nature biotechnology

High-throughput engineering of the mouse genome coupled with high-resolution expression analysis

David M Valenzuela[1], Andrew J Murphy[1,3], David Frendewey[1,3], Nicholas W Gale[1,3], Aris N Economides[1,3], Wojtek Auerbach[1], William T Poueymirou[1], Niels C Adams[1], Jose Rojas[1], Jason Yasenchak[1], Rostislav Chernomorsky[1], Marylene Boucher[1], Andrea L Elsasser[1], Lakeisha Esau[1], Jenny Zheng[1], Jennifer A Griffiths[1], Xiaorong Wang[1], Hong Su[1], Yingzi Xue[1], Melissa G Dominguez[1], Irene Noguera[1], Richard Torres[1], Lynn E Macdonald[1], A Francis Stewart[2], Thomas M DeChiara[1] & George D Yancopoulos[1]

One of the most effective approaches for determining gene function involves engineering mice with mutations or deletions in endogenous genes of interest. Historically, this approach has been limited by the difficulty and time required to generate such mice. We describe the development of a high-throughput and largely automated process, termed VelociGene, that uses targeting vectors based on bacterial artificial chromosomes (BACs). VelociGene permits genetic alteration with nucleotide precision, is not limited by the size of desired deletions, does not depend on isogenicity or on positive–negative selection, and can precisely replace the gene of interest with a reporter that allows for high-resolution localization of target-gene expression. We describe custom genetic alterations for hundreds of genes, corresponding to about 0.5–1.0% of the entire genome. We also provide dozens of informative expression patterns involving cells in the nervous system, immune system, vasculature, skeleton, fat and other tissues.

Now that the human genome has been sequenced[1,2], the next challenge is to assign function to the newly identified genes. Perhaps the most effective approach for determining gene function involves deliberately engineering gene mutations in mouse embryonic stem (ES) cells, and then generating mice harboring the corresponding genetic changes[3–6]. The resulting phenotypes often provide insight into the function of the altered genes. Unfortunately, this approach is dependent upon highly skilled scientists employing time-consuming methodologies to produce custom alterations on a gene-by-gene basis[7]. The two limiting steps are the generation of 'gene targeting vectors' and the subsequent selection of rare ES cell clones in which the targeting vector has correctly altered the gene. To bypass the difficulties of generating mice with custom genetic alterations, alternative approaches have focused on generating random mutations, using either chemical mutagenesis[8,9] or gene-trap methodologies[10–12]. Approaches that can deliberately target and manipulate a gene (or genes) of interest have obvious advantages, if they can compete with the efficiency of random-mutagenesis methods.

Here we describe VelociGene, a high-throughput, largely automated approach that uses BAC-based targeting vectors to alter many genes in parallel. Similar BAC-based vectors have been used to generate random transgenic insertions in mice[13–16], but not to target endogenous genes. VelociGene is at least equivalent in throughput to random gene-trapping approaches[11,12] yet allows the deliberate modification of any gene of interest.

RESULTS AND DISCUSSION

Comparison to current targeting technologies

To produce a desired genetic alteration in ES cells, one must first introduce the alteration into a targeting vector that is subsequently used to replace the native gene in ES cells by homologous recombination[5,6]. The homologous recombination occurs between DNA segments flanking the alteration on the targeting construct and the homologous DNA segments in the native gene. This process is inefficient; most targeting vectors introduced into ES cells insert randomly throughout the genome, often with less than 0.1–1.0% integrating homologously[17]. Several factors seem to improve the rate of homologous replacement: longer homologous flanking sequences on the targeting vector (though lengths up to only about 10–20 kb have been previously explored)[17–20]; homologous flanking sequences that are completely isogenic to those in the ES cell[21,22]; limiting the size of the deletion[17]; and use of positive–negative selection to bias against random integrants[23].

Although almost any type of genetic alteration can theoretically be introduced via targeting constructs, simple deletions spanning key parts of the gene of interest are most common as it is often difficult to produce highly customized genetic alterations with large DNA targeting vectors. For example, precise replacements of a gene's coding region with a reporter gene[24–26] are not routinely exploited because they have been difficult to engineer in targeting vectors with long flanking regions[7]. Finally, the process of identifying the rare ES cells in which homologous recombination has occurred involves time-consuming assays including Southern blotting or long PCR[7]. Such assays depend on finding ES clones in which sequences within the native locus (but outside of the flanking regions shared by the targeting vector) have been linked to sequences unique to the targeting vector; such linking of sequences

[1]Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill Road, Tarrytown, New York 10591, USA. [2]BIOTEC, Technische Universitat Dresden, Dresden, Germany 01307. [3]These authors contributed equally to this work. Correspondence should be addressed to G.D.Y. (velocigene@regeneron.com).

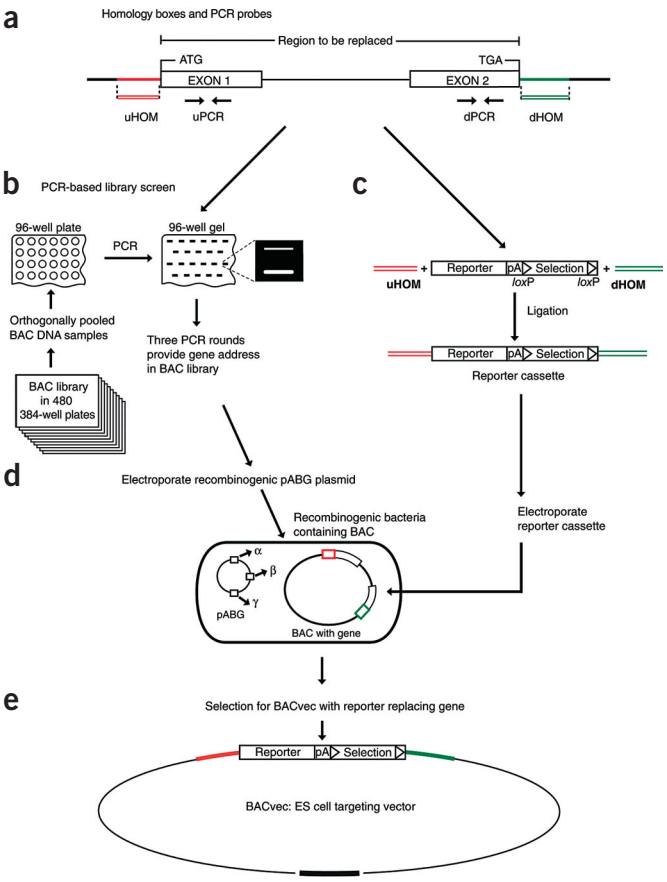

Figure 1 Schematic representation of approach used to generate BAC-based targeting vectors (BACvecs). (**a**) Gene of interest and oligonucleotides that must be synthesized for the gene. (**b**) Approach used to isolate BAC containing gene of interest from an orthogonally matrixed BAC library. (**c**) Generation of reporter-selection cassette flanked by homology boxes (uHOM and dHOM) that will target homologous recombination in bacteria. (**d**) Recombinogenic bacteria (containing the pABG plasmid encoding recombination enzymes, as well as the BAC containing the gene of interest), into which the reporter-selection cassette is introduced. (**e**) Final BACvec in which gene of interest is precisely replaced with reporter and selection genes; after linearization, this BACvec is ready for introduction into ES cells.

only occurs in correctly targeted ES cells. Currently, the design of the targeting vector is limited by the assay used to screen for correctly targeted ES cell clones[7], effectively limiting the use of longer homology arms on the targeting vectors simply because they would further complicate the screening assay.

