Functional analysis of secreted and transmembrane proteins critical to mouse development

Kevin J. Mitchell[1,2*], Kathy I. Pinson[1*], Olivia G. Kelly[1], Jane Brennan[¶], Joel Zupicich[1], Paul Scherz[1], Philip A. Leighton[2], Lisa V. Goodrich[2], Xiaowei Lu[2], Brian J. Avery[1], Peri Tate[1], Kariena Dill[1], Edivinia Pangilinan[1], Paul Wakenight[1], Marc Tessier-Lavigne[2] & William C. Skarnes[1]

** These authors contributed equally to the work.*

We describe the successful application of a modified gene-trap approach, the secretory trap, to systematically analyze the functions *in vivo* of large numbers of genes encoding secreted and membrane proteins. Secretory-trap insertions in embryonic stem cells can be transmitted to the germ line of mice with high efficiency and effectively mutate the target gene. Of 60 insertions analyzed in mice, one-third cause recessive lethal phenotypes affecting various stages of embryonic and postnatal development. Thus, secretory-trap mutagenesis can be used for a genome-wide functional analysis of cell signaling pathways that are critical for normal mammalian development and physiology.

Introduction

Gene-trapping potentially offers a powerful tool to generate numerous insertional mutations in mice and to perform phenotypic screens for developmental mutants[1–4]. A large number of gene-trap insertions can be isolated in mouse embryonic stem (ES) cells and are immediately accessible to molecular characterization using high-throughput methods to generate sequence tags for each mutated gene[5–8]. New mutations can be generated in mice at a rate far exceeding that of conventional gene targeting and, in contrast to classical *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis[9,10], recessive lethal phenotypes can be identified at all developmental stages and can be easily maintained. In addition, various criteria can be used to pre-select genes of interest[11–14], and their functional analysis can proceed rapidly based on a confluence of sequence, mapping, expression, and phenotype information.

Despite the obvious potential of this technology, the paucity of mutant phenotypes generated by gene-trapping raises genuine concerns about the utility of this approach for large-scale mutagenesis. Efforts to generate libraries of ES cell lines harboring gene-trap insertions are now underway[7,8], amply demonstrating that thousands of genes are accessible using this approach. The value of these resources is still, however, unproven, particularly with regard to the efficiency by which the insertions induce loss-of-function mutations in mice. Several reported examples of gene-trap insertions that fail to mutate the target gene reinforce the perception that gene-trap insertions are not effective mutagens[15–18]. In the few cases in which phenotypes were observed, lethality generally occurred at early embryonic stages[2,19–21]. This raises an additional concern that, because of the requirement for the expression of the trapped gene in ES cells, gene-trapping will be biased against genes that function later in development.

We are interested in defining new signaling pathways that are critical for normal mammalian development and physiology. To this end, we are using a modified gene-trapping strategy to isolate insertional mutations specifically in secreted and membrane-spanning proteins[13]. This class of proteins is especially relevant to the study of developmental and homeostatic processes such as cell–cell signaling mechanisms, cell proliferation, and cell migration. Secreted and membrane proteins may be of particular importance in mammalian development, as suggested by a large expansion of new extracellular protein architectures in humans compared with worms and flies[22]. Thus, we cannot rely on genetic approaches in other model organisms for the analysis of a significant fraction of cell-surface proteins in mammals.

In this study, sequence tags corresponding to the trapped gene were obtained for over 500 secretory-trap cell lines, from which 60 insertions were selected for phenotypic analysis in mice. One-third of the insertions induced recessive lethal phenotypes with examples that perturbed all stages of embryonic and postnatal development. We have assigned essential functions to nine known genes not previously analyzed in mice as well as to five novel membrane proteins with no known function in any system. Our results demonstrate that the secretory trap is well suited to large-scale functional analysis of secreted and membrane-spanning proteins and can be used to isolate a wealth of new and interesting phenotypes in mice.

Results

The original secretory trap vector, pGT1.8TM (ref. 13), and several modified versions engineered in different reading frames (see Methods and ref. 23) were electroporated into feeder-independent ES cell lines. These plasmid-based vectors contain a type II membrane-spanning domain fused to the amino terminus of the β-geo reporter,

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [2]Howard Hughes Medical Institute and Department of Anatomy, University of California, San Francisco, California, 94143 USA. [¶]Present address: Department of Molecular & Cell Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA. Correspondence should be addressed to W.C.S. (e-mail: skarnes@socrates.berkeley.edu).

Table 1 • Genes associated with productive insertions of the secretory-trap vector

Known mouse genes (59) (gene symbol, name, accession number)

Transmembrane		
Adam19	a disintegrin and metalloprotease domain 19	NM_009616
Adam23	a disintegrin and metalloprotease domain 23	NM_011780
Aplp2	amyloid-beta (A4) precursor-like protein-2	NM_009691
Atrn	attractin	NM_009730
Cdh1	cadherin 1	NM_009864
Cdh3	cadherin 3	X06340
Crim1	cysteine-rich motor neuron 1	AF168680
Ddx26	DEAD/H box polypeptide 26 (Notch2-like)	NM_008715
Emb	embigin	NM_010330
Eng	endoglin	NM_007932
Epha2	Eph receptor A2	NM_010139
Epha4	Eph receptor A4	NM_007936
Fath	fat tumor suppressor homolog (*Drosophila*)	AJ250768
Fgfr1	fibroblast growth factor receptor 1	NM_010206
Gpa33	glycoprotein A33 (transmembrane)	NM_021610
Gpc1	glypican-1	NM_016696
Gpc3	glypican-3	NM_016697
Gpc4	glypican-4	NM_008150
Icam1	intercellular adhesion molecule	M31585
Ifnar	interferon (α and β) receptor	NM_010508
Igf1r	insulin-like growth factor-I receptor	AF05618
Igf2r	insulin-like growth factor-2 receptor	U04710
Il6st	interleukin-6 signal transducer	NM_010560
Itga5	integrin-α-5	NM_010577
Itga6	integrin-α-6	NM_008397
Itgb1	integrin-β-1 (fibronectin receptor-β)	Y00769
Jag1	jagged 1	NM_013822
Jcam1	junction cell adhesion molecule -1	U89915
Kit	kit oncogene	NM_021099
Ldlr	low-density lipoprotein receptor	NM_010700
Lifr	leukemia inhibitory factor receptor	NM_013584
Lrp2	low-density lipoprotein receptor-related protein-2	L34049
Lrp6	low-density lipoprotein receptor-related protein	NM_008514
Mdu1	antigen identified by monoclonal antibodies 4F2	NM_008577
Mfge8	milk fat globule-EGF factor-8 protein	NM_008594
Mme	membrane metalloendopeptidase	NM_008604
Neo1	neogenin	NM_008684
Notch1	Notch gene homolog-1 (*Drosophila*)	NM_008714
Notch2	Notch gene homolog-2 (*Drosophila*)	D32210
Notch3	Notch gene homolog-3 (*Drosophila*)	NM_008716
Nrp2	neuropilin-2	AF022861
Plxn1	plexin A1	NM_008881
Ptprf	protein tyrosine phosphatase, receptor-type, F	AF300943
Ptprg	protein tyrosine phosphatase, receptor-type, G	NM_008981
Ptprk	protein tyrosine phosphatase, receptor-type, K	NM_008983
Ptprs	protein tyrosine phosphatase, receptor type, S	X82288
Pvs	poliovirus sensitivity	NM_008990
Rnp24-pending	coated vesicle membrane protein	Y17793
Robo1	roundabout homolog 1	NM_019770
Sdc4	syndecan 4	NM_011521
Transmembrane (continued.)		
Sdfr1	stromal cell-derived factor receptor-1	NM_009145
Sdfr2	stromal cell-derived factor receptor 2	NM_009146
Selel	selectin, endothelial cell, ligand	NM_009149
Sema4b	semaphorin-4b	X85992
Sema6a	semaphorin-6a	NM_018744
Sorl1	sortilin-related receptor, LDLR class A repeats-containing	AF031816
Stim1	stromal interaction molecule 1	NM_009287
Tgfbr1	transforming growth factor, β receptor-I	NM_009370
Tmeff1	transmembrane protein with EGF-like and 2 follistatin-like domains 1	AJ400622
ER/Golgi/Lysosomal (30)		
Asph	aspartate-β-hydroxylase	AF289487
Atf6	activating transcription factor-6	AK020270
Atp6k	ATPase, H+ transporting lysosomal (vacuolar proton pump)	AA792941
B4galt1	UDP-Gal:β-GlcNac β- 1,4-galactosyltransferase, 1	NM_022305
B4galt6	UDP-Gal:glucosylceramide β-1,4-galactosyltransferase	NM_019737
B3gnt	β-1,3-N-acetylglucosaminyltransferase 1	NM_016888
Bgl	β galactosidase complex	NM_009752
Cai	calcium binding protein, intestinal	J05186
Canx	calnexin	NM_007597
Cd3612	CD36 antigen-like-2	NM_007644
Ctsc	cathepsin C	NM_009982
Dnajb11	DnaJ (Hsp40) homolog, subfamily B, member 11	NM_026400
Ero1l-pending	ERO1-like (*S. cerevisiae*)	NM_015774
Ext1	exostoses (multiple) 1	NM_010162
Gcnt2	glucosaminyl (N-acetyl) transferase 2, I-branching enzyme	NM_008105
Gfpt2	glutamine fructose-6-phosphate transaminase-2	NM_013529
Grp58	glucose-regulated protein, 58kD	NM_007952
Hs6st1	heparan sulfate 6-O-sulfotransferase-1	NM_015818
Lamp2	lysosomal membrane glycoprotein-2	NM_010685
Man2a1	mannosidase 2, α 1	NM_008549
Manba	mannosidase, beta A, lysosomal	AF306557
Mbtps1	membrane-bound transcription factor protease, site 1	NM_019709
P4hb	prolyl 4-hydroxylase, β-polypeptide	X06453
Pcsk3	proprotein convertase subtilisin/kexin type 3	NM_011046
Plod2	procollagen lysine, 2-oxoglutarate 5-dioxygenase 2	NM_011961
Ppib	peptidylprolyl isomerase B	NM_011149
Ptdss2	phosphatidylserine synthase 2	NM_013782
Rpn2	ribophorin II	NM_019642
Stch	stress 70 protein chaperone, microsome-associated, 60 kD	AI226600
Tor1b	torsin family 1, member B	AJ297743
Secreted/extracellular matrix (26)		
Agrn	agrin	U84407
Bmp8a	bone morphogenetic protein-8a	NM_007558
Ctsl	cathepsin L	NM_009984
Cyr61	cysteine-rich protein-61	NM_010516
Fbln1	fibulin-1	NM_010180
Fn1	fibronectin-1	X93167
Fut8	fucosyltransferase-8	NM_016893
Grn	granulin	NM_008175
Hspg2	perlecan (heparan sulfate proteoglycan-2)	NM_008305
Igfbp2	insulin-like growth factor binding protein-2	NM_008342

