Identification and validation of PDGF transcriptional targets by microarray-coupled gene-trap mutagenesis

Weisheng V Chen[1,2], Jeffrey Delrow[3], Philip D Corrin[1], Jason P Frazier[1] & Philippe Soriano[1]

We developed a versatile, high-throughput genetic screening strategy by coupling gene mutagenesis and expression profiling technologies. Using a retroviral gene-trap vector optimized for efficient mutagenesis and cloning, we randomly disrupted genes in mouse embryonic stem (ES) cells and amplified them to construct a cDNA microarray. With this gene-trap array, we show that transcriptional target genes of platelet-derived growth factor (PDGF) can be efficiently and reliably identified in physiologically relevant cells and are immediately accessible to genetic studies to determine their *in vivo* roles and relative contributions to PDGF-regulated developmental processes. The same platform can be used to search for genes of specific biological relevance in a broad array of experimental settings, providing a fast track from gene identification to functional validation.

Targeted mutagenesis studies using null alleles, conditional deletions and point mutations of ligand and receptor genes have shown that PDGF signaling has essential roles in diverse developmental contexts[1,2], but crucial downstream effectors have not yet been identified in PDGF-responsive cells and tissues. Conventionally, transcriptional target genes are identified by stimulating quiescent cells in culture and then monitoring expression changes using differential display[3], serial analysis of gene expression[4] or microarrays[5]. Some PDGF-responsive genes identified in these screens are involved in specific regulated cellular responses, including proliferation, adhesion, chemotaxis and survival[6–8]. One limitation of these RNA-based methods is that physiological validation of the identified targets requires independent genetic approaches, which are typically laborious and time-consuming.

To systemically explore PDGF targets and their *in vivo* functions, we carried out gene-trap mutagenesis, in which a promoterless reporter is introduced into ES cells to randomly tag and mutate genes, integrating expression monitoring, molecular cloning and functional analysis[9–11]. Transcriptionally responsive genes can be identified by induction trapping based on the altered expression of the gene-trap reporter on stimulation[12–15]. In our initial attempts, however, we did not identify any PDGF-regulated genes, probably owing to the low abundance of PDGF receptors in ES cells. In addition, because these

cells are not normal targets of growth factor signaling in the developing embryo, responsive genes identified in this manner may have little *in vivo* relevance.

Therefore, whereas subsequent functional dissection necessitates gene trapping to be done in ES cells, other physiologically relevant responsive cells are required to identify PDGF transcriptional targets. We circumvented this limitation by coupling gene trapping with microarray technology. Using a retroviral vector optimized for efficient cloning, we amplified trapped transcripts from individual ES cell clones to construct a cDNA microarray. The availability of this
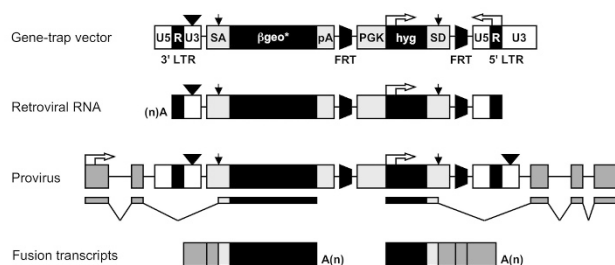

Figure 1 The ROSAFARY gene-trap vector. The gene-trap vector contains a promoter trap module (SAβgeo*pA) and a poly-A trap module (PGKhygSD) in the same orientation in the retroviral backbone between two long terminal repeats (LTRs). The poly-A trap module is flanked with FRT sites, which can be removed with Flp recombinase. The viral promoter and enhancer are both deleted in the U3 region of the 3′ long terminal repeat (inverted triangle) to avoid potential interference with gene expression in the resulting provirus. The splice acceptor (SA) and the splice donor (SD) are derived from the intron 1–exon 2 and the exon 1–intron 1 boundaries of the adenoviral type 2 major late transcription unit, respectively. The intron-exon junction of each splice site is indicated with an arrow. After inserting into an intron of an endogenous gene at a permissive site and in the correct orientation, the promoter trap module and the poly-A trap module can be activated to form fusion transcripts with the 5′ or 3′ exons, respectively. Only the exon portion of each splice site is included in fusion transcripts. Gene-trapped clones were obtained with the promoter trap module by G418 selection, and trapped genes were cloned with the poly-A trap module by 3′ RACE for higher cloning efficiency and longer sequences. Boxed arrows indicate the transcription start and direction.

[1]Program in Developmental Biology, Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [2]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [3]Genomics Resource, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. Correspondence should be addressed to P.S. (psoriano@fhcrc.org).

gene-trap array allowed us to identify PDGF targets in physiologically relevant cells and to generate mutant mice directly from the corresponding gene-trapped ES cell clones. The same platform can be used to simultaneously identify and validate differentially expressed genes in a broad array of experimental settings, providing higher specificity and efficiency than conventional gene-trap mutagenesis screens.

RESULTS

A microarray-compatible gene-trap vector

We previously developed a Mo-MuLV-based retroviral gene-trap vector, ROSAβgeo, which transduces a *lacZ-neo* fusion gene downstream of an adenoviral splice acceptor[10]. On integration into an intron, the promoterless reporter functions as an artificial 3′ terminal exon to intercept and terminate transcription from the endogenous promoter. ROSAβgeo and ROSAβgeo*, which contains a more sensitive reporter[14], are used widely and are highly mutagenic, with most gene-trap insertions analyzed occurring at the 5′ end of genes. This may be attributable to preference of retrovirally mediated insertions in transcriptional start regions[16], strength of the adenoviral splice acceptor and instability of longer fusion transcripts or proteins. One drawback of the insertion site preference is that cloning of the trapped transcripts with 5′ RACE[17] can be technically challenging due to the higher G+C content in start regions[18], and cDNA fragments amplified are often too short to be suitable for expression studies.

To increase cloning efficiency and obtain longer cDNAs for array fabrication, we constructed a new retroviral gene-trap vector, called reverse orientation splice acceptor for array (ROSAFARY), by incorporating a poly-A trap cassette into ROSAβgeo* (**Fig. 1**). The poly-A trap cassette comprises a PGK promoter–driven hygromycin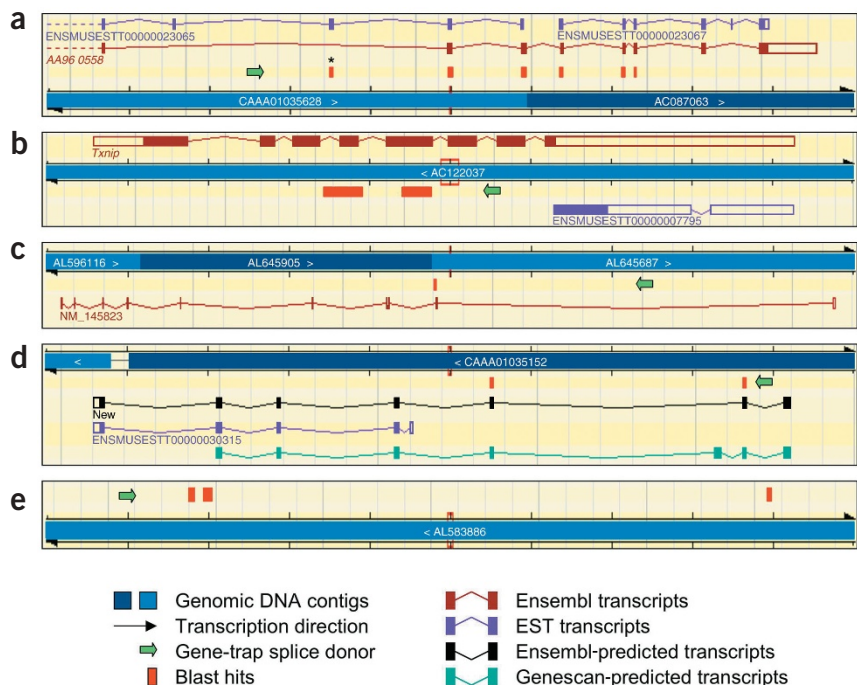 gene and an adenoviral splice donor and can thus function as an artificial 5′ terminal exon to initiate transcription from the insertion site. Gene-trap events are obtained with the promoter trap module by G418 selection alone, and we used the poly-A trap module only to efficiently amplify longer, 3′-enriched cDNAs of trapped genes by 3′ RACE.

