

**Fig 2** We calculated the ratio of  $r^2$  for each pair of SNPs in samples of isolated and outbred populations. The average of these ratios is plotted separately for the Finnish, Sardinian and Ashkenazi samples, and for pairs at distances exceeding or less than 200 kb. For the Finnish/CEPH and Sardinian/CEPH, we made 18 and 21 comparisons for distances exceeding and less than 200 kb, respectively; for Ashkenazi/CEPH populations, we made 12 and 5 comparisons. The sample sizes of CEPH, Finnish, Sardinian and Ashkenazi populations were 92, 100, 150 and 100, respectively.

populations is the applicability of the case-control design. The case-control design promises to be the most efficient paradigm for the association of genes with complex phenotypes<sup>9</sup>. One of the disadvantages of the case-control paradigm is its susceptibility to false-positive results through population stratification, if the cases and the controls are not well matched. With isolated populations, this is less of a risk because of the known and homogeneous genetic background of all individuals.

Sagiv Shifman<sup>1</sup> & Ariel Darvasi<sup>1,2</sup>

<sup>1</sup>Department of Evolution Systematics and Ecology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. <sup>2</sup>IDgene Pharmaceuticals Ltd, Beit Ofer – 5 Heftzadi Street, PO Box 34478, Jerusalem 91344, Israel. Correspondence should be addressed to A.D. (e-mail: ariel@idgene.com).

- Abbott, A. *Nature* **406**, 340–342 (2000).
- Boehnke, M. *Nature Genet.* **25**, 246–247 (2000).
- Eaves, I.A. *et al.* *Nature Genet.* **25**, 320–323 (2000).
- Taillon-Miller, P. *et al.* *Nature Genet.* **25**, 324–328 (2000).
- Jorde, L.B. *Am. J. Hum. Genet.* **56**, 11–14 (1995).
- Ott, J. *Proc. Natl. Acad. Sci. USA* **97**, 2–3 (2000).
- Collins, A., Lonjou, C. & Morton, N.E. *Proc. Natl. Acad. Sci. USA* **96**, 15173–15177 (1999).
- Kruglyak, L. *Nature Genet.* **22**, 139–44 (1999).
- Risch, N.J. *Nature* **405**, 847–856 (2000).
- Roa, B.B., Boyd, A.A., Volcik, K. & Richards, C.S. *Nature Genet.* **14**, 185–187 (1996).
- Frankel, W.N. *Trends Genet.* **11**, 471–477 (1995).

Additional work is necessary to estimate with accuracy the advantage of an isolated population owing to reduction in genetic heterogeneity. It is, however, worth noting that reduction in genetic heterogeneity directly affects the GRR, which is roughly proportional to the square root of the sample size required to identify a given gene<sup>9</sup>. Consequently, even a moderate effect on the GRR may have a significant effect on the required sample size.

The differences in LD observed between isolated and outbred populations on one

hand, and the expected difference in the extent of genetic heterogeneity on the other, are independent and thus have a multiplicative effect on the required sample size. For example, if, because of differences in LD, the sample size needs to be increased threefold, and if, because of the differences in GRR another threefold increase is required, the sample size will need to be increased ninefold with outbred populations to achieve power comparable to that of an isolated population.

An additional advantage of using isolated

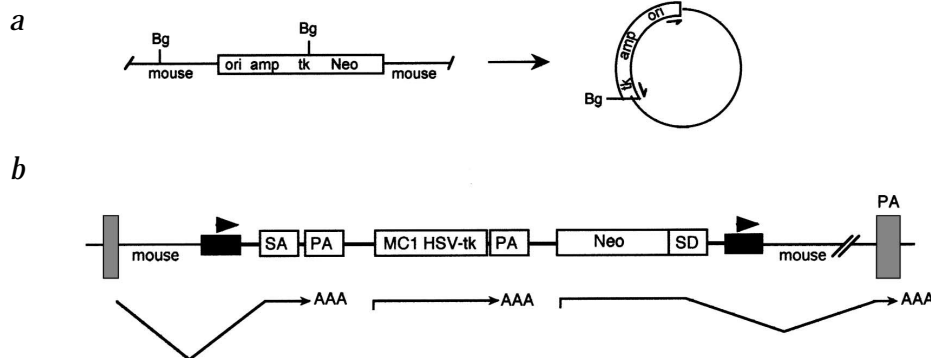
## DelBank: a mouse ES-cell resource for generating deletions