Table 1 • (continued)

Secreted/extracellular matrix (continued.)			**New genes with human homologs[a] (21)**		
Lama1	laminin, α 1	NM_008480	*C12orf8*	chromosme 12 open reading frame 8	NM_006917
Lama5	laminin, α 5	U37501	CDA08[b]	hypothetical protein CDA08	NM_030790
Lamb1-1	laminin B1 subunit 1	M15525	CH1[b]	membrane protein CH1	NM_014283
Lamc1	laminin, γ-1	J03484	CTL1[b]	transporter-like protein	NM_022109
Leftb	left–right determination, factor B	D83921	*ENPP3*	ectonucleotide pyrophosphatase/phosphodiesterase 3	NM_005021
Lman1	lectin, mannose-binding, 1	AK011495	*GNS*	N-acetylglucosamine-6-sulfatase	NM_002076
Nid2	nidogen-2	NM_008695	*GOLPH2*	golgi phosphoprotein 2	NM_016548
Nodal	nodal	X70514	*GPR48*	G protein-coupled receptor-48	NM_018490
Ntn1	netrin-1	NM_008744	*GS3786*	predicted osteoblast protein	NM_014888
Pdgfa	platelet-derived growth factor, α	NM_008808	HSCP1	serine carboxypeptidase-1	NM_021626
Pdir-pending	protein disulfide isomerase-related	AK012415	HSGP25L2G[b]	p25L2 protein	X90872
Pros1	protein S (α)	Z25469	HUGT1	UDP-glucose glycoprotein glucosyltransferase-1	NM_020120
Spint2	serine protease inhibitor, Kunitz type 2	AF099020	N33[b]	putative prostate cancer tumor suppressor	NM_006765
Tdgf1	teratocarcinoma-derived growth factor	NM_011562	*NLGN3*	neuroligin 3	NM_018977
Tfpi	tissue factor pathway inhibitor	AF004833	*NPC2*	Niemann-Pick disease, type C2 gene	NM_006432
Tnrc5	trinucleotide repeat containing 5	AK005532	*PIGK*	phosphatidylinositol glycan, class K	AF022913
Multiple membrane-spanning (7)			PM5[b]	pM5 protein	NM_014287
Gabbr1	gamma-aminobutyric acid (GABA) B receptor, 1	AF114168	*PRCP*	prolylcarboxypeptidase	NM_005040
Kcnk1	potassium channel, subfamily K, member 1	NM_008430	*PTK7*	protein tyrosine kinase-7	NM_002821
Kcnk5	potassium channel, subfamily K, member 5	NM_021542	*RPN1*	ribophorin-I	NM_002950
Lisch7-pending	liver-specific bHLH-Zip transcription factor	U49507	*SLC28A1*	solute carrier family-28, member 1	NM_004213
Mdu1	antigen identified by monoclonal antibodies 4F2	NM_008577			
Slc2a1	solute carrier family 2 (facilitated glucose transporter) 1	M23384			
Tm9sf2	transmembrane 9 superfamily member 2	AK010325			

[a]Symbols representing the human genes are indicated, [b]Temporary gene symbol.

which is flanked by a splice acceptor and polyadenylation signal. In theory, activation of the β-geo reporter depends on the production of fusion protein products that incorporate an N-terminal signal sequence (or a type II transmembrane domain) from the gene at the site of insertion.

In a previous study, we found that the secretory trap enriched, but did not absolutely select for, insertions in secreted and membrane-spanning proteins. As many as half of the G418-resistant, β–galactosidase (βgal) positive colonies carried insertions that were inefficiently spliced or spliced to the 5′ untranslated exons of genes that encoded non-secreted proteins[24]. It was, however, noted that pre-selecting cell lines showing a distinctive pattern of βgal activity (perinuclear and multiple cytoplasmic dots of staining) significantly increased the proportion of properly spliced insertions in *bona fide* secreted or membrane proteins[24]. The insertions reported in this study are compiled from cell lines that display the 'secretory' pattern of βgal expression.

Spectrum of trapped genes

Sequence tags were obtained for a total of 528 secretory-trap insertions using direct solid-phase sequencing of 5′ RACE reactions[5]. One-quarter (126) of the sequences were identified as non-productive insertions, that is, insertions in 5′ UTRs or structural RNAs, and were eliminated from the screen. Of the remaining 402 sequence tags, 371 correspond to genes that encode known or putative secreted and membrane-spanning proteins. Because of multiple independent insertion events, mainly in a small number of loci (see below), this set of sequences represents a total of 187 individual genes (Table 1).