Cloning and analysis of trapped transcripts

We used a high-throughput 3′ RACE procedure to isolate trapped transcripts from 2,880 individual gene-trapped ES cell clones, of which 2,160 (75%) yielded high quality products and were reamplified in preparative PCR reactions for array printing. To assess the spectrum of trapped genes, we analyzed RACE products from 607 randomly selected clones (**Supplementary Table 1** online). Homology searching showed that 440 gene-trap sequence tags (72%) matched known mouse genes, 60 (10%) matched mouse expressed-sequence tags (ESTs) and 107 (18%) did not match any known mouse cDNA. The trapped genes were distributed on all chromosomes except the Y, suggesting that a large number of genes across the genome could be accessed by this vector. Some loci seem to be more readily trapped than others, as also found by other large-scale gene trapping efforts using different vectors[19–22], and 40 genes were trapped more than once. This phenomenon may be due to recombination hot spots, more permissive integration sites available in larger genes or a limited pool of genes that could be trapped in the genome. Most trapped sequences that matched known genes aligned with the sense strand (**Fig. 2a**), but 14 (3%) aligned with the reverse strand of the corresponding transcript, which might indicate trapping of new, naturally occurring antisense genes (**Fig. 2b**).

Of the 107 new transcripts identified, 91 mapped to the available sequences on the mouse genome. Several sequences matched

Figure 2 Genome mapping of selected gene-trap sequence tags. (**a**) A trapped sequence aligned with the sense strand of a known gene (*AA960558*). The BLAST matches have a splicing pattern that correlates with annotated exons of the transcript. The slight differences in exon boundaries result from homologies in flanking sequences. An alternative exon (marked with an asterisk) is included in the trapped transcript, which corresponds to an EST transcript (ENSMUSESTT00000023065) of the same cluster. (**b**) A trapped sequence aligned with the antisense strand of a known gene (*Txnip*). The splicing pattern indicated by BLAST matches differs considerably from that of the sense transcript, suggesting distinct exon boundaries and thus trapping in an antisense gene. A neighboring EST transcript (ENSMUSESTT00000007795) is also antisense to *Txnip*, which may belong to the same gene that encodes the trapped transcript. (**c**) A trapped new sequence aligned with an annotated intron of a known gene (NM_145823), which might have originated from alternative or cryptic splicing. (**d**) A trapped new sequence aligned with new exons predicted by both Ensembl and Genescan. This may correspond to an insertion in the gene encoding the EST transcript ENSMUSESTT00000030315, which has identical 3′ splicing pattern to the computer predictions. (**e**) A trapped new sequence not aligned with any known or predicted exons but showing a splicing pattern. The precise positions of the gene-trap splice donors within introns are unknown.

annotated introns of known genes (**Fig. 2c**), probably representing previously undiscovered splice variants, although the possibility of cryptic splicing cannot be excluded. Other sequences overlapped with predicted exons (**Fig. 2d**). Most of the new transcripts aligned with regions without any known or predicted genes in the vicinity, and some had splicing patterns (**Fig. 2e**). These observations, and the microarray experiments (described below) showing expression of most of these transcripts, suggested that the transcripts probably represent bona fide new genes. Therefore, in addition to including many known genes or ESTs, the gene-trap array may identify new genes that are difficult to capture by conventional cDNA cloning.

We assessed the putative gene-trap insertion site in the endogenous transcript in 392 gene traps that occurred in known genes with full-length sequence available (**Fig. 3**). The trapping frequency gradually declined from 5′ to 3′, with most insertions (328, 83%) occurring at the 5′ untranslated region (UTR; 123, 31%) and the first half of the coding sequence (205, 52%). The actual insertion site may lie further upstream, owing to 5′ sequence quality trimming, under-representation of 5′ sequences in current databases and potential alternative splicing that could skip the immediate next exon. Therefore, most ROSAFARY insertions should result in little or no endogenous protein attached to the gene-trap reporter and probably constitute null alleles.

Evaluation of the 2K gene-trap array set

To evaluate the performance of the 2K gene-trap array set, we generated expression profiles of adult mouse brain and testis by paired hybridizations with fluorescently labeled cDNA targets (**Fig. 4a**). After filtering out low intensity and low quality data, a total of 1,705 (79%) genes remained. Pairwise comparison showed that 254 genes (15%) were differentially expressed by a factor of 2 or more, with 80 (5%) and 174
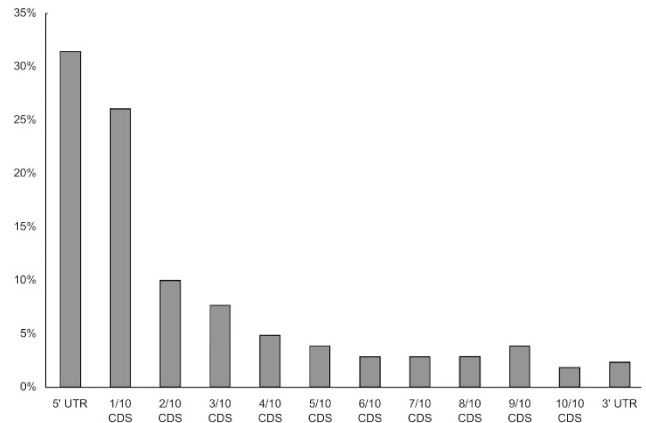

Figure 3 Insertion site preference. The gene-trap insertion site was determined by aligning each trapped sequence with the corresponding endogenous transcript, and the percentages of insertions in the 5′ UTR, each portion of the coding sequence (CDS) and the 3′ UTR were calculated and plotted. Most insertions occurred in the 5′ UTR and the beginning of coding sequence.