Chromosomal deletions are valuable reagents for identifying and mapping genetic function. To simplify the creation of deletions in mice, we developed a collection of embryonic stem (ES) cell clones called DelBank. DelBank is based upon a technology in which radiation-induced deletions of herpes simplex virus

thymidine kinase (*tk*) cassettes, inserted by homologous recombination into F<sub>1</sub> hybrid ES cells, are selected with the antiviral drug 1-2'-deoxy-2'-fluoro-β-D-arabinofuranosyl-5-iodouracil (FIAU)<sup>1,2</sup>. The deletions range in size from less than 1 cM to over 20 cM. Cells of the F<sub>1</sub> generation retain germline competence follow-

ing irradiation, and the heterozygosity enables the molecular mapping of deletion endpoints.

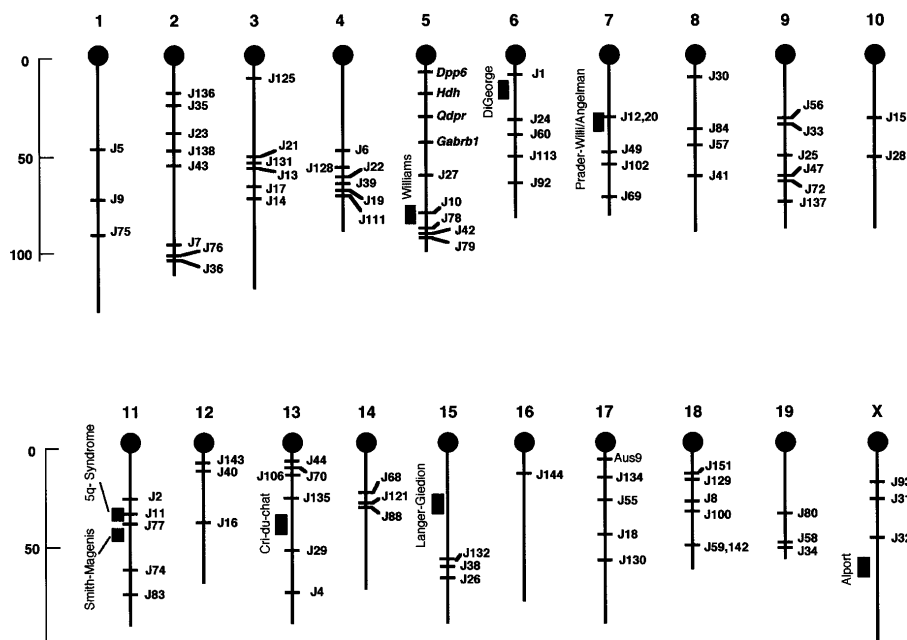
DelBank consists of more than 90 F<sub>1</sub> hybrid ES cell clones, each containing a randomly integrated *tk*-containing vector. We generated clones using either a plasmid vector (pBanTKcass<sup>+</sup>; Fig. 1a) or modified retroviral poly(A) trap vector<sup>3</sup> (Fig. 1b). These were introduced into v6.4 (C57BL/6J×129S4/SvJae) or v17.2 (BALB/cJ×129S4/SvJae) ES cells<sup>1,4</sup> by electroporation (plasmid) or retroviral infection (gene trap), followed by selection for



**Fig. 1** DelBank vector strategies. **a**, Plasmid rescue. DNA from transformed ES cell clones was cleaved with *Bgl*I, circularized and transformed into bacteria. The sequence of the 'rescued' mouse DNA was determined to generate amplimers that identify SSCPs. **b**, Gene trapping. PA, poly(A) signal; SD, splice donor. Shown is a productive gene trap, where transcripts from an endogenous gene (left) undergo splicing into a vector splice acceptor (SA) and are terminated at vector PA, while the neo gene acquires a host PA. Arrowheads indicate *LoxP* sites.