All classes of secreted and membrane-spanning proteins are represented. Three-quarters match known genes in mice (122) or are highly related to human genes (21). The remainder (44) of the sequence tags correspond to novel genes that are currently represented only by mammalian ESTs and/or genomic sequences in the public data base. Conceptual translation of assembled full-length sequences (see Methods) predicts that they all encode secreted and membrane-spanning proteins (Fig. 1). Many contain novel combinations of conserved extracellular domains and have predicted homologs of unknown function in other organisms. A small number of the target genes seem to be specific to vertebrates.

Non-random integration of the secretory-trap vector

Certain loci appear to be preferred sites for insertion of the secretory-trap vector. Twenty-seven independent insertions were recovered in the laminin γ 1 gene, followed by six genes that were isolated 10–14 times (Eph receptor A2, cadherin 1, protein tyrosine phosphatase receptor type K, *fat* tumor supressor homolog, perlecan, and low-density lipoprotein receptor). Insertions in these seven loci make up over a quarter of the total insertions in known genes. Another 38 loci have been trapped between 2 and 6 times, whereas 68 genes are represented by a single insertion. A similar nonrandom distribution of gene-trap insertions was reported previously[7,8]. The discontinuous distribution of repeat events into at least three discrete classes makes it difficult to estimate the number of genes accessible by this approach. The proportion of genes represented by a single insertion has, however, remained constant at one in three throughout these experiments, indicating that the screen is far from saturation.

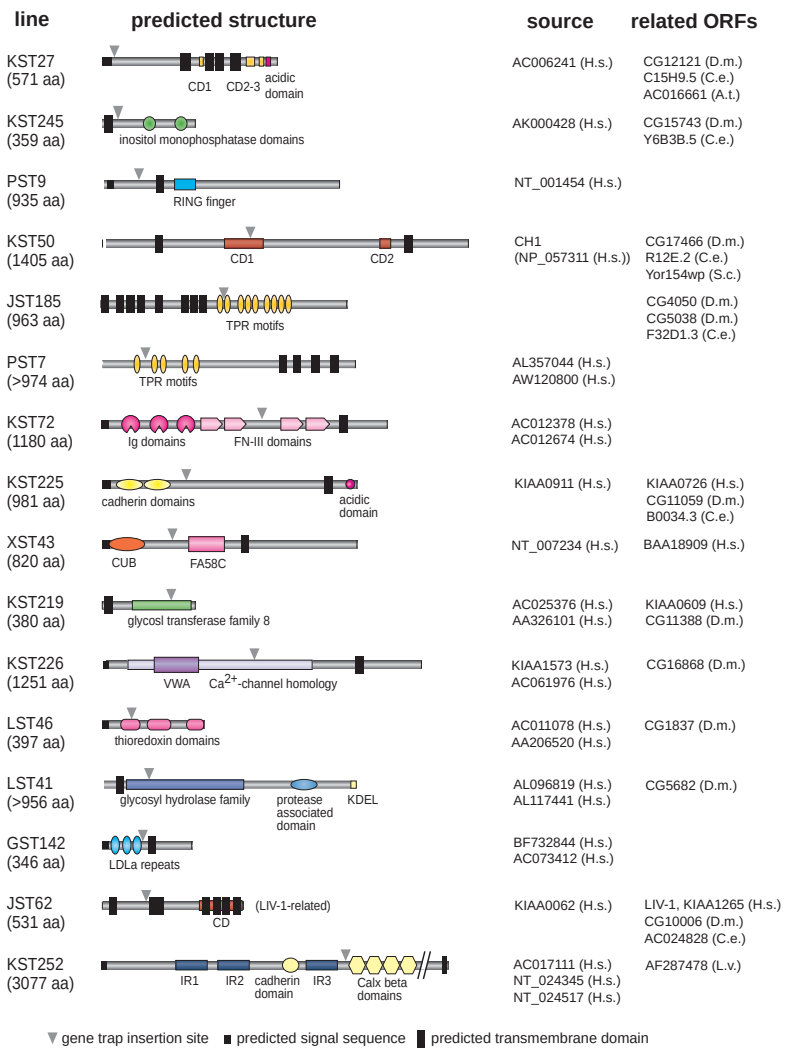

Fig. 1 Predicted structures of 16 new membrane or secreted proteins. Schematics show the predicted protein structures of 16 trapped genes, based on the presumed human orthologs. Sequences were compiled from human genomic, cDNA, and EST sequences in the public database ("Source"). JST185 structure was predicted from the mouse cDNA (J.B., unpublished data). Accession numbers/gene identifiers for related genes or predicted ORFs from humans, flies, worms, or other organisms are listed on the right. The black boxes represent putative signal sequences and transmembrane domains whereas triangles mark the insertion site. Known protein motifs and novel conserved domains are shown in color. Domain abbreviations: TPR, tetratricopeptide repeat ; Ig, immunoglobulin domain; FN-III, fibronectin type III repeat; FA58C, coagulation factor 5/8 C-terminal domain; VWA, von Willebrand factor type A domain; KDEL, endoplasmic reticulum retention sequence; LDLa, low-density lipoprotein receptor domain class A. IR denotes an internal repeat and CD a conserved domain. Species abbreviations: A.t., *Arabidopsis thaliana*; C.e., *Caenorhabditis elegans*; D.m., *Drosophila melanogaster*; H.s., *Homo sapiens*; L.t., *Lytechinus variegatus* (sea urchin); S.c., *Saccharomyces cerevisiae*.

We find that the frequency of repeat events at a particular locus does not correlate with the expression level of the target gene based on βgal staining in ES cells or northern blot analysis (data not shown). In addition, repeat events were often found in multiple introns of the same gene, arguing against a direct effect of specific sequences at the site of insertion or within the vector in promoting integration. Finally, there is no evidence for the clustering of insertions on chromosomes: insertions in 89 genes mapped in the mouse were evenly distributed across the genome, the average number per chromosome roughly correlating with chromosome length.

Screening for recessive lethal phenotypes

Insertions that disrupt the coding sequence of known or putative cell surface proteins were selected for blastocyst injection. We observed a high rate of germline transmission with cell lines derived from two feeder-independent ES cell lines (CGR8.8 and E14Tg2a-clone 4). In our hands, a minimum of 90% of our cell lines have the potential to contribute to the germ line of mice by the injection of fewer than 40 blastocyts. Our experiments demonstrate that favorable rates of germline transmission can be achieved with feeder-independent ES cells, at a great saving of cost and labor.

To identify genes essential for development, offspring from intercrosses of F_2 heterozygotes are genotyped at weaning to detect the absence of homozygous mutants. At the same time, these animals are examined for obvious morphological or behavioral defects. For lines that produce no or fewer than expected homozygous mutant animals, embryos are collected and genotyped at various developmental stages to determine when lethality occurs and to look for gross developmental abnormalities. In a concurrent screen, lines were also examined for defects in nervous system wiring[23].

Secretory-trap insertions generate null alleles

Two lines of evidence suggest that insertions of the secretory-trap vector induce phenotypic null alleles. First, a comparison of 11 secretory trap insertions with corresponding gene knockouts shows, in all but one case (Ext1, see below), that our insertion alleles produce similar if not identical phenotypes (see Table 2 and refs. 23, 25–27). Second, molecular analysis of several secretory-trap mutant lines has detected negligible levels of wild-type transcripts or protein[13,25–29]. We have extended this analysis using sensitive and quantitative RNase protection assays in three additional lines (*Sema6a, Notch3,* and *Epha4*) and found only trace amounts (<1%) of wild-type transcripts (Fig. 2). We conclude that secretory-trap insertions consistently generate null or very strong hypomorphic alleles.

