Cell

A Mouse for All Reasons

The International Mouse Knockout Consortium[1,2,*]
[1]Communicating authors: Francis S. Collins, National Human Genome Research Institute, National Institutes of Health, USA; Janet Rossant, Hospital for Sick Children, University of Toronto, Canada; Wolfgang Wurst, GSF-National Research Center for Environment and Health, Technical University, Munich, Germany
[2]For a full list of International Mouse Knockout Consortium members, please see http://www.cell.com/cgi/content/full/128/1/9/DC1/.
*Correspondence: francisc@mail.nih.gov
DOI 10.1016/j.cell.2006.12.018

Three major mouse knockout programs are underway worldwide, working together to mutate all protein-encoding genes in the mouse using a combination of gene trapping and gene targeting in mouse embryonic stem (ES) cells. Although the current emphasis is on production of this valuable resource, there are significant efforts to facilitate program coordination, to enhance the availability of this resource, and to plan for future efforts in mouse genetics research.

The rapid advances and scale up of projects in DNA sequencing during the past 15 years have produced complete genome sequences of dozens of species. The great challenge facing biologists today is to ascribe function to the thousands of genes discovered through these efforts. Answering this challenge requires description of the functions of each gene in normal physiology and development, as well as the contribution of mutant alleles to inherited diseases. One powerful approach to help decipher the functions of genes is the manipulation of genes in intact animals. The ability to perform gene knockouts in the laboratory mouse via ES cells has been a particularly successful research tool for nearly two decades, using two major methodologies: gene trapping and gene targeting (see Note for definitions).

A survey of the Mouse Genome Informatics database (MGI; http://www.informatics.jax.org) and the International Gene Trap Consortium (IGTC; http://www.genetrap.org) reveals that nearly 4000 targeted knockouts of genes have already been made in mice and that another 7000 unique genes have been trapped in ES cells. Given the overlap between these sets, about 9000 mouse genes in total have been knocked out. In contrast to the gene-trapping efforts, which have produced readily available ES cell libraries, the gene-targeting efforts have thus far been accomplished mostly as a cottage industry with many investigators contributing one or a few knockout mice engineered using a wide variety of approaches, and fewer than 900 knockouts are actually available from a repository.

Participants at a 2003 meeting at the Banbury Center of the Cold Spring Harbor Laboratory proposed that the systematic mutagenesis of all protein-encoding genes in the mouse would provide a standardized resource that would save time and expense (Austin et al., 2004). At the same time, pan-European discussions sponsored by the European Commission (EC) endorsed similar goals (Auwerx et al., 2004). Elements that supported these proposals included the availability of a highly accurate mouse genome sequence indicating a tractable number of genes; the high degree of homology between the mouse and human genomes; the excellent track record of the mouse as a model for human diseases and traits; and the sophisticated genetic tools and resources available for the mouse. The first step of the proposals has now come to fruition, and major mouse knockout programs are underway worldwide. Here, we describe the status of these efforts, the production goals, the methods for generation of research reagents, the plan to achieve close coordination among the programs, and the future direction of this research.

Three International Consortia

The complexity and cost of such an undertaking required the deployment and coordination of resources on a global scale. Three funding agencies, the U.S. National Institutes of Health (NIH), the EC, and Genome Canada and its partners have responded by funding such programs (see Supplemental Data available with this article online). These programs will use gene targeting and trapping to complete a resource of mutated ES cells and mice that will be comprehensive and will establish and support distribution centers and web-based data dissemination to deliver these resources to the biomedical community.