(10%) more highly expressed in testis or brain, respectively (**Fig. 4b**). Sequence analysis of selected differentially expressed genes identified several genes known to be abundant in one tissue or the other, as expected (data not shown). Among the 607 annotated genes, the proportion and intensity distribution of expressed known genes (327 of 440, 74%), ESTs (44 of 60, 73%) and new transcripts (73 of 107, 68%) were similar (**Fig. 4b**), experimentally supporting the idea that most new sequences were transcribed from genes not yet identified.
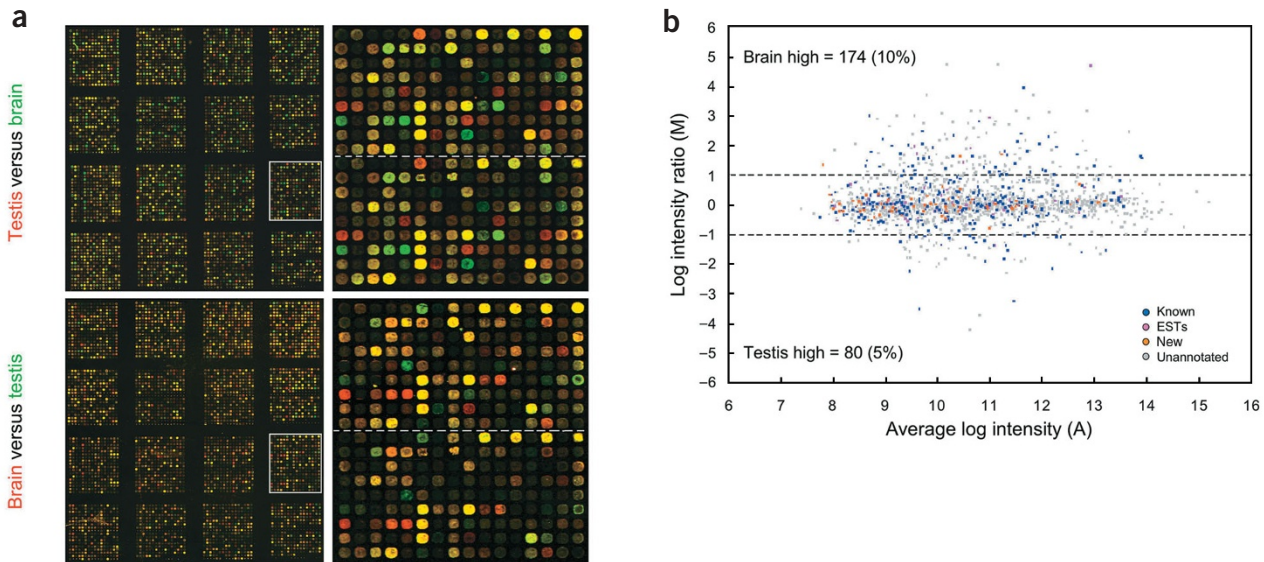

Figure 4 Expression profiling of adult brain and testis tissues with the 2K gene-trap array set. Cy3- (green) or Cy5-labeled (red) cDNA targets were prepared from total RNA isolated from the two tissues and hybridized in pairs onto arrays. (**a**) Global and enlarged views of array images of hybridizations with dye reversal as indicated. Each block contains duplicated spots of RACE products from the same clone arrayed evenly on both sides of the dotted line. Red or green spots indicated that the corresponding genes are more abundantly expressed in one tissue or the other, whereas yellow spots indicated that the expression levels are similar in both tissues. Signal intensities between duplicated spots and hybridizations were very similar. (**b**) Scatter plot analysis. For each gene, average signal intensities were calculated from duplicated spots of both hybridizations. Spots corresponding to sequence-annotated traps in known genes, ESTs and new transcripts are highlighted in different colors as indicated. The intensity distribution between categories was similar. M, $\log_2$ (R/G); A, $(\log_2 R + \log_2 G)/2$.

Gene-trap array screening for transcriptional targets

After confirming the functionality of ROSAFARY and the gene-trap array, we sought to use the platform to search for trapped PDGF targets. Although a number of established cell lines can be used for this purpose, we chose to use low-passage primary mouse embryonic fibroblasts (MEFs) to identify physiologically relevant targets. We first profiled transcriptional responses at multiple time points in wild-type MEFs by stimulating them with PDGF-BB, which interacts with all three forms of PDGF receptor dimers (αα, αβ and ββ). The pairwise comparison of control over reference ('same versus same') hybridizations showed a low level of technical variation. Accordingly, we first set the arbitrary cut-off to 1.5 times greater than baseline (>8 s.d.). Fifty-two genes were up- or downregulated within the 4-h observation window and clustered into different groups (**Fig. 5a**). Most genes showed a gradual change in expression kinetics, providing internal validation for differential expression observed at individual time points.

Next, we examined responses of mutant MEFs derived from PDGFRα- or PDGFRβ- deficient embryos to PDGF-BB after 1 h and 4 h. Because they have only homodimers of one receptor, profiling PDGF-BB responses in these mutant cells not only independently validates the general targets found in wild-type cells but also may identify genes preferentially activated through each receptor homodimer. To control for inherent variances in steady-state expression levels of target genes, we carried out paired hybridizations using cDNA targets prepared from stimulated and unstimulated cells of the same genotype. Hierarchical clustering of induced and repressed genes in MEFs of all three genotypes showed markedly similar expression patterns (**Fig. 5b**).

Although most genes showed the greatest change in expression in wild-type cells, suggesting an additive effect, several were more robustly induced or repressed in one or the other mutant. This phenomenon could be due to different baseline expression levels, interchip variance or noise associated with the low cut-off. PDGF responses in general seemed much weaker in PDGFRβ-deficient than in PDGFRα-deficient cells, prompting further investigation described below.

Identification and verification of target genes

Twenty-nine PDGF transcriptional targets were identified as being induced or repressed by a factor of more than 1.5 at two or more data points in the merged data set (**Fig. 5c** and **Supplementary Table 2** online). These include 25 unique genes of various functional categories, including transcriptional factors, metabolic enzymes, membrane proteins, cytoplasmic signaling molecules and cytoskeletal proteins. Some of these genes are regulated by PDGF or other growth factor signaling pathways, are involved in PDGF-regulated cellular responses or developmental processes, or are implicated in tumorigenesis commonly associated with PDGF overactivity (**Table 1**). Each pair of the four genes obtained from recurring gene-trap events (*Cdh11*, *Bteb1*, *Lmna* and *AA960558*) was clustered next to each other pair (**Fig. 5c**), showing high reproducibility and stringency. The induction or repression of most target genes was resistant to treatment with cycloheximide, indicating that they do not require additional protein synthesis and are probably involved in the primary PDGF transcriptional response (data not shown).

We carried out northern-blot analysis using an independent batch of wild-type MEFs, which verified that 23 of the 25 genes were induced or repressed by PDGF-BB with similar kinetics to those observed from the array analysis (**Fig. 6**). Expression of the other two genes was below the level of detection (**Table 1**). Because PDGF-BB binds to PDGFRβ with higher affinity than it does to PDGFRα, which might result in lower responses in PDGFRβ-deficient MEFs, we treated wild-type MEFs in parallel with PDGF-AA, which only recognizes the αα receptor homodimer. Most genes examined were responsive to this ligand, but the extent of induction or repression was generally much less than with PDGF-BB (**Fig. 6**). This was in agreement with the array results and preliminary observations that PDGF-BB stimulation of wild-type MEFs resulted in a higher level of MAPK activation than PDGF-AA treatment (data not shown). Therefore, our results indicated that most PDGF target genes are preferentially regulated by PDGFRβ in MEFs, consistent with evidence from kinase domain swapping studies that PDGFRβ, but not PDGFRα, can functionally substitute for the other receptor *in vivo*[23].

