Gene targeting using a promoterless gene trap vector (''targeted trapping'') is an efficient method to mutate a large fraction of genes

Roland H. Friedel*†, Andrew Plump*†‡, Xiaowei Lu*†, Kerri Spilker*†, Christine Jolicoeur*†, Karen Wong*†, Tadmiri R. Venkatesh§, Avraham Yaron*†, Mary Hynes*, Bin Chen*, Ami Okada*, Susan K. McConnell*, Helen Rayburn*†, and Marc Tessier-Lavigne*†¶

*Department of Biological Sciences, †Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305; and §Department of Biology, City College, City University of New York, New York, NY 10031

Contributed by Marc Tessier-Lavigne, July 11, 2005

A powerful tool for postgenomic analysis of mammalian gene function is gene targeting in mouse ES cells. We report that homologous recombination using a promoterless gene trap vector (''targeting trapping'') yields targeting frequencies averaging above 50%, a significant increase compared with current approaches. These high frequencies appear to be due to the stringency of selection with promoterless constructs, because most random insertions are silent and eliminated by drug selection. The promoterless design requires that the targeted gene be expressed in ES cells at levels exceeding a certain threshold (which we estimate to be ≈1% of the transferrin receptor gene expression level, for the secretory trap vector used here). Analysis of 127 genes that had been trapped by random (nontargeted) gene trapping with the same vector shows that virtually all are expressed in ES cells above this threshold, suggesting that targeted and random trapping share similar requirements for expression levels. In a random sampling of 130 genes encoding secretory proteins, about half were expressed above threshold, suggesting that about half of all secretory genes are accessible by either targeted or random gene trapping. The simplicity and high efficiency of the method facilitate systematic targeting of a large fraction of the genome by individual investigators and large-scale consortia alike.

ES cells | gene trapping

Gene trapping by random insertion of a vector into the genome of mouse ES cells has allowed the generation of a large collection of mutations that are an invaluable resource for the study of mammalian biology (1, 2). In a typical 5′ gene trapping strategy, the trapping vector comprises a 5′ splice acceptor site and a promoterless antibiotic resistance gene [such as a gene for neomycin phosphotransferase (neo)] fused to a marker gene (such as a gene for β-gal). When the vector inserts in-frame in an intron of a gene expressed in ES cells, it traps the upstream exons, resulting in a transcript that directs the expression of a fusion protein comprising the sequences encoded in the trapped exons and those in the vector. This allows for isolation of productive insertions by selection with the antibiotic and subsequent screening for β-gal-positive colonies, provided the trapped gene is expressed at a sufficiently high level in ES cells. The advantage of random gene trapping is the extreme ease with which a large number of mutations can be generated. In addition, specific vector designs can enrich for insertions in specific classes of genes; thus, a ''secretory gene trap vector'' was designed to target genes encoding proteins with a signal sequence (secreted and membrane anchored proteins) (3). The disadvantage of random gene trapping is that the specific genes that are trapped cannot be specified in advance, so there is no guarantee that any particular gene of interest will be mutated. In addition, to date, it has been unclear what fraction of the genes in the genome are expressed at a sufficient level in ES cells to be accessible by this method.

In contrast to random gene trapping, gene targeting provides a directed strategy that exploits homologous recombination to modify a specific gene locus. Traditional vectors used for homologous recombination usually contain an antibiotic resistance gene driven by a constitutive promoter (such as the *Pgk* promoter). The advantage of this method is that a particular gene of interest can be targeted regardless of whether it is expressed in ES cells. The disadvantage is that the vector can confer antibiotic resistance to transfected ES cells, whether it has recombined into the targeted locus or has inserted in a nonspecific way. As a consequence of this lack of stringent selection, the rate of correctly targeted ES cell clones (the ''targeting frequency'') in a typical gene-targeting experiment is usually only low to moderate, in the range of 0.5- 5%, making it a more labor-intensive procedure than random gene trapping.

An alternative strategy for gene targeting is to use promoterless gene-targeting vectors, the kind used in random gene trapping, and to rely on the endogenous promoter of the targeted gene to drive expression of drug resistance. Because the targeting vector does not have a promoter to drive antibiotic resistance, most nonspecific insertions are eliminated, so that this approach should in principle be able to reach much higher targeting frequencies than conventional gene targeting. The utility of promoterless gene targeting (an approach we call ''targeted trapping'') was recognized in the early years of gene targeting and was successfully applied to mutate several genes (4–10). For the method to work, however, the targeted gene must be expressed at a sufficiently high level in ES cells to confer antibiotic resistance. In these early studies, no attempt was made to assess the level at which a gene has to be expressed in ES cells to make it a successful target for a promoterless construct, and the predicted improvement in targeting frequencies expected from the approach was not systematically explored (4–10), potentially explaining why this approach has not been widely adopted by other laboratories in the intervening decade.

We therefore set out to define the level of gene expression required for promoterless gene targeting (targeted trapping) to be successful and to assess whether this method can be used to access a significant fraction of genes in the genome. Our results, focused primarily on using the secretory gene trap vector,

Freely available online through the PNAS open access option.