One caveat is that the fusion proteins generated by this approach could retain some wildtype function or exert dominant activities as a result of the expression of a portion of the extracellular domain of the trapped gene fused to the membrane-tethered β-geo reporter. It is important to note, however, that βgal activity and protein in these insertions are concentrated in the endoplasmic reticulum and in multiple inclusion bodies[13] (probably lysosomes). Fusions with secreted and plasma-membrane proteins are thus sequestered intracellularly, thereby minimizing any effect that they may have at the cell surface. In contrast, fusions to proteins that normally reside in an intracellular membrane compartment such as the endoplasmic reticulum may be more likely to retain residual or dominant activities. The insertion in *Ext1*, a gene encoding a glycosyltransferase that localizes to the endoplasmic reticulum[30], may represent such an example. This allele produces a fusion protein that retains the first 320 amino acids of Ext1, eliminates wildtype mRNA transcripts, but causes a less severe phenotype than the reported gene knockout[31].

Table 2 • Mutant phenotypes induced by secretory-trap insertions

Line	Gene name	Symbol	Lethality	Phenotype
PST115	site-1 protease	*Mbtps1*	embryonic	peri-implantation
Ex160	cell division cycle 42 homolog[a]	*Cdc42*	embryonic	peri-implantation[b]
Ex183	laminin, gamma-1	*Lamc1*	embryonic	peri-implantation[b]
KST272	serine protease inhibitor, Kunitz type 2	*Spint2*	embryonic	gastrulation defect (clefting of the epiblast)
KST27	novel multi-TM		embryonic	gastrulation defect (CNS and tailbud defects)
Ex65	transformed mouse 3T3 cell double minute 2[a]	*Mdm2*	embryonic	gastrulation defect[b]
Ex180	fibronectin-1	*Fn1*	embryonic	gastrulation defect[b]
LST64	exostoses (multiple) 1	*Ext1*	embryonic	cyclopia, abnormal limbs[c]
Ex187	low-density lipoprotein receptor-related protein-6	*Lrp6*	embryonic	axis truncation, limb defects[28]
PST9	novel TM		neonatal	cleft palate, delay in lung development
KST245	novel TM (similar to human FLJ20421)		neonatal	shortened limbs and body axis
ST629	netrin-1	*Ntn1*	neonatal	axon-guidance defects[39]
Ex136	glypican-3	*Gpc3*	neonatal	embryonic overgrowth, cystic kidneys[25,b]
Ex192	agrin	*Agrn*	neonatal	neuromuscular junction defects[26,b]
Ex54	baculoviral IAP repeat-containing-6[a]	*Birc6*	neonatal	no obvious defects
KST50	novel TM (similar to human CH1)		neonatal	no obvious defects
GST70	fat tumor suppressor homolog (*Drosophila*)	*Fath*	neonatal	no obvious defects
KST229	gamma-aminobutyric acid (GABA) B receptor-1	*Gabbr1*	neonatal	no obvious defects
JST185	novel multi-TM		neonatal	respiratory distress, lung pathology
KST21	a disintegrin and metalloprotease domain-23	*Adam23*	postnatal	tremor, ataxia (die by P14)
PST38	Eph receptor-A4	*Epha4*	viable	abnormal gait, axon-guidance defects
KST69	semaphorin-6a	*Sema6a*	viable	axon-guidance defects[23,b]
GST39	neuropilin-2	*Nrp2*	viable	axon-guidance defects[27,b]
PST112	attractin	*Atrn*	viable	coat color (mahogany)[b]
ST534	protein tyrosine phosphatase, receptor-type, F	*Ptprf*	viable	mammary gland branching and lactation defects[b]

[a]Insertions in coding sequences of non-secreted genes. [b]Phenocopy targeted knockouts. [c]Does not phenocopy targeted knockout.

Phenotype analysis of secretory-trap insertions

A total of 60 lines of mice were generated and bred for phenotype analysis, of which 20 were found to cause recessive lethal phenotypes, five additional lines showing visible adult phenotypes (Table 2). Selected phenotypes affecting various stages of embryonic and postnatal development are described below (Fig. 3).

Insertions causing embryonic lethality

Seven insertions cause lethal phenotypes at or before midgestation (Fig. 3*a*–*g*). Embryos carrying a mutation in site-1 protease, an enzyme involved in the regulation of sterol biosynthesis[32], do not form a normal epiblast and die at implantation (data not shown). Mutant blastocysts grown in culture fail to maintain the inner cell mass as indicated by the absence of *Oct4*-expressing cells (Fig. 3*a*,*b*), demonstrating the requirement for site-1 protease in early post-implantation development. Insertions in laminin gamma 1 and Cdc42 (a non-secreted protein) are also presumed to cause peri-implantation lethal phenotypes since homozygous embryos are present at E3.5, but not at E7.5 (data not shown).

An insertion in *Spint2,* encoding hepatocyte growth factor activator inhibitor-2, (a member of the Kunitz serine protease inhibitor family and a negative regulator of hepatocyte growth factor (HGF) signaling)[33], results in lethality at gastrulation. Mutant embryos show an unusual and severe clefting of the embryonic ectoderm at E7.5 (Fig. 3*c*,*d*), coinciding with high expresssion of *Spint2* in the epiblast at this stage (Fig. 3*e*). Although the primitive streak and embryonic mesoderm are clearly visible (data not shown), *Spint2* mutant embryos do not progress to the headfold stage. Gastrulation defects are also observed with an insertion in the gene encoding KST27, a predicted multipass membrane protein, (Fig. 1). Mutant embryos at E9.5 do not turn properly, a process that is absolutely necessary for continued development, and show obvious morphologic abnormalities in the neural tube and tailbud (Fig. 3*f*). In addition, somitogenesis is perturbed as revealed by *in situ* hybridization with the somite marker *Mox1* (Fig. 3*g*). Mutant embryos for insertions in the genes encoding fibronectin and Mdm2 (a non-secreted protein) are resorbed by E10.5 (data not shown). The time of lethality is consistent with the reported gene knockouts of fibronectin[34] and Mdm2 (ref. 35).

Severe developmental defects leading to embryonic lethality at later stages are associated with insertions in exostoses (multiple) 1 (*Ext1*), a tumor-suppressor gene in humans[36], and *Lrp6*, a new member of the low-density lipoprotein receptor gene family[37]. Embryos homozygous for the insertion in *Ext1* show a variable loss of ventral midline structures (manifested as cyclopia in Fig. 3*i*) and delayed limb development (data not shown). *Lrp6* mutant embryos also show a number of severe developmental abnormalities that include truncation of the body axis, loss of distal limb structures, microophthalmia and craniofacial defects (Fig. 3*j*). A detailed analysis of the insertional mutation in *Lrp6* is presented elsewhere[28].

Insertions causing neonatal lethality

Eight insertional mutations cause neonatal lethality, of which three show obvious morphologic abnormalites at birth. Hemizygous male pups carrying an insertion in the X-linked proteoglycan glypican-3 are larger than normal and show kidney dysplasia (data not shown), as was reported for the targeted null mutation[38]. Neonates homozygous for an insertion in a novel membrane protein with similarity to inositol monophosphatases (KST245) display craniofacial defects and shortened limbs and body axis (Fig. 3*k*,*l*). Consistent with a role in osteogenesis, the reporter gene was expressed in the growth plate, cartilage, and periosteum of the long bones (Fig. 3*m*). An insertion in a novel

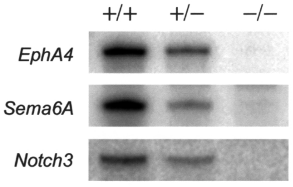

Fig. 2 The secretory-trap vector induces null or near null alleles. RNase protection assays were performed on RNA purified from three independent lines of mice (*Epha4*, *Sema6a*, and *Notch3*) carrying PLAP secretory-trap vector insertions. One sample is shown for each genotype: wild type (+/+), heterozygous (+/–), and homozygous mutant (–/–). All the probes were designed to encompass the secretory-trap vector insertion site in the mRNA for each gene. All samples were normalized by hybridization to a control actin probe (data not shown), and a measure of the signal remaining in the homozygous mutant samples was obtained: for *Epha4*, 0.9%; *Sema6a*, 0.8%; and *Notch3*, 1%.