KOMP (*K*nock*O*ut *M*ouse *P*roject, http://www.knockoutmouse.org), funded by the NIH, encompasses a number of efforts that all aim to enhance the availability and utility of mouse knockout strains. The first effort involved the acquisition of 251 knockout strains, and extensive associated phenotype data, from Deltagen and Lexicon Genetics. These strains were chosen from a list of approximately 2000 commercially derived gene knockouts by an NIH committee comprised of scientists representing areas of human disease and all major fields of biology. The strains have been deposited, as cryopreserved embryos, at two repositories and the data are available from the MGI database (http://www.informatics.jax.org/external/ko/). A second effort of KOMP aims to support the deposition

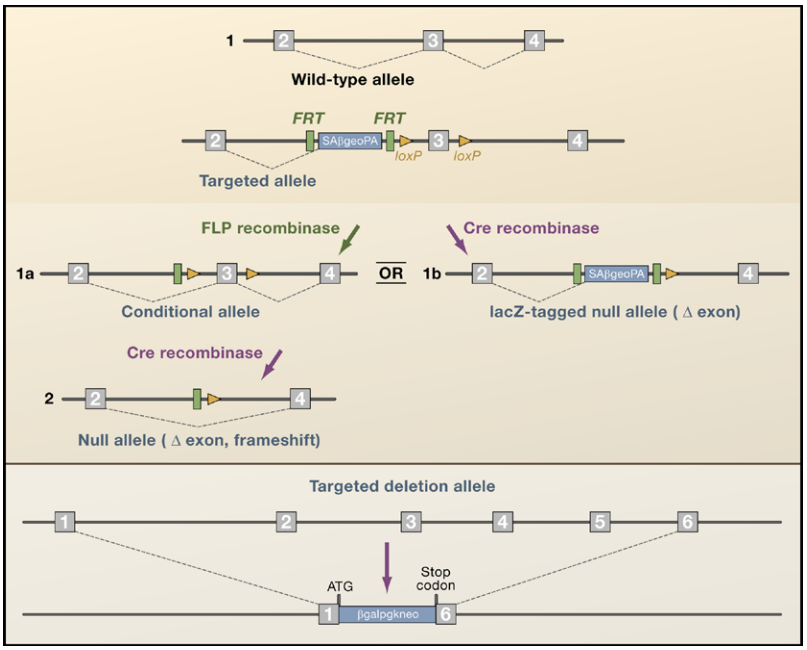

Figure 1. Different Approaches for Gene Targeting

(Top) Targeted trapping of conditional alleles as carried out by KOMP and EUCOMM. The diagram depicts a model gene before and after homologous recombination with the targeted-trapping system. The abbreviation SAβgeoPA refers to an SV-40 splice acceptor site (SA), a promoterless β-galactosidase reporter gene followed by a promoterless neomycin-resistance gene (βgeo), and a polyadenylation signal (PA). A promoter driven neo cassette will be used to target genes not expressed in mouse ES cells. In both cases, the resultant allele is predicted to be null due to a strong splice acceptor in the expression cassette. The combination of FRT and loxP sites allows further manipulation to either remove the reporter cassette (1a) to restore wild-type expression or to delete the critical exon (1b). The allele in (1a) can be used in conjunction with Cre transgenic mice to produce conditional knockouts (2).

(Bottom) Allele deletion carried out by KOMP's Regeneron. This diagram demonstrates that the deletion, driven by homologous recombination with a recombineered BAC construct, removes all coding sequences of the gene and inserts an expression cassette driven off the endogenous gene's promoter. The abbreviation βgalpgkneo refers to a promoterless β-galactosidase reporter gene (βgal) and a phosphoglycerate kinase promoter driving the neomycin-resistance gene cassette, which includes a polyadenylation signal (pgkneo).

