

A public gene trap resource for mouse functional genomics

To the editor:

Gene trapping is a high-throughput approach that can be used to introduce insertional mutations across the genome in mouse embryonic stem (ES) cells. Gene trap vectors simultaneously mutate and report the expression of the endogenous gene at the site of insertion and provide a DNA tag for the rapid identification of the disrupted gene. The generation of mutant mice from a large collection of ES cell lines carrying gene trap insertions could be applied to large-scale functional analysis of the ~30,000 mammalian genes. The overall impact of gene trap resources will rest on the fraction of the genome that is accessible with this technology, the efficiency relative to other competing technologies and the availability of such a resource to the academic community. Lexicon Genetics, a US-based biotechnology company, was the first to implement a genome-wide gene trapping program¹ and has developed OmniBank (<http://www.lexicon-genetics.com>), the largest library of mutant ES cell lines. A parallel effort was initiated in the public domain by several academic groups in the International Gene Trap Consortium (IGTC; <http://www.igtc.ca>). The recent release of OmniBank sequence tags to GenBank² has made it possible to compare the size and efficiency of the existing gene trap libraries.

We confirm that Lexicon achieved close to 60% coverage of the genome from 200,000 OmniBank sequence tags deposited in GenBank (Fig. 1). Our analysis, supported independently by Lexicon³, indicates that the rate of trapping new genes was not linear but declined within the first 100,000 tags to a rate at which 1 new gene was added every 35 tags, comparable to the efficiency of high-throughput gene targeting methods⁴. To date, the IGTC has attained 32% genome coverage in 27,000 tags; trapping

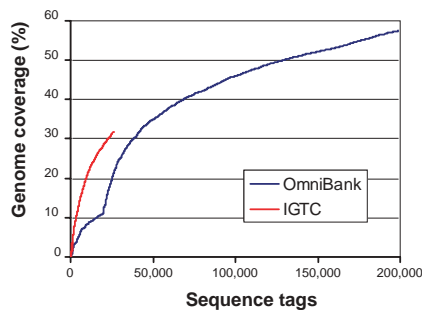


Figure 1 Comparison of the rates of trapping of the IGTC and OmniBank resources. Unique Ensembl genes were identified using MAPTAG (<http://www.sanger.ac.uk/Software/MAPTAG>), an automated annotation program that identifies short, almost perfect matches to gene features in Ensembl. An additional 10% of the trapped genes were identified from BLAST searches ($\leq E^{-04}$) of the RefSeq database. Genome coverage was calculated as the fraction of 8,000 full-length 'sentinel' genes² trapped in each resource.

is likewise nonlinear, but the initial rate seems to be somewhat faster than Lexicon's (Fig. 1). The seemingly higher efficiency may relate to the diversity of plasmid and retroviral vector designs used by the IGTC that could help overcome insertion site preferences of any single vector⁵; further studies are needed to fully understand how vector design and other experimental factors influence the efficiency of gene trapping. One-fifth of the genes trapped by the IGTC were not represented in the sequence tags released by Lexicon (Supplementary Tables 1–3 online). Thus, the two efforts together have trapped nearly two-thirds of all genes in mice. We conclude that gene trapping is an effective strategy to mutate a substantial fraction of the genes in mice that compares favorably with gene-targeting approaches. Furthermore, we continue to refine the technology, particularly in developing strategies for postinsertional modification

of the trapped loci to create a wide range of desired alleles. The IGTC will provide an important public resource of new mutations in mice that will accelerate the pace of functional annotation of the mammalian genome.

Gene trap cell lines generated by the IGTC are available without restriction (<http://baygenomics.ucsf.edu>; <http://www.genetrapp.de>; <http://www.esccells.ca>; <http://www.sanger.ac.uk/genetrapp>; <http://www.fhcrc.org/labs/soriano/GTdb>; <http://www.cmhd.ca>) and all sequence tags are mapped on the Ensembl mouse genome browser http://www.ensembl.org/Mus_musculus/; select DAS Source 'GeneTrap').

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

International Gene Trap Consortium

William C Skarnes¹, Harald von Melchner², Wolfgang Wurst³, Geoff Hicks⁴, Alex S Nord¹, Tony Cox¹, Stephen G Young⁵, Patricia Ruiz⁶, Phil Soriano⁷, Marc Tessier-Lavigne⁸, Bruce R Conklin⁵, William L Stanford⁹ & Janet Rossant¹⁰

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK. ²Laboratory for Molecular Hematology, University of Frankfurt Medical School, Frankfurt am Main, Germany.

³Institute of Developmental Genetics, GSF Research Center, Neuherberg, Germany, and Max Planck Institute of Psychiatry, Munich, Germany. ⁴Manitoba Institute of Cell Biology, Department of Biochemistry & Medical Genetics, Faculty of Medicine, University of Manitoba, Winnipeg, Canada. ⁵Gladstone Institute of Cardiovascular Disease, Department of Medicine, University of California San Francisco, California, USA.

⁶Max-Planck Institute for Molecular Genetics, Center for Cardiovascular Research, Berlin, Germany. ⁷Program in Developmental Biology, Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. ⁸Genetech, South San Francisco,

California, USA. ⁹Institute of Biomaterials & Bioengineering and ⁸Samuel Lunenfeld Research Institute, Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Canada. Correspondence

should be addressed to W.C.S. (skarnes@sanger.ac.uk).

1. Zambrowicz, B.P. et al. *Nature* 392, 608–611 (1998).
2. Zambrowicz, B.P. et al. *Proc. Natl. Acad. Sci. USA* 100, 14109–14114 (2003).

3. Zambrowicz, B.P. et al. *Proc. Natl. Acad. Sci. USA* 101, 4332–4333 (2004).
4. Valenzuela, D.M. et al. *Nat. Biotechnol.* 21, 652–659 (2003).
5. Hansen, J. et al. *Proc. Natl. Acad. Sci. USA* 100, 9918–9922 (2003).