RING finger gene (PST9) results in cleft palate (Fig. 3*n*,*o*) and a delay in lung maturation in homozygous embryos (Fig. 3*p*,*q*). These pups die several hours after birth, exhibiting respiratory distress and swollen abdomens caused by swallowing air.

Based on external morphology as well as histologic sections of mutant pups at birth, no gross abnormalities are associated with four neonatal lethal insertions in: *fat* tumor suppressor homolog; gamma-aminobutyric acid (GABA) B receptor 1; baculovirus IAP repeat-containing 6 protein; and KST50, a novel membrane protein (data not shown). Animals homozygous for insertions in agrin and netrin-1 also die at birth of neurologic defects that have been described elsewhere[26,39].

Insertions causing postnatal phenotypes

We observe postnatal defects in seven lines, resulting in lethality in two cases. An insertion in *Adam23* causes severe tremors and ataxia in homozygous pups (Fig. 3*r*) and death by 2 weeks of age. *Adam23* encodes a member of the metalloproteinase-disintegrin family that regulates cell adhesion through binding to integrins[40]. Based on reporter gene expression, *Adam23* is highly expressed in Purkinje cells in the cerebellum (Fig. 3*s*) and in the basal ganglia[23], areas that are important in feedback control of movement. Mice homozygous for JST185, an insertion in a predicted multiple membrane-spanning protein with tetratricopeptide repeat domains, generally die at birth of respiratory distress; a few pups, however, survive beyond weaning and gradually develop a severe lung pathology (J.B and W.C.S., manuscript in preparation). By 8–10 weeks, homozygous adults show signs of labored breathing, and histologic sections of the lungs reveal a dramatic increase in elastin deposition, fibrosis, and inflammation (Fig. 3*t*,*u*).

Five other insertions are viable but display subtle adult phenotypes. These include mammary branching and lactation defects (protein tyrosine phosphatase receptor type F), a coat color phenotype (attractin), and nervous system wiring defects (neuropilin 2, Eph receptor A4, and semaphorin 6a), two of which were identified in a concurrent screen[23]. We expect that many of the viable insertions will prove to have subtle defects that would be revealed using more comprehensive phenotyping protocols akin to those developed for the analysis of ENU-induced mutations [9,10].

Discussion

In this study, we used the secretory-trap approach to generate a large number of insertional mutations in genes encoding cell surface proteins and carried out a phenotypic analysis of 60 insertional mutations in mice. Our work demonstrates that the secretory-trap vector can access all classes of proteins targeted to the secretory pathway and that insertions in these genes can be efficiently transmitted to the germ line of mice. Most impor-

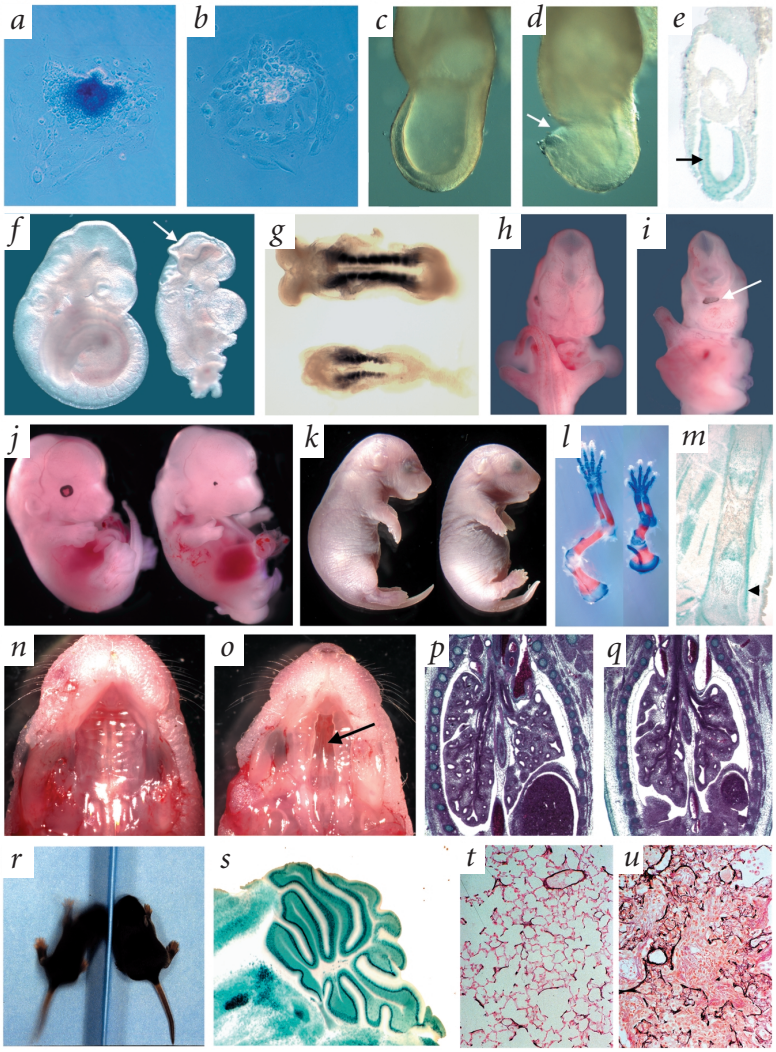

Fig. 3 Survey of embryonic, neonatal, and postnatal phenotypes. ***a***,***b***, Site-1 protease. Four-day outgrowth cultures of wildtype (*a*) and mutant (*b*) blastocysts stained for *Oct4* expression showing the absence of inner cell mass cells in the mutant embryos. ***c***–***e***, Spint2. E7.5 mutant epiblasts (*d*) show abnormal clefting (arrow) near the embryonic/extra-embryonic junction (compare with wild type, *c*). *e*, Spint2 is expressed in the embryonic ectoderm (arrow) of E7.5 embryos (7 μm section of βgal-stained embryo). ***f***,***g***, KST27 (novel, predicted multi-TM protein). E9.5 mutant embryos (*f*, right; compare with wild- type, left) fail to turn properly, show severe defects in the developing CNS (arrow shows an open neural tube at mid/hindbrain level), and form smaller, disorganized somites, as revealed by *Mox1* expression at E8.5 (*g*, bottom; compare with wild type, top, ventral view, anterior left). ***h***,***i***, Ext1. Mutant embryos (*i*; compare with wild type, *h*) show variable loss of ventral midline structures, often leading to cyclopia (*i*; arrow shows the single eye). ***j***, Lrp6. Lateroventral view of a E12.5 mutant embryo (right; compare with wild type, left) showing severe fore- and hindlimb defects, no tail, spina bifida, craniofacial defects, and microophthalmia. ***k***–***m***, KST245 (novel, predicted TM protein). Mutant pups at birth show a shortening of the limbs (*k*, right; compare with wild type, left). A skeletal prep of mutant forelimb shows a reduction in the length of the long bones of the limb (*l*, right; compare with wild type, left). *m*, KST245 is expressed in the periosteum (arrowhead), as well as in the cartilage of the developing E16.5 tibia (30 μm section stained for βgal activity). ***n***–***q***, PST9 (novel, predicted TM protein). View of the roof of the mouth of wild-type (*n*) and mutant (*o*) neonates shows a full cleft of the secondary palate in the mutant (arrow). Histologic section of E14.5 wild-type (*p*) and mutant (*q*) embryos shows delayed lung development in the mutant. ***r***,***s***, Adam23. *r*, Wild-type (right) and mutant (left) pups at P12. The mutant animal is noticeably smaller than its littermate, is ataxic and displays a strong tremor (resulting in a blurred image in this 1-s exposure). *s*, Adam23 is strongly expressed in Purkinje cells in the cerebellum (200 μm section stained for βgal activity). ***t***,***u***, JST185 (novel multi-TM protein). Histologic sections of adult lung stained for elastin. A comparison of 10-week-old wild-type (*t*) and mutant (*u*) lungs shows a dramatic increase in elastin (black staining), inflammation, and fibrosis. E, embryonic day; TM, transmembrane.

tantly, these insertions effectively mutate the trapped gene to create null or severe hypomorphic alleles. Recessive phenotypes were observed in 40% of the mutant lines, causing lethality over a range of developmental stages as well as subtle phenotypes in adult mice. Our results provide the first clear demonstration that gene-trapping is an efficient tool for the functional analysis of genes in mice and is not limited to the study of genes that act early in development.