In an attempt to streamline the production of targeting vectors, provide flexibility in customizing genetic alterations, and allow constructs of optimal length, we exploited bacterial homologous recombination together with BACs[13,15,16,27–32]. BACs offer several advantages: they have high stability, may propagate up to 300 kb of DNA, and can be easily produced and isolated. BAC engineering by bacterial homologous recombination involves the transient and controlled expression of genes that mediate homologous recombination in *Escherichia coli*, thereby allowing the bacteria to mediate recombination between a 'modification cassette' and a BAC sharing short homologous stretches[15,16,29–32]. We applied simple variations of this approach to generate almost any type of gene alteration, including conditional alleles, point mutations and gene swaps.

Our approach to generate targeting vectors in which the coding region of a gene of interest is precisely replaced by a reporter gene involves the following steps. First, a series of oligonucleotides are synthesized for the target gene (**Fig. 1a**): a 50–200 base pair (bp) double-stranded oligonucleotide exactly matching the native sequence just upstream from the ATG initiation codon of the gene of interest (uHOM); a 50–200 bp double-stranded oligonucleotide exactly matching the native sequence just downstream from the TGA termination codon of the gene of interest (dHOM); and two pairs of 17–24 nucleotide (nt) gene-specific single-stranded oligonucleotides just inside the two deletion points (uPCR and dPCR primers, as indicated).

Second, BACs containing the gene of interest are identified using either a BAC end sequence database (such as that at http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html) or a preconstructed BAC map of the mouse genome (http://www.ncbi.nlm.nih.gov/genome/guide/mouse; http://www.ensembl.org/Mus_musculus/cytoview)[33]. Alternatively, the PCR primer sets, uPCR and dPCR, are used to generate two short gene-specific PCR products for screening formatted BAC libraries (**Fig. 1b**).

Third, the double-stranded uHOM and dHOM oligonucleotides are ligated to the ends of a linear reporter cassette, in which a reporter gene (for example, one encoding β-galactosidase) is in tandem with a selection gene (such as one encoding neomycin resistance) flanked by

Table 1 Summary information on 10 of 200 targeted alleles

VG No.[a]	Gene name	Accession no.[b]	Protein class	Focus area	Deletion size (kb)	Homologous arm size (kb)[c]	Percent targeting[d]
100	C5L2	NM_176912	7TM	Inflammation and immunity	1	110	4.2
101	C5aR	NM_007577	7TM	Inflammation and immunity	1	105	4.2
102	CNTFR	NM_016673	GPI	Obesity and metabolism	17	71	4.2
103	IL17R	NM_008359	1TM	Inflammation and immunity	19	94	2.1
104	IL18	NM_008360	Secreted	Inflammation and immunity	10	110	8.3
106	CLF-1	NM_018827	Secreted	Obesity and metabolism	10	26	1.4
107	EphA2	NM_010139	1TM	Unknown	8	135	3.1
108	EphA2-am2	NM_010139	1TM	Unknown	8	135	5.2
109	PIGF	NM_008827	Secreted	Vascular	8	60	3.5
110	NaDC-1	AF201903	12TM	Aging	25	75	2.1

[a]Targeted alleles, 200 (see **Supplementary Table 1** online for list of all 200 targeted alleles). [b]NCBI GenBank. [c]Average homology (200 alleles), 112 kb. [d]Average targeting frequency (200 alleles), 3.8%.

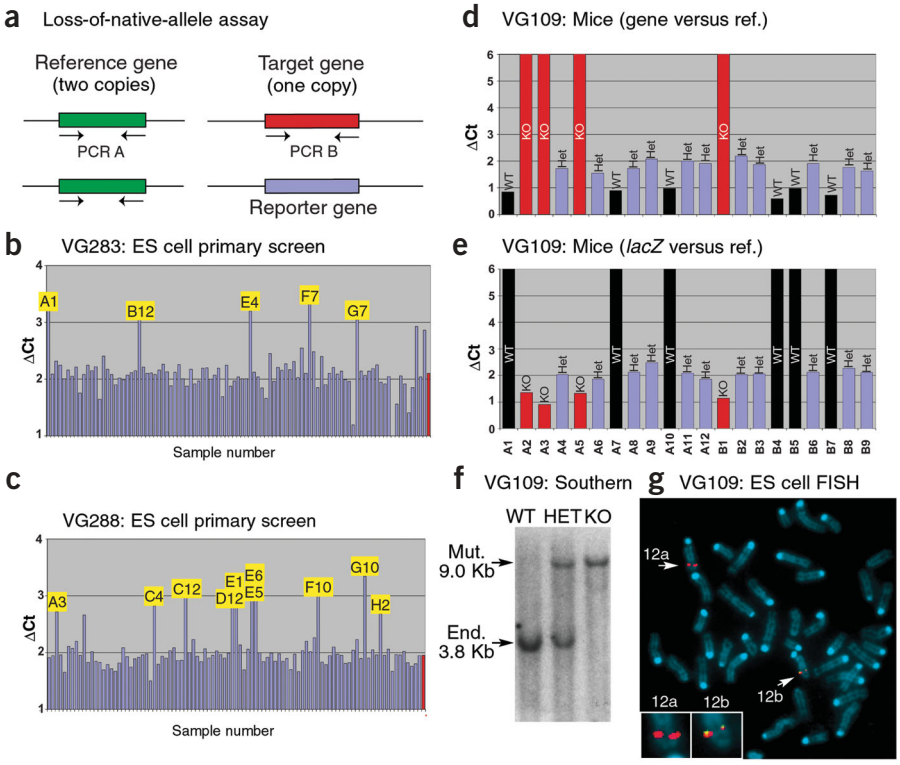

Figure 2 Loss-of-native-allele assay. (**a**) The target gene is reduced from two copies to one copy in the genome, in contrast to the reference gene. (**b**,**c**) Two primary screens of 96 ES cell clones, in each of which a single uPCR or dPCR probe is compared to a single reference gene to yield a ΔC_T value; yellow boxtops indicate correctly targeted ES cell clones and red bar on right indicates median value for 96 samples. (**d**, **e**) Screen of tail DNAs, demonstrating that mice wild-type (WT), heterozygous (Het) and homozygous (KO) with respect to the genetic alteration can be unequivocally distinguished by two parallel locus quantification assays—KOs are completely missing the native allele and thus exhibit an infinite ΔCt when quantifying the gene of interest versus a reference gene, whereas WTs lack the introduced reporter *lacZ* gene and thus exhibit an infinite ΔCt when quantifying *lacZ* versus a reference gene; Het mice are unique in that they are intermediate in both assays. (**f**) Southern analysis to confirm that mice derived from putatively targeted ES clones contain predicted genetic alterations. End., endogenous native allele. Mut., modified mutant allele. (**g**) FISH confirmation of correctly targeted ES clones identified by 'loss-of-native allele' assay. Hybridization with the parental BAC probe (red) identified two sets of chromatids on both chromosome homologs (12a and 12b), as expected; hybridization with the inserted reporter (*lacZ*) probe (green), which appears yellow (12b) when overlapping the red BAC probe signal) identifies the chromatids on only one of the chromosome 12 homologs (12b), as expected, corresponding to the endogenous allele into which the reporter has replaced the native sequence.

loxP sites that allows for positive selection in both bacterial and mammalian cells (using a dual promoter system combining the eukaryotic phosphoglycerate kinase (PGK) promoter with the prokaryotic EM7 promoter) (**Fig. 1c**).