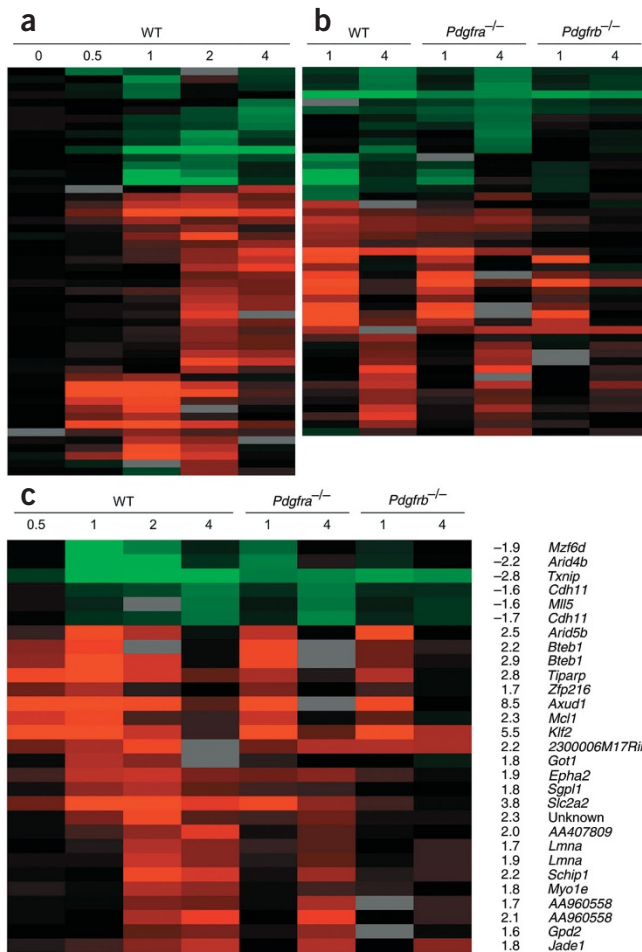
Figure 5 Hierarchical clustering of PDGF transcriptional target genes. (**a**) PDGF-BB transcriptional responses in wild-type (WT) MEFs. Only genes present at 75% or more data points and showing a change of expression by a factor 1.5 or more at one or more data points were selected. Genes were grouped based on the similarity of their expression kinetics in the time course. Red, induced; green, repressed; black, no change; gray, not expressed or low-quality data. Induction or repression was reproducible at neighboring time points (h). (**b**) PDGF-BB transcriptional responses in wild-type (WT), PDGFRα-deficient (*Pdgfra*−/−) and PDGFRβ-deficient (*Pdgfrb*−/−) MEFs. Genes were selected and clustered according to the same criteria as above. The expression kinetics and differences in levels of induction or repression between cells of different genotypes were reproducible. (**c**) Cluster analysis of PDGF-regulated genes using combined data and higher stringency. Only genes present at 75% or more data points and showing a change of expression by a factor 1.5 or more at two or more data points were selected. The highest relative change and sequence identity for each gene are indicated.

Table 1 PDGF transcriptional target genes in MEFs identified in gene-trap array screens

Gene symbol	Gene name, classification, function[a]	GenBank accession number	PDGF[b]
Arid4b	AT rich interactive domain 4B (Rbp1 like) Putative transcription factor with AT-rich interaction domain (ARID) Upregulated in cancers	NM_194262	Rep[c]
Arid5b	AT rich interactive domain 5B (Mrf1 like) Transcription factor with AT-rich interaction domain (ARID) Implicated in smooth muscle cell differentiation and proliferation KO causes retarded growth, abnormal development of reproductive and lymphoid organs	NM_023598	Ind
Axud1	AXIN1 up-regulated 1 Nuclear protein Upregulated by AXIN1 and TGFβ; down-regulated in cancers; implicated in apoptosis	NM_153287	Ind
Bteb1	Basic transcription element binding protein 1 Transcription factor with zinc finger and Sin3-interacting domain (SID) Upregulated by thyroid hormone (T3); regulates cell growth and gene transcription KO causes behavioral defects	NM_010638	Ind
Cdh11	Cadherin 11 Cadherin family of calcium-dependent cell adhesion molecules Upregulated by FGF2, EGF, TGFβ1, BMP2; downregulated by TGFα; implicated in cancers Restrains cranial neural crest migration and influences neural crest specification in Xenopus KO causes decrease in bone density	NM_009866	Rep
Epha2	Eph receptor A2 Eph family of receptor tyrosine kinases Downregulated by estrogen, myc, p53, E-cadherin, Hoxa1, Hoxb1; implicated in cancers Dominant negative expression results in abnormal neural crest cell migration in Xenopus Overexpression induces apoptosis and inhibits adhesion and migration KO shows bifurcated caudal notochord and short, kinky tail	NM_010139	Ind
AA407809	Expressed sequence AA407809 Enzyme involved in nucleotide transport and metabolism	NM_030724	Ind
AA960558	Expressed sequence AA960558 Pleckstrin homology (PH) domain containing protein Putatively involved in PI3K signaling	NM_133942	Ind
Got1	Glutamate oxaloacetate transaminase 1, soluble Enzyme involved in amino acid transport and metabolism	NM_010324	Ind
Gpd2	Glycerol phosphate dehydrogenase 2, mitochondrial Enzyme involved in NADH shuttle and glycerol metabolism KO causes reduced viability, body weight, and adiposity	NM_010274	Ind
Jade1	Jade1 PHD finger protein Putative transcription factor with PHD finger domain Interacts with tumor suppressor von Hippel-Lindau syndrome protein KO viable and fertile	NM_172303	Ind
Mzf6d	Kruppel associated box containing gene Putative transcription factor with KRAB zinc finger domains (KRAB box A and C)	AY149175	Rep
Klf2	Kruppel-like factor 2 (lung) Transcription factor with zinc finger domain Upregulated by fluid shear stress; downregulated by TNFα; negative regulator of adipogenesis KO causes blood vessel destabilization and reduction in pericytes; essential for lung development	NM_008452	Ind
Lmna	Lamin A Intermediate filament protein of nuclear lamina Upregulated by TPA; developmentally regulated (vs. lamin B) KO causes growth retardation, muscular dystrophy, and axonal neuropathy	NM_019390	Ind
Mcl1	Myeloid cell leukemia sequence 1 Bcl2 family of apoptosis regulator proteins Upregulated by EGF, VEGF, GM-CSF, IL-3, IL-6, IFNα, Activin A; implicated in cancers Inhibits apoptosis in many cell types KO causes pre-implantation embryonic lethality	NM_008562	Ind
Mll5	Myeloid/lymphoid or mixed-lineage leukemia 5 Putative transcription factor with PHD finger domain and SET domain Implicated in myeloid leukemia	BC036286	Rep[c]
Myo1e	Myosin IE SH3 domain containing unconventional myosin Putatively involved in signal transduction and cytoskeletal organization	NM_181072	Ind
2300006M17Rik	RIKEN cDNA 2300006M17 gene Unknown	XM_127388	Ind
Schip1	Schwannomin interacting protein 1 Coiled-coil domain containing protein Specifically associates with schwannomin, a membrane-cytoskeleton linker and tumor suppressor	NM_013928	Ind
Slc2a2	Solute carrier family 2 (facilitated glucose transporter), member 2 Enzyme involved in glucose transport and metabolism Upregulated by IGF-1 and IGF-2; downregulated by TNFα and IL-1β KO causes non-insulin-dependent diabetes mellitus and inverted α/β cell ratio in pancreatic islets	NM_031197	Ind

Table 1 (continued)

Gene symbol	Gene name, classification, function[a]	GenBank accession number	PDGF[b]
Sgpl1	Sphingosine phosphate lyase 1[d] Enzyme involved in amino acid transport and metabolism Breaks down sphingosine-1-phosphate, a second messenger modulating PDGF responses in normal and cancer cells, such as proliferation, motility, apoptosis and invasiveness RNA interference causes multiple defects in *C. elegans*	NM_009163	Ind
Tiparp	TCDD-inducible poly(ADP-ribose) polymerase WWE domain and poly(ADP-ribose) polymerase (PARP) domain containing protein Upregulated by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) Putatively regulating protein ubiquitination and ADP ribosylation	NM_178892	Ind
Txnip	Thioredoxin interacting protein[e] Enzyme involved in cellular redox state regulation (thioredoxin inhibitor) Downregulated by PDGF; upregulated by vitamin D3; upregulated in tumors Hypomorphic mice show hyperlipidemia	NM_023719	Rep
Unknown	Unknown		Ind
Zfp216	Zinc finger protein 216 Zinc finger domain containing protein	NM_009551	Ind

[a]Literature references were not cited due to space limit. [b]Transcriptional response to PDGF treatment: Ind, induced; Rep, repressed. [c]Expression of these two genes was undetected by northern-blot hybridization. [d]The RACE product seemss to be a fusion between *Sgpl1* and *Fosl1*, a known PDGF target gene. [e]The trapped gene is antisense to *Txnip*, as shown in **Figure 2b**.