Gene-trap vector design is critical for optimal results

We identified two features of the technology that were essential to the success of our screen. In theory, the pool of genes accessible to gene-trapping is limited by the requirement for some level of expression in ES cells to confer drug resistance. The sensitivity of the drug selection is therefore of paramount importance for trapping genes expressed at low levels in ES cells. One essential improvement we made was to correct a missense mutation in the β-geo reporter to restore wildtype neomycin phosphotransferase activity[13]. Consequently, we are able to recover G418-resistant colonies that express the trapped gene in ES cells at levels below detection by northern blot analysis. Vectors based on this 'wild-type' β-geo reporter should permit access to a larger pool of target genes including those which function at later stages in development.

Prescreening cell lines for insertions that splice efficiently to coding sequences is also critical for ensuring the optimal generation of mutant alleles. In a previous study, we described two classes of nonproductive secretory-trap events that do not appear to be mutagenic: insertions that are inefficiently spliced and those which splice into the 5′ UTR of the trapped gene[24]. In contrast, our present analysis of 60 productive insertions (that is, properly spliced to coding exons of the trapped gene) shows that one-third of these events cause lethal phenotypes. Taken together with our molecular analysis of several mutant lines and the phenotypic similarities between our insertion alleles and the corresponding targeted null mutations, we conclude that the secretory trap insertions that disrupt coding sequence are highly mutagenic.

Gene-based screens

Gene trapping combines the advantages of gene-based approaches and classic forward genetics. We have shown that a large number of insertional mutations can be isolated in ES cells, characterized by 5′ RACE and efficiently transmitted to the germ line of mice. Thus, gene-trapping provides a rapid method for addressing the function of known genes with presumed biochemical properties.

We identified essential functions for nine known genes with defined or presumed biochemical activities. For several of these genes, a comparison of the mutant phenotypes with those of genes in the same biochemical pathways is highly informative. For example, the site-1 protease enzyme is required for the cleavage of sterol response element binding proteins (Srebp1 and -2) to release these transcription factors to the nucleus in response to a low sterol level[32]. The insertion in the gene encoding site-1 protease results in a failure to form the epiblast in early embryos. This early phenotype is more severe than that caused by the knockout of Srebp1, in which over half the homozygous embryos survive past E11 and appear normal, and of Srebp2, which is lethal late in embryonic development[41]. Mutation of site-1 protease presumably inhibits the function of both Srebp1 and Srebp 2, which may partially compensate for one another in the single knockouts. Alternatively, the increased severity of the phenotype could reflect a wider function for site-1 protease.

A crucial role in early development is also revealed for another enzyme, serine protease inhibitor, Kunitz type 2 (*Spint2*), a negative regulator of HGF signaling[33]. Null mutations in the genes encoding HGF itself or its receptor, c-met, cause lethality by E14.5 (refs. 42–44). In contrast, *Spint2* mutant embryos die by E7.5 with obvious morphologic defects in the epiblast, suggesting that the hyperactivation of HGF has a more deleterious effect on embryonic development than the loss of HGF activity.

Ext1 encodes a glycosyltransferase involved in the synthesis of heparan sulfate. Mutations in *tout-velu*, the Drosophila homolog of *Ext1*, phenocopy hedgehog segment-polarity defects, and *tout-velu* has been shown to be required for the diffusion of the Hedgehog morphogen[45,46]. Mutants homozygous for the insertion in *Ext1* (Fig. 3*h*,*i*) display ventral midline defects in the central nervous system that are reminiscent of defects in Sonic hedgehog mutant mice[47], providing support for a conserved role of EXT1 in regulating hedgehog signaling. Interestingly, this phenotype differs from that of a targeted null mutation in which embryos fail to gastrulate and lack organized mesoderm and extra-embryonic tissues[31]. The reasons for this discrepancy are not clear. Although wildtype transcripts are not detected in embryos homozygous for our insertion allele, it is possible that the fusion protein generated by the insertion in *Ext1* retains some activity in the endoplasmic reticulum.

Finally, the phenotype of the insertion in *Adam23* is very similar to that reported for the knockout of the highly related gene, *Adam22* (ref. 48). These two genes may therefore function together in a nonredundant fashion, possibly in the cerebellum, where both are highly expressed. In addition, ADAM23 has been shown to promote cell adhesion through the binding of integrin-αvβ3 (ref. 40). The knockout of the integrin-β3 gene is, however, viable[49], suggesting that ADAM23 function can also be mediated through other binding partners.

Phenotype-driven screens

Gene-trapping is particularly well suited to the study of novel genes where there is no incentive or rationale for generating a targeted mutation. In addition, since a reasonably large number of genes can be analyzed for function using this approach, phenotype-driven screens are feasible and will uncover functions for known genes that could not have been predicted *a priori*. For example, the spectrum of phenotypes (Fig. 3*j*) caused by an insertion in the low-density lipoprotein receptor-related protein, Lrp6, revealed an unexpected role for this membrane protein in transducing multiple Wnt signals[28].

The power of the secretory-trap approach is best illustrated by the phenotypic analysis of four genes encoding novel transmembrane proteins that display specific developmental defects affecting gastrulation (KST27), bone growth (KST245), palate formation (PST9), and lung maturation (PST9, JST185). These proteins define new gene families in mammals (Fig. 1), three of which have predicted homologs in invertebrates and one of which, PST9, appears to be unique to vertebrates. A comparison between phenotypes caused by these insertional mutations and a growing collection of characterized mouse mutants is useful in generating testable hypotheses regarding the roles of these novel proteins in specific pathways or biologic processes. We note, for example, striking similarities between the phenotypes of PST9, KST245, and JST185 mutant mice and those with mutations in transforming growth factor (TGF) β3 (ref. 50), aggrecan[51], and double mutations in fibroblast growth factor (FGF) receptors 3 and 4 (ref. 52), respectively.

A resource of gene-trap mutations

The value of gene-trapping extends well beyond simply generating null mutations in mice. First, the βgal reporter gene provides reliable, high-resolution expression information for the trapped gene, obviating the need to perform time-consuming expression studies. Second, the inclusion of other reporter genes in the vector provides additional versatility and can be tailored to specific

biologic questions. The secretory vector, for example, was modified to include an axonal marker, placental alkaline phosphatase, and was used successfully in a concurrent study to screen for mutations in cell surface receptors that cause brain wiring defects[23]. Thus, gene-trapping provides an efficient and versatile tool for the functional analysis of genes in mammals, taking full advantage of the wealth of sequence and mapping information from the human and mouse genome projects.

Having validated the secretory-trap approach for the functional analysis of secreted and membrane-spanning proteins in mice, we are continuing our efforts to build a library of insertional mutations in this important class of proteins. Because ES cells can be frozen indefinitely and easily disseminated within the scientific community, a library of insertional mutations in ES cells solves the archiving and distribution problems inherent in preserving mouse strains. ES cell lines and mice are freely available to the academic community, and complete information on this resource is available on the WorldWide Web (http://www.genetrap.org). In parallel, we are extending our screen to include nonsecreted genes through the use of more conventional gene-trap vectors that are optimally designed to enrich for insertions in coding sequences of the target gene (BayGenomics consortium, http://baygenomics.ucsf.edu).

