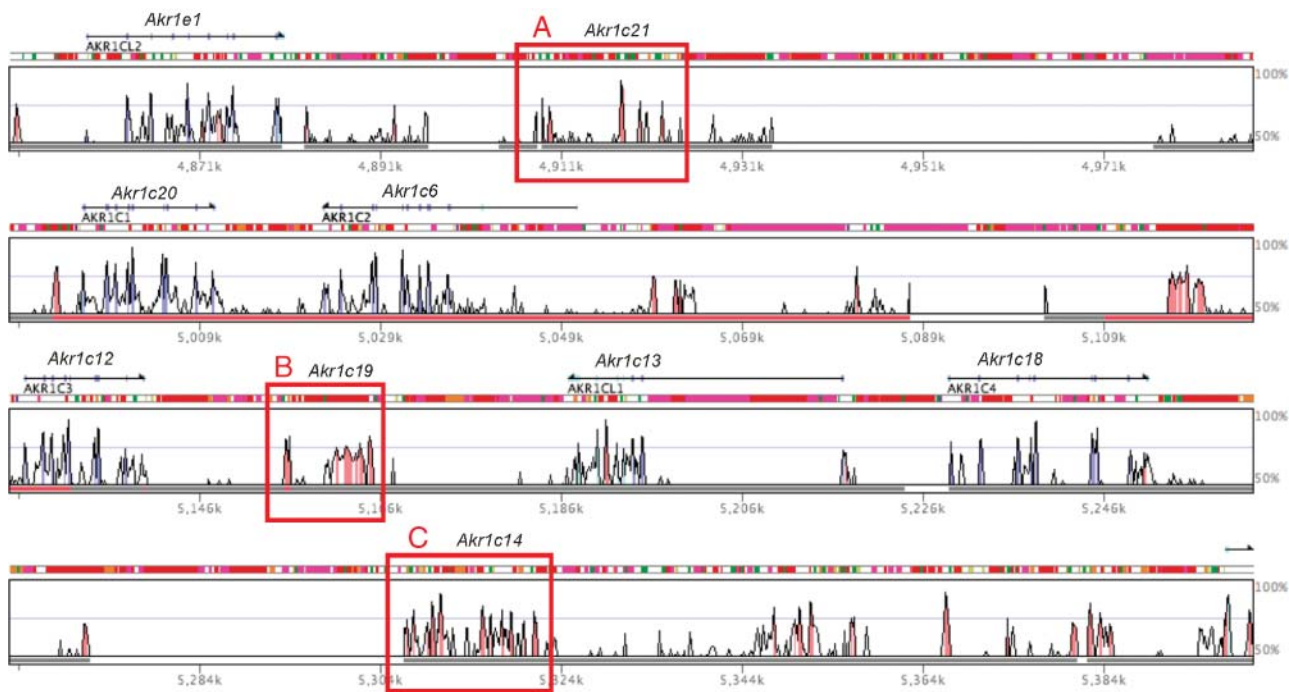
Fig. 2. Identification of new human aldo-keto reductase (AKR) genes through comparative sequence analysis. The VISTA alignment tool was used to align a region of human chromosome 10 (positions 4,850,500–5,400,000) with the syntenic region of mouse chromosome 13, containing members of the AKR gene family. The graphs show the percentage nucleotide similarity between mouse and human (50–100%). The six known human AKR genes (*AKR1CL2*, *AKR1C1*, *AKR1C2*, *AKR1C3*, *AKR1CL1*, *AKR1C4*) are indicated by black arrows above the alignment. The corresponding mouse genes are shown above these in italics. Boxed regions A–C denote putative new human AKR genes, which we propose to be the orthologs of mouse *Akr1c21*, *Akr1c19*, and *Akr1c14*.

sequences covering a 1.6 kb region near the 5′ end of the apo[a] gene in 18 different primates, it was possible to identify a novel regulatory sequence found only in Old World monkeys and hominids. This technique is a powerful tool to identify primate-specific regulatory motifs, which cannot be detected through human-mouse sequence comparisons.

With additional sequence information that will become available through projects such as ENCODE (for Encyclopedia of DNA Elements), which will sequence 1% of the orthologous region of the genome in >20 mammalian species, and new tools developed for the analysis of both evolutionarily close and distant genomes, comparative genomics will become an even more powerful approach for deciphering information not only about protein-coding genes but also about noncoding microRNA genes and *cis*-regulatory sequence (16, 17).

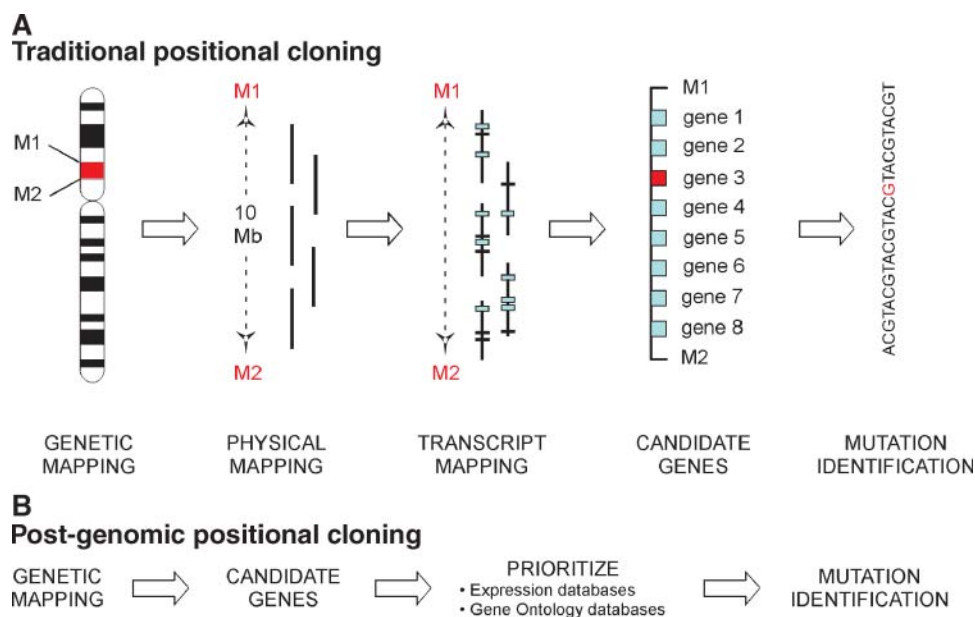## IDENTIFICATION OF LIPID METABOLISM GENES BY POSITIONAL CLONING

Identification of genetic determinants of common diseases such as atherosclerosis, obesity, and diabetes has been a popular, but often elusive, goal. A key difficulty is the inherent complexity of these disorders, which involve interactions between the environment and multiple genes, each of which makes a relatively small contribution to overall susceptibility. On the other hand, the identification of genes underlying Mendelian disorders in lipid

metabolism is more straightforward and has provided valuable information about the physiological role of many genes that may also exhibit genetic variations in more common lipid metabolism phenotypes. One method for gene identification is the candidate gene approach, in which genes suspected of contributing to a disease or trait are tested for association between the disease and genetic polymorphisms in humans, or by generating transgenic or knockout mouse models. A limitation of this approach is that only known genes previously implicated in aspects of the disease are considered, precluding the discovery of new genes or new roles for known genes. These limitations can be overcome using approaches such as positional cloning, in which a disease gene is identified based on its position in the genome and segregation with the disease, with no assumptions about function.

Positional cloning strategy

The positional cloning approach traditionally involves genetic mapping of the disease phenotype, delimiting the physical stretch of DNA sequence spanning the region, identifying all genes residing there, and testing these candidates for involvement in the disease by sequence analysis in affected versus unaffected individuals (**Fig. 3A**). One of the first successful positional cloning efforts, to isolate the gene underlying Huntington's disease, required almost 10 years of intensive effort by an international consortium to identify the gene, even after the identification of a closely linked genetic marker (18). Fortunately, the

A **Traditional positional cloning**

GENETIC MAPPING   PHYSICAL MAPPING   TRANSCRIPT MAPPING   CANDIDATE GENES   MUTATION IDENTIFICATION

B **Post-genomic positional cloning**

GENETIC MAPPING   CANDIDATE GENES   PRIORITIZE
• Expression databases
• Gene Ontology databases   MUTATION IDENTIFICATION

Fig. 3. Traditional versus postgenomic positional cloning strategies. A: Before the availability of the genome sequence, positional cloning involved several labor-intensive steps. After genetic mapping to a chromosomal region, the physical portion of the genome was isolated on large insert DNA clones, often requiring dozens of clones to cover the region. Genes residing on the large insert clones were identified experimentally. Once genes were identified, they were evaluated for potential involvement in the disease by gene expression analysis and DNA sequencing to identify the mutation. B: The availability of the genome sequence streamlines the positional cloning approach, supplanting experimental techniques with in silico analysis of physical map position and gene content. Additional resources, such as gene ontology and gene expression databases, help prioritize candidate genes for mutation analysis.

availability of genomic resources, including genome sequences, polymorphisms, and gene expression information, has streamlined this process in several respects. As a result, several important lipid metabolism genes, underlying both Mendelian and complex disorders, have recently been identified, as shown in **Table 2** and described below.

The initial step in positional cloning, genetic mapping, has been accelerated by the identification of thousands of microsatellite and single nucleotide polymorphism (SNP) markers, which are distributed throughout the mouse and human genomes and available in the databases listed in Table 1. This facilitates rapid, high-resolution genetic mapping, leading to efficient narrowing of the critical gene region. The identification of haplotype blocks in both the human and mouse genomes has further increased mapping efficiency, particularly when searching for common genetic variants that underlie complex traits (reviewed in 19–21). Haplotype blocks in the human genome are regions of linkage disequilibrium that typically span several megabases. The functional implication is that a single marker within a haplotype block can suffice to capture all information within that stretch when searching for linkage (22). Representative SNP markers for specific regions of the genome have been defined and are accessible in the HapMap database (23). The mouse genome also exhibits a type of haplotype block structure attributable to the derivation of hundreds of different inbred strains from a few ancestral species, whose genomes have been distributed in distinct combinations in each of the modern-day strains (24). This feature can be exploited in mapping the location of a trait that differs between two mouse strains, in that underlying genetic differences may be more likely to reside in regions that carry different ancestral alleles (20). This is primarily applicable to mapping common genetic variations rather than spontaneous single gene mutations.

The availability of whole genome sequences has virtually eliminated the two labor-intensive intermediate steps of the positional cloning process, physical mapping and identification of genes in the region (Fig. 3). As recently as a few years ago, once a candidate region was delimited by genetic markers, it was necessary to define the physical map region and genes contained therein experimentally by isolating large insert DNA clones followed by exon trapping or cDNA selection (25, 26). This tedious process has largely been replaced by simple database queries. The genetic markers that delimit the candidate region can be submitted to a genome browser, and with a few keystrokes, the entire physical region, including sequence and known and predicted genes, is accessible (Table 1). This effectively converts the positional cloning approach into a "positional candidate" approach (Fig. 3B). It should be noted, however, that identification of novel genes by this method still offers challenges, as many uncharacterized genes are still annotated incorrectly, or not predicted at all, in current versions of the genome sequence (e.g., *Diet1*, described below).