Fourth, the selected bacteria sample, harboring the BAC containing the gene of interest, is prepared for the homologous recombination step by introduction of a plasmid (pABG) encoding the recombination activities under the control of a transiently inducible promoter (**Fig. 1d**).

Fifth, the linear 'uHOM-reporter cassette-dHOM' product derived from step 3 above is introduced by electroporation into the recombination-competent bacteria harboring the BAC as well as the pABG plasmid encoding recombination enzymes (**Fig. 1c**,**d**). The recombination enzymes are transiently induced by adding arabinoside, and bacteria in which the BAC has precisely incorporated the reporter cassette are selected using the drug-selectable marker incorporated into the reporter cassette (**Fig. 1e**). The modified BAC is confirmed by sequencing and restriction enzyme analysis.

Screening ES cells targeted with BAC-based vectors

The above procedures allow rapid generation of very large BAC-based targeting vectors (BACvecs). We generated 200 BACvecs containing, on average, over 100 kb of flanking DNA arms and allowing deletions of at least 75 kb (**Table 1** and **Supplementary Table 1** online). BAC-based vectors have not been used previously to homologously target genes in ES cells, primarily because their long flanking arms are not compatible with current approaches (involving Southern blotting and long PCR across such arms) to screen for correctly targeted ES cells. Instead, bacterial homologous recombination has been used to generate more conventionally sized target vectors, by starting with smaller initial genomic clones or by trimming BAC-based constructs in subsequent steps[30]. In addition to requiring additional custom modifications for each vector, such approaches also lose the potential advantages afforded by very long flanking arms.

To use BACvecs to target genes in ES cells, we devised an alternate screening strategy amenable to automation, which we term a 'loss-of-native-allele' assay. Native gene sequences in ES cells are normally present in diploid copy number, except for genes on the X and Y chromosomes, which are represented once. Thus, an ES cell in which the native gene sequence was successfully altered loses one copy of the sequence, while ES cells with unsuccessful targeting events (for example, random integration of the BACvec) retain all copies of the sequence. We developed an optimized variation of real-time PCR that reliably detects ES clones in which loss of one native allele has occurred (**Fig. 2a**). The assay employs the same uPCR and dPCR amplification primers (marking the ends of the desired replacement, **Fig. 1a**) used to select the BAC in step 2 above, but the primers are amplified in combination with a fluorescent probe that is quantitatively activated as amplified product accumulates. The assay is performed robotically, using ES clones formatted in a 96-well format, and the data are analyzed digitally.

The assay not only screens for correctly targeted ES clones (**Fig. 2b**,**c**) but can also be used to screen the resulting mice—in automated fashion it distinguishes mice that are wild type, heterozygous or homozygous for the desired genetic alteration (**Fig. 2d**,**e**). Southern blotting can be used to validate the mice derived from ES clones selected by the 'loss-of-native-allele' assay, as mice homozygous for a gene deletion would completely lose the native restriction enzyme fragment encompassing the gene deletion (**Fig. 2f**). The 'loss-of-

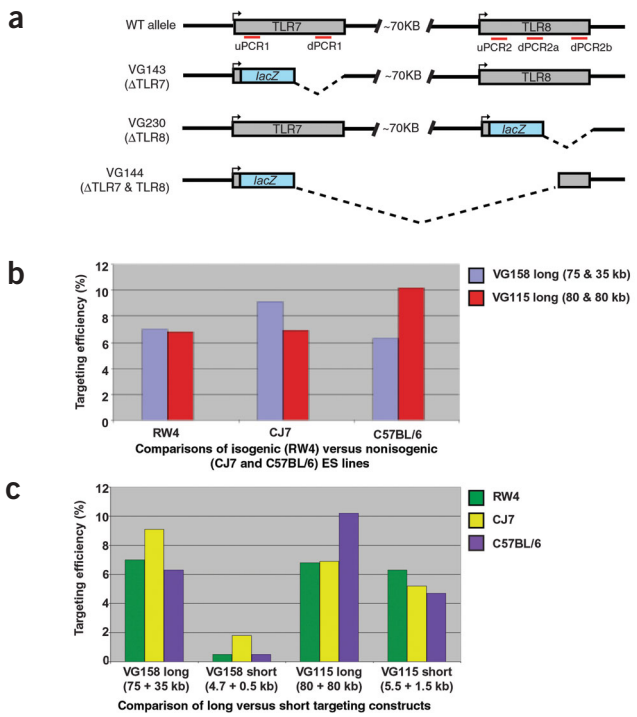

Figure 3 Generation of complex alleles and large deletions, and targeting efficiencies. (**a**)Three altered alleles involving two adjacent related genes (Toll receptor 7, TLR7, and Toll receptor 8, TLR8) are depicted. In the first altered allele (VG143), only TLR7 is deleted and replaced with the *lacZ* reporter; in the second altered allele (VG230), only TLR8 is deleted and replaced with the *lacZ* reporter; whereas in the third allele (VG 144), both TL7 and TLR8 are deleted, with *lacZ* replacing TLR7. (**b**) BACvecs do not show isogenicity dependence, as demonstrated by the fact that two different BACvecs target with similar frequencies into isogenic ES cells (129X1/SvJ-derived RW4 cells), into closely related but nonisogenic ES cells (129S1/Sv-derived CJ7 cells) and into distantly related ES cells (C57Bl/6J-derived ES cells). (**c**) BACvecs did not seem to offer the advantage of greater targeting frequency as compared to shorter targeting vectors when the latter already targeted at high rates (see VG115; compare long and short targeting constructs), but they did show substantially greater targeting frequency than did shorter targeting vectors that targeted at low or undetectable rates (see VG158; compare long and short targeting constructs).

native-allele' screening approach works well for autosomal genes (**Fig. 2a–f**) as well as for genes on the X and Y chromosomes (**Fig. 3a** and **Table 2**). The latter screen shows how this approach works to delete either of a pair of adjacent related genes separately (replacing the deleted gene with a reporter), while maintaining the integrity and copy number of the remaining adjacent gene; it also demonstrates the technique's ability to remove both genes simultaneously through a very large (>70-kb) deletion (**Fig. 3a** and **Table 2**). Extensive analysis of five targeted alleles, including the recloning, restriction mapping, and sequence analysis of BACs representing the targeted allele, indicated that the alleles contained only the desired genetic alterations without unwanted deletions, duplications, or events corresponding to single homologous recombination (data not shown). No single homologous recombination events, which would be detected by our 'loss-of-native-allele' assay as a discordance between uPCR and dPCR results, were seen in the 200 targeted alleles described; our data suggest that the use of long homology arms decreases the frequency of such aberrant targeting events.