Alternatively, although these MEFs express both receptors and respond to both ligands, the expression levels of each receptor or their downstream signaling effectors may vary, which could contribute to the observed differences in transcriptional responses.

Mutant generation and analysis

The most important advantage of the gene-trap array over conventional microarrays is that mice with mutations in target genes can be directly generated without *de novo* gene targeting. In a secondary, unbiased phenotype-driven screen, we started systematic blastocyst injections of the corresponding clones. So far, we have obtained high-percentage chimeras from most clones, and 10 of 12 test-bred lines have resulted in germline transmission. One line, ROSA71, was established in early phase of the study and has been characterized in some detail. The trapped gene encodes serine-threonine kinase receptor associated protein (STRAP), which inhibits TGFβ signaling by recruiting SMAD7 to TGFβRI[24]. We identified *Strap* as a PDGF-BB-inducible gene in a pilot gene-trap array screen using MEFs serum-deprived for 24 h rather than 48 h (**Fig. 7**). Notably, transcripts of inhibitory SMADs, SMAD6 and SMAD7, are both upreg-

ulated by epidermal growth factor[25], suggesting a possible mechanism by which engagement of a receptor tyrosine kinase antagonizes TGFβ signaling. The gene-trap insertion results in a null allele and recessive embryonic lethality between embryonic day (E) 10.5 and E12.5. Homozygous mutant embryos had defects in angiogenesis, cardiogenesis, somitogenesis, neural tube closure and embryonic turning. The phenotypes observed in the original line were indistinguishable from those found in a subline in which the poly-A trap module was deleted by breeding to ROSA26[Flper] mice[26]. We are further investigating the molecular mechanisms underlying mutant phenotypes and their relevance to TGFβ and PDGF signaling pathways *in vivo*.

DISCUSSION

Here we show that PDGF transcriptional targets can be efficiently and reliably identified in physiologically relevant cells with the gene-trap array, and that mutant mice can be directly generated from the corresponding gene-trap clones to determine their *in vivo* functions. Subsequent genetic analysis may help assess the relevance of the identified targets to the phenotypes observed in PDGF mutants. Although we are just starting to analyze functionally the uncovered PDGF targets, the available mutants of a subset of these genes have already shown that this is a promising approach. For example, disrupting *Klf2* in the mouse results in blood vessel destabilization and reduction in the number of vascular smooth muscle cells[27], a crucial target of signaling through PDGFRβ. Mutating *Arid5b* leads to defects in testis development, including small undescended testis, reduced interstitium and arrested spermatogenesis[28], indicative of testosterone deficiency and Leydig cell loss as observed in *Pdgfa* and *Pdgfra* mutants[1]. More focused searches could identify target genes specifically expressed in PDGF-regulated cell populations, and genetic epistasis analysis could be done by crossing the gene-trap lines with PDGFR mutants to place target genes in the genetic
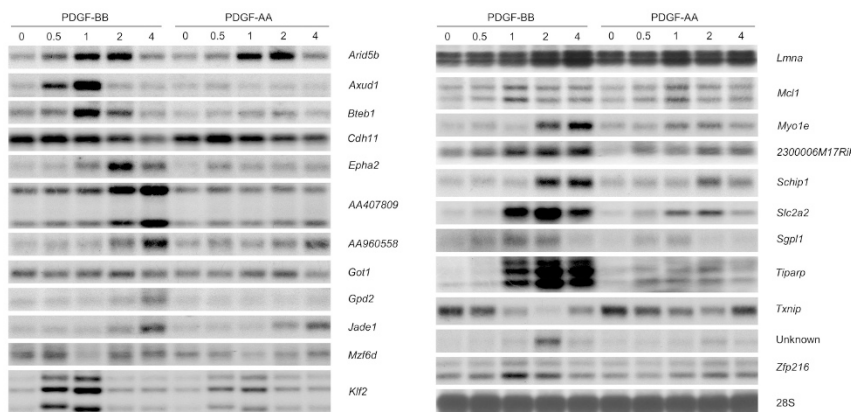

Figure 6 Northern-blot hybridizations of selected PDGF target genes identified in the gene-trap array screens. Independently isolated wild-type MEFs were grown in low serum medium for 48 h and induced with PDGF-BB or PDGF-AA for indicated times (h). Most genes were responsive to both growth factors, but the magnitude of induction or repression was always higher with PDGF-BB than with PDGF-AA, consistent with the array data.
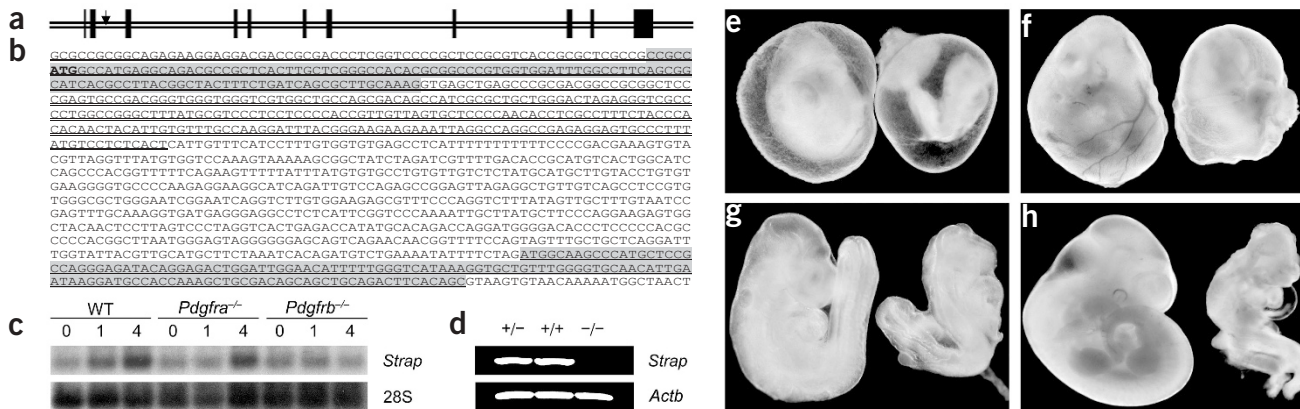
Figure 7 Characterization of the ROSA71 gene-trap mutation. (**a**) Schematic representation of the *Strap* locus disrupted in ROSA71. Arrow indicates the gene-trap insertion site. (**b**) Partial genomic DNA sequence of *Strap* encompassing intron 1–intron 3. Exon 2 and exon 3 are shaded, and the initiator ATG codon is highlighted in bold. Partial 3′ RACE and 5′ genomic anchoring PCR sequences immediately flanking the gene-trap vector sequences are singly and doubly underlined, respectively. (**c**) MEFs of indicated genotypes were grown in low-serum medium for 24 h and induced with 30 ng ml⁻¹ of PDGF-BB for 1 h and 4 h before northern-blot analysis. *Strap* was induced 4 h after treatment in both wild-type (WT) and PDGFRα-deficient (*Pdgfra⁻/⁻*) cells. (**d**) RT-PCR analysis of *Strap* and *Actb* expression in MEFs isolated from wild-type (+/+), heterozygous (+/–) and homozygous (–/–) ROSA71 embryos. Full-length *Strap* transcript was not detected in homozygotes. (**e–h**) Gross morphology of wild-type (left) and ROSA71-*Strap⁻/⁻* (right) embryos at E9.5 (**e,g**) and E10.5 (**f,h**). Mutants had underdeveloped yolk sac vasculature, arrested neural tube closure and embryonic turning, as well as abnormal hearts and somites.