Finally, whole genome data are valuable in screening candidate genes within a mapped region, which can be

TABLE 2. Lipid metabolism genes identified by positional cloning (1997–2006)

Mutation	Type	Causative Gene	Protein	Lipid Phenotype	References
Mouse					
add, adrenocortical lipid depletion	Recessive	Soat1	Acyl-CoA cholesterol acyltransferase (ACAT)	Cholesteryl ester depletion in adrenal cortex at puberty	132
Ath1, atherosclerosis 1	QTL	Tnfsf4	Tumor necrosis factor (ligand) superfamily, member 4	Increased susceptibility to diet-induced aortic lesions, reduced HDL cholesterol levels	75
Atherosclerosis resistance	QTL	Alox5	5-Lipoxygenase	Resistance to aortic lesion formation despite hypercholesterolemia	74
cld, combined lipase deficiency	Recessive	Unpublished	Unpublished	Chylomicronemia associated with low levels of lipoprotein lipase and hepatic lipase, leading to neonatal lethality	Doolittle and Peterfy, personal communication
cpk, congenital polycystic kidney	Recessive	Cys1	Cystin 1	Renal cysts, altered glycolipid and sulfolipid content in kidney	133
Diet1, resistance to atherogenic diet	Recessive	Diet1	Diet1 (novel protein)	Resistance to diet-induced hypercholesterolemia and aortic lesions	Vergnes and Reue, unpublished data
fld, fatty liver dystrophy	Recessive	Lpin1	Lipin	Adipose tissue deficiency, transient fatty liver during neonatal period, persistent peripheral neuropathy	39
Hyplip1, hyperlipidemia 1	Recessive	Txnip	Thioredoxin-interacting protein	Increased cholesterol, triglycerides, apolipoprotein B	47
hypl, hypolipidemic	Recessive	Angptl3	Angiopoietin-like 3	Hypolipidemia in the presence of hyperinsulinemia and hyperglycemia	59
jvs, juvenile visceral steatosis	Recessive	Slc22a5	Solute carrier family 22 (organic cation transporter), member 5	Fatty liver, hypoglycemia, increased urinary acylcarnitine excretion	134, 135
lcsd, lysosomal cholesterol storage disorder	Recessive	Npc1	Niemann-Pick type C1	Lysosomal storage of unesterified cholesterol, foam cells in visceral tissues, myelin depletion, death	136
pcy, polycystic kidney disease	Recessive	Nphp3	Nephrophthisis 3	Renal cysts, abnormal fatty acid and phospholipid content in kidney, increased serum cholesterol and phospholipids	137
spm, sphingomyelinosis	Recessive	Npc1	Niemann-Pick type C1	See lcsd	136
T2dm2, type 2 diabetes mellitus 2	QTL	Sorcs1	VPS10 domain receptor protein, SorCS 1	Impaired insulin secretion in response to glucose challenge and disrupted pancreatic islet morphology on obese background	76
wrf, wrinkle-free	Recessive	Slc27a4	Solute carrier family 27, member 4; fatty acid transporter 4 (FATP4)	Thick skin, lipid droplets in perinatal epidermis, impaired epidermal lipid homeostasis	138
Human					
ARH, autosomal recessive hypercholesterolemia	Recessive	ARH	ARH (LDL receptor adaptor protein)	Hypercholesterolemia, xanthomas, premature atherosclerosis, impaired hepatic LDL receptor function	38
FCHL, familial combined hyperlipidemia	QTL	USF1	USF-1, upstream stimulatory factor-1	Increased plasma cholesterol and/or triglycerides, potentially with low HDL cholesterol, small dense LDLs, increased apolipoprotein B, and premature coronary heart disease	91
Obesity	QTL	SLC6A14	Solute carrier family 6, member 14	Obesity (defined as body mass index > 30 kg/m²)	94
Sitosterolemia	Recessive	ABCG5 or ABCG8	ATP binding cassette transporter family member (ABCG5 or ABCG8)	Hyperabsorption of dietary sitosterols (plant sterols) and cholesterol, hypercholesterolemia, xanthomas, premature coronary artery disease	31
Tangier disease	Recessive	ABCA1	ATP binding cassette transporter family member (ABCA1)	HDL deficiency, enlarged tonsils attributable to cholesteryl ester accumulation, hepatosplenomegaly, peripheral neuropathy	35–37
Other					
Shumiya cataract rat	Two semidominant alleles	Lss and Fdft1	Lanosterol synthase and farnesyl diphosphate farnesyltransferase 1 (squalene synthase)	Cholesterol deficiency-associated cataracts at 11 weeks of age, onset associated with the combination of two hypomorphic mutant genes involved in cholesterol biosynthesis	34
Fat-free zebrafish	Recessive	ffr	Fat-free	Impaired intestinal lipid absorption associated with altered Golgi structure and vesicular trafficking, leading to lethality	33

QTL, quantitative trait locus.

prioritized using a combination of in silico and experimental data. For example, candidate genes can be investigated in the Gene Ontology database (Table 1), in which genes are classified on the basis of all known information with respect to their cellular compartment (e.g., nucleus, endoplasmic reticulum, etc.), biological process (e.g., signal transduction, fat cell differentiation, etc.), and molecular function (e.g., catalytic activity, receptor, etc.) (27). Analysis of the Gene Ontology classification for all candidate genes at a particular locus may indicate high-priority candidates based on attributes that relate to the phenotype. Another criterion for inclusion or exclusion of candidate genes is the tissue expression pattern. For example, a gene that underlies abnormal very low density lipoprotein production may be expected to have high expression levels in liver, whereas a gene that affects adipocyte differentiation is expected to be expressed in adipose tissue. Rather than perform Northern blot or RT–PCR experiments to determine the expression pattern of a set of candidate genes, expression information for the vast majority of known and predicted genes can be obtained through DNA microarray databases (28). For example, the Genome Institute of Novartis (Table 1) provides expression data for nearly 80 human tissues, including 21 brain subregions and several classes of B- and T-cells in addition to standard tissue types, as well as for >60 mouse tissues, including seven stages of embryonic development. These in silico tools help prioritize candidate genes to analyze further using experimental approaches.

Although the isolation of genes via positional cloning typically follows a standard strategy, the details leading to the identification of specific gene mutations are unique to each mutant gene. However, a few generalities have emerged. In the mutations described here, the most fruitful approach to screen a series of candidate genes has been to investigate altered gene expression levels in the mutant. This presupposes that the mutation is of a type that will lead to altered mRNA transcription or stability rather than affecting the function of a protein that is otherwise expressed at normal levels. Interestingly, in the recently cloned lipid metabolism mutations, altered expression levels have been prevalent, suggesting that such mutations are quite common or that this type of mutation is more easily identified and therefore overrepresented in the mutant genes identified to date. The causative mutations include gene rearrangements, small insertions or deletions, and point mutations that lead to premature stop codons, so that the mRNA transcripts produced from such genes are degraded, presumably by nonsense-mediated mRNA decay (29). The methods for screening gene expression alterations range from gene-by-gene analysis via Northern blot or RT-PCR to characterization of global mRNA expression patterns by DNA microarray hybridization. The use of microarray analysis to compare gene expression patterns in wild-type versus mutant tissues was instrumental in the identification of the *Cd36* gene mutation in the spontaneous hypertensive rat and in the identification of the *ABCG5* and *ABCG8* mutations in human sitosterolemia (30, 31). More recently, microarray analysis has been combined with linkage analysis in segregating mouse and rat populations to identify genes underlying complex traits, as described below.

Identification of mutations underlying single gene disorders

Coincidental with the availability of the genome resources described above, there has been notable progress in the positional cloning of genes underlying lipid metabolism disorders in mouse and other experimental organisms, as well as in human (see Table 2 for lipid-related genes identified by positional cloning since 1997). The majority of these have been single gene disorders, but successes in complex disorders have also been reported. Ten years ago, we reviewed about a dozen naturally occurring single gene mouse mutations affecting lipid metabolism, many of which had been genetically mapped but the genes not identified (32). Interestingly, aside from three mutations in extinct mouse strains, all of these genes have now been identified. These include genes that are now quite familiar in the lipid field, such as the *Npc1* (Niemann Pick type C1) gene, identified as the causative mutation underlying both the lysosomal cholesterol storage disorder (*lcsd*) and sphingomyelinosis (*spm*) mouse mutations, and the *Soat1* gene encoding sterol *O*-acyltransferase 1 (more commonly known as ACAT-1), responsible for the causative mutation in the *ald* (adrenocortical lipid depletion) mouse (Table 2 and references therein). In addition to the mouse, positional cloning in model organisms has identified lipid metabolism genes such as *fat-free* (*ffr*), required for normal intestinal lipid absorption in zebrafish, and mutations in two separate cholesterol synthetic genes as a cause of cataracts in the Shumiya rat (33, 34).

In humans, positional cloning and/or positional candidate methods have been used to identify mutations underlying single gene disorders in three separate members of the ABC transporter family: mutations responsible for sitosterolemia were identified in *ABCG5* and *ABCG8*, whereas Tangier disease and familial HDL deficiency, characterized by virtually absent circulating HDL, were attributed to mutations in *ABCA1* (31, 35–37). Positional cloning was also used to identify mutations underlying human autosomal recessive hypercholesterolemia in a novel gene encoding a putative LDL receptor adaptor protein (38). Notably, a few genes underlying multigenic lipid disorders have also been isolated via positional cloning very recently in both mouse and human, as described below. As an illustration of one of the strengths of the positional cloning approach, many of the recently identified genes have been completely novel genes without any previous characterization (e.g., *Lpin1*, *Diet1*) or have been known genes not previously implicated in lipid metabolism (e.g., *Txnip*, *Angptl3*). Below, we describe a few of the recently positionally cloned mouse genes that are particularly relevant to dyslipidemias; for more examples, see Table 2.

The fld mutation. In 2001, positional cloning of the fatty liver dystrophy (*fld*) mouse mutation identified a novel

gene, lipin (*Lpin1*), which defined a new family of three related genes in mammals (*Lpin1*, *Lpin2*, and *Lpin3*) (39). Interestingly, single lipin homologs also exist in invertebrates and single-celled eukaryotes, suggesting a fundamental cellular role for lipin. The *fld* mutation leads to a rearrangement of the *Lpin1* gene, producing a null mutation. Lipin-deficient mice exhibit fatty liver and hypertriglyceridemia during the neonatal period, followed by peripheral neuropathy beginning at ∼3 weeks of age, and complete lack of adipose tissue and secondary insulin resistance throughout their lifetime (reviewed in 40). Lipin is expressed predominantly in adipose tissue, skeletal muscle, and testis, with lower level expression detectable in several other tissues. Studies with lipin-deficient *fld* mice and lipin transgenic mice have revealed that lipin is required for normal adipocyte development and that it plays a role in energy metabolism in muscle (41–43).

Whereas lipin-deficient mice are lipodystrophic as a result of impaired fat cell differentiation, lipin transgenic mice become obese, suggesting that subtle genetic variations that influence lipin expression levels could influence fat content and glucose homeostasis in humans. Indeed, this appears to be the case, as an association study in which polymorphisms spanning the human *LPIN1* gene were typed in Finnish dyslipidemic families and in obese and lean subjects revealed significant associations with body mass index and insulin levels (44). It is of interest to determine the molecular function of lipin and its family members. Two recent studies in the yeast *Saccharomyces cerevisiae* provide interesting clues. In one study, the yeast lipin homolog was shown to have a role in regulating phospholipid biosynthesis and nuclear growth during the cell cycle (45). Another group studying yeast determined that lipin acts as a magnesium-dependent phosphatidic acid phosphatase, an enzyme with a key role in triacylglycerol synthesis through the conversion of phosphatidic acid to diacylglycerol (46). It remains to be determined whether mammalian lipin has similar properties and what specific roles the individual lipin family members play.