Abbreviations: Trfr, transferrin receptor; TM, transmembrane domain; neo, neomycin phosphotransferase; GO, Gene Ontology.

See Commentary on page 13001.

‡Present address: Merck Research Laboratories, 126 East Lincoln Avenue, Rahway, NJ 07065.

¶To whom correspondence should be sent at the present address: Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080. E-mail: marctl@gene.com.

establish that this method is likely to be useful for a large fraction of genes and define an approximate level of gene expression in ES cells that is required for successful targeting with this strategy.

Methods

Targeting Constructs. The vectors used for targeted trapping of genes encoding secretory proteins (pTT0TM, pTT1TM, and pTT2TM for the three reading frames, respectively) were derived from the gene trap secretory vector pGT0-2TMpfs (11) by flanking the vector cassette with AscI sites. A version of the targeted trapping vector without transmembrane domain (TM) (pTT0, pTT1, and pTT2), which was used for targeted trapping of genes encoding cytosolic/nuclear proteins, was derived from pTT0-2TM by introducing the ScaI/ClaI fragment of pGT0-2 (12).

Targeting constructs were designed by using the University of California, Santa Cruz, and Celera mouse genome databases. For genes encoding secretory proteins, the insertion of the pTT0-2TM secretory trap cassette was usually positioned into the middle of the second-to-last intron upstream of the exon encoding the type I TM or the GPI signal, respectively. For type II TM proteins, the insertion was positioned downstream of the exon encoding the TM. For cytosolic/nuclear genes, the insertion of the pTT0-2 cassette was placed in one of the first introns.

Homology arms flanking the insertion site were typically 5 and 3 kb in length (for the 5′ and 3′ arms, respectively) and were generated by PCR (Expand High Fidelity PCR kit, Roche Applied Science, Indianapolis) from genomic DNA of E14Tg2A.4 ES cells. Suitable restriction sites were added to the primers to allow successive cloning of the homology arms and the targeted trap pTT0-2TM/pTT0-2 AscI-cassette into pBluescript (Stratagene). Exons on the 5′ homology arm were sequenced to control for nonsense mutations. A detailed protocol of the cloning procedures is available on request.

ES Cell Culture/Electroporation. The feeder independent ES cell line E14Tg2a.4 was cultured as described (13). For the targeted trapping constructs, 5×10^7 cells were electroporated with 50 μg of linearized DNA. Cells were seeded at a density of 4×10^6 cells per 10-cm dish. Colonies were picked after 10 days of selection with 125 μg/ml G418 (GIBCO no. 11811-031). ES cell clones were analyzed for correct targeting by Southern blot analysis with a 1-kb 5′ external probe. Before blastocyst injection, selected clones were confirmed with an internal *neo* probe and by genomic PCR with a 3′ external primer.

Gene Expression Analysis. Real-time RT-PCR analysis of gene expression in ES cells was performed with SYBR green chemistry on Stratagene Mx3000P and Applied Biosystems 5700 systems. cDNA was synthesized from DNaseI- (Roche Applied Science) treated total RNA of E14Tg2a.4 cells with oligo(dT) primers and Superscript II Reverse Transcriptase (Invitrogen).

Primers were designed as 20-mers with 50–60% GC content, spanning one of the last introns, and amplifying a product of 100–120 bp. PCR reactions were run as triplicates on 96-well plates, with each reaction containing cDNA derived from 16 ng of total RNA, 7.5 pmol of each primer, and 1× SYBR green reaction mix (Applied Biosystems) in a 25-μl volume. The temperature profile was 10 min at 95°C and then 40 cycles at 95°C for 15 sec, 60°C for 30 sec, and 72°C for 30 sec.

We measured expression levels relative to a standard curve of the expression of the transferrin receptor (Trfr) primers: TGGGAACAGGTCTTCTGTTG and TGCAGTCCAGCTGGCAAAGA. The final expression levels were the averages obtained from measurements from three independent preparations of E14Tg2a.4 ES cells.

The list of genes for previously randomly trapped secretory genes was based on Table 1 in Mitchell *et al.* (14). The RIKEN genes of control set 1 were chosen by alphanumerical order from the web site (http://fantom.gsc.riken.go.jp) from the following gene ontology (GO) categories: GO:0005887, integral plasma membrane; GO:0005789, endoplasmatic reticulum membrane; GO:0005788, endoplasmatic reticulum lumen, GO:0005764, lysosome; and GO:0005576, extracellular. The genes for the list of randomly trapped transcription factors were randomly picked from cell lines at www.baygenomics.ucsf.edu from the GO category GO:0005634, nucleus.