Table 2 'Loss-of-native-allele' quantification for X-linked gene

	uPCR1	dPCR1	uPCR2	dPCR2a	dPCR2b
WT allele	1	1	1	1	1
VG142 (ΔTLR7)	0	0	1	1	1
VG230 (ΔTLR8)	1	1	0	0	0
VG144 (ΔTLR7 and ΔTLR8)	0	0	0	0	1

The table summarizes the loss-of-allele quantification, showing that the desired alterations were obtained; because these genes are on the X chromosome, they are normally present in one copy, and no copies remain after deletion.

Because Southern blot confirmation of targeting using BACvecs can be done only on homozygous mutant mice (**Fig. 2f**), we used a different assay to verify ES cells selected by the initial 'loss-of-native-allele' assay. This secondary assay involves fluorescence *in situ* hybridization (FISH) of chromosomes in metaphase spreads from the candidate ES clones, using as probes both the parental BAC representing the gene of interest and a probe specific for the reporter gene. Correct targeting would result in cells in which the parental BAC probe identifies the same location on the paired chromatids of two different chromosomes (and as the location of most mouse genes is now known based on their sequence, this BAC-labeled chromosomal locus should correspond to the known position), whereas the reporter probe precisely double-labels only one of these loci (**Fig. 2g**). In 61 of 61 ES clones examined and predicted to be correctly targeted based on the 'loss-of-native-allele' assay, confirmation was obtained with the FISH-based assay. This assay is not of sufficiently high throughput to serve as a primary screen but is useful for confirming candidate ES clones. In addition, the assay allows for karyotypic characterization of ES cell candidates, enabling selection of euploid clones known to generate higher frequencies of germline transmission than do ES cells with chromosomal abnormalities[17].

Use of non-isogenic targeting vectors and large deletions

We used Velocigene to generate hundreds of targeted genetic alterations in ES cells and mice (200 of which are summarized in **Table 1** and **Supplementary Table 1** online). Deletions were as large as 70 kb. The average confirmed targeting frequency observed was 3.8%, and very few targeting vectors produced rates below 1%. These average targeting rates are notably better than would be expected from conventional targeting vectors that do not use positive–negative selection, suggesting that the very long homology arms on these vectors may be aiding the targeting process, consistent with previous studies showing that increasing arm sizes up to 20 kb is beneficial[17–20].

Conventional targeting vectors seem to be limited by the extent of homology between the vector flanking arms and the native gene, so that completely isogenic arms may be required for optimal targeting efficiency[21,22]. To explore whether our very large BAC-based targeting vectors may remove this requirement, we compared the targeting frequencies of several of our vectors in three types of cells: isogenic 129X1/SvJ-derived 'RW4' ES cells; nonisogenic but closely related 129S1/Sv-derived 'CJ7' ES cells; and distantly related 'C57BL/6' ES cells. BACvecs do not seem to require isogenicity, presumably because of their very long homology arms (**Fig. 3b**). Although we show that the BACvecs need not be isogenic with the ES cell DNA, it is likely that their homology arms contain long sequences identical with the non-isogenic target ES DNA, and that recombination is occurring within these regions of sequence identity.

Of the first 225 genes targeted in nonisogenic ES cells, only 5 (CD22, NM_009845; Fli, NM_008026; IL-1R4, NM_010743; PROX1, NM_008937; SPRY3, XM_142331) did not exhibit targeting in the first 288 ES clones screened. No further clones were screened for these genes, and no further determination was made as to why targeting was not successful with these genes, or whether targeting would have occurred at higher frequency in isogenic ES cells for these rare first-pass failures.

As compared with conventional, shorter vectors targeting the same genes (**Fig. 3c** and data not shown), BACvecs did not have a higher targeting frequency when the shorter vectors had high targeting rates (**Fig. 3c**, VG115, long and short), but did have higher targeting frequencies compared with shorter targeting vectors with low targeting rates (**Fig. 3c**, VG158, long and short).

Finally, BACvecs enable very large chromosomal deletions to be engineered as efficiently as smaller deletions. For example, within the same locus (**Fig. 3a**), deletions of over 70 kb were generated with similar frequency to those of a few kilobases (see targeting frequencies for VG nos. 143 and 144 in **Supplementary Table 1** online). Note that the rates reported herein for VG158 are distinct from those reported in **Supplementary Table 1** online, as they are the results of newer and simultaneous transfection of the three ES lines with the long and short versions of the construct. This was deemed the fairest way to carry out the comparison because targeting rates can vary depending on the batch and status of the ES cells; once again, no advantage of isogenicity was observed even for short constructs.