The Hyplip1 mutation. In 2002, the gene underlying the *Hyplip1* mouse mutation was identified by positional cloning and found to reside in the gene for thioredoxin-interacting protein (*Txnip*), also previously known as the tumor suppressor, vitamin D_3-upregulated protein (VDUP-1) (47). The *Hyplip1* mutant mouse exhibits metabolic abnormalities similar to familial combined hyperlipidemia (FCHL), with increased cholesterol, triglyceride, and apoB levels and increasing severity of the hyperlipidemia with age (48). *Hyplip1* mice carry a nonsense mutation in the *Txnip* gene, leading to dramatically reduced mRNA levels (47). TXNIP appears to have a role in modulating the cellular redox state, with important metabolic effects. In liver, these include the regulation of biochemical branch points for gluconeogenesis and lipogenesis, such that TXNIP-deficient animals divert substrates such as phosphoenolpyruvate to lipogenesis rather than gluconeogenesis, contributing to hyperlipidemia in these animals (49, 50). Interestingly, it was determined that TXNIP-deficient

mice appear to have similar levels of thioredoxin activity as wild-type mice, raising the possibility that TXNIP deficiency may exert its effects in part by other mechanisms (51). Indeed, in pancreatic β-cells, *Txnip* expression is highly induced by glucose and *Txnip* overexpression promotes β-cell apoptosis (52). This negative effect of TXNIP on cell proliferation is consistent with its initially proposed role as a tumor-suppressor gene and with the reduced expression observed in tumor cell lines and human tumors (53–56). In agreement with the role of TXNIP as a tumor suppressor, TXNIP deficiency in the *Hyplip1* mouse strain was recently shown to be sufficient to initiate hepatocellular carcinoma (57). The discovery that a gene initially characterized as a tumor suppressor is the underlying cause of hyperlipidemia illustrates the independence of positional cloning from preconceived ideas regarding gene function.

The hypl mutation. The KK/San mouse carries a spontaneous mutation, *hypl*, that renders these animals hypolipidemic. KK/San mice are derived from the KK mouse strain, a multigenic model of moderate obesity and type 2 diabetes characterized by hyperlipidemia, hyperinsulinemia, and hyperglycemia (58). In a group of KK mice maintained at a laboratory in Japan, a spontaneous mutation occurred producing animals that are hypolipidemic despite the maintenance of high body weight, insulin, and glucose levels (59). Most pronounced were reductions in triglycerides, but cholesterol and nonesterified fatty acids were also reduced. Positional cloning identified the mutation in the angiopoietin-like 3 (*Angptl3*) gene, resulting in a premature stop codon and a dramatic reduction in mRNA levels (59). ANGPTL3 is a member of the vascular endothelial growth factor family, but it lacks the ability to bind to the Tie2 receptor, suggesting that it is unlikely to have a role in blood vessel formation, as do other family members (60). ANGPTL3 is produced in liver and secreted into the circulation. In contrast to the reduced ANGPTL3 levels observed in KK/San mice, overexpression or administration of ANGPTL3 leads to increased plasma lipids (59). Several studies have provided insight into the mechanisms by which ANGPTL3 levels influence lipid levels. ANGTPL3 was found to inhibit lipoprotein lipase activity, leading to an 85% increase in peripheral disposal of exogenous triglycerides and accounting for the reduced triglyceride levels in the ANGPTL3-deficient KK/San mouse (59, 61). Interestingly, ANGPTL3 appears to have a different role in adipocytes, in which it promotes lipolysis, thus explaining the reduced levels of circulating fatty acids in KK/San mice (62).

Angptl3 genetic variations have also been associated with altered atherosclerosis susceptibility in mice and humans. Transfer of the *hypl* mutation onto the apoE null mouse strain significantly reduces aortic lesion size (63). Additionally, a difference in ANGPTL3 amino acid sequence has been proposed as the causative variation underlying the *Ath8* atherosclerosis susceptibility locus (64). A genetic association between the human *ANGPTL3* gene and atherosclerotic lesions has also been reported (64). Be-

cause both ANGPTL3 and its close relative ANGPTL4 are secreted proteins that inhibit lipoprotein lipase, there is interest in these factors as targets for the treatment of dyslipidemias and the metabolic syndrome (reviewed in 65, 66). In this vein, it is of interest to determine which factors regulate ANGPTL3 levels. It has been shown that *Angptl3* gene expression is induced by the liver X receptor and negatively regulated by thyroid hormone, suggesting a potential mechanism for the hypotriglyceridemic effects of thyroid receptor β agonists (67, 68).

The Diet1 mutation. The widely studied C57BL/6J mouse strain is among the most susceptible to hypercholesterolemia and aortic lesions when fed a diet containing cholesterol, fat, and cholate (69). Therefore, it was highly unexpected when the nearly genetically identical C57BL/6ByJ substrain was found to be resistant to hypercholesterolemia and aortic lesions, despite similar levels of food intake and dietary cholesterol absorption (70). The resistance to increased plasma cholesterol levels in C57BL/6ByJ mice is associated with altered hepatic gene expression leading to the enhanced conversion of cholesterol to bile acids and subsequent excretion in the urine and feces (71, 72). Resistance to hypercholesterolemia cosegregated with high bile acid levels in a cross between C57BL/6ByJ and an unrelated mouse strain, indicating that a single locus controls both traits (71). Via positional cloning, we recently identified the causative mutation in a novel gene, *Diet1*, that has undergone rearrangement in the C57BL/6ByJ genome (our unpublished data). In this case, the information available in genome databases had incorrectly predicted *Diet1* as three separate genes, and no expressed sequence tags were available, calling the expression of this gene into question. Our subsequent analysis revealed that *Diet1* is prominently expressed in intestine and illustrates that some tissues may be underrepresented in the expressed sequence tag databases. Given these issues, it is highly unlikely that *Diet1* would have been identified as a candidate gene for hypercholesterolemia resistance through any methods that rely on assumptions about function. Analysis of *Diet1* function is under way, but it appears to have its primary role in bile acid metabolism in intestine, with secondary effects on the regulation of bile acid synthesis in liver (our unpublished data). It will be interesting to determine how *Diet1* is regulated and whether it provides a potential target for cholesterol lowering through the modulation of bile acid metabolism.

Positional cloning of genes underlying complex traits

In addition to mutations underlying Mendelian disorders, the identification of mutations contributing to complex traits, such as atherosclerosis and diabetes, is now becoming a reality. Genes contributing to complex traits are first mapped through whole genome linkage screens, which often detect several quantitative trait loci (QTLs) that contribute to the phenotype. In the mouse, individual QTLs contributing to a complex phenotype may be isolated on a defined genetic background by the production of congenic strains (73). Congenic strains typically contain a relatively large genome segment carrying the mutation or QTL, which may harbor hundreds of potential candidate genes. Thus, for successful evaluation of candidates, the original congenic region may be subdivided through further breeding to produce subcongenic strains containing a tractable number of candidate genes.

The strategy of genetic mapping, congenic strain production, and positional candidate gene screening has recently met with success to identify genes underlying complex traits associated with lipid metabolism in the mouse. These include *Alox5* (5-lipoxygenase) and *Tnfsf4* (encoding the OX40 ligand) as genes influencing atherosclerosis susceptibility and the *Sorcs1* (a member of the sortillin, SorLA, and SorCS1-3 family) mutation in type 2 diabetes (74–76). The functions of the proteins encoded by these genes in atherosclerosis and diabetes remain to be clarified. Nevertheless, what is known about these genes is consistent with a role in disease. For example, the OX40 ligand is thought to enhance T-cell function, and it has been shown previously that T-lymphocytes promote atherosclerosis (77). Consistent with this, mice carrying targeted mutations in *Tnfsf4* have smaller atherosclerotic lesions, whereas transgenic overexpression resulted in larger lesions (75). Furthermore, two independent human populations exhibited an association between an SNP within the gene and the risk of myocardial infarction (75). In the case of 5-lipoxygenase, it is well established that this enzyme has a role in acute inflammation through the production of leukotrienes from arachidonic acid (reviewed in 78). In atherogenesis, it is possible that 5-lipoxygenase may function in processes ranging from monocyte chemotaxis and proliferation to lesion destabilization and rupture. The role of SorCS1 in type 2 diabetes is unclear at present, but it may be related to *Sorcs1* expression in pancreatic islets. Mice with altered *Sorcs1* expression levels exhibit decreased insulin secretion and disrupted islet morphology (76). One hypothesis is that SorCS1, through the binding of platelet-derived growth factor, has a role in stabilizing the microvasculature in islets, which could influence islet growth, survival, and insulin secretion.

A promising approach for the identification of genes underlying complex traits is known as "genetical genomics," which integrates linkage analysis and gene expression data for progeny in a genetic cross. This technique has only recently become feasible with the advent of affordable whole genome linkage and expression analysis via microarray technology. It has already been embraced by several groups for the analysis of complex traits in mouse and rat and promises to revolutionize the identification of genes and gene networks controlling lipid metabolism and cardiovascular disease (reviewed in 79–82). In this approach, animals in an experimental cross are each typed for a genome-wide panel of genetic markers as well as clinical phenotypes (e.g., lipid levels, fat mass, or atherosclerotic lesion scores) and gene expression levels of thousands of genes using DNA microarrays. The combination of gene expression levels with genetic linkage data allows the mapping of loci that determine transcript abundance, known as "expression QTLs," in the same way

as is traditionally used to map a lipid or disease phenotype. The mapping of these loci immediately reveals whether the control point for determining the levels of a transcript occurs at the gene itself (i.e., acting in *cis*) or is physically located elsewhere in the genome, implying regulation by some other gene(s) (acting in *trans*). If a locus for *cis*-regulated gene expression coincides with that for a clinical trait, the corresponding gene serves as a likely candidate for determining the trait and can be interrogated for DNA sequence variations that might be responsible. Thus, the integration of linkage and gene expression data with clinical traits will provide a powerful tool to prioritize candidate genes mapping to QTLs in complex traits. An example is the recent identification of *Insig2* as a susceptibility gene for plasma cholesterol levels (83). Interestingly, this approach also detected *Alox5* as a gene that influences susceptibility to obesity and bone density (84). This approach is being applied to identify gene variations that influence complex traits such as obesity, plasma lipid levels, atherosclerosis, and diabetes and will no doubt lead to candidate genes and increased understanding of the gene networks that determine these complex traits (85–88).

In humans, numerous potential susceptibility genes for traits such as hyperlipidemia, obesity, and coronary artery disease have been reported through typing genetic polymorphisms in case-control studies. There are, however, caveats in interpreting results obtained using this approach, as false-positives are common as a result of factors such as population admixture or selection bias. Here, we will consider only susceptibility genes for complex lipid-related disorders that have been identified by whole genome linkage and positional cloning. A striking example is the identification of a gene contributing to FCHL, which is characterized by increased serum cholesterol and triglyceride levels and is the most common familial lipid disorder predisposing to premature heart disease (89). Using whole genome scans in a Finnish population, a QTL for FCHL was identified on chromosome 1q21-23 (90). Coincidentally, this region of the human genome corresponds to a region of mouse chromosome 3 that harbors the *Txnip* mutation causing combined hyperlipidemia, which raised the possibility that the same gene was responsible for this disorder in mouse and human (48, 90). However, subsequent high-resolution mapping and positional candidate gene analysis revealed that the human FCHL susceptibility gene is *USF1*. *USF1* encodes upstream stimulatory factor-1, a transcription factor that regulates several genes involved in lipid and glucose homeostasis, making it an attractive fit for the observed phenotype of FCHL (91). Although the *USF1* gene variants identified (two synonymous SNPs) were strongly associated with FCHL in Finnish subjects, it is unclear whether these are the two causative variants. However, recent confirmation of the associations between *USF1* and FCHL in other populations provides strong evidence that *USF1* is a susceptibility gene for FCHL (92, 93).