Table 1. Targeting frequencies of targeted trapping constructs

Targeted gene	Targeting frequency, %*	Expression level, % (% of *Trfr*)†
Genes encoding secreted or membrane anchored proteins‡		
Plexin-B2	95 (112/118)	35 (±3.6)
Coxadr	88 (102/116)	13 (±0.5)
Kremen-1	83 (100/120)	5.8 (±0.6)
Ryk	74 (83/112)	109 (±9.3)
Plexin-B1	69 (83/120)	14 (±1.8)
SynCAM	68 (67/99)	59 (±5.8)
Robo-1	61 (36/59)	17 (±2.5)
Nectin-3	58 (63/108)	73 (±1.8)
Teneurin-4	58 (84/144)	21 (±5.4)
Sema4C	55 (40/73)	6.4 (±1.9)
CDO	55 (55/100)	5.1 (±0.8)
Sema7A	52 (61/116)	2.0 (±0.3)
PTP mu	46 (30/65)	2.1 (±0.5)
BOC	45 (65/144)	2.0 (±0.1)
Neto2	17 (15/88)	4.6 (±0.6)
TEM7	15 (11/75)	2.1 (±0.3)
8D6 antigen	6.3 (6/94)	46 (±7.5)
PTP λ	0.8 (1/120)§	3.5 (±0.6)
Protocadherin-21	0 (0/120)§	4.8 (±1.8)
Sema3B	0 (0/132)§	4.1 (±2.0)
Sema5A	0 (0/99)§	3.9 (±0.6)
Plexin-A4	0 (0/87)§	1.0 (±0.1)
Teneurin-1	Low colonies¶	0.1 (±0.04)
Netrin-G1	Low colonies¶	0.02 (±0.005)
Genes encoding cytosolic or nuclear proteins		
Etv5	97 (104/107)	419 (±157)
Grg4	66 (79/120)	37 (±36)
Grg3	65 (68/105)	17.0 (±5.7)
MBTL	5.0 (5/100)	5.0 (±1.9)
SmarcD3	0.9 (1/108)	5.2 (±3.2)

*Ratio of correct recombination events as determined by Southern blot with a 5′ external probe. Numbers in parentheses are positive clones vs. clones screened.

†The expression of genes in the E14 ES cell line was determined by semiquantitative real-time RT-PCR relative to the expression level of the transferrin receptor gene. Data were averaged from three independent RNA preparations.

‡These genes encode type I TM proteins, except Kremen-1 and Teneurin-1 and -4, which are type II TM, and Sema-7A and Netrin-G1, which are GPI-linked.

§Sum of two electroporation experiments.

¶Fewer than 10 colonies per electroporation, not screened.

Results

High Targeting Frequencies with a Gene Trap Vector. To test whether gene targeting with a gene trap vector ("targeted trapping") would systematically increase targeting efficiencies and to determine how widely applicable targeted trapping might be, we performed targeted trapping with a secretory gene trap vector (see Fig. 1*A*) that we previously used in collaboration with W. Skarnes in a random screen for axon guidance molecules (11). This vector has three useful properties: it enriches for genes

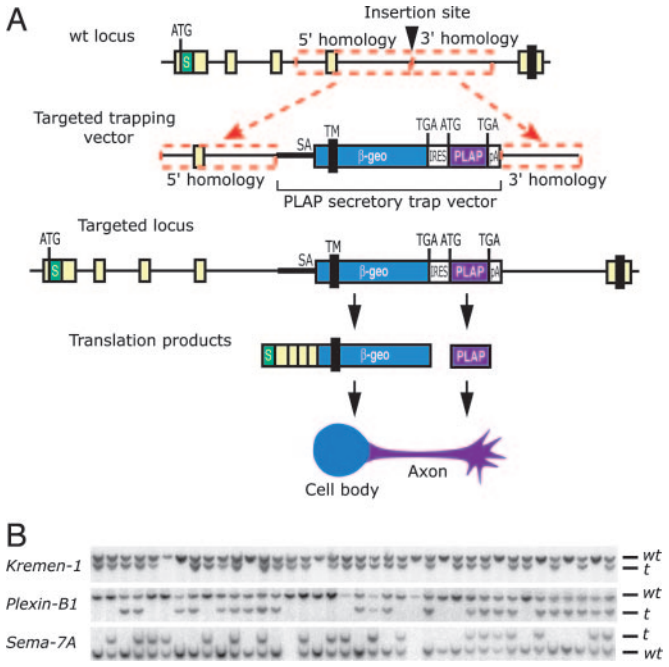

Fig. 1. The targeted trapping strategy for secretory genes and examples of the high targeting frequencies observed using this strategy. (*A*) The placental alkaline phosphatase (PLAP) secretory trap vector was flanked by homology arms to facilitate insertion into an intron of a hypothetical targeted gene (exons shown in yellow). The resulting bicistronic transcript encodes two proteins: a fusion of the endogenous protein with β-geo (the fusion protein is retained in an intracellular compartment) and placental alkaline phosphatase, which is localized on axonal membranes. S, signal sequence; SA, splice acceptor. (*B*) Three sections of Southern blots of randomly picked ES cell clones from electroporation experiments using the Kremen-1, Plexin-B1, and Sema-7A targeted trapping constructs. Double bands indicate successful targeting, and single bands represent wild-type clones. The overall targeting frequencies were 83% for Kremen-1 (100 positive clones of 120 clones analyzed), 69% for Plexin-B1 (83/120), and 52% for Sema-7A (61/116). wt, wild-type allele; t, targeted allele.

encoding secreted and membrane-anchored proteins ("secretory" genes); it is highly mutagenic, because of both the strong 5′ splice acceptor (which prevents splicing over the insertion) and intracellular targeting of the resulting fusion protein; and, in addition to containing β-geo (a fusion of neo and β-gal), it contains a second reporter, placental alkaline phosphatase, which labels axons and allows their visualization (3, 11, 14). Use of this vector on a large scale has allowed us to randomly trap numerous secretory genes (see www.baygenomics.ucsf.edu). Nonetheless, many genes of interest that possess appropriate introns for insertion of the vector have so far failed to be trapped by this method. This failure could arise from, first, insufficient expression of the genes in ES cells; second, a bias in the insertion of the vector to particular loci [which is known to occur to some extent, because there are insertional "hot-spots" (14)]; or third, the failure to apply the approach on a large enough scale. We reasoned that use of the targeted trapping strategy could help distinguish between these possibilities, particularly in determining whether expression level influences the ability of a gene to be targeted.