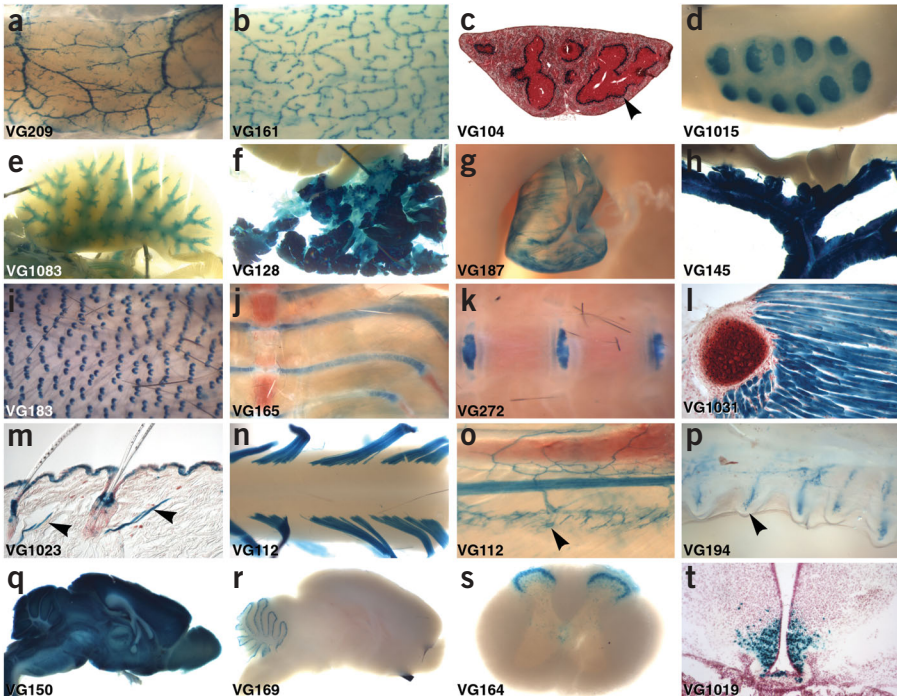

Figure 4 High-throughput screening of reporter gene expression (*lacZ* reporter, in blue) in various VelociGene-generated mice (in whole-mount views, unless indicated otherwise). (**a**) Blood vessels, intestinal surface. (**b**) Lymphatic vessels, intestinal surface (note distinctive vessel patterns differing from those in **a**). (**c**) Expression only at marginal zones between red and white pulp, noted in spleen sections. (**d**) Peyer's patches, intestine. (**e**) Bronchial epithelium of lungs. (**f**) Pancreas. (**g**) Gall bladder wall. (**h**) Mesenteric fat surrounding mesenteric vessels. (**i**) Hair follicles in skin. (**j**) Cartilaginous portions of ribs. (**k**) Intervertebral discs. (**l**) Large skeletal muscles, cross-sectional view of lower limb. (**m**) Tiny arrector pili muscles (arrowheads) at base of hair shafts, in sebaceous glands, and epidermis, viewed in skin sections. (**n**) Expression restricted to large nerve roots as they exit spinal cord. (**o**) Nerves and terminal fibers as they branch and synapse into muscle fibers, high-power image of rib cage. (**p**) Expression only in nerve terminals in teeth (arrowhead), high-power image of teeth. (**q**) High-level expression throughout brain. (**r**) Expression restricted to cerebellum. (**s**) Expression restricted to dorsal laminae of gray matter of spinal cord, corresponding to neuronal layer specifically involved in receiving afferent pain input, noted in sections of spinal cord. (**t**) Expression restricted to arcuate nucleus, corresponding to neurons involved in regulation of food intake, noted in sections of brain in hypothalamic area.

High-resolution expression profiling

As a first step towards understanding the function of the 200 genes listed in **Table 1** and **Supplementary Table 1** online, we performed high-resolution expression analyses of the reporter genes in mice derived from the BACvecs, using both whole-mount and cross-sectional methodologies optimized for high throughput. Many specific and informative expression patterns were noted (**Figs. 4** and **5**). Genes specific to either blood or lymphatic vasculature were noted (**Fig. 4a**,**b**), to different immune structures such as the spleen and Peyer's patches (**Fig. 4c**,**d**), and to structures such as the bronchi of the lungs, the pancreas, the gall bladder, the mesenteric fat surrounding the mesenteric vessels, the hair follicles, the ribs and the vertebral discs (**Fig. 4e–k**). This approach provides increasing resolution in terms of sites of expression, as is demonstrated by an example of a gene expressed highly in large skeletal muscles (**Fig. 4l**) as well as a gene expressed in the smallest muscles of the body—the *arrector pili* that make hair shafts stand on end (**Fig. 4m**). The power of increasing resolution is also highlighted by a gene specifically expressed in large nerve roots exiting the spinal cord (**Fig. 4n**), but also in the smallest motor nerve terminals as they branch to make terminal synapses with muscle (**Fig. 4o**). Another gene is expressed in a different set of nerve terminals, those found in teeth (**Fig. 4p**). Examples were seen of genes expressed broadly throughout the brain (e.g. **Fig. 4q**), as well as in very discrete regions, such as the cerebellum (**Fig. 4r**), the dorsal laminae of the gray matter in the spinal cord specifically involved in pain sensation (**Fig. 4s**), and the arcuate nucleus of the hypothalamus involved in regulation of food intake (**Fig. 4t**).

Detailed expression surveys of many genes simultaneously revealed multiple genes specific to each of the locations noted above (**Fig. 5**). The subdistributions within these locations often differed markedly, hinting at related but distinct roles for the gene products in these tissues. For example, not only did we observe multiple genes specific to Peyer's patches, but these genes could be distinguished based on whether they were expressed in the periphery of germinal centers, within the germinal centers themselves or within the lymphatic vessels surrounding and draining into the patches (**Fig. 5a**). Similarly, not only did we observe multiple genes specific to the spleen, but some of the genes were in immune cells at the marginal zone between the red and white pulp, whereas others

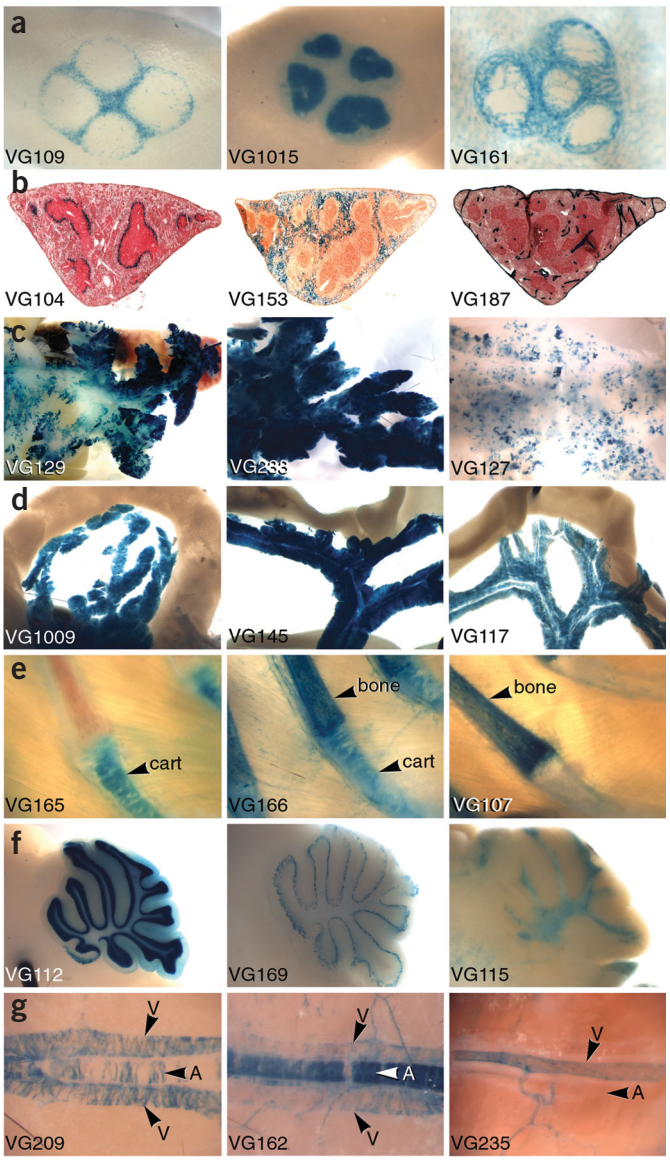

Figure 5 Examples of genes expressed in similar locations with distinctive subdistributions (in whole-mount views, unless indicated otherwise). (**a**) Three genes with differing patterns in Peyer's patches: left, at periphery of germinal zones; middle, in germinal zones themselves; right, in lymphatic vessels surrounding germinal zones. (**b**) Three genes with differing patterns in spleen (cross-sectional analysis): left, in immune cells at the marginal zone between the red and white pulp; middle, broadly expressed throughout the white pulp; right, not expressed in either red or white pulp, but expressed in splenic myotrabeculae and capsule. (**c**) Three genes with differing patterns in pancreas. (**d**) Three genes expressed similarly in mesenteric fat, surrounding mesenteric vessels. (**e**) Three genes expressed in ribs with differing patterns: left, restricted to cartilaginous portion of rib; middle, expressed in both cartilaginous and bony portion of rib; right, expressed only in bony part of rib. (**f**) Three genes expressed with differing sub-localizations in cerebellum: left, expressed in granule cell layer; middle, expressed in Purkinje cells; right, expressed in white matter in oligodendrocytes. (**g**) Three genes expressed with differing sub-localizations in blood vessels (high-power views of blood vessels in intercostal region): left, restricted to circumferential smooth muscle cells coating both arterial, A, and venous, V, vessels; middle, preferentially expressed in circumferential smooth muscle cells encircling arterial as opposed to venous vessels; right, restricted to the endothelium of veins.