A positional cloning/candidate approach was also used to identify an obesity susceptibility gene, *SLC6A14* (94). This gene encodes an amino acid transporter that is thought to regulate tryptophan availability for serotonin synthesis and hence may influence appetite control. Susceptibility genes for coronary artery disease were recently also identified using positional candidate or genome-wide association studies. These were reviewed recently and include the MEF2A transcription factor, *LTA* (lymphotoxin-α), *ALOX5AP* (5-lipoxygenase-activating protein), and *PDE4D* (phosphodiesterase 4D) (reviewed in 95). These recent successes in the identification of genes underlying not only monogenic but also multigenic lipid disorders indicate that optimism is warranted for an eventual understanding of key genetic factors in common human lipid disorders.

FUNCTIONAL GENOMICS FOR THE ELUCIDATION OF GENE FUNCTION IN LIPID METABOLISM

In the postgenomic era, attention has turned to the annotation and characterization of gene function. One way to understand the function of a gene is to overexpress it or disrupt it in the mouse or other experimental model organism. Since the generation of the first transgenic mouse in 1981, there has been a rapid evolution of technology, and it is now routine to produce mouse models with tissue- and temporal-specific knockouts or knockins at a specific gene of interest (reviewed in 96, 97). Nevertheless, these gene-specific approaches are generally laborious and time-consuming and require the generation of unique gene constructs for each mutation produced. Thus, there is now a significant effort to apply methods that allow high-throughput gene mutation, with the eventual goal of generating mutations in every gene in the mouse genome (98, 99). Whole genome mutagenesis approaches have already been applied to model organisms such as *S. cerevisiae* and *Caenorhabditis elegans* via gene deletion and RNA interference (RNAi), respectively (100, 101). In *C. elegans*, one RNAi screen aimed at the identification of fat regulatory genes revealed potential lipid metabolism candidate genes for further analysis in mammals (102). Although RNAi, which selectively targets mRNA for degradation, has been applied to knock down gene expression in cultured mammalian cells, technical problems have prevented its widespread use in the mouse, although new refinements are being developed that may overcome these problems (103, 104). To date, the most widespread approaches for genome-wide mutagenesis in the mouse are chemical mutagenesis and gene-trap mutagenesis.

Ethylnitrosourea mutagenesis

Although not a new technique, chemical mutagenesis of mice recently experienced a renaissance as a method for functional genome analysis on a large scale (105–108). The chemical of choice is ethylnitrosourea (ENU), an alkylating agent that causes mutation at a frequency of 300–1,200 times greater than the spontaneous mutation rate (109). ENU is administered to male mice in a series of

injections, which mutagenizes spermatogonial stem cells. The mutagenized males are mated to wild-type females, and the resulting offspring are screened directly to detect dominant mutations or mated for two additional generations to detect recessive mutations. Mutations induced by ENU are typically point mutations, which may lead to missense, nonsense, and splice-site mutations. Thus, unlike gene knockouts, which result in complete ablation of gene function, ENU treatment results in a variety of mutation types, including loss-of-function, hypomorphic, which reduce but do not abolish gene function, and antimorphic, which antagonize gene function. The analysis of an alleleic series consisting of several distinct mutations within a gene can provide valuable information about gene and protein function that is not necessarily revealed by the analysis of knockout mutations, which represents an advantage of the ENU mutagenesis approach. A disadvantage of the approach is the necessity to produce and perform phenotypic assays on hundreds or thousands of live mice to reveal phenotypes of interest. Furthermore, once a phenotype is detected, it must be test bred to demonstrate that it is indeed heritable, followed by genetic mapping and positional cloning to identify the mutant gene. Positional cloning of mutant loci has been a bottleneck in this process, but the recent development of techniques for "gene-driven" ENU screens may alleviate this problem (discussed briefly below).

Among the clinical phenotypes frequently assessed in ENU-mutagenized mice are traits that are relevant to lipid metabolism, obesity, and diabetes. These include blood glucose and lipid levels, heart rate, blood pressure, body weight, and body composition. ENU mutagenesis programs based in the United States, Germany, and England have all identified ENU mutants with alterations in several of these traits (110–113) (Table 1). For example, the Munich ENU Mouse Mutagenesis Project has screened for increased plasma cholesterol levels in 15,000 mutagenized mice (113). More than 100 mutant mice were detected with hypercholesterolemia, and from these, nine distinct hypercholesterolemic mouse lines have been developed. Likewise, the UK ENU Mutagenesis Program from Harwell reported mutants with low plasma total cholesterol, low HDL cholesterol, and increased triglycerides (112). Several mutagenesis projects make animals with various phenotypes available to the public for further characterization. The Mouse Heart, Lung, Blood, and Sleep Disorders Center at the Jackson Laboratory has identified mice with reduced or increased total cholesterol, HDL cholesterol, triglyceride, and glucose levels, hypertension, abnormal heart rate, or electrocardiogram measurements, anemia, coagulation abnormalities, and obesity (see Table 1 for website). Some of the mutations have already been mapped to specific chromosomal regions, including mutations causing high total cholesterol, low HDL cholesterol, and hyperglycemia. Mutants characterized in this screen are made available to interested investigators for a nominal fee through a web-based interface for a limited time period.

Upon identification of an interesting phenotype in a mutagenized mouse, the task to identify the underlying mutation by positional cloning is similar to that described above. As most of these mutagenesis projects are relatively recent, to date there have been few lipid metabolism genes isolated from ENU-mutagenized mice. One interesting example is the identification of a mutation in *Cd36*, previously shown to be a physiologic receptor for oxidized LDL, in an ENU-mutagenized mouse with an immunodeficiency phenotype (114). This finding revealed a role for CD36 as a sensor of microbial diacylglycerides as a requirement for mounting an immune response against numerous bacteria, fungi, and protozoa.

In the future, alternatives to the standard phenotypic screening followed by positional cloning may increase the throughput and utility of ENU mutagenesis. With the availability of mouse genome sequence and high-throughput mutation screening via PCR coupled with high-temperature gradient capillary electrophoresis, it is now possible to screen cryopreserved sperm from ENU-mutagenized mice for mutations in any gene of interest (115–117). Sperm containing mutations of interest are then used to generate live mice by in vitro fertilization. In the future, this approach may enable high-throughput, gene-driven, rather than phenotype-driven, ENU screens of an archive containing thousands of sperm samples with mutations in a large proportion of the genes.

Gene-trap mutagenesis

Gene-trap mutagenesis involves the mutation of mouse embryonic stem cells by random insertion of trapping vectors into genes or gene regulatory sequences, leading to the impairment or ablation of gene function (reviewed in 118). The site of insertion can be determined through PCR and sequencing using the vector as a tag, and the results can be collated into databases that can be searched to locate mutations in specific genes. This is a true gene-driven approach, as live mice are not generated until a mutation of interest is identified in the database of embryonic stem cells. Gene-trap vectors typically contain a reporter gene (e.g., β-galactosidase) and a selectable marker gene (e.g., neomycin^R) or a combination of both functions (e.g., β-geo). Integration of a gene-trap vector into a gene intron leads to the production of a fusion mRNA transcript containing upstream exons from the endogenous gene joined to the reporter gene sequences by virtue of a splice acceptor site incorporated into the vector (**Fig. 4A**). The translation of the fusion transcript results in the expression of the reporter gene driven by the endogenous gene regulatory sequences in a pattern that mimics the expression of the endogenous gene. Staining for β-galactosidase reporter expression provides a tool to learn about the expression of the gene under study. Mutations arising from the insertion of gene-trap vectors have a high likelihood of causing loss-of-function or hypomorphic mutations, depending on the position of insertion within the gene.

A key question arises concerning the extent to which gene-trapping can be used to produce genome-wide mutations. Although the insertion of gene-trap vectors is
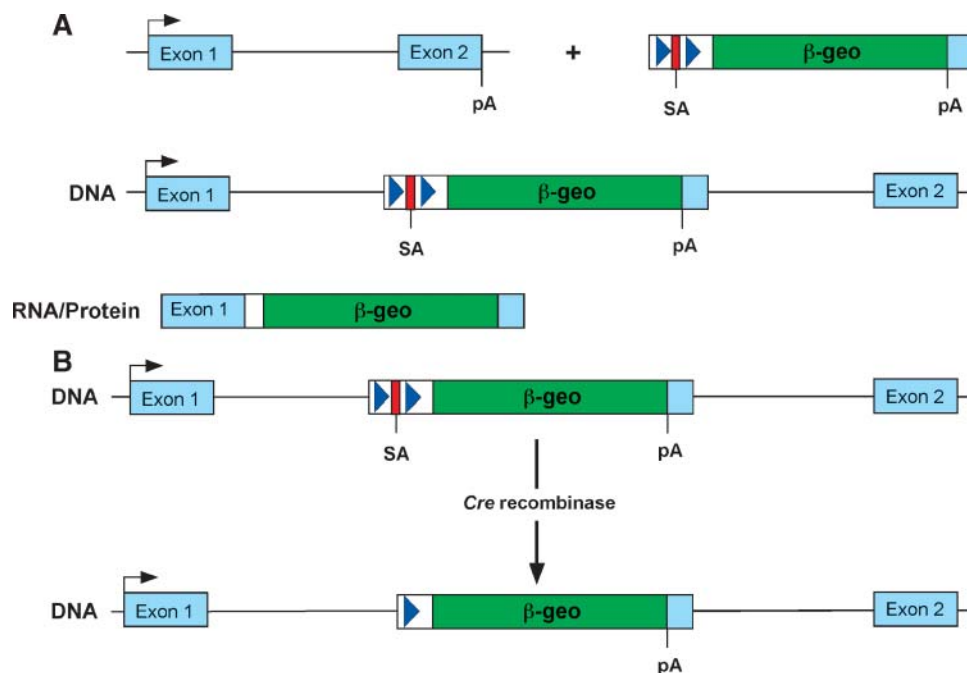
Fig. 4. Standard and conditional gene-trapping approaches. A: A typical gene-trap vector consists of a splice acceptor site (SA) upstream of a promoterless *β-geo* gene, which is a fusion of β-galactosidase reporter and neomycin resistance genes. The gene-trap vector is introduced into mouse embryonic stem cells and inserts randomly into the genome. If insertion occurs within an intron, a fusion transcript consisting of the up-stream exons of the endogenous gene fused to *β-geo* will be produced through utilization of the gene-trap splice acceptor site. Because the *β-geo* gene is promoterless, expression occurs only when inserted within a gene. The insertion within the gene serves to disrupt normal gene function and provides a reporter that can be exploited to study tissue and temporal patterns of gene expression. B: The incorporation of *lox*P sites (triangles) flanking the splice acceptor site in the gene-trap vector allows conditional splice acceptor de-letion from existing gene-trap cell lines by treatment with *Cre* recombinase. Deletion of the splice acceptor site may allow the transcription of some wild-type mRNA through the use of the endogenous splice accep-tor site in exon 2. Wild-type splicing efficiency is typically reduced as a result of the large insertion and therefore may result in the production of a hypomorphic allele. Hypomorphic alleles are particularly useful for the rescue of embryonic lethal mutations. pA, polyadenylation site.

theoretically random, in practice, some genes have sus-tained several independent hits but others have not yet been trapped. Some factors that clearly influence trapping efficiency are gene size (with larger genes providing a better target) and expression levels in embryonic stem cells (with a threshold level being necessary for the expression of the selectable marker). Nevertheless, a substantial proportion of mouse genes have sustained gene-trap insertions. Through the efforts of several international public gene-trapping projects, which have recently been combined into a single database (Table 1), >45,000 embryonic stem cell lines rep-resenting >40% of the mouse genes have been deposited (119, 120). A commercial gene-trap effort by Lexicon Genetics has independently produced 200,000 embryonic stem cell lines, which cover 60% of mouse genes (119). These represent substantial resources, especially compared with the 3,600 gene knockout and knockdown mutations produced by standard gene-targeting techniques over more than two decades. As gene-trap resources have only been available for a few years, generation of the correspond-ing mouse models is still in the early stages. Below, we provide examples of gene-trap mutations that influence lipid metabolism.