In a first targeted trapping experiment using the secretory trap vector, we set out to target 24 genes encoding membrane-anchored proteins, mostly type I TM proteins, but also some type II and glycosylphosphatidylinositol-anchored proteins, that were of interest to us for their possible involvement in axon guidance (Table 1). In this initial experiment, we designed our homology arms to mimic the events that occur in random gene traps, aiming to insert the vector in an appropriate intron of the target gene, i.e., upstream of the exon encoding the TM for type I TM proteins but downstream of the exon encoding the TM for type II TM proteins (Fig. 1*A*; see *Methods* for details). We typically used homology arms obtained by genomic PCR that were ≈5 and ≈3 kb in length for the 5′ and 3′ arms, respectively. In a second experiment, we designed targeting constructs aimed at introducing an insertion of the gene trap vector together with a major deletion in the targeted locus (ranging from 0.8 to 27 kb in size) into a set of seven secretory genes (Table 2, which is published as supporting information on the PNAS web site).

Consistent with the idea that the promoterless design would reduce the number of false-positive ES cell colonies arising from random insertions, in the first experiment with the 24 insertional targeted trapping constructs, we found that electroporation experiments yielded on average 10- to 100-fold fewer colonies than with conventional *Pgk-neo* targeting vectors. The number of G418-resistant colonies in successful experiments varied from 50 to several hundred, and we typically picked 120 colonies for analysis (Table 1). [In these experiments, we used a G418 concentration of 125 μg/ml, the same as for random gene trapping (13), and about the lowest practicable level for selection with this antibiotic (data not shown)].

Of the 24 targeting constructs, 18 produced correct gene targeting events, as assessed by Southern blot analyses using 5′ external probes; the approach was successful for all three types of membrane-anchored proteins (type I, type II, and GPI-anchored) (Figs. 1*B* and 2; Table 1). Remarkably, the targeting frequencies of these 18 successful constructs averaged >50%. Two of the successful constructs revealed moderate targeting frequencies of 0.8% and 6.3%, but for the remaining 16 constructs, frequencies of 15–95% were obtained. These targeting frequencies are significantly higher than those of the most efficient gene-targeting methods reported to date, including bacterial artificial chromosome (BAC)-mediated targeting (average targeting frequency of 3.8%) (15). Six constructs either failed to generate correctly targeted clones, or did not produce enough colonies for analysis (Table 1). Thus, in this initial experiment, 18 of 24 constructs (75%) yielded appropriate insertions. This success rate, although also high, is lower than that reported for BAC-mediated targeting (98%) (15), but it is likely to be at similar high levels if the targeted genes are chosen to have a high enough expression level in ES cells (see below). In our second experiment, using constructs designed to introduce the gene trap insertion together with a major deletion into a set of seven genes, only one construct resulted in successful targeting. Interestingly, two of the genes that failed to be targeted with deletion constructs (*BOC* and *CDO*) were successfully targeted in parallel with a construct designed to introduce only an insertion, raising the possibility (although by no means proving) that a deletion strategy is less efficient than a pure insertional strategy (Table 2).

Expression Levels in ES Cells Required for Targeted Trapping. Although the majority of insertional constructs were targeted successfully with high targeting frequencies, some gave lower frequencies, and some failed to give recombinants. We hypothesized that the targeting frequencies of the constructs might correlate with the expression levels of the targeted genes in ES cells, because this would determine the level of expression of the antibiotic resistance gene. To test for a potential correlation, we measured the expression levels of the targeted genes in E14Tg2A.4 ES cells [used for both our random (11) and targeted trapping experiments] with a semiquantitative real-time RT-PCR assay, using the expression level of the *Trfr*, a moderately expressed housekeeping gene, as a reference. Expression relative to *Trfr* was measured in three independent experiments (starting from three separate ES cell mRNA isolates) and the average of