were more broadly expressed throughout the white pulp, and yet others were found in splenic myotrabeculae (**Fig. 5b**). Multiple genes were found specific to the pancreas (**Fig. 5c**), fat (**Fig. 5d**), ribs (**Fig. 5e**), cerebellum (**Fig. 5f**) and blood vasculature (**Fig. 5g**). Once again, most could be distinguished by their differing subdistributions; in the ribs, genes were expressed only within the cartilage, in both cartilage and bone, or in bone alone (**Fig. 5e**), whereas genes in the cerebellum were noted in the granule cells, Purkinje cells or oligodendrocytes (**Fig. 5f**). The vasculature provided enormous diversity at very high resolution: genes were seen expressed in circumferential smooth muscle cells encircling both arteries and veins, preferentially in smooth muscle cells encircling arterial vessels, or only in the endothelium of veins (**Fig. 5g**). We are pursuing biological analyses based on such initial data. For example, we have recently shown that two genes specifically expressed in muscle are required for the muscle atrophy response to injury[34], whereas other genes regulate food intake and obesity, cartilage development, reproductive function, immune function and other important biological processes (data not shown).

Conclusion

Just as development of the automated DNA sequencer enabled large-scale efforts that succeeded in sequencing entire genomes, high-throughput approaches are necessary to assign functions to the entire set of genes within the genome. Although some of the bottlenecks in generating genetically engineered mice have been addressed by recent advances, such as techniques to speed the generation of targeting vectors[30,35], major limitations remain. We have described a high-throughput approach that allows for the rapid generation of mice bearing almost any genetic alteration in the gene of interest, including conditional alleles, point mutations or swaps of related genes. The approach delivers nucleotide precision, allows deletion of regions of at least 70 kb, and does not require isogenic targeting vectors or positive–negative selection.

We focused here on examples in which the genetic alteration involved precise replacement of the gene of interest with a reporter gene, allowing for high-resolution localization of the gene of interest *in vivo*. Although our screen of two hundred genes provided valuable insights into their potential roles, detailed biological efforts will be required to follow up on the initial observations. However, our initial screen facilitates educated choices, triages the genes into likely areas of biological relevance and expertise, and provides mice genetically altered for the gene of interest as tools for subsequent studies.

An effort of the scale described here requires comprehensive data management. The VelociGene process uses an instrumental database solution that tracks molecular constructs, ES cell targeting, expression analysis and imaging, genotyping, phenotyping and mouse colony management. As an additional aid in phenotyping, tissues from the knockout mice are being subjected to high-throughput microarray-based gene expression profiling to determine the molecular consequences of each genetic manipulation in all affected tissues (C. Lin *et al.*, personal communication). The information obtained is being used to generate a comparative database that links each genetic alteration with its effects on gene expression and relates these to gene expression changes seen in models of human disease.

METHODS

Construction of pABG plasmid and the reporter-selection cassette. The lambda phage red exo (alpha), red bet (beta) and gam (gamma) coding sequences were cloned into the vector pBAD-TOPO (Invitrogen) to generate pABG, which

allows controlled expression of the three lambda proteins from an arabinose promoter as a polycistronic message in the order exo, bet, gam. The reporter-selection cassette contained the following in sequence: the *E. coli lacZ* gene, the SV40 late polyadenylation signal and a unit flanked by *lox*P sites that consists of the promoter of the gene encoding mouse PGK, the *E. coli* EM7 promoter, the neomycin-kanamycin resistance gene and the mouse PGK gene polyadenylation signal. This reporter and selection cassette was ligated to homology boxes ranging in size from 50 to 200 bp that were synthesized by PCR.

Construction of BAC-based targeting vectors. We used Easy-to-Screen DNA Pools (Incyte Genomics) to isolate target BACs from an *E. coli* DH10B library of 184,320 clones generated from the 129X1/SvJ-derived RW4 mouse ES cell line inserted in the pBeloBAC11 vector. A nested matrix strategy identified the target BAC clone in three rounds of PCRs analyzed with the E-gel 96 High-Throughput Agarose Electrophoresis System (Invitrogen). Bacterial clones harboring the target BAC were made electrocompetent by resuspension in cold 10% glycerol and then electroporated with pABG in a 0.1 ml cuvette (Bio-Rad), followed by plating on Luria broth (LB) agar containing chloramphenicol (12.5 μg/ml) and ampicillin (50 μg/ml). Colonies containing both the pABG plasmid and the target BAC were picked and inoculated into 125 ml of LB plus chloramphenicol and ampicillin until the culture reached an OD_{600} of 0.2; arabinose was then added. We made the bacteria electrocompetent, as above, mixed 50 μl of the cell suspension with 0.3 μg of the reporter-selection cassette and electroporated as above. We then cultured the cells for 3–4 h in liquid LB medium and plated them on LB agar containing kanamycin (25 μg/ml) and chloramphenicol (12.5 μg/ml). After overnight incubation, colonies were picked and checked for accurate homologous recombination by PCR with primers flanking the homology boxes. BACvecs were purified from positive clones and electroporated into DH10B (Electromax, Invitrogen), then selected on LB agar containing kanamycin and chloramphenicol to eliminate the pABG plasmid. Colonies were picked and grown in large cultures for purification of the BACvec DNA (Qiagen). To determine the sizes of the deletion and the homology arms and to rule out rare unwanted deletions, we analyzed the BACvecs by diagnostic restriction enzyme digestion in parallel with the parental BACs. In addition, we confirmed the precision of cassette insertion by sequencing the BACvecs in the region of the homology boxes. BACvecs that passed these quality control tests were linearized by digestion with *Not*I, purified by extraction with (25:24:1 vol/vol/vol) phenol/chloroform/isoamyl alcohol (Sigma) and ethanol precipitation, dissolved in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA and stored at 4 °C.