ATP citrate lyase mutation. A gene-trap mutation was used to examine the physiological function of ATP citrate lyase, one of two cytosolic enzymes that synthesize acetyl-CoA, a precursor of both triglycerides and cholesterol. An embry-onic stem cell line containing a gene-trap insertion in the ATP citrate lyase gene (*Acly*) was identified in the gene-trap database and used to generate mice (121). The homozygous *Acly* mutation was embryonic lethal, but as a result of the insertion of the β-galactosidase reporter gene, it was possi-ble to histologically characterize in detail the in vivo expres-sion pattern of ATP citrate lyase in heterozygotes carrying one copy of the gene-trap allele. These studies revealed ubiquitous expression of ATP citrate lyase, but particularly high levels in tissues undergoing high rates of de novo lipo-genesis, including liver of mice fed a high-carbohydrate diet, and in developing brain. Interestingly, although *Acly* ex-pression was much lower in adult compared with developing brain, it remained particularly pronounced in cholinergic neurons, because of a requirement for ATP citrate lyase in acetylcholine synthesis.

Phosphatidylserine decarboxylase mutation. The role of the phosphatidylserine decarboxylase gene (*Psd*) in phospho-

lipid biosynthesis was recently elucidated in vivo through the characterization of a gene-trap mutant (122). Phosphatidylethanolamine synthesis occurs in the endoplasmic reticulum and in the mitochondria and is controlled by *Psid* specifically in mitochondria. Mice homozygous for the disruption of *Psid* died during development, revealing that phosphatidylethanolamine synthesis in the endoplasmic reticulum cannot substitute for the mitochondrial pathway. Furthermore, fibroblasts isolated from *Psid*$^{-/-}$ mice revealed that the reduction in phosphatidylethanolamine levels resulted in abnormally shaped and fragmented mitochondria. Heterozygous mutant mice exhibited an up-regulation of a key enzyme in the endoplasmic reticulum pathway for phosphatidylethanolamine synthesis and, hence, normal tissue phospholipid content.

Phosphatidylserine synthase 2 mutation. An additional gene with a role in phospholipid synthesis, phosphatidylserine synthase 2 (*Pss2*), has been investigated in gene-trap mice. *Pss1* and *Pss2* encode two distinct serine-exchange enzymes involved in phosphatidylserine synthesis. Mice homozygous for gene-trap-induced PSS2 deficiency had normal development and tissue phospholipid content, although ~10% of the animals exhibited testicular atrophy (123). Analysis of the reporter gene revealed high *Pss2* expression in testis as well as brown adipose tissue, neurons, and myometrium. Characterization of serine-exchange activity in tissues from *Pss2*$^{-/-}$ mice revealed a compensatory mechanism to maintain normal phosphatidylserine content involving both reduced phosphatidylserine degradation and increased PSS1 activity (124).

Acylglycerolphosphate acyltransferase 6 mutation. One approach to use the gene-trap database is to search for insertions in genes of interest, as illustrated in the examples above. Another approach is to examine the sequences of genes of unknown function that have sustained gene-trap insertions. Using the latter approach, we identified an insertion in a gene (*Agpat6*) with sequence similarity to a family of 1-acylglycerol-3-phosphate *O*-acyltransferase (AGPAT) enzymes. AGPATs are involved in triacylglycerol synthesis, catalyzing the transfer of acyl groups to lysophosphatidic acid. The most well-characterized member of the family, AGPAT2, has been shown to have a regulatory role in adipocyte differentiation, and null mutations result in congenital lipodystrophy (125, 126). Mice carrying a gene-trap mutation in *Agpat6* revealed important roles for AGPAT6 in triglyceride accumulation in white and brown adipose tissue and in milk (127, 128). Using the β-galactosidase reporter gene, we determined that *Agpat6* is expressed in several tissues, with extremely high expression in brown adipose tissue and substantial levels in white adipose tissue, testis, and specific regions of the brain (**Fig. 5A, B**). *Agpat6*$^{-/-}$ mice were viable, but offspring of AGPAT6-deficient mothers died within a few days unless transferred to a foster mother (127). This prompted an analysis of *Agpat6* expression in mammary gland, where β-galactosidase staining was detected in epithelial cells. Mammary gland from lactating *Agpat6*$^{-/-}$ mice had underdeveloped alveoli and ducts and reduced triglyceride accumulation. AGPAT6-deficient mice also had 25% lower body weight attributable to reduced triglyceride storage in several adipose tissue depots (128).
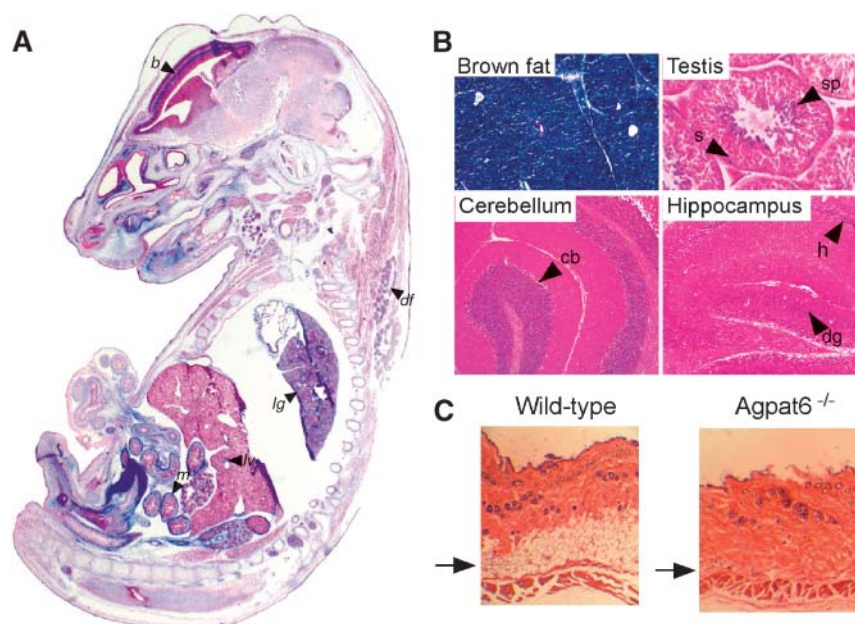
Fig. 5. Characterization of *Agpat6* gene-trap mice. A: Detection of *Agpat6* expression via β-galactosidase staining in a longitudinal section of a 15 day *Agpat6*$^{+/-}$ mouse embryo. Arrowheads indicate expression in brain (b), dorsal brown fat (df), lung (lg), liver (lv), and mesenchyme (m) of the gut. B: *Agpat6* expression in adult mouse tissues. Arrowheads indicate Sertoli cells (s), spermatids (sp), cerebellar lobule (cb), hippocampus (h), and dentate gyrus (dg). C: Cross-section of skin from wild-type and *Agpat6*$^{-/-}$ mice. The prominent subdermal fat layer indicated by the arrow in wild-type animals is absent from the mutant. Images are from references 127 and 128, used with permission.

Adipose tissue in the subdermal region was reduced dramatically (Fig. 5C). $Agpat6^{-/-}$ mice exhibited resistance to both diet-induced and genetically induced obesity, which was associated with reduced triglyceride content in adipose tissue and increased energy expenditure. These studies revealed that $Agpat6$ plays a unique physiological role that cannot be substituted for by other members of this gene family.

Future of gene trapping. As with traditional gene knockouts, many gene-trap mutants are lethal in the homozygous state. The analysis of reporter gene expression in heterozygotes has allowed the study of in vivo gene expression, but information about physiological gene function is still limited in these cases. One potential way to overcome this limitation is the use of conditional gene-trap vectors. Because insertion of a gene-trap vector disrupts gene function by introducing an inappropriate splice acceptor, the conditional elimination of the splice acceptor may allow the partial restoration of gene function at a level that permits normal development yet still causes a phenotype attributable to reduced expression (i.e., a hypomorphic allele). This can be accomplished by engineering the gene-trap vector to contain *lox*P sites flanking the splice acceptor such that treating the embryonic stem cells with Cre recombinase, or crossing the mutant mice with Cre-expressing animals, will remove the splice acceptor (Fig. 4B). Because the vector remains inserted into the gene, cryptic splice events may occur, leading to reduced efficiency of normal splicing and low levels of wild-type transcripts. Documentation of this strategy has not yet been presented in the literature, but preliminary results in our laboratory suggest that expression of Cre recombinase can rescue an otherwise embryonic lethal mutation in a portion of the animals so that the role of the affected gene in adult mice can be studied (our unpublished results).

Another promising addition to the gene-trapping repertoire is the technique of "targeted trapping." As mentioned above, gene-trap insertion is not fully random, and the technique is therefore likely to reach a plateau without achieving gene-trap insertions in all protein-coding genes (119). To produce mutations in the remaining genes, a high-efficiency method of targeted mutation would be valuable. The targeted trapping strategy, which involves homologous recombination with a promoterless gene-trapping vector, produces targeted mutations in specific genes with >50% efficiency, a substantial improvement over the 1–5% efficiency of standard homologous recombination (129). The primary limitation of this technique is that the target gene must be expressed to some degree in embryonic stem cells so that the selectable marker present in the promoterless construct will be expressed. Recent studies have defined a threshold expression level corresponding to 1% of that of the transferrin receptor (129). Targeted trapping may be applicable to perhaps half of the genes remaining after the saturation of standard gene-trapping techniques, and other methods may be required for the remaining genes. It is possible that the use of different families of trapping vectors coupled with various delivery methods (e.g., electroporation vs. retroviral infection) may also allow gene-trap insertions in additional genes. Additional techniques for the high-throughput generation of targeting vectors, such as bacterial artificial chromosome recombineering, may also increase the efficiency of targeted trapping in the future (130, 131).