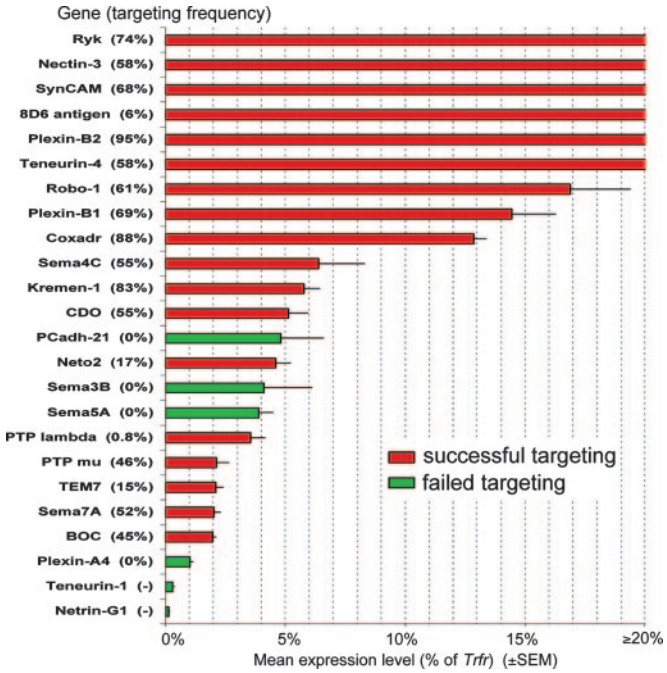

Fig. 2. Correlation of the expression levels and targeting frequencies of the 24 secretory genes targeted by targeted trapping. Data are plotted from Table 1, with genes sorted by expression level in E14 ES cells (relative to the expression level of the Transferrin receptor gene, *Trfr*). Numbers in parentheses denote targeting frequencies for each gene. A threshold for successful targeting with this promoterless targeted trapping construct appears to lie between 1% and 2% *Trfr* expression. Expression levels are cut off at 20% for clarity (see Table 1 for values >20%).

the three values was taken. The expression levels of successfully targeted genes ranged from 2.0% to 109% relative to *Trfr*, including all genes that were expressed higher than 5% of *Trfr* expression (Fig. 2 and Table 1). Of the six genes that failed targeting, three had the lowest expression levels of all 24 genes (*Plexin-A4*, *Teneurin-1*, and *Netrin-G1* with levels of 0.02–1%), and the other three had expression levels at the lower range of the successfully targeted genes (*PCadh-21*, *Sema3B*, and *Sema5A* with levels between 3.9% and 4.8%). The apparent correlation between failed targeting and lower expression levels in ES cells is consistent with the idea that a minimum expression level is required for successful targeted trapping. The data further suggested that this minimum threshold may be ≈1% of *Trfr* expression, and that targeted trapping has a high success rate for genes expressed at levels higher than 5% of *Trfr* expression.

To obtain more definitive information on the relationship between the expression level of a gene in ES cells and the possibility of trapping with our gene trap vector, we analyzed the expression levels of a set of 127 genes (all of which are secretory) that had been previously trapped in our random gene trap screen using the same secretory trap vector (14) (the "trapped set"). To ascertain whether these genes showed a bias toward high expression in ES cells, we compared their expression levels with those of two control sets of secretory genes. The first set (control set 1) contained a random selection of 55 secretory genes obtained from the RIKEN cDNA collection; we attempted to use an unbiased approach in selecting these genes (see *Methods*). The second (control set 2) was comprised of 75 genes encoding members of several secreted and membrane-anchored protein families implicated in axon guidance or regeneration (Netrins, Slits, Semaphorins, Ephrins, Nogos, and their receptors). Many of these ligand or receptor families are large (e.g., there are 19 Semaphorins and 11 Semaphorin receptors) and spread throughout the genome, and we reasoned that there was no particular reason for the members of these families to be expressed systematically at high or low levels in ES cells.

We analyzed the expression levels of these three sets of genes, again as a percentage of the *Trfr* expression using real-time RT-PCR. Little variability in expression level was observed among experiments, as evidenced by the small standard deviations (on average <28% of the mean expression level). Strikingly, we observed a dramatic difference in the distributions of expression levels between the trapped set and the two control sets (Fig. 3 and Tables 3–5, which are published as supporting information on the PNAS web site). In the case of the trapped set, 97% of the genes had expression levels above 1% of the *Trfr* expression level, and 87% of the genes had levels above the 5% *Trfr* level. The range of expression above this level did not show particular trends or clustering but rather formed a smooth distribution, with the most highly expressed gene expressed at 465% of *Trfr*. In contrast, the two control sets differed markedly from the test set but were very similar to one another, with just 55% and 57% of genes expressed above 1% *Trfr* in control sets 1 and 2 and 45% and 38% of genes expressed above 5% *Trfr*, respectively. Interestingly, when we examined a third control set of 145 genes encoding nonsecretory proteins (transcription factors selected by arbitrary rather than random criteria), we found that their expression levels were distributed similarly to those of the two sets of control secretory genes, with ≈44% expressed above 1% *Trfr* (Table 6, which is published as supporting information on the PNAS web site).