ES cell growth, electroporation and genomic DNA isolation. We used standard culture conditions[36] to grow and maintain the following mouse ES cells lines: CJ7 (129S1/Sv-derived[37]), F1H4 (a 129B6/F1-derived ES line established by previously described methods[38]), RW4 (129X1/SvJ-derived; Incyte Genomics) and an unnamed C57Bl/6J-derived ES cell line[39]. Except where indicated, the CJ7 and F1H4 lines were used for most experiments. ES cells were electroporated with linearized BACvec and selected with G418. Up to 288 neomycin-resistant colonies were picked and expanded in 96-well plates. We prepared ES cell genomic DNA by an automated procedure on a Biomek FX robotic work station (Beckman-Coulter) by washing with phosphate-buffered saline lacking calcium and magnesium (Irvine Scientific), lysing with 0.125 ml of SV RNA lysis buffer (Promega), capturing DNA on a Wizard SV96 DNA binding plate (Promega) and washing with Wizard SV wash solution (Promega). The bound DNA was eluted with 0.125 ml of buffer AE (Qiagen) into a 96-well culture plate and stored at 4 °C. DNA concentrations were determined by the PicoGreen method (Molecular Probes) on a Victor 1420 plate reader (Wallac).

Screening of ES cell clones using 'loss-of-native-allele' assay. To screen ES cell clones, we developed a 'loss-of-native-allele' assay in which we use TaqMan (Applied Biosystems) real-time quantitative PCR (qPCR) to count the number of intact copies of the targeted allele compared with copies of reference genes (**Fig. 2a**). We designed two target-specific primer sets (forward and reverse PCR primers (Gene Link) and a TaqMan probe that recognizes the PCR product) corresponding to the upstream (uPCR) and downstream (dPCR) end of each targeted deletion (**Fig. 1a**) and compared each of these target-specific qPCRs to six reference qPCRs for a total of 12 assays per construct for each ES clone. We routinely screen up to 288 ES clones per targeting construct. A Biomek FX robotic work station assembled the qPCRs in optical 384-well clear reaction plates (Applied Biosystems). The qPCRs were cycled in an ABI PRISM 7900HT Sequence Detection System (SDS). For each qPCR, the SDS software determined the threshold cycle (C_T), the point in the PCR at which the fluorescence signal reached the preset threshold. We calculated the difference in the C_T values (ΔC_T) between the target-specific uPCRs and dPCRs and each of the reference PCRs. Because most of the ES cell clones are not expected to have sustained a targeted loss of allele, we screened for ES cell clones whose ΔC_T values were consistently greater than the median ΔC_T for the population by approximately one cycle, as expected for a clone that has lost one of the two targeted alleles. For targeted loci on the X or Y chromosomes, which are single copy in our male ES cells, we screened for ES cell clones whose C_T values for the targeted allele indicated no amplification (no target detected) compared with normal C_T values for the reference loci. To confirm that our correctly targeted ES cell clones had a single copy of the targeting vector, we performed qPCR assays similar to the 'loss-of-native-allele' assay with primer and probe sets that recognized the *lacZ* and neomycin-resistance gene sequences in the targeting vectors and compared their ΔC_T values to those obtained with DNAs extracted from confirmed heterozygous (one copy) and homozygous knockout (two copies) animals obtained with similar targeting vectors.

Automated genotyping of animals. Tail samples were digested for 16 h at 55 °C in 0.1 M Tris-HCl (pH 7.6), 0.2 M NaCl, 5 mM EDTA, 0.2% (wt/vol) SDS and 0.3 mg/ml proteinase K (Roche). Tissue debris was removed by centrifugation, and genomic DNA was isolated from the lysate on a Kingfisher magnetic particle processor (ThermoLabsytems) with a genomic DNA kit (ThermoLabsystems). We used the 'loss-of-native-allele' and copy number quantitative PCR assays described above to determine the genotypes of F_1 and F_2 mice derived from the targeted ES cell clones. The 'loss-of-native-allele' assay scores 2 (wild type), 1 (heterozygous) and 0 (knockout) copies of the targeted allele, whereas the copy number assay for the *lacZ-neo* sequences in the targeting vector confirms these results—0 copies (wild type), 1 copy (heterozygous) and 2 copies (knockout).

Other methods. For a number of projects, we confirmed that ES cell clones were correctly targeted by standard fluorescence *in situ* hybridization (FISH) procedures[40–43] using the parental BAC and the expression-selection cassette as probes labeled by nick translation (Invitrogen) with the fluorescently labeled dUTP nucleotides (Vysis) Spectrum Orange for the BAC and Spectrum Green for the expression-selection cassette. β-galactosidase activity was detected histochemically by X-gal staining, as described previously[44].

Materials. Requests for ES cells or mice presented in this paper can be submitted to velocigene@regeneron.com.

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

ACKNOWLEDGMENTS

We thank Margaret Karow for intellectual input, Petra Kraus for analysis of gene expression patterns and Fred W. Alt, Craig Bassing, Richard Flavell, Mike Brown, Joe Goldstein, Eric Shooter and Kornelia Polyak for suggesting genes to target. We thank Amalia Dutra for advice and assistance with FISH. We would also like to thank Li Pan, Joyce McClain, Virginia Hughes, Jeffrey Vercollone, Kethi Mullei, Jorge Bermudez, Jennifer Leung, Ning Yuan, Samantha Park, Ruth Santos, Collen Correa, Huan Jiang, Jinsop Om, Youngli Chang, Haige Zhang and Jose Rodriguez for technical assistance, as well as Vicki Lan and Scott Staton for graphics support.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 22 January; accepted 28 March 2003
Published online 5 May 2003; corrected 30 May 2003 (details online);
doi:10.1038/nbt822

1. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).