PROSPECTS

The study of naturally occurring mutations in genes has proved to be one of the most fruitful approaches to understanding both disease pathogenesis and normal protein function. A classic example is the delineation of LDL receptor mutations in familial hypercholesterolemia, which led to an understanding not only of that disease but also of many aspects of cholesterol homeostasis and receptor-mediated endocytosis. In the postgenomic era, tools are available to allow more rapid identification of gene mutations and to create models with any mutation of interest, so that the rate-limiting step will move beyond the technical issues of gene discovery to studies of the biology and physiology. Many lipid metabolism genes remain to be identified, as do genes underlying complex disorders of lipid metabolism. As illustrated by some of the examples discussed here, the expanding repertoire of genomic resources has facilitated gene identification, and the pace will likely accelerate as current techniques evolve to take full advantage of these resources. The mouse model will continue to be central to these efforts, as a full understanding of lipid metabolism requires the study of interactions between cells, tissues, organs, and environment that can only be achieved in a whole organism. Sometime in the not too distant future, it may be possible to find desired mutations in nearly any gene of interest in a catalog of mouse embryonic stem cells. The tools are available to then transfer findings in the mouse to humans and, potentially, to the clinic.▪

REFERENCES

1. International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature.* **409:** 860–921.
2. International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature.* **431:** 931–945.
3. International Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature.* **420:** 520–562.
4. Hardison, R. C. 2003. Comparative genomics. *PLoS Biol.* **1:** 156–160.
5. Nobrega, M. A., and L. A. Pennacchio. 2004. Comparative genomic analysis as a tool for biological discovery. *J. Physiol.* **554:** 31–39.

6. Pennacchio, L. A., and E. M. Rubin. 2003. Comparative genomic tools and databases: providing insights into the human genome. *J. Clin. Invest.* **111:** 1099–1106.

7. Tautz, D. 2000. Evolution of transcriptional regulation. *Curr. Opin. Genet. Dev.* **10:** 575–579.

8. Pennacchio, L. A., M. Olivier, J. A. Hubacek, J. C. Cohen, D. R. Cox, J. C. Fruchart, R. M. Krauss, and E. M. Rubin. 2001. An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science.* **294:** 169–173.

9. Talmud, P. J., E. Hawe, S. Martin, M. Olivier, G. J. Miller, E. M. Rubin, L. A. Pennacchio, and S. E. Humphries. 2002. Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. *Hum. Mol. Genet.* **11:** 3039–3046.

10. Vergnes, L., J. Phan, A. Stolz, and K. Reue. 2003. A cluster of eight hydroxysteroid dehydrogenase genes belonging to the aldo-keto reductase supergene family on mouse chromosome 13. *J. Lipid Res.* **44:** 503–511.

11. Bauman, D. R., S. Steckelbroeck, and T. M. Penning. 2004. The roles of aldo-keto reductases in steroid hormone action. *Drug News Perspect.* **17:** 563–578.

12. Zheng, P., L. A. Pennacchio, W. Le Goff, E. M. Rubin, and J. D. Smith. 2004. Identification of a novel enhancer of brain expression near the apoE gene cluster by comparative genomics. *Biochim. Biophys. Acta.* **1676:** 41–50.

13. Lawn, R. M., K. Schwartz, and L. Patthy. 1997. Convergent evolution of apolipoprotein(a) in primates and hedgehog. *Proc. Natl. Acad. Sci. USA.* **94:** 11992–11997.

14. Boffelli, D., J. McAuliffe, D. Ovcharenko, K. D. Lewis, I. Ovcharenko, L. Pachter, and E. M. Rubin. 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science.* **299:** 1391–1394.

15. Ovcharenko, I., D. Boffelli, and G. G. Loots. 2004. eShadow: a tool for comparing closely related sequences. *Genome Res.* **14:** 1191–1198.

16. Loots, G. G., and I. Ovcharenko. 2005. Dcode.org anthology of comparative genomic tools. *Nucleic Acids Res.* **33:** W56–W64.

17. ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) project. *Science.* **306:** 636–640.

18. The Huntington's Disease Collaborative Research Group. 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell.* **72:** 971–983.

19. Crawford, D. C., and D. A. Nickerson. 2005. Definition and clinical importance of haplotypes. *Annu. Rev. Med.* **56:** 303–320.

20. DiPetrillo, K., X. Wang, I. M. Stylianou, and B. Paigen. 2005. Bioinformatics toolbox for narrowing rodent quantitative trait loci. *Trends Genet.* **21:** 683–692.

21. Cuppen, E. 2005. Haplotype-based genetics in mice and rats. *Trends Genet.* **21:** 318–322.

22. International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature.* **437:** 1299–1320.

23. Thorisson, G. A., A. V. Smith, L. Krishnan, and L. D. Stein. 2005. The International HapMap Project Web site. *Genome Res.* **15:** 1592–1593.

24. Wade, C. M., E. J. Kulbokas, 3rd, A. W. Kirby, M. C. Zody, J. C. Mullikin, E. S. Lander, K. Lindblad-Toh, and M. J. Daly. 2002. The mosaic structure of variation in the laboratory mouse genome. *Nature.* **420:** 574–578.

25. Herman, G. E. 1998. Physical mapping of the mouse genome. *Methods.* **14:** 135–151.

26. Boehm, T. 1998. Positional cloning and gene identification. *Methods.* **14:** 152–158.

27. The Gene Ontology Consortium. 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* **25:** 25–29.

28. Su, A. I., T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, et al. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA.* **101:** 6062–6067.

29. Behm-Ansmant, I., and E. Izaurralde. 2006. Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. *Genes Dev.* **20:** 391–398.

30. Aitman, T. J., T. Gotoda, A. L. Evans, H. Imrie, K. E. Heath, P. M. Trembling, H. Truman, C. A. Wallace, A. Rahman, C. Dore, et al. 1997. Quantitative trait loci for cellular defects in glucose and fatty acid metabolism in hypertensive rats. *Nat. Genet.* **16:** 197–201.

31. Berge, K. E., H. Tian, G. A. Graf, L. Yu, N. V. Grishin, J. Schultz, P. Kwiterovich, B. Shan, R. Barnes, and H. H. Hobbs. 2000. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science.* **290:** 1771–1775.

32. Reue, K., and M. H. Doolittle. 1996. Naturally occurring mutations in mice affecting lipid transport and metabolism. *J. Lipid Res.* **37:** 1387–1405.

33. Ho, S. Y., K. Lorent, M. Pack, and S. A. Farber. 2006. Zebrafish fat-free is required for intestinal lipid absorption and Golgi apparatus structure. *Cell Metab.* **3:** 289–300.

34. Mori, M., G. Li, I. Abe, J. Nakayama, Z. Guo, J. Sawashita, T. Ugawa, S. Nishizono, T. Serikawa, K. Higuchi, et al. 2006. Lanosterol synthase mutations cause cholesterol deficiency-associated cataracts in the Shumiya cataract rat. *J. Clin. Invest.* **116:** 395–404.

35. Brooks-Wilson, A., M. Marcil, S. M. Clee, L. H. Zhang, K. Roomp, M. van Dam, L. Yu, C. Brewer, J. A. Collins, H. O. Molhuizen, et al. 1999. Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat. Genet.* **22:** 336–345.

36. Bodzioch, M., E. Orso, J. Klucken, T. Langmann, A. Bottcher, W. Diederich, W. Drobnik, S. Barlage, C. Buchler, M. Porsch-Ozcurumez, et al. 1999. The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat. Genet.* **22:** 347–351.

37. Rust, S., M. Rosier, H. Funke, J. Real, Z. Amoura, J. C. Piette, J. F. Deleuze, H. B. Brewer, N. Duverger, P. Denefle, et al. 1999. Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat. Genet.* **22:** 352–355.

38. Garcia, C. K., K. Wilund, M. Arca, G. Zuliani, R. Fellin, M. Maioli, S. Calandra, S. Bertolini, F. Cossu, N. Grishin, et al. 2001. Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science.* **292:** 1394–1398.

39. Peterfy, M., J. Phan, P. Xu, and K. Reue. 2001. Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nat. Genet.* **27:** 121–124.

40. Reue, K., and J. Donkor. 2006. Lipin: a determinant of adiposity, insulin sensitivity and energy balance. *Future Lipidol.* **1:** 91–101.

41. Phan, J., M. Peterfy, and K. Reue. 2004. Lipin expression preceding peroxisome proliferator-activated receptor-gamma is critical for adipogenesis in vivo and in vitro. *J. Biol. Chem.* **279:** 29558–29564.

42. Peterfy, M., J. Phan, and K. Reue. 2005. Alternatively spliced lipin isoforms exhibit distinct expression pattern, subcellular localization, and role in adipogenesis. *J. Biol. Chem.* **280:** 32883–32889.

43. Phan, J., and K. Reue. 2005. Lipin, a lipodystrophy and obesity gene. *Cell Metab.* **1:** 73–83.

44. Suviolahti, E., K. Reue, R. M. Cantor, J. Phan, M. Gentile, J. Naukkarinen, A. Soro-Paavonen, L. Oksanen, J. Kaprio, A. Rissanen, et al. 2006. Cross-species analyses implicate Lipin 1 involvement in human glucose metabolism. *Hum. Mol. Genet.* **15:** 377–386.

45. Santos-Rosa, H., J. Leung, N. Grimsey, S. Peak-Chew, and S. Siniossoglou. 2005. The yeast lipin Smp2 couples phospholipid biosynthesis to nuclear membrane growth. *EMBO J.* **24:** 1931–1941.

46. Han, G. S., W. I. Wu, and G. M. Carman. 2006. The *Saccharomyces cerevisiae* Lipin homolog is a Mg^{2+}-dependent phosphatidate phosphatase enzyme. *J. Biol. Chem.* **281:** 9210–9218.

47. Bodnar, J. S., A. Chatterjee, L. W. Castellani, D. A. Ross, J. Ohmen, J. Cavalcoli, C. Wu, K. M. Dains, J. Catanese, M. Chu, et al. 2002. Positional cloning of the combined hyperlipidemia gene Hyplip1. *Nat. Genet.* **30:** 110–116.

48. Castellani, L. W., A. Weinreb, J. Bodnar, A. M. Goto, M. Doolittle, M. Mehrabian, P. Demant, and A. J. Lusis. 1998. Mapping a gene for combined hyperlipidaemia in a mutant mouse strain. *Nat. Genet.* **18:** 374–377.

49. Hui, T. Y., S. S. Sheth, J. M. Diffley, D. W. Potter, A. J. Lusis, A. D. Attie, and R. A. Davis. 2004. Mice lacking thioredoxin-interacting protein provide evidence linking cellular redox state to appropriate response to nutritional signals. *J. Biol. Chem.* **279:** 24387–24393.

50. Donnelly, K. L., M. R. Margosian, S. S. Sheth, A. J. Lusis, and E. J. Parks. 2004. Increased lipogenesis and fatty acid reesterification contribute to hepatic triacylglycerol stores in hyperlipidemic $Txnip^{-/-}$ mice. *J. Nutr.* **134:** 1475–1480.

51. Sheth, S. S., L. W. Castellani, S. Chari, C. Wagg, C. K. Thipphavong, J. S. Bodnar, P. Tontonoz, A. D. Attie, G. D. Lopaschuk, and A. J. Lusis. 2005. Thioredoxin-interacting protein deficiency disrupts the fasting-feeding metabolic transition. *J. Lipid Res.* **46:** 123–134.

52. Minn, A. H., C. Hafele, and A. Shalev. 2005. Thioredoxin-interacting protein is stimulated by glucose through a carbohydrate response element and induces beta-cell apoptosis. *Endocrinology.* **146:** 2397–2405.

53. Chen, K. S., and H. F. DeLuca. 1994. Isolation and characterization of a novel cDNA from HL-60 cells treated with 1,25-dihydroxyvitamin D-3. *Biochim. Biophys. Acta.* **1219:** 26–32.

54. Ikarashi, M., Y. Takahashi, Y. Ishii, T. Nagata, S. Asai, and K. Ishikawa. 2002. Vitamin D3 up-regulated protein 1 (VDUP1) expression in gastrointestinal cancer and its relation to stage of disease. *Anticancer Res.* **22:** 4045–4048.