networks regulated by one or the other receptor. Individual targets may also be subject to regulation by other signaling pathways, however, and a null mutation could lead to early embryonic lethality that could complicate the assessment of their contributions to PDGF functions. This limitation, inherent to any general transcription profiling and gene ablation approach, may be circumvented in the future by constructing gene-trap arrays using vectors that allow conditional mutagenesis.

One technical limitation of gene-trap mutagenesis is that mutations generated in ES cells must be transmitted through the germ line before functional analysis. Because large-scale phenotype-driven gene-trap screens are prohibitively laborious and costly, preselection of gene-trap events of interest is highly desirable. A number of expression-driven screening strategies have been designed to search for genes that are restricted to specific lineages, are responsive to external stimuli or encode secreted proteins[29]. But all these screens are confined to ES cells or their differentiated derivatives and rely on reporter expression. Sequence-driven screens based on the identity of trapped genes[19–22,30] are intrinsically biased towards highly annotated genes, and limited functional cues can be deduced from the trapped sequences representing ESTs or new genes. With the gene-trap array, virtually any cell type can now be screened for differentially expressed genes. Moreover, the versatility of microarrays allows for diverse experimental designs to find genes of specific biological relevance, including a substantial proportion of new genes currently unavailable elsewhere. Although standard microarrays and available gene-trap lines can be used in combination for the same purpose, reciprocal BLAST analysis comparing the most current Affymetrix mouse array with public gene-trap resources found limited overlap, and many matches through the short gene-trap sequence tags were unreliable (**Supplementary Fig. 1** online). As each spot on our array is directly correlated to a physical clone in which the corresponding gene is disrupted, the gene-trap array possesses the combined power of expression profiling and gene mutagenesis, providing a fast track from gene discovery to functional analysis.

METHODS

Gene-trap retrovirus. We assembled the poly-A trap cassette PGKhygSD by three-way ligation of a *Bam*HI-*Not*I fragment containing the plasmid backbone, the PGK promoter and a KOZAK-optimized initiator ATG derived from PGKβgeo; a *Bgl*II-*Xba*I fragment containing the hygromycin coding sequence amplified from pCEP4 (Invitrogen); and an *Xba*I-*Not*I fragment containing the splice donor sequence amplified from pIVTQ (a gift from S. Berget, Baylor College of Medicine, Houston, Texas). The splice donor sequence corresponds to nucleotides 6,025–6,195 of the adenovirus type 2 genome (accession number J01917), which includes the entire leader exon 1 and part of intron 1 of the major late transcription unit. To prevent potential activation of trapped genes due to translational readthrough or frameshifting, we introduced amber codons in each reading frame in front of the splice donor. After adding flanking FRT sites, we inserted the poly-A trap cassette into pSAβgeo* using a *Sal*I linker to place it at the same orientation as the promoter trap cassette. We constructed the retroviral vector pGep⁻ like pGen⁻ (ref. 32), except that the 3′ long terminal repeat did not include the gene *supF* and we deleted a fragment (*Nhe*I-*Ecl*136II) that removes the viral enhancer and promoter as well as a cryptic splice site[31]. We then excised the fragment containing both promoter trap and poly-A trap cassettes and inserted it into the unique *Xho*I site of pGep⁻. We linearized the resulting retroviral gene-trap vector, ROSAFARY, with *Dra*I and electroporated it into GP+E86 cells for packaging[32]. The titer of the gene-trap retrovirus was about 10⁴ colony forming units per ml as assayed on NIH 3T3 cells for G418-resistant colonies.

ES cell culture. We maintained AK7.1 ES cells derived from 129S4 in Dulbecco's modified Eagle medium (DMEM) with 15% fetal bovine serum, 50 U ml⁻¹ penicillin, 50 g ml⁻¹ streptomycin and 0.1 mM β-mercaptoethanol on γ-irradiated SNL76/7 (G418^R) feeder cell layers. To obtain gene-trapped clones, we infected log-phase ES cells with medium containing virus at low multiplicity and selected cells in medium containing 200 g ml⁻¹ G418 for 10–12 d (ref. 33). We picked resistant colonies individually for expansion in 96-well plates and selected quality clones for further culture. We kept frozen stocks in 96-format microtubes (CBL, 2600.mini) to facilitate the retrieval of individual clones.

We used G418 selection to obtain promoter trap events, and our screens are limited to genes expressed in ES cells. Due to the high sensitivity of the βgeo* reporter, however, many weakly expressed genes are also accessible to this vector, as indicated by the lack of X-Gal staining in ~40% of trapped clones. We did not isolate poly-A trap events, as hygromycin expression is driven by an internal promoter and can occur regardless of the insertion site in a gene. Because shorter fusion transcripts tend to be more easily stabilized and most mammalian genes

have a long 3′ UTR[34], poly-A trapping may enrich for insertions at the 3′ end of genes. In addition, in initial tests of the ROSAFARY vector, we obtained roughly equal numbers of resistant clones with G418 and hygromycin selection, but fewer than 20% of clones were resistant to both. This suggests that functional proteins of one or the other selectable marker are usually either not produced or not stabilized, probably because most gene-trap insertions would interfere with the endogenous *cis*-acting elements required for efficient splicing[35] and thus compromise the functionality of one or the other trap module. Therefore, poly-A trapping may not effectively disrupt the trapped genes but instead create hypomorphic or neutral mutations, as have been observed in as many as 30% of mutant lines obtained with a retroviral poly-A trap[36].

MEF isolation and culture. We obtained wild-type, *Pdgfra*−/− and *Pdgfrb*−/− embryos from intercrosses of *Pdgfra*+/− or *Pdgfrb*+/− mice in a hybrid 129S4 C57BL/6J genetic background and genotyped them as previously described[37,38]. We isolated primary MEFs from individual E12.5 embryos and maintained them in DMEM with 10% fetal calf serum, 50 U ml−1 penicillin and 50 g ml−1 streptomycin. For PDGF induction, we pooled 6–8 individual MEF cultures of the same genotype at passage 3 and expanded them to subconfluence on 15-cm plates. After growing them in low serum medium (0.5% fetal calf serum in DMEM) for 48 h, we treated MEFs with 30 ng ml−1 PDGF-BB or PDGF-AA for desired time periods and collected them for RNA preparation. We used MEFs in untreated plates of the same batch to prepare control or reference RNA at time 0. Wild-type MEFs isolated in this manner express both PDGF receptors at high levels, respond to both PDGF-AA and PDGF-BB and show the expected kinetics in immediate early gene induction after PDGF stimulation, suggesting that they are an appropriate model for studying PDGF signaling through each receptor.