Methods

Secretory-trap vectors. The pGT1.8TM vector[13] was engineered in each of the three reading frames (designated pGT0tm, pGT1tm, and pGT2tm) by deleting 135 bp from the 3′ end of the mouse engrailed 2 (*En2*) exon sequence and introducing *Bgl*II linkers. The vectors pGT0TMpfs, pGT1TMpfs, and pGT2TMpfs are based on pGT0-2tm and contain the following additional elements: 1) an IRES/placental alkaline phosphatase reporter cassette[23] between β-geo and the SV40 polyA signal, 2) FRT sites on either site of β-geo, and 3) a *Sal*I site in place of the *Hind*III site at the 5′ end of the *En2* splice acceptor to permit linearization of the vector prior to electroporation.

ES cell culture and electroporation. Two feeder-independent mouse ES cell lines (CGR8.8 and E14Tg2a.4, derived from the 129/Ola strain) were electroporated with the secretory-trap vector DNA as described[24]. Colonies were picked, plated in duplicate wells, and assayed for βgal activity by X-gal staining[24]. Cell lines exhibiting the "secretory" pattern[13] were plated in duplicate wells for RNA isolation and for freezing.

5′ RACE–PCR and direct sequencing. 5′ RACE was carried out using total RNA from ES cell lines as described[5,24]. Oligonucleotides for 5′ RACE of ES cell lines generated with the pGT0-2TMpfs vectors were as follows. First strand cDNA 5′–CGCCAGGGTTTTCCCAGT (in lacZ); second strand cDNA (T-tailed anchor oligo) 5′–GGTTGTGAGCTCTTCTAGATGGT$_{(17)}$; first PCR 5′–GGTTGTGAGCTCTTCTAGATGGT (anchor oligo) and 5′–GATCATCGCTCCCATATATGAG (at CD4/lacZ junction); second PCR 5′–biotin–GGTTGTGAGCTCTTCTAGATGG (biotinylated anchor oligo) and 5′–AATAGGATGCAGAGCCCCGT (in CD4); sequencing primer 5′–Cy5–AAGAAGGAGCCTTCTCTGCC (in CD4). Cycle-sequencing reactions using dye primer chemistry (USB kit #US78500) were run on an ALF Express automated fluorescent sequencing machine.

Bioinformatics. Sequence tags generated by 5′ RACE were searched against the nonredundant (nr) and expressed sequence tag (dbest) databases at the NCBI using BLAST. Predicted full-length ORFs were assembled for the presumed human orthologs of novel mouse genes using BLAST and GENSCAN programs. Signal sequences, transmembrane segments, defined protein domains, and internal repeats were identified using the TMPred, SMART and Pfam programs.

ES cell injections, breeding and genotyping. 129/Ola ES cell-derived chimeras were generated by the injection of C57BL/6 blastocyts. The resulting male chimeras were bred to C57BL/6 females to test for germline transmission. Agouti offspring were genotyped by the dot-blot hybridization of tail biopsies[53] using vector DNA lacking the *En2* sequences as a probe. Animals heterozygous for the insertion were backcrossed one generation to C57BL/6 animals prior to setting up intercrosses. Intercross litters at weaning were genotyped by dot-blot hybridization in which heterozygous and homozygous animals were distinguished by signal strength. Homozygous viable lines were confirmed by breeding, and homozygous lethal lines were further analyzed by timed father:daughter matings of heterozygous intercross animals. Embryos were genotyped by X-gal staining or dot-blot hybridization. The target gene for each of the lethal lines was confirmed by 5′ RACE of RNA prepared from the tissues of heterozygous animals.

RNase protection assays. RNase protections were carried out as described[27]. cDNA probes spanning the site of insertion were as follows: *Epha4*, a 550 bp *Bam*HI–*Bgl*II fragment; *Sema6A*, a 470 bp *Xho*I–*Pst*I fragment; and *Notch3*, a 337 bp fragment generated by RT–PCR. All probes were cloned into pBluescript II KS (Stratagene) for *in vitro* transcription.

Blastocyst outgrowth culture and *in situ* hybridization. Blastocysts were flushed from the uterus at E3.5 and were cultured in ES media on gelatin for 3–4 days before being stained with X-gal to genotype. The whole-mount *in situ* hybridization of mouse embryos was carried out essentially as described[54].

Skeletal preparation and histology. Skeletal preparations were performed by standard methods[55]. X-gal staining was carried out on whole embryos (prior to E10.5) and sectioned in wax, or carried out directly on frozen tissue sections (after E12.5). Histology of wax tissue sections was performed as described[28]. For elastin staining of JST185 lung, sections were dewaxed in xylene and 100% ethanol, stained in Miller's elastin for 2 h, rinsed in ethanol and water, and stained in Van Geison's stain for 5 min. Sections were dehydrated and mounted using DPX.

Acknowledgments

We thank A. Smith for providing us with the feeder-independent ES cell lines and the Oct4 probe, C. Wright for the Mox1 probe, J. Gladden for his help on the PST9 phenotype analysis, Sheila Avery for help with formatting Fig. 1, and A. Jeske and S. Elson for technical support. Funding for this project was provided by grants to W.C.S. and M.T-L. by a Program in Genomics Applications from the National Heart, Lung, and Blood Institute (NHLBI), grants to W.C.S. from the BBSRC (UK), the Chicago Community Trust, the NICHD, and the March of Dimes, and to M.T.-L. from the NIMH. The ongoing screen is also supported by a Program in Genomics Applications from the NHBLI. K.J.M. was supported by a fellowship from the Jane Coffin Childs Memorial Fund for Medical Research, O.G.K. by a fellowship from the NSF, L.V.G by a fellowship from the Helen Hay Whitney Foundation.. P.A.L. was a Howard Hughes Medical Institute Fellow, and P.T. is a Burroughs Wellcome Fellow of the Life Sciences Research Foundation. X.L. is a postdoctoral associate and M.T.L. an Investigator of the Howard Hughes Medical Institute. W.C.S. is a 1998 Searle Scholar.

Received 6 March; accepted 24 May 2001.

1. Gossler, A., Joyner, A.L., Rossant, J. & Skarnes, W.C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463–465 (1989).
2. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1515–1523 (1991).
3. Skarnes, W.C., Auerbach, B.A. & Joyner, A.L. A gene trap approach in mouse embryonic stem cells: the lacZ reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev.* **6**, 903–918 (1992).
4. von Melchner, H. *et al*. Selective disruption of genes expressed in totipotent embryonal stem cells. *Genes Dev.* **6**, 919–927 (1992).
5. Townley, D.J., Avery, B.J., Rosen, B. & Skarnes, W.C. Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. *Genome Res.* **7**, 293–298 (1997).
6. Hicks, G.G. *et al*. Functional genomics in mice by tagged sequence mutagenesis. *Nature Genet.* **16**, 338–344 (1997).
7. Zambrowicz, B.P. *et al*. Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611 (1998).
8. Wiles, M.V. *et al*. Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. *Nature Genet.* **24**, 13–14 (2000).
9. Nolan, P.M. *et al*. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nature Genet.* **25**, 440–443 (2000).
10. Hrabe de Angelis, M. *et al*. Genome-wide, large-scale production of mutant mice