3. Goldstein, J.L. Laskers for 2001: knockout mice and test-tube babies. *Nat. Med.* **7**, 1079–1080 (2001).
4. Evans, M.J. The cultural mouse. *Nat. Med.* **7**, 1081–1083 (2001).
5. Capecchi, M.R. Generating mice with targeted mutations. *Nat. Med.* **7**, 1086–1090 (2001).
6. Smithies, O. Forty years with homologous recombination. *Nat. Med.* **7**, 1083–1086 (2001).
7. Cheah, S.S. & Behringer, R.R. Contemporary gene targeting strategies for the novice. *Mol. Biotechnol.* **19**, 297–304 (2001).
8. Hrabe de Angelis, M.H. *et al.* Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat. Genet.* **25**, 444–447 (2000).
9. Nolan, P.M. *et al.* A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat. Genet.* **25**, 440–443 (2000).
10. Zambrowicz, B.P. *et al.* Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611 (1998).
11. Mitchell, K.J. *et al.* Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nat. Genet.* **28**, 241–249 (2001).
12. Leighton, P.A. *et al.* Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**, 174–179 (2001).
13. Yang, X.W., Model, P. & Heintz, N. Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat. Biotechnol.* **15**, 859–865 (1997).
14. Yang, X.W., Wynder, C., Doughty, M.L. & Heintz, N. BAC-mediated gene-dosage analysis reveals a role for Zipro1 (Ru49/Zfp38) in progenitor cell proliferation in cerebellum and skin. *Nat. Genet.* **22**, 327–335 (1999).
15. Lee, E.C. *et al.* A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56–65 (2001).
16. Heintz, N. BAC to the future: the use of bac transgenic mice for neuroscience research. *Nat. Rev. Neurosci.* **2**, 861–870 (2001).
17. Joyner, A.L. (ed.) *Gene Targeting; a Practical Approach,* edn. 2 (Oxford University Press, New York, 1999).
18. Hasty, P., Rivera-Perez, J. & Bradley, A. The length of homology required for gene targeting in embryonic stem cells. *Mol. Cell Biol.* **11**, 5586–5591 (1991).
19. Deng, C. & Capecchi, M.R. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell Biol.* **12**, 3365–3371 (1992).
20. Zhang, H., Hasty, P. & Bradley, A. Targeting frequency for deletion vectors in embryonic stem cells. *Mol. Cell Biol.* **14**, 2404–2410 (1994).
21. Waldman, A.S. & Liskay, R.M. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell Biol.* **8**, 5350–5357 (1988).
22. te Riele, H., Maandag, E.R. & Berns, A. Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl. Acad. Sci. USA* **89**, 5128–5132 (1992).
23. Mansour, S.L., Thomas, K.R. & Capecchi, M.R. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352 (1988).
24. Mansour, S.L., Thomas, K.R., Deng, C.X. & Capecchi, M.R. Introduction of a *lacZ* reporter gene into the mouse int-2 locus by homologous recombination. *Proc. Natl. Acad. Sci. USA* **87**, 7688–7692 (1990).
25. Puri, M.C., Rossant, J., Alitalo, K., Bernstein, A. & Partanen, J. The receptor tyrosine kinase TIE is required for integrity and survival of vascular endothelial cells. *EMBO J.* **14**, 5884–5891 (1995).
26. Miquerol, L., Gertsenstein, M., Harpal, K., Rossant, J. & Nagy, A. Multiple developmental roles of VEGF suggested by a *lacZ*-tagged allele. *Dev. Biol.* **212**, 307–322 (1999).
27. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**, 123–128 (1998).
28. Muyrers, J.P., Zhang, Y., Testa, G. & Stewart, A.F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* **27**, 1555–1557 (1999).
29. Narayanan, K., Williamson, R., Zhang, Y., Stewart, A.F. & Ioannou, P.A. Efficient and precise engineering of a 200 kb β-globin human/bacterial artificial chromosome in *E. coli* DH10B using an inducible homologous recombination system. *Gene Ther.* **6**, 442–447 (1999).
30. Angrand, P.O., Daigle, N., van der Hoeven, F., Scholer, H.R. & Stewart, A.F. Simplified generation of targeting constructs using ET recombination. *Nucleic Acids Res.* **27**, e16 (1999).
31. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97**, 5978–5983 (2000).
32. Muyrers, J.P., Zhang, Y. & Stewart, A.F. Techniques: recombinogenic engineering–new options for cloning and manipulating DNA. *Trends Biochem. Sci.* **26**, 325–331 (2001).
33. Gregory, S.G. *et al.* A physical map of the mouse genome. *Nature* **418**, 743–750 (2002).
34. Bodine, S.C. *et al.* Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science* **294**, 1704–1708 (2001).
35. Testa, G. *et al.* Large BAC-based targeting constructs present new options for genomic engineering. *Nat. Biotechnol.* **21**, 443–447 (2003).
36. Matise, M., Auerbach, W. & Joyner, A.L. in *Gene Targeting: A Practical Approach* (ed. Joyner, A.L.) 101–132 (Oxford University Press, New York, 2000).
37. Swiatek, P.J. & Gridley, T. Perinatal lethality and defects in hindbrain development in mice homozygous for a targeted mutation of the zinc finger gene Krox20. *Genes Dev.* **7**, 2071–2084 (1993).
38. Auerbach, W. *et al.* Establishment and chimera analysis of 129/SvEv- and C57BL/6-derived mouse embryonic stem cell lines. *Biotechniques* **29**, 1024–1028, 1030, 1032 (2000).
39. Schuster-Gossler, K. *et al.* Use of coisogenic host blastocysts for efficient establishment of germline chimeras with C57BL/6J ES cell lines. *Biotechniques* **31**, 1022–1024, 1026 (2001).
40. Lundsteen, C. & Lind, A.M. A test of a climate room for preparation of chromosome slides. *Clin. Genet.* **28**, 260–262 (1985).
41. Pinkel, D., Straume, T. & Gray, J.W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* **83**, 2934–2938 (1986).
42. Lichter, P., Cremer, T., Borden, J., Manuelidis, L. & Ward, D.C. Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridization using recombinant DNA libraries. *Hum. Genet.* **80**, 224–234 (1988).
43. Pecker, I. *et al.* Identification and chromosomal localization of Atm, the mouse homolog of the ataxia-telangiectasia gene. *Genomics* **35**, 39–45 (1996).
44. Suri, C. *et al.* Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis. *Cell* **87**, 1171–1180 (1996).

Identification of anti-repressor elements that confer high and stable protein production in mammalian cells

Ted HJ Kwaks, Phil Barnett, Wieger Hemrika, Tjalling Siersma, Richard GAB Sewalt, David PE Satijn, Janynke F Brons, Rik van Blokland, Paul Kwakman, Arle Kruckeberg, Angèle Kelder & Arie P Otte
Nat. Biotechnol. **21, 553–558 (2003)**

The accession numbers supplied by the authors in Table 1, p. 555, were incorrect. The correct ones are given below.

Old number	Corrected number
505648	AY190749
505650	AY190750
505652	AY190751
509911	AY190752
505656	AY190753
505658	AY190754
505660	AY190755
505662	AY190756
509915	AY190757
505666	AY190758

High-throughput engineering of the mouse genome coupled with high-resolution expression analysis

David M Valenzuela, Andrew J Murphy, David Frendewey, Nicholas W Gale, Aris N Economides, Wojtek Auerbach, William T Poueymirou, Niels C Adams, Jose Rojas, Jason Yasenchak, Rostislav Chernomorski, Marylene Boucher, Andrea L Elsasser, Lakeisha Esau, Jenny Zheng, Jennifer Griffiths, Xiaorong Wang, Hong Su, Yingzi Xue, Melissa G Dominguez, Irene Noguera, Richard Torres, Lynn E Macdonald, A Francis Stewart, Thomas M DeChiara & George D Yancopoulos
Nat. Biotechnol. **21, 652–659 (2003)**

In the author list, the name of author Rostislav Chernomorsky was misspelled Rostislav Chernomorski. In addition, an acknowledgment was omitted: the authors would like to acknowledge Panos Ioannou for help with early bacterial homologous recombination experiments. The corrected acknowlegments should read: We thank Margaret Karow for intellectual input, Petra Kraus for analysis of gene expression patterns and Fred W. Alt, Craig Bassing, Richard Flavell, Mike Brown, Joe Goldstein, Eric Shooter and Kornelia Polyak for suggesting genes to target. We thank Andrea Elsasser for excellent database design, implementation and support, and Dr. Amalia Dutra for advice and assistance with FISH. We would also like to thank Li Pan, Joyce McClain, Virginia Hughes, Jeffrey Vercollone, Kethi Mullei, Jorge Bermudez, Jennifer Leung, Ning Yuan, Samantha Park, Ruth Santos, Collen Correa, Huan Jiang, Jinsop Om and Youngli Chang for technical assistance, Vicki Lan and Scott Staton for graphics support and Panos Ioannou for help with early bacterial homologous recombination experiments.