**Fig. 2** Genomic distribution of DelBank insertions. The location of each clone and the way in which we mapped it are indicated on the DelBank web site. Approximate presumed locations of human contiguous gene syndromes (CGS) or homologous regions are shown, based on data interpreted from OMIM, the Mouse Genome database and the U.C. Davis Man/mouse synteny maps. 'Aus9' is an abbreviation for *D17A10*.



transformants in G418. We recovered DNA flanking the integration sites of single-copy insert pBanTKcass clones by plasmid rescue (Fig. 1). We sequenced the inserts, designed PCR primers to identify single-strand conformation polymorphisms (SSCPs) between *Mus spretus* and C57BL/6J and then mapped the insertion

sites on the BSS interspecific backcross mapping panel ([www.jax.org/resources/documents/cmdata/](http://www.jax.org/resources/documents/cmdata/)). With the gene trap, the vector-directed transcription of neo requires the acquisition of a host poly(A) signal to confer G418 resistance upon transfected ES cells. We cloned these insertion sites by 3'-RACE. Some trapped genes have known map locations. Otherwise, insertion sites were determined using the T31 radiation hybrid mapping panel ([www.jax.org/resources/documents/cmdata/](http://www.jax.org/resources/documents/cmdata/)).

DelBank insertion sites are distributed throughout the genome, with at least one per mouse chromosome (Fig. 2). We include some insertions that were targeted by homologous recombination to chromosomes 5 and 17 (refs. 1,2). Several DelBank integrations lie near regions syntenic to those deleted in human contiguous gene syndromes.

As DelBank clones are useful only if they retain germline colonizing capability, 10 parental lines were evaluated by us and DelBank users, and at least eight produced germline chimeras. To explore the extent of DelBank's utility, we successfully induced deletions and produced germline chimeras with a clone containing an inser-

tion [Tg(tkneo)20Jcs] near the mouse counterpart of the Prader-Willi syndrome (PWS) critical region on chromosome 7.

We are expanding DelBank to reach a goal of over 200 clones. Many DelBank clones remain unmapped, but the insertion sites will become known as genome data become available. Assuming an average maximum viable deletion size in mice of approximately 5 cM (refs. 2,5,6), each clone could potentially yield deletions covering 10 cM, for a DelBank total of 2,000 cM, slightly more than the mouse genome size of approximately 1,500 cM. We maintain a searchable DelBank website (<http://lena.jax.org/~jcs/Delbank.html>) containing sequence and map positions of insertion sites and technical information on the construction and use of DelBank. Clones are freely available to academic researchers.

The ability to rapidly create deletion complexes along chromosomes will facilitate systematic functional analyses of the mammalian genome. Deletions can be used for multiple applications, including region-directed *N*-ethyl-*N*-nitrosourea saturation mutagenesis, modeling human deletion syndromes<sup>7,8</sup>, mapping existing mutations<sup>9</sup> or defining quantitative trait loci<sup>10</sup>.

#### Acknowledgments

This work was primarily supported by a grant from the Merck Genome Research Institute (to JCS). NCG was supported by a Training Grant (HD07065) from NIH.

Neal C. Goodwin<sup>1</sup>, Yasumasa Ishida<sup>2</sup>, Suzanne Hartford<sup>1</sup>, Cate Wnek<sup>1</sup>, Rebecca A. Bergstrom<sup>1</sup>, Philip Leder<sup>3</sup> & John C. Schimenti<sup>1</sup>

<sup>1</sup>The Jackson Laboratory, Bar Harbor, Maine 04609, USA. <sup>2</sup>Department of Experimental Pathology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan.

<sup>3</sup>Department of Genetics, Harvard Medical School, Howard Hughes Medical Institute, Boston, Massachusetts, USA. Correspondence should be addressed to J.C.S. (e-mail: [jcs@jax.org](mailto:jcs@jax.org)).

1. You, Y. *et al.* *Nature Genet.* **15**, 285–288 (1997).
2. Schimenti, J. *et al.* *Genome Res.* **10**, 1043–1050 (2000).
3. Ishida, Y. & Leder, P. *Nucleic Acids Res.* **27**, e35 (1999).
4. You, Y. *et al.* *Mamm. Genome* **9**, 232–234 (1998).
5. Su, H., Wang, X. & Bradley, A. *Nature Genet.* **24**, 92–95 (2000).
6. LePage, D.F., Church, D.M., Millie, E., Hassold, T.J. & Conlon, R.A. *Proc. Natl. Acad. Sci. USA* **97**, 10471–10476 (2000).
7. Naf, D. *et al.* *Hum. Mol. Genet.* **10**, 91–98 (2001).
8. Lindsay, E.A. *et al.* *Nature* **401**, 379–383 (1999).
9. Bergstrom, R.A., You, Y., Erway, L.C., Lyon, M.F. & Schimenti, J.C. *Genetics* **150**, 815–822 (1998).
10. Planchart, A., You, Y. & Schimenti, J.C. *Genetics* **155**, 803–812 (2000).