by ENU mutagenesis. *Nature Genet.* **25**, 444–447 (2000).
11. Wurst, W. *et al.* A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. *Genetics* **139**, 889–899 (1995).
12. Forrester, L.M. *et al.* An induction gene trap screen in embryonic stem cells: identification of genes that respond to retinoic acid in vitro. *Proc. Natl Acad. Sci. USA* **93**, 1677–1682 (1996).
13. Skarnes, W.C., Moss, J.E., Hurtley, S.M. & Beddington, R.S. Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl Acad. Sci. USA* **92**, 6592–6596 (1995).
14. Tate, P., Lee, M., Tweedie, S., Skarnes, W.C. & Bickmore, W.A. Capturing novel mouse genes encoding chromosomal and other nuclear proteins. *J. Cell Sci.* **111**, 2575–2585 (1998).
15. Gasca, S., Hill, D.P., Klingensmith, J. & Rossant, J. Characterization of a gene trap insertion into a novel gene, cordon-bleu, expressed in axial structures of the gastrulating mouse embryo. *Dev. Genet.* **17**, 141–154 (1995).
16. Voss, A.K., Thomas, T. & Gruss, P. Compensation for a gene trap mutation in the murine microtubule-associated protein 4 locus by alternative polyadenylation and alternative splicing. *Dev. Dyn.* **212**, 258–266 (1998).
17. McClive, P. *et al.* Gene trap integrations expressed in the developing heart: insertion site affects splicing of the PT1-ATG vector. *Dev. Dyn.* **212**, 267–276 (1998).
18. Sam, M. *et al.* Aquarius, a novel gene isolated by gene trapping with an RNA-dependent RNA polymerase motif. *Dev. Dyn.* **212**, 304–317 (1998).
19. DeGregori, J.D. *et al.* A murine homolog of the yeast RNA1 gene is required for postimplantation development. *Genes Dev.* **8**, 265–276 (1994).
20. Chen, Z., Friedrich, G.A. & Soriano, P. Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev.* **8**, 2293–2301 (1994).
21. Friedrich, G.A., Hildebrand, J.D. & Soriano, P. The secretory protein Sec8 is required for paraxial mesoderm formation in the mouse. *Dev. Biol.* **192**, 364–374 (1997).
22. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
23. Leighton, P.A. *et al.* Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**, 174–179 (2001).
24. Skarnes, W.C. Gene trapping methods for the identification and functional analysis of cell surface proteins. *Methods Enzymol.* **328**, 592–615 (2000).
25. Paine-Saunders, S., Viviano, B.L., Zupicich, J., Skarnes, W.C. & Saunders, S. Glypican-3 controls cellular responses to BMP4 in limb patterning and skeletal development. *Dev. Biol.* **225**, 179–187 (2000).
26. Burgess, R.W., Skarnes, W.C. & Sanes, J.R. Agrin isoforms with distinct amino termini: differential expression, localization, and function. *J. Cell Biol.* **151**, 41–52 (2000).
27. Chen, H. *et al.* Neuropilin-2 regulates the development of selective cranial and sensory nerves and hippocampal mossy fiber projections. *Neuron* **25**, 43–56 (2000).
28. Pinson, K.I., Brennan, J., Monkley, S., Avery, B.J. & Skarnes, W.C. An LDL receptor-related protein mediates Wnt signaling in mice. *Nature* **407**, 535–538 (2000).
29. Yeo, T.T. *et al.* Deficient LAR expression decreases basal forebrain cholinergic neuronal size and hippocampal cholinergic innervation. *J. Neurosci. Res.* **47**, 348–360 (1997).
30. Cormick, C. *et al.* The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate. *Nature Genet.* **19**, 158–161 (1998).
31. Lin, X. *et al.* Disruption of gastrulation and heparan biosynthesis in EXT1-deficient mice. *Dev. Biol.* **224**, 299–311 (2000).
32. Brown, M.S. & Goldstein, J.L. A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. *Proc. Natl. Acad. Sci. USA* **96**, 11041–11048 (1999).
33. Kawaguchi, T. *et al.* Purification and cloning of hepatocyte growth factor activator inhibitor type 2, a Kunitz-type serine protease inhibitor. *J. Biol. Chem.* **272**, 27558–27564 (1997).
34. George, E.L., Georges-Labouesse, E.N., Patel-King, R.S., Rayburn, H. & Hynes, R.O. Defects in mesoderm, neural tube and vascular development in mouse embyros lacking fibronectin. *Development* **119**, 1079–1091 (1993).
35. Jones, S.N., Roe, A.E., Donehower, L.A. & Bradley, A. Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53. *Nature* **378**, 206–208 (1995).
36. Ahn, J. *et al.* Cloning of a putative tumour suppressor gene for hereditary multiple exostoses (EXT1). *Nature Genet.* **11**, 137–143 (1995).
37. Brown, S.D. *et al.* Isolation and characterization of LRP6, a novel member of the low density lipoprotein receptor gene family. *Biochem. Biophys. Res. Comm.* **248**, 879–888 (1998).
38. Cano-Gauci, D.F. *et al.* Glypican-3-deficient mice exhibit developmental overgrowth and some of the abnormalities typical of Simpson–Golabi–Behmel syndrome. *J. Cell Biol.* **146**, 255–264 (1999).
39. Serafini, T. *et al.* Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001–1014 (1996).
40. Cal, S., Freije, J.M., Lopez, J.M., Takada, Y. & Lopez-Otin, C. ADAM23/MDC3, a human disintegrin that promotes cell adhesion via interaction with the alphavbeta3 integrin through an RGD-independent mechanism. *Mol. Biol. Cell* **4**, 1457–1469 (2000).
41. Shimano, H. *et al.* Elevated levels of SREBP-2 and cholesterol synthesis in livers of mice homozygous for a targeted disruption of the SREBP-1 gene. *J. Clin. Invest.* **100**, 2115–2124 (1997).
42. Schmidt, C. *et al.* Scatter factor/hepatocyte growth factor is essential for liver development. *Nature* **373**, 699–702 (1995).
43. Uehara, Y. *et al.* Placental defect and embryonic lethality in mice lacking hepatocyte growth factor/scatter factor. *Nature* **373**, 702–705 (1995).
44. Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A. & Birchmeier, C. Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. *Nature* **376**, 768–771 (1995).
45. The, I., Bellaiche, Y. & Perrimon, N. Hedgehog movement is regulated through tout velu-dependent synthesis of a heparan sulfate proteoglycan. *Mol. Cell* **4**, 633–639 (1999).
46. Bellaiche, Y., The, I. & Perrimon, N. Tout-velu is a Drosophila homologue of the putative tumour suppressor EXT-1 and is needed for Hh diffusion. *Nature* **394**, 85–88 (1998).
47. Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**, 407–413 (1996).
48. Sagane, K., Yamazaki, K., Mizui, Y. & Tanaka, I. Cloning and chromosomal mapping of mouse ADAM11, ADAM22 and ADAM23. *Gene* **236**, 79–86 (1999).
49. Hodivala-Dilke, K.M. *et al.* Beta3-integrin-deficient mice are a model for Glanzmann thrombasthenia showing placental defects and reduced survival. *J. Clin. Invest.* **103**, 229–238 (1999).
50. Kaartinen, V. *et al.* Abnormal lung development and cleft palate in mice lacking TGF-beta 3 indicates defects of epithelial-mesenchymal interaction. *Nature Genet.* **11**, 415–421 (1995).
51. Watanabe, H. *et al.* Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene. *Nature Genet.* **7**, 154–157 (1994).
52. Weinstein, M., Xu, X., Ohyama, K. & Deng, C.X. FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development* **125**, 3615–3623 (1998).
53. Brennan, J. & Skarnes, W.C. Gene trapping in mouse embryonic stem cells. *Methods Mol. Biol.* **97**,123–138 (1999).
54. Wilkinson, D.G. & Nieto, M.A. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**, 361–367 (1993).
55. Parr, B.A. & McMahon, A.P. Dorsalizing signal Wnt-7a is required for normal polarity of D-V and A-P axes of mouse limb. *Nature* **374**, 350–353 (1995).