55. Nishinaka, Y., A. Nishiyama, H. Masutani, S. Oka, K. M. Ahsan, Y. Nakayama, Y. Ishii, H. Nakamura, M. Maeda, and J. Yodoi. 2004. Loss of thioredoxin-binding protein-2/vitamin D3 up-regulated protein 1 in human T-cell leukemia virus type I-dependent T-cell transformation: implications for adult T-cell leukemia leukemogenesis. *Cancer Res.* **64:** 1287–1292.

56. Tome, M. E., D. B. Johnson, L. M. Rimsza, R. A. Roberts, T. M. Grogan, T. P. Miller, L. W. Oberley, and M. M. Briehl. 2005. A redox signature score identifies diffuse large B-cell lymphoma patients with a poor prognosis. *Blood.* **106:** 3594–3601.

57. Sheth, S. S., J. S. Bodnar, A. Ghazalpour, C. K. Thipphavong, S. Tsutsumi, A. D. Tward, P. Demant, T. Kodama, H. Aburatani, and A. J. Lusis. 2006. Hepatocellular carcinoma in Txnip-deficient mice. *Oncogene.* **25:** 3528–3536.

58. Nakamura, M., and K. Yamada. 1967. Studies on a diabetic (KK) strain of the mouse. *Diabetologia.* **3:** 212–221.

59. Koishi, R., Y. Ando, M. Ono, M. Shimamura, H. Yasumo, T. Fujiwara, H. Horikoshi, and H. Furukawa. 2002. Angptl3 regulates lipid metabolism in mice. *Nat. Genet.* **30:** 151–157.

60. Camenisch, G., M. T. Pisabarro, D. Sherman, J. Kowalski, M. Nagel, P. Hass, M. H. Xie, A. Gurney, S. Bodary, X. H. Liang, et al. 2002. ANGPTL3 stimulates endothelial cell adhesion and migration via integrin alpha vbeta 3 and induces blood vessel formation in vivo. *J. Biol. Chem.* **277:** 17281–17290.

61. Kaplan, R., T. Zhang, M. Hernandez, F. X. Gan, S. D. Wright, M. G. Waters, and T. Q. Cai. 2003. Regulation of the angiopoietin-like protein 3 gene by LXR. *J. Lipid Res.* **44:** 136–143.

62. Shimamura, M., M. Matsuda, S. Kobayashi, Y. Ando, M. Ono, R. Koishi, H. Furukawa, M. Makishima, and I. Shimomura. 2003. Angiopoietin-like protein 3, a hepatic secretory factor, activates lipolysis in adipocytes. *Biochem. Biophys. Res. Commun.* **301:** 604–609.

63. Ando, Y., T. Shimizugawa, S. Takeshita, M. Ono, M. Shimamura, R. Koishi, and H. Furukawa. 2003. A decreased expression of angiopoietin-like 3 is protective against atherosclerosis in apoE-deficient mice. *J. Lipid Res.* **44:** 1216–1223.

64. Korstanje, R., P. Eriksson, A. Samnegard, P. G. Olsson, K. Forsman-Semb, S. Sen, G. A. Churchill, J. Rollins, S. Harris, A. Hamsten, et al. 2004. Locating Ath8, a locus for murine atherosclerosis susceptibility and testing several of its candidate genes in mice and humans. *Atherosclerosis.* **177:** 443–450.

65. Oike, Y., M. Akao, Y. Kubota, and T. Suda. 2005. Angiopoietin-like proteins: potential new targets for metabolic syndrome therapy. *Trends Mol. Med.* **11:** 473–479.

66. Li, C. 2006. Genetics and regulation of angiopoietin-like proteins 3 and 4. *Curr. Opin. Lipidol.* **17:** 152–156.

67. Inaba, T., M. Matsuda, M. Shimamura, N. Takei, N. Terasaka, Y. Ando, H. Yasumo, R. Koishi, M. Makishima, and I. Shimomura. 2003. Angiopoietin-like protein 3 mediates hypertriglyceridemia induced by the liver X receptor. *J. Biol. Chem.* **278:** 21344–21351.

68. Fugier, C., J. J. Tousaint, X. Prieur, M. Plateroti, J. Samarut, and P. Delerive. 2006. The lipoprotein lipase inhibitor ANGPTL3 is negatively regulated by thyroid hormone. *J. Biol. Chem.* **281:** 11553–11559.

69. Paigen, B., A. Morrow, C. Brandon, D. Mitchell, and P. Holmes. 1985. Variation in susceptibility to atherosclerosis among inbred strains of mice. *Atherosclerosis.* **57:** 65–73.

70. Mouzeyan, A., J. Choi, H. Allayee, X. Wang, J. Sinsheimer, J. Phan, L. W. Castellani, K. Reue, A. J. Lusis, and R. C. Davis. 2000. A locus conferring resistance to diet-induced hypercholesterolemia and atherosclerosis on mouse chromosome 2. *J. Lipid Res.* **41:** 573–582.

71. Phan, J., T. Pesaran, R. C. Davis, and K. Reue. 2002. The Diet1 locus confers protection against hypercholesterolemia through enhanced bile acid metabolism. *J. Biol. Chem.* **277:** 469–477.

72. Vergnes, L., J. Phan, M. Strauss, S. Tafuri, and K. Reue. 2003. Cholesterol and cholate components of an atherogenic diet induce distinct stages of hepatic inflammatory gene expression. *J. Biol. Chem.* **278:** 42774–42784.

73. Silver, L. M. 1995. Mouse Genetics. Concepts and Applications. Oxford University Press, New York.

74. Mehrabian, M., H. Allayee, J. Wong, W. Shi, X. P. Wang, Z. Shaposhnik, C. D. Funk, and A. J. Lusis. 2002. Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice. *Circ. Res.* **91:** 120–126.

75. Wang, X., M. Ria, P. M. Kelmenson, P. Eriksson, D. C. Higgins, A. Samnegard, C. Petros, J. Rollins, A. M. Bennet, B. Wiman, et al. 2005. Positional identification of TNFSF4, encoding OX40 ligand, as a gene that influences atherosclerosis susceptibility. *Nat. Genet.* **37:** 365–372.

76. Clee, S. M., B. S. Yandell, K. M. Schueler, M. E. Rabaglia, O. C. Richards, S. M. Raines, E. A. Kabara, D. M. Klass, E. T. Mui, D. S. Stapleton, et al. 2006. Positional cloning of Sorcs1, a type 2 diabetes quantitative trait locus. *Nat. Genet.* **38:** 688–693.

77. Hansson, D. K., P. Libby, U. Schonbeck, and Z. Q. Yean. 2002. Innate and adaptive immunity in the pathogenesis of atherosclerosis. *Circ. Res.* **91:** 281–291.

78. Mehrabian, M., and H. Allayee. 2003. 5-Lipoxygenase and atherosclerosis. *Curr. Opin. Lipidol.* **14:** 447–457.

79. Ghazalpour, A., S. Doss, X. Yang, J. Aten, E. M. Toomey, A. Van Nas, S. Wang, T. A. Drake, and A. J. Lusis. 2004. Thematic review series: The pathogenesis of atherosclerosis. Toward a biological network for atherosclerosis. *J. Lipid Res.* **45:** 1793–1805.

80. Drake, T. A., E. E. Schadt, and A. J. Lusis. 2006. Integrating genetic and gene expression data: application to cardiovascular and metabolic traits in mice. *Mamm. Genome.* **17:** 466–479.

81. Petretto, E., J. Mangion, M. Pravanec, N. Hubner, and T. J. Aitman. 2006. Integrated gene expression profiling and linkage analysis in the rat. *Mamm. Genome.* **17:** 480–489.

82. Williams, R. W. 2006. Expression genetics and the phenotype revolution. *Mamm. Genome.* **17:** 496–502.

83. Cervino, A. C., G. Li, S. Edwards, J. Zhu, C. Laurie, G. Tokiwa, P. Y. Lum, S. Wang, L. W. Castellini, A. J. Lusis, et al. 2005. Integrating QTL and high-density SNP analyses in mice to identify Insig2 as a susceptibility gene for plasma cholesterol levels. *Genomics.* **86:** 505–517.

84. Mehrabian, M., H. Allayee, J. Stockton, P. Y. Lum, T. A. Drake, L. W. Castellani, M. Suh, C. Armour, S. Edwards, J. Lamb, et al. 2005. Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nat. Genet.* **37:** 1224–1233.

85. Schadt, E. E., S. A. Monks, T. A. Drake, A. J. Lusis, N. Che, V. Colinayo, T. G. Ruff, S. B. Milligan, J. R. Lamb, G. Cavet, et al. 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* **422:** 297–302.

86. Wang, S., N. Yehya, E. E. Schadt, H. Wang, T. A. Drake, and A. J. Lusis. 2006. Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genet.* **2:** 148–159.

87. Lan, H., M. Chen, J. B. Flowers, B. S. Yandell, D. S. Stapleton, C. M. Mata, E. T. Mui, M. T. Flowers, K. L. Schueler, K. F. Manly, et al. 2006. Combined expression trait correlations and expression quantitative trait locus mapping. *PLoS Genet.* **2:** 51–61.

88. Lum, P. Y., Y. Chen, J. Zhu, J. Lamb, S. Melmed, S. Wang, T. A. Drake, A. J. Lusis, and E. E. Schadt. 2006. Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. *J. Neurochem.* **97(Suppl. 1):** 50–62.

89. Castro Cabezas, M. 2003. Postprandial lipaemia in familial combined hyperlipidaemia. *Biochem. Soc. Trans.* **31:** 1090–1093.

90. Pajukanta, P., J. S. Bodnar, R. Sallinen, M. Chu, T. Airaksinen, Q. Xiao, L. W. Castellani, S. S. Sheth, M. Wessman, A. Palotie, et al. 2001. Fine mapping of Hyplip1 and the human homolog, a potential locus for FCHL. *Mamm. Genome.* **12:** 238–245.

91. Pajukanta, P., H. E. Lilja, J. S. Sinsheimer, R. M. Cantor, A. J. Lusis, M. Gentile, X. J. Duan, A. Soro-Paavonen, J. Naukkarinen, J. Saarela, et al. 2004. Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nat. Genet.* **36:** 371–376.

92. Huertas-Vazquez, A., C. Aguilar-Salinas, A. J. Lusis, R. M. Cantor, S. Canizales-Quinteros, J. C. Lee, L. Mariana-Nunez, R. M. Riba-Ramirez, A. Jokiaho, T. Tusie-Luna, et al. 2005. Familial combined hyperlipidemia in Mexicans: association with upstream transcription factor 1 and linkage on chromosome 16q24.1. *Arterioscler. Thromb. Vasc. Biol.* **25:** 1985–1991.

93. Putt, W., J. Palmen, V. Nicaud, D. A. Tregouet, N. Tahri-Daizadeh, D. M. Flavell, S. E. Humphries, and P. J. Talmud. 2004. Variation in USF1 shows haplotype effects, gene:gene and gene:environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II. *Hum. Mol. Genet.* **13:** 1587–1597.

94. Suviolahti, E., L. J. Oksanen, M. Ohman, R. M. Cantor, M. Ridderstrale, T. Tuomi, J. Kaprio, A. Rissanen, P. Mustajoki, P. Jousilahti, et al. 2003. The SLC6A14 gene shows evidence of association with obesity. *J. Clin. Invest.* **112:** 1762–1772.