Targeted Trapping with a Nonsecretory Trap Vector. The expression values described above apply specifically to the use of a particular vector, the secretory trap vector, in targeted trapping; however, it is expected that this strategy may be applied to any vector that has been used successfully in gene trapping. Indeed, at least two genes were previously successfully targeted with other gene trap vectors (9, 10), although the targeting frequencies in those experiments were not reported. However, one might expect that the gene expression threshold in ES cells for gene trapping (whether by random or targeted trapping) might differ for different vectors, because each vector can have a different selectable marker, and even if the same selectable marker is used, the stability or localization of the fusion protein resulting from the insertion can differ among targets. To extend our experiments to an alternative vector, we performed a pilot experiment in which a vector designed to target nonsecretory genes (12) was used in targeted trapping. Five genes encoding transcription factors, each of which was expressed above 5% *Trfr* in ES cells, were chosen as targets. As for the secretory trap vector, we constructed homology arms ≈5 and 3 kb in size, but in these cases, we targeted the vector into one of the first introns to keep the size of the endogenous domain of the fusion protein as short as possible. All five targeting attempts were successful, with three genes giving very high (>60%) targeting frequencies and the other two respectable frequencies of 5% and 0.9% (Table 1). This experiment confirms, as expected, that vectors other than the secretory trap vector can be used in targeted trapping and can yield very high targeting frequencies, although the threshold expression level required for successful targeting with this particular vector cannot be discerned from our limited sample.

Discussion

Several conclusions appear warranted from this analysis, some of which are more definitive than others. First, the genes trapped at random with the secretory trap vector are expressed in ES cells at higher levels than control populations of similar genes. This conclusion relies on the assumption that our control sets are representative of secretory genes in general, but this assumption

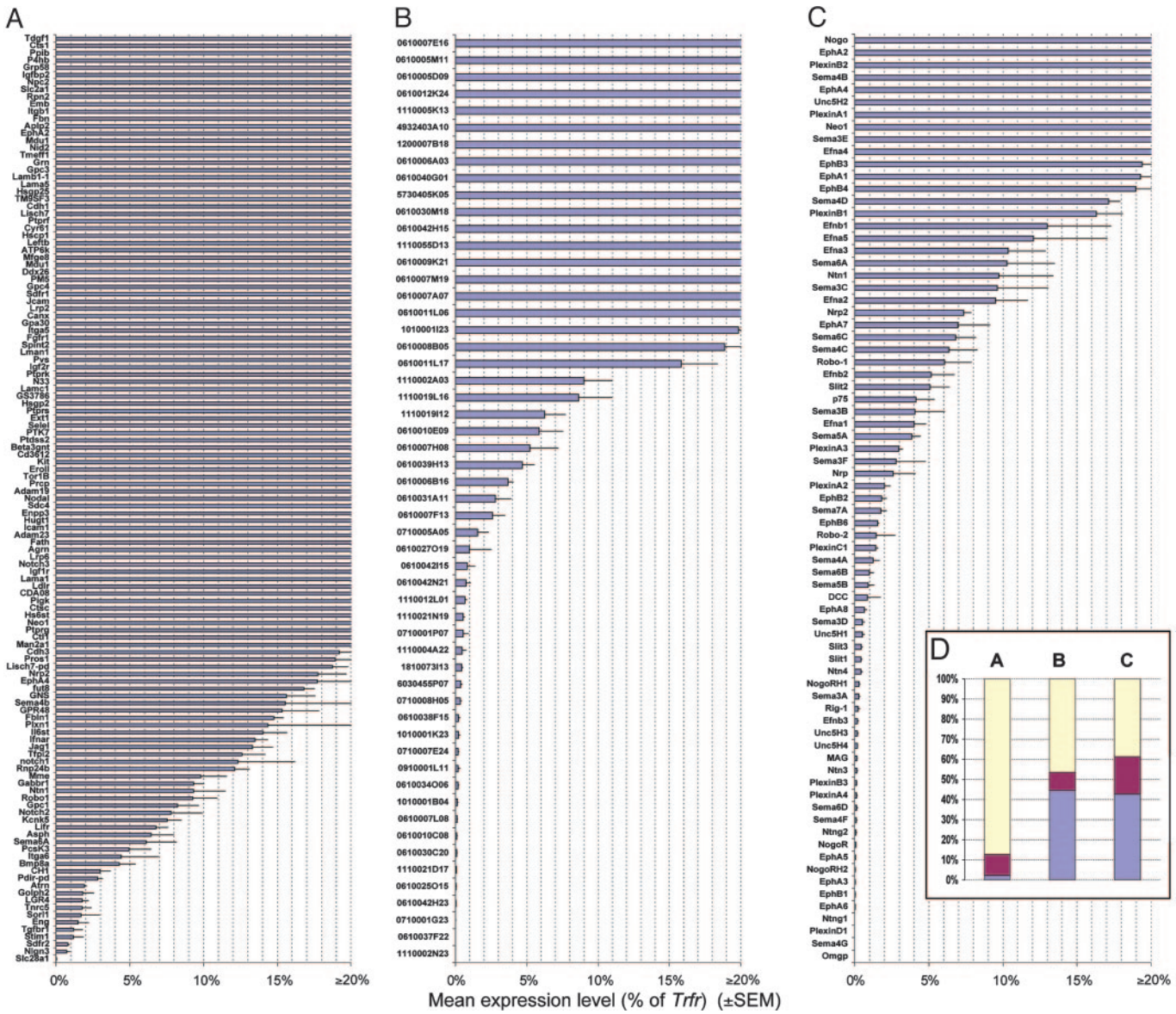

Fig. 3. Distribution of expression levels in ES cells (relative to *Trfr* expression level) for three groups of secretory proteins. Expression levels are cut off at 20% for clarity (see Tables 3–5 for values >20%). (*A*) Genes that were trapped previously by random secretory gene trapping ($n = 127$). (*B*) Control set 1: Arbitrary selection of RIKEN clones that encode secretory proteins ($n = 55$). (*C*) Control set 2: Selection of genes encoding known or suspected axon guidance/regeneration ligands and receptors ($n = 75$). (*D*) Summary of the mean expression levels in *A–C*, showing the fraction of genes predicted to be above the threshold for targeted trapping (yellow, >5%), at or near the threshold (maroon, 1–5%), and below threshold (blue, <1%).