3′ RACE. We designed a simple and economical procedure to directly isolate poly(A)+ RNA from gene-trapped ES cell clones in a 96-well format. We lysed confluent cultures of ES cell clones in a 96-well plate in 100 l of lysis buffer (4 M guanidine thiocyanate, 0.5% lauroylsarcosinate, 1% β-mercaptoethanol, 100 mM Tris-HCl, pH 7.2) per well. We then mixed the cell lysate in each well with 20 g MPG streptavidin-biotinylated oligo(dT)$_{25}$ complex (CPG) diluted in 200 l of binding buffer (400 mM LiCl, 20 mM EDTA, 100 mM Tris-HCl, pH 7.2) and transferred it into the corresponding well in a Multiscreen-HV filter plate (Millipore) fitted into a vacuum manifold. After incubating at room temperature for 10 min, we filtered the mixtures and washed them twice in 200 l of wash buffer I (150 mM LiCl, 1 mM EDTA, 0.1% lauroylsarcosinate, 10 mM Tris-HCl, pH 8.0) and 200 l of wash buffer II (150 mM LiCl, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0). We then eluted the bound poly(A)+ RNA in each well in 40 l of elution buffer (0.1 mM EDTA, 1 mM Tris-HCl, pH 8.0) after incubating at 55 ºC for 15 min and filtered it into a 96-well plate.

We carried out reverse transcription and 3′ RACE PCR reactions essentially as described[33]. We primed 10 l of denatured poly(A)+ RNA from each clone with anchoring oligo QT in a 20- l reverse transcription reaction containing 12.5 U of RNaseOUT RNase inhibitor and 50 U of Superscript II Reverse Transcriptase (Invitrogen). We diluted the resulting cDNA pools to 1:5 and 1:25 in two sets of 96-well plates. We then carried out paired primary 3′ RACE PCR reactions in a volume of 20 l using a high-fidelity PCR system with mixed KlenTaq1 (Ab Peptides) and *pfu* (Stratagene) polymerases and diluted the products to 1:100 in 96-format microtubes for diagnostic nested PCR. We used the hygromycin-specific primer HYGF and the SD exon-specific primer SDEXF in the primary and nested PCR reactions, respectively, paired with anchoring primers QA and QB. Both PCR reactions were done on GeneAmp 9700 thermocyclers using the same cycling parameters as described[33]. Primer sequences are available on request.

We separated diagnostic nested PCR products in 1% agarose gel, blotted them and hybridized them with an SD intron-specific oligo to detect products resulting from aberrant splicing events or genomic DNA contamination. We compared products from paired nested PCR reactions to select for the diluted template that led to more robust amplification and less complex banding than the other, and we abandoned clones that did not amplify in either nested PCR reactions or yielded only SD intron-positive products. We reassembled the selected microtubes containing the diluted primary PCR products into 96-well format for preparative nested PCR in a volume of 100 l using the same primers and cycling conditions, with 10- l templates in standard PCR buffers and mixes of *Taq* (Invitrogen) and *pfu* polymerases.

We observed robust amplification in >90% of the selected clones. 3′ RACE products mostly ranged from 0.5 kb to 2.0 kb in size, comparable to the average size of amplicons used to make conventional cDNA arrays. Most clones yielded only one prominent product, but ~20% of clones yielded multiple bands (usually two to three). Sequencing of the differently sized products from representative clones indicated that these mostly resulted from alternative splicing or nonstringent priming during reverse transcription reactions, to a lesser extent from coamplification of unspliced products and rarely from clonal contamination. The adenoviral splice donor was correctly used in fusion transcripts, as indicated by the replacement of vector sequence by the trapped sequence precisely at the exon-intron junction.

Sequencing and bioinformatics. RACE products were either gel-purified before sequencing or directly sequenced without purification in 96-well format with ABI Big-Dye terminator cycle sequencing in 10- l reactions using 5 pmol SDEXF as primer. Sequence quality analysis, trimming and masking were done with Phred, Sequencher and RepeatMasker, respectively. We considered masked sequence tags with consecutive sequences shorter than 50 nucleotides to be uninformative and excluded them from further analysis (47 of 654 in the random set). We used quality tags to search against the NR and EST databases at National Center for Biotechnology Information with BLASTN to determine homology to known transcripts for gene identification. An expect value of e−20 or less was considered significant, and transcripts of best BLAST matches were taken to annotate the corresponding tags. We determined putative gene-trap insertion sites in known candidate genes by aligning the sequence tags to the best-matching transcripts with full-length sequence information available. Genome mapping of masked sequence tags was done using the Ensembl genome browser, with BLASTN and expect cut-off of e−10.

Array construction, expression profiling and data analysis. We constructed the 2K gene-trap array set using standard cDNA microarray methodologies as described elsewhere[39]. We purified RACE products from the preparative nested PCR reactions with the Multiscreen-PCR filtration system (Millipore) and mechanically spotted them in duplicate onto polylysine-coated microscope slides using an OmniGrid high-precision robotic gridder (GeneMachines). We isolated total RNA from tissues or cultured cells with TriZol reagent (Invitrogen) according to the instructions provided by the manufacturer and assessed RNA quality with a 2100 Bioanalyzer (Agilent). We labeled and hybridized targets essentially as described[40]. We generated cDNA targets using a standard aminoallyl labeling protocol, in which 30 g total RNA was coupled to either Cy3 or Cy5 fluorophores. Pairs of labeled cDNA targets were cohybridized to microarrays for 16 h at 63 °C and sequentially washed at room temperature in 1 sodium saline citrate (SSC) and 0.03% SDS for 2 min, 1 SSC for 2 min, 0.2 SSC with agitation for 20 min and 0.05 SSC with agitation for 10 min. We immediately centrifuged arrays until dry and scanned them using a GenePix 4000 scanner (Axon Instruments). We analyzed images using GenePix Pro 3.0.

We carried out hybridizations in duplicate with fluorophore reversal to compensate for dye bias. Signal intensities from duplicated spots and duplicated hybridizations were quantified, corrected over background and normalized. To account for sequence-dependent fluorophore incorporation biasing, we generated expression profiles twice for each sample in which both fluorophore-labeled orientations were represented. For each array, we filtered spot intensity signals and removed those values that did not exceed 3 s.d. above the background signal in at least one signal channel and those spots flagged as questionable by the GenePix Pro software. Spot-level ratios (Cy5/Cy3) were log$_2$ transformed and a loess normalization ($f = 0.67$) strategy was applied using S-Plus (MathSoft) to correct for observed intra-array intensity-dependent ratio biasing. For each sample comparison, we analyzed data by first averaging the spot-level, normalized log ratios of the two reverse-complement arrays and then taking the average of the intra-array, gene-level duplicate features. At the spot-level averaging step, we analyzed only those features possessing both fluorophore-orientation data points. Under this analysis strategy, paired 'same versus same' comparisons using independently cultured and labeled samples resulted in the global array mean log$_2$ (Cy5/Cy3) = 0.00 ± 0.07 (1 s.d.). We carried out average linkage clustering as described[41] using the GeneCluster/TreeView software package. To exclude genes expressed at very low levels, we included only genes represented at >75% of all data points in the cluster analysis.