95. Wang, Q. 2005. Molecular genetics of coronary artery disease. *Curr. Opin. Cardiol.* **20:** 182–188.

96. van der Weyden, L., D. J. Adams, and A. Bradley. 2002. Tools for targeted manipulation of the mouse genome. *Physiol. Genomics.* **11:** 133–164.

97. Glaser, S., K. Anastassiadis, and A. F. Stewart. 2005. Current issues in mouse genome engineering. *Nat. Genet.* **37:** 1187–1193.

98. Austin, C. P., J. F. Battey, A. Bradley, M. Bucan, M. Capecchi, F. S. Collins, W. F. Dove, G. Duyk, S. Dymecki, J. T. Eppig, et al. 2004. The knockout mouse project. *Nat. Genet.* **36:** 921–924.

99. Auwerx, J., P. Avner, R. Baldock, A. Ballabio, R. Balling, M. Barbacid, A. Berns, A. Bradley, S. Brown, P. Carmeliet, et al. 2004. The European dimension for the mouse genome mutagenesis program. *Nat. Genet.* **36:** 925–927.

100. Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science.* **285:** 901–906.

101. Sugimoto, A. 2004. High-throughput RNAi in *Caenorhabditis elegans*: genome-wide screens and functional genomics. *Differentiation.* **72:** 81–91.

102. Ashrafi, K., F. Y. Chang, J. L. Watts, A. G. Fraser, R. S. Kamath, J. Ahringer, and G. Ruvkun. 2003. Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature.* **421:** 268–272.

103. Cullen, L. M., and G. M. Arndt. 2005. Genome-wide screening for gene function using RNAi in mammalian cells. *Immunol. Cell Biol.* **83:** 217–223.

104. Yu, J., and A. P. McMahon. 2006. Reproducible and inducible knockdown of gene expression in mice. *Genesis.* **44:** 252–261.

105. Williams, R. W., L. Flaherty, and D. W. Threadgill. 2003. The math of making mutant mice. *Genes Brain Behav.* **2:** 191–200.

106. Guenet, J. L. 2004. Chemical mutagenesis of the mouse genome: an overview. *Genetica.* **122:** 9–24.

107. Clark, A. T., D. Goldowitz, J. S. Takahashi, M. H. Vitaterna, S. M. Siepka, L. L. Peters, W. N. Frankel, G. A. Carlson, J. Rossant, J. H. Nadeau, et al. 2004. Implementing large-scale ENU mutagenesis screens in North America. *Genetica.* **122:** 51–64.

108. Cordes, S. P. 2005. N-Ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express. *Microbiol. Mol. Biol. Rev.* **69:** 426–439.

109. Justice, M. J., J. K. Noveroske, J. S. Weber, B. Zheng, and A. Bradley. 1999. Mouse ENU mutagenesis. *Hum. Mol. Genet.* **8:** 1955–1963.

110. Nolan, P. M., J. Peters, M. Strivens, D. Rogers, J. Hagan, N. Spurr, I. C. Gray, L. Vizor, D. Brooker, E. Whitehill, et al. 2000. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat. Genet.* **25:** 440–443.

111. Hrabe de Angelis, M. H., H. Flaswinkel, H. Fuchs, B. Rathkolb, D. Soewarto, S. Marschall, S. Heffner, W. Pargent, K. Wuensch, M. Jung, et al. 2000. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat. Genet.* **25:** 444–447.

112. Hough, T. A., P. M. Nolan, V. Tsipouri, A. A. Toye, I. C. Gray, M. Goldsworthy, L. Moir, R. D. Cox, S. Clements, P. H. Glenister, et al. 2002. Novel phenotypes identified by plasma biochemical screening in the mouse. *Mamm. Genome.* **13:** 595–602.

113. Mohr, M., M. Klempt, B. Rathkolb, M. H. de Angelis, E. Wolf, and B. Aigner. 2004. Hypercholesterolemia in ENU-induced mouse mutants. *J. Lipid Res.* **45:** 2132–2137.

114. Hoebe, K., P. Georgel, S. Rutschmann, X. Du, S. Mudd, K. Crozat, S. Sovath, L. Shamel, T. Hartung, U. Zahringer, et al. 2005. CD36 is a sensor of diacylglycerides. *Nature.* **433:** 523–527.

115. Quwailid, M. M., A. Hugill, N. Dear, L. Vizor, S. Wells, E. Horner, S. Fuller, J. Weedon, H. McMath, P. Woodman, et al. 2004. A gene-driven ENU-based approach to generating an allelic series in any gene. *Mamm. Genome.* **15:** 585–591.

116. Michaud, E. J., C. T. Culiat, M. L. Klebig, P. E. Barker, K. T. Cain, D. J. Carpenter, L. L. Easter, C. M. Foster, A. W. Gardner, Z. Y. Guo, et al. 2005. Efficient gene-driven germ-line point mutagenesis of C57BL/6J mice. *BMC Genomics.* **6:** 1–18.

117. Augustin, M., R. Sedlmeier, T. Peters, U. Huffstadt, E. Kochmann, D. Simon, M. Schoniger, S. Garke-Mayerthaler, J. Laufs, M. Mayhaus, et al. 2005. Efficient and fast targeted production of murine models based on ENU mutagenesis. *Mamm. Genome.* **16:** 405–413.

118. Stanford, W. L., J. B. Cohn, and S. P. Cordes. 2001. Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.* **2:** 756–768.

119. Skarnes, W. C., H. von Melchner, W. Wurst, G. Hicks, A. S. Nord, T. Cox, S. G. Young, P. Ruiz, P. Soriano, M. Tessier-Lavigne, et al. 2004. A public gene trap resource for mouse functional genomics. *Nat. Genet.* **36:** 543–544.

120. Nord, A. S., P. J. Chang, B. R. Conklin, A. V. Cox, C. A. Harper, G. G. Hicks, C. C. Huang, S. J. Johns, M. Kawamoto, S. Liu, et al. 2006. The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse. *Nucleic Acids Res.* **34:** D642–D648.

121. Beigneux, A. P., C. Kosinski, B. Gavino, J. D. Horton, W. C. Skarnes, and S. G. Young. 2004. ATP-citrate lyase deficiency in the mouse. *J. Biol. Chem.* **279:** 9557–9564.

122. Steenbergen, R., T. S. Nanowski, A. Beigneux, A. Kulinski, S. G. Young, and J. E. Vance. 2005. Disruption of the phosphatidylserine decarboxylase gene in mice causes embryonic lethality and mitochondrial defects. *J. Biol. Chem.* **280:** 40032–40040.

123. Bergo, M. O., B. J. Gavino, R. Steenbergen, B. Sturbois, A. F. Parlow, D. A. Sanan, W. C. Skarnes, J. E. Vance, and S. G. Young. 2002. Defining the importance of phosphatidylserine synthase 2 in mice. *J. Biol. Chem.* **277:** 47701–47708.

124. Steenbergen, R., T. S. Nanowski, R. Nelson, S. G. Young, and J. E. Vance. 2006. Phospholipid homeostasis in phosphatidylserine synthase-2-deficient mice. *Biochim. Biophys. Acta.* **1761:** 313–323.

125. Agarwal, A. K., E. Arioglu, S. De Almeida, N. Akkoc, S. I. Taylor, A. M. Bowcock, R. I. Barnes, and A. Garg. 2002. AGPAT2 is mutated in congenital generalized lipodystrophy linked to chromosome 9q34. *Nat. Genet.* **31:** 21–23.

126. Gale, S. E., A. Frolov, X. Han, P. E. Bickel, L. Cao, A. Bowcock, J. E. Schaffer, and D. S. Ory. 2006. A regulatory role for 1-acylglycerol-3-phosphate-O-acyltransferase 2 in adipocyte differentiation. *J. Biol. Chem.* **281:** 11082–11089.

127. Beigneux, A. P., L. Vergnes, X. Qiao, S. Quatela, R. Davis, S. M. Watkins, R. A. Coleman, R. L. Walzem, M. Philips, K. Reue, et al. 2006. Agpat6—a novel lipid biosynthetic gene required for triacylglycerol production in mammary epithelium. *J. Lipid Res.* **47:** 734–744.

128. Vergnes, L., A. P. Beigneux, R. Davis, S. M. Watkins, S. G. Young, and K. Reue. 2006. Agpat6 deficiency causes subdermal lipodystrophy and resistance to obesity. *J. Lipid Res.* **47:** 745–754.

129. Friedel, R. H., A. Plump, X. Lu, K. Spilker, C. Jolicoeur, K. Wong, T. R. Venkatesh, A. Yaron, M. Hynes, B. Chen, et al. 2005. Gene targeting using a promoterless gene trap vector ("targeted trapping") is an efficient method to mutate a large fraction of genes. *Proc. Natl. Acad. Sci. USA.* **102:** 13188–13193.

130. Warming, S., N. Costantino, D. L. Court, N. A. Jenkins, and N. G. Copeland. 2005. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33:** 1–12.

131. Testa, G., Y. Zhang, K. Vintersten, V. Benes, W. W. Pijnappel, I. Chambers, A. J. Smith, A. G. Smith, and A. F. Stewart. 2003. Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nat. Biotechnol.* **21:** 443–447.

132. Meiner, V. L., C. L. Welch, S. Cases, H. M. Myers, E. Sande, A. J. Lusis, and R. V. Farese, Jr. 1998. Adrenocortical lipid depletion gene (ald) in AKR mice is associated with an acyl-CoA:cholesterol acyltransferase (ACAT) mutation. *J. Biol. Chem.* **273:** 1064–1069.

133. Hou, X., M. Mrug, B. K. Yoder, E. J. Lefkowitz, G. Kremmidiotis, P. D'Eustachio, D. R. Beier, and L. M. Guay-Woodford. 2002. Cystin, a novel cilia-associated protein, is disrupted in the cpk mouse model of polycystic kidney disease. *J. Clin. Invest.* **109:** 533–540.

134. Lu, K., H. Nishimori, Y. Nakamura, K. Shima, and M. Kuwajima. 1998. A missense mutation of mouse OCTN2, a sodium-dependent carnitine cotransporter, in the juvenile visceral steatosis mouse. *Biochem. Biophys. Res. Commun.* **252:** 590–594.

135. Nezu, J., I. Tamai, A. Oku, R. Ohashi, H. Yabuuchi, N. Hashimoto, H. Nikaido, Y. Sai, A. Koizumi, Y. Shoji, et al. 1999. Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter. *Nat. Genet.* **21:** 91–94.

136. Loftus, S. K., J. A. Morris, E. D. Carstea, J. Z. Gu, C. Cummings, A. Brown, J. Ellison, K. Ohno, M. A. Rosenfeld, D. A. Tagle, et al. 1997. Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene. *Science.* **277:** 232–235.

137. Omran, H., K. Haffner, S. Burth, C. Fernandez, B. Fargier, A. Villaquiran, H. G. Nothwang, S. Schnittger, H. Lehrach, D. Woo, et al. 2001. Human adolescent nephronophthisis: gene locus synteny with polycystic kidney disease in pcy mice. *J. Am. Soc. Nephrol.* **12:** 107–113.

138. Moulson, C. L., D. R. Martin, J. J. Lugus, J. E. Schaffer, A. C. Lind, and J. H. Miner. 2003. Cloning of wrinkle-free, a previously uncharacterized mouse mutation, reveals crucial roles for fatty acid transport protein 4 in skin and hair development. *Proc. Natl. Acad. Sci. USA.* **100:** 5274–5279.