appears valid because similar distributions were observed for two control sets selected using completely different criteria. It was predicted, based on how gene trap vectors work, that we would find a threshold of gene expression required for random trapping, and our results both bear out this prediction and establish an approximate threshold in random trapping experiments. We cannot, however, assume that this threshold marks an absolute cutoff for all genes, because we know little about how the fusion of the N-terminal part of a protein to β-geo might influence the efficiency of neo resistance.

Second, the results of our targeted trapping experiments suggest that targeted gene trapping with a promoterless vector requires a gene expression threshold similar to that found in random trapping experiments using the same vector, i.e., a threshold ≈1% of *Trfr* for the secretory trap vector used in this study. This conclusion must be considered more tentative, given the limited number of genes (twenty-four) that we attempted to target and the fact that the average expression level of the genes we chose to target was higher than that of randomly selected genes (compare Fig. 2 with Fig. 3). Further studies will be required to determine the extent to which the expression level thresholds for targeted trapping and random trapping are indeed similar. It should be noted that we did not necessarily expect the thresholds for random and targeted trapping to be the same, both because there could be different constraints for random insertions compared with homologous recombination events, and because the selection in targeted trapping was only for antibiotic resistance, whereas in random trapping, the ES cells were selected first for antibiotic resistance and second for the pattern of β-gal staining (3). This secondary screen could, in principle, have biased the selection toward higher levels of gene expression. [Arguing against this possibility, however, is our finding that a set of genes trapped at random using a nonsecretory trap vector (for which no secondary β-gal screen was used)

had a distribution toward higher expression levels similar to those trapped by secretory trapping using both antibiotic and β-gal expression selection (Table 7, which is published as supporting information on the PNAS web site)].

Third, the first two conclusions together suggest that a targeted trapping strategy might be successfully applied to at least half of all secretory genes, because about half are expressed above the 1% threshold level. Again, this conclusion must be considered tentative until more genes are targeted.

Collectively, these studies demonstrate that targeted trapping facilitates highly efficient gene targeting (i.e., it provides a high rate of successful targeting by constructs, and for each successful construct, it provides a high targeting frequency) for genes that are expressed above a minimum threshold level. The high targeting frequency presumably arises from the promoterless design of the vector, which eliminates drug resistance arising from most random insertion events. Little is known about the precise factors that influence targeting efficiency. The secondary structure of the construct and the chromatin status of the targeted locus might play important roles, potentially explaining why we failed to successfully target three genes with expression levels above 1%. We have generated our homology arms by genomic PCR, which proved to be a very fast and efficient strategy. However, we cannot rule out that PCR mutations may, in some cases, have affected the success of our targeting experiments. The use of annotated genomic clones [e.g., the MICER library (16)] as a source for homology arms could provide an efficient strategy that avoids PCR.

Our data suggest that most or all secretory genes with expression levels above 5% *Trfr* can be targeted with the secretory trap vector in the E14 ES cell line, and that some fraction, possibly sizeable, of those with expression levels of 1–5% *Trfr* can be targeted as well. If so, it would not be unreasonable to expect that ≈50% of secretory genes can be targeted with this vector in this ES cell line. Several approaches can be envisaged to extend the spectrum of genes accessible with targeted trapping, such as the use of alternative ES cell lines with gene expression profiles that differ from those of E14 ES cells.

The 5′ trapping vectors used in our study all rely on endogenous expression of the trapped gene in ES cells. A gene trap vector that was designed to be independent of target gene expression level is the polyA trap vector (17, 18). This vector, which contains a constitutively expressed resistance cassette followed by a splice donor, lacks a polyadenylation signal and should, in theory, depend on a 3′ splice event to achieve proper polyadenylation and hence a stable mRNA. However, a recent large-scale gene trap screen with a polyA trap vector revealed this strategy is less efficient than 5′ gene trapping (19), suggesting that the polyA trap selection lacks sufficient stringency. We attempted targeted trapping with a polyA trap vector by replacing the secretory trap cassette in two of our successful targeting constructs with a polyA trap cassette, but in both cases, we failed to obtain positive ES cell colonies (data not shown), and we did not pursue this strategy further. The potential of targeted trapping using polyA trap vectors remains to be explored more fully.

Finally, recombination sites, such as loxP and FRT, can be added to the targeted trapping vectors, and this should greatly facilitate the generation of alleles that can be conditionally mutated by activity of appropriate recombinases. Gene trap vectors that are flanked by recombination sites are already being used for random trapping. The insertion of such sites into the homology arms of targeted trapping constructs can, in the same way, be used to generate alleles that can be conditionally inactivated.