Northern-blot hybridization. We separated 5 g of total RNA in 1.2% denaturing agarose gel, transferred it onto Hybond N+ membranes and fixed it by ultraviolet crosslinking. We prepared radioactive probes from gel-purified RACE products and used a mouse 28S rDNA probe as a loading control. We carried out hybridization in modified Church and Gilbert buffer (7% SDS, 10 mM EDTA, 0.5 M NaH_2PO_4–Na_2HPO_4, pH 7.2) for 16 h and washed membranes in 1 SSC, 0.1% SDS for 2 h before exposing to films.

Mutant mouse generation. We injected ES cells carrying the desired gene-trap mutation into C57BL/6J blastocysts to generate chimeras according to standard procedures. The poly-A trap module could be deleted by breeding with ROSA26[Flper] mice[26]. Mice were housed in microisolator racks in a facility accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care and experimentation was reviewed by the Hutchinson Center Institutional Review Committee.

Molecular analysis of ROSA71 gene-trap mutation. We carried out 5′ genomic anchoring PCR as described using gene trap–specific primers SASP and U5SP[33]. We genotyped ROSA71 adults and embryos with three-primer PCR using a forward gene-specific primer 5′ to the insertion site, a reverse gene-specific primer 3′ to the insertion site and a β-gal-specific reverse primer. Wild-type, heterozygous and homozygous MEFs were derived from E9.5 embryos and subjected to RT-PCR analysis using *Strap* primers corresponding to the 5′ and 3′ cDNA sequences flanking the insertion site. *Actb* primers were described elsewhere[42]. Primer sequences are available on request.

URLs. Further information regarding ROSA-series gene-trap vectors can be found at http://www.fhcrc.org/labs/soriano/trap.html. Supplementary materials and the complete microarray data sets are available at http://parma.fhcrc.org/GTA. The gene-trap array set is now being fully annotated, and the gene-trap sequence tags can be found at Genome Survey Sequences Database at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/dbGSS). The gene-trap arrays and corresponding ES cell clones are an additional resource of the International Gene Trap Consortium (http://www.igtc.ca).

Note: Supplementary information is available on the Nature Genetics website.

1. Betsholtz, C., Karlsson, L. & Lindahl, P. Developmental roles of platelet-derived growth factors. *Bioessays* **23**, 494–507 (2001).
2. Hoch, R.V. & Soriano, P. PDGF roles in animal development. *Development* **130**, 4769–4784 (2003).
3. Liang, P. & Pardee, A.B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967–971 (1992).
4. Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, K.W. Serial analysis of gene expression. *Science* **270**, 484–487 (1995).
5. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
6. Rollins, B.J. & Stiles, C.D. Serum-inducible genes. *Adv. Cancer Res.* **53**, 1–32 (1989).
7. Herschman, H.R. Primary response genes induced by growth factors and tumor promoters. *Annu. Rev. Biochem.* **60**, 281–319 (1991).
8. Fambrough, D., McClure, K., Kazlauskas, A. & Lander, E.S. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* **97**, 727–741 (1999).
9. 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).
10. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513–1523 (1991).
11. von Melchner, H. *et al.* Selective disruption of genes expressed in totipotent embryonal stem cells. *Genes Dev.* **6**, 919–927 (1992).
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. Bonaldo, P., Chowdhury, K., Stoykova, A., Torres, M. & Gruss, P. Efficient gene trap screening for novel developmental genes using IRES beta geo vector and in vitro preselection. *Exp. Cell Res.* **244**, 125–136 (1998).
14. Komada, M., McLean, D.J., Griswold, M.D., Russell, L.D. & Soriano, P. E-MAP-115, encoding a microtubule-associated protein, is a retinoic acid-inducible gene required for spermatogenesis. *Genes Dev.* **14**, 1332–1342 (2000).
15. Komada, M. & Soriano, P. [Beta]IV-spectrin regulates sodium channel clustering through ankyrin-G at axon initial segments and nodes of Ranvier. *J. Cell Biol.* **156**, 337–348 (2002).
16. Wu, X., Li, Y., Crise, B. & Burgess, S.M. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749–1751 (2003).
17. Frohman, M.A., Dush, M.K. & Martin, G.R. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002 (1988).
18. Waterston, R.H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
19. Hicks, G.G. *et al.* Functional genomics in mice by tagged sequence mutagenesis. *Nat. Genet.* **16**, 338–344 (1997).
20. Wiles, M.V. *et al.* Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. *Nat. Genet.* **24**, 13–14 (2000).
21. Mitchell, K.J. *et al.* Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nat. Genet.* **28**, 241–249 (2001).
22. Hansen, J. *et al.* A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc. Natl. Acad. Sci. USA* **100**, 9918–9922 (2003).
23. Klinghoffer, R.A., Mueting-Nelsen, P.F., Faerman, A., Shani, M. & Soriano, P. The two PDGF receptors maintain conserved signaling in vivo despite divergent embryological functions. *Mol. Cell* **7**, 343–354 (2001).
24. Datta, P.K. & Moses, H.L. STRAP and Smad7 synergize in the inhibition of transforming growth factor beta signaling. *Mol. Cell. Biol.* **20**, 3157–3167 (2000).
25. Afrakhte, M. *et al.* Induction of inhibitory Smad6 and Smad7 mRNA by TGF-beta family members. *Biochem. Biophys. Res. Commun.* **249**, 505–511 (1998).
26. Farley, F.W., Soriano, P., Steffen, L.S. & Dymecki, S.M. Widespread recombinase expression using FLPeR (flipper) mice. *Genesis* **28**, 106–110 (2000).
27. Kuo, C.T. *et al.* The LKLF transcription factor is required for normal tunica media formation and blood vessel stabilization during murine embryogenesis. *Genes Dev.* **11**, 2996–3006 (1997).
28. Lahoud, M.H. *et al.* Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. *Genome Res.* **11**, 1327–1334 (2001).
29. Stanford, W.L., Cohn, J.B. & Cordes, S.P. Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.* **2**, 756–768 (2001).
30. Zambrowicz, B.P. *et al.* Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. *Proc. Natl. Acad. Sci. USA* **100**, 14109–14114 (2003).
31. Medico, E., Gambarotta, G., Gentile, A., Comoglio, P.M. & Soriano, P. A gene trap vector system for identifying transcriptionally responsive genes. *Nat. Biotechnol.* **19**, 579–582 (2001).
32. Soriano, P., Friedrich, G. & Lawinger, P. Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells. *J. Virol.* **65**, 2314–2319 (1991).
33. Chen, W.V. & Soriano, P. Gene trap mutagenesis in embryonic stem cells. *Methods Enzymol.* **365**, 367–386 (2003).
34. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
35. Maniatis, T. & Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236–243 (2002).
36. Abuin, A., Holt, K.H., Platt, K.A., Sands, A.T. & Zambrowicz, B.P. Full-speed mammalian genetics: in vivo target validation in the drug discovery process. *Trends Biotechnol.* **20**, 36–42 (2002).
37. Soriano, P. Abnormal kidney development and hematological disorders in PDGF beta-receptor mutant mice. *Genes Dev.* **8**, 1888–1896 (1994).
38. Hamilton, T.G., Klinghoffer, R.A., Corrin, P.D. & Soriano, P. Evolutionary divergence of platelet-derived growth factor alpha receptor signaling mechanisms. *Mol. Cell. Biol.* **23**, 4013–4025 (2003).
39. DeRisi, J.L., Iyer, V.R. & Brown, P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
40. Pritchard, C.C., Hsu, L., Delrow, J. & Nelson, P.S. Project normal: defining normal variance in mouse gene expression. *Proc. Natl. Acad. Sci. USA* **98**, 13266–13271 (2001).
41. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
42. Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072–2075 (1998).