Conclusion

Our results provide insight into the size of the pool of genes accessible by random gene trapping and indicate that targeted trapping is a complementary and potentially more efficient way of targeting some or all of the accessible pool. Because of its simplicity and high efficiency, the method will dramatically facilitate gene knockouts by small laboratories; at the same time, the approach is scalable (e.g., by using genomic libraries to generate the targeting constructs), making it amenable to high-throughput knockouts of a large fraction of the mammalian genome.

We thank W. Skarnes (Sanger Center, Hinxton, U.K.), W. Stanford (University of Toronto, Toronto), D. Rotin (University of Toronto), and members of the M.T.-L. and S.K.M. laboratories for plasmids, advice, and helpful discussions; and D. Lewis and T. Giffon for help with the real-time PCR system. Funding for this project was provided by grants to M.T.-L. from the National Institutes of Mental Health (Grant RO1MH60612, Identifying Brain Wiring Mechanisms by Gene Trapping) and from the National Heart, Lung, and Blood Institute (NHLBI) (Grant U01HL66600, the NHLBI–Bay Area Functional Genomics Consortium). T.R.V. was supported by National Institutes of Health Grants S06GM008168 and 5G12RR03060. R.H.F. was supported by a fellowship of the Deutsche Forschungsgemeinschaft; X.L., by a fellowship of the Damon Runyon Foundation; and M.T.-L., by the Howard Hughes Medical Institute.

1. Stanford, W. L., Cohn, J. B. & Cordes, S. P. (2001) *Nat. Rev. Genet.* **2,** 756–768.
2. Skarnes, W. C., von Melchner, H., Wurst, W., Hicks, G., Nord, A. S., Cox, T., Young, S. G., Ruiz, P., Soriano, P., Tessier-Lavigne, M., *et al.* (2004) *Nat. Genet.* **36,** 543–544.
3. Skarnes, W. C., Moss, J. E., Hurtley, S. M. & Beddington, R. S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6592–6596.
4. Schwartzberg, P. L., Goff, S. P. & Robertson, E. J. (1989) *Science* **246,** 799–803.
5. Jeannotte, L., Ruiz, J. C. & Robertson, E. J. (1991) *Mol. Cell. Biol.* **11,** 5578–5585.
6. Stanton, B. R., Reid, S. W. & Parada, L. F. (1990) *Mol. Cell. Biol.* **10,** 6755–6758.
7. te Riele, H., Maandag, E. R., Clarke, A., Hooper, M. & Berns, A. (1990) *Nature* **348,** 649–651.
8. Charron, J., Malynn, B. A., Fisher, P., Stewart, V., Jeannotte, L., Goff, S. P., Robertson, E. J. & Alt, F. W. (1992) *Genes Dev.* **6,** 2248–2257.
9. Lanier, L. M., Gates, M. A., Witke, W., Menzies, A. S., Wehman, A. M., Macklis, J. D., Kwiatkowski, D., Soriano, P. & Gertler, F. B. (1999) *Neuron* **22,** 313–325.
10. Wallace, M. J., Batt, J., Fladd, C. A., Henderson, J. T., Skarnes, W. & Rotin, D. (1999) *Nat. Genet.* **21,** 334–338.
11. Leighton, P. A., Mitchell, K. J., Goodrich, L. V., Lu, X., Pinson, K., Scherz, P., Skarnes, W. C. & Tessier-Lavigne, M. (2001) *Nature* **410,** 174–179.
12. Tate, P., Lee, M., Tweedie, S., Skarnes, W. C. & Bickmore, W. A. (1998) *J. Cell Sci.* **111,** 2575–2585.
13. Skarnes, W. C. (2000) *Methods Enzymol.* **328,** 592–615.
14. Mitchell, K. J., Pinson, K. I., Kelly, O. G., Brennan, J., Zupicich, J., Scherz, P., Leighton, P. A., Goodrich, L. V., Lu, X., Avery, B. J., *et al.* (2001) *Nat. Genet.* **28,** 241–249.
15. Valenzuela, D. M., Murphy, A. J., Frendewey, D., Gale, N. W., Economides, A. N., Auerbach, W., Poueymirou, W. T., Adams, N. C., Rojas, J., Yasenchak, J., *et al.* (2003) *Nat. Biotechnol.* **21,** 652–659.
16. Adams, D. J., Biggs, P. J., Cox, T., Davies, R., van der Weyden, L., Jonkers, J., Smith, J., Plumb, B., Taylor, R., Nishijima, I., *et al.*. (2004) *Nat. Genet.* **36,** 867–871.
17. Salminen, M., Meyer, B. I. & Gruss, P. (1998) *Dev. Dyn.* **212,** 326–333.
18. Zambrowicz, B. P., Friedrich, G. A., Buxton, E. C., Lilleberg, S. L., Person, C. & Sands, A. T. (1998) *Nature* **392,** 608–611.
19. Zambrowicz, B. P., Abuin, A., Ramirez-Solis, R., Richter, L. J., Piggott, J., BeltrandelRio, H., Buxton, E. C., Edwards, J., Finch, R. A., Friddle, C. J., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 14109–14114.