of additional mouse knockout strains, produced by academic researchers, into repositories. Toward that end, the Mutant Mouse Regional Resource Centers (http://www.mmrrc.org) are supported to "repatriate" and archive 320 mouse strains for broad distribution. A prioritized list was generated from the set of all published mouse knockouts, and the investigators that published the strains are currently being requested to deposit their mice. The centerpiece of the KOMP knockout effort, however, consists of two programs that aim to create 8500 targeted mutations in ES cells in genes that have not yet been knocked out. The first is a consortium comprising The Children's Hospital Oakland Research Institute, the Wellcome Trust Sanger Institute, and the University of California at Davis ("CSD"). The other group consists of investigators at Regeneron Pharmaceuticals. Both groups have established high-throughput pipelines to target genes in mouse ES cells. The CSD group will create "null-first conditional-ready" alleles (Figure 1). It is clear, however, that not all genes will be suitable for this allele design. Currently, compact genes and genes composed of complex transcript variants will not be readily amenable to this approach due to DNA sequence requirements to engineer the loxP sites flanking a critical exon that upon removal would ensure a null mutation. The use of BacVectors (targeting vectors with on average over 100 kb of homology), as developed by Regeneron, provides an alternate approach that is not dependent on gene structure, gene expression in ES cells, or a single exon that is critical for function. Regeneron will create null mutant alleles by deleting all or most of the exons in given genes, preferably from the initiator ATG codon to the stop codon (Valenzuela et al., 2003; Figure 1). Additionally, as part of a quality-control effort, about 500 mutant mouse strains will be generated by KOMP and made available through repositories (Table 1).

EUCOMM (*EU*ropean *C*onditional *M*ouse *M*utagenesis Program, http://www.eucomm.org), funded by the FP6 program of the EC, combines the expertise of 11 research institutions across Europe and is coordinated by the GSF–National Research Center for Environment and Health and the Wellcome Trust Sanger Institute. The goals of the EUCOMM project are to produce 12,000 conditional gene trap mutations and 8000 conditional targeted mutations in mouse ES cells (Table 1). From these, EUCOMM will generate up to 320 mutant mouse strains for genes of biological and medical relevance, mainly for quality-control reasons but also to determine gene function in vivo. The gene trapping and targeting vectors are designed to allow the generation of a null mutation as well as a conditional mutation (Schnütgen et al., 2005, Figure 1). In support of these efforts, the EUCOMM consortium will also produce 20 ligand-inducible, Cre-recombinase-expressing transgenic mice that will be available to the research community as well.

NorCOMM (*Nor*th American *C*onditional *M*ouse *M*utagenesis Project, http://norcomm.phenogenomics.ca/index.htm) is a Canadian project, funded by Genome Canada and partners, whose mission is to contribute to the generation of a mouse ES cell resource with characterized mutations in every gene in the genome. Three centers in Canada are contributors to this initiative—the University of Manitoba, the University of Toronto, and the University of British Columbia. The project's goals include the generation of 10,000 gene trap insertion clones in ES cells derived from the 129 mouse strain using conditional-ready gene trap

Table 1. International Mouse Gene Knockout Programs

Type of Resource	Type of Knockout	2006	2007	2008	2009	2010	Totals
KOMP							
ES Cell	Targeted Deletion	175	500	941	942	942	3500
Mouse	Targeted Deletion	50	50	50	50	50	250
ES Cell	Targeted Conditional	1000	1000	1000	1000	1000	5000
Mouse	Targeted Conditional	50	50	50	50	50	250
EUCOMM							
ES Cell	Trapped Conditional	3000	6000	3000			12,000
ES Cell	Targeted Conditional	1000	3000	4000			8000
Mouse	Mixed	20	100	200			320
NorCOMM							
ES Cell	Trapped Conditional	1000	4000	3000	2000		10,000
ES Cell	Targeted Conditional	100	400	750	750		2000
Mouse	Mixed	25	25	25	25		100
Cumulative for All Programs							
ES Cell	Trapped (Conditional)	4000	14000	20000	22000		22,000[a]
ES Cell	Targeted (Deletion)	175	675	1616	2558	3500	3500
ES Cell	Targeted (Conditional)	2100	6500	12,250	14,000	15,000	15,000
Mouse	Mixed	125	400	715	830	920	920

Production timelines and totals are shown here. This table shows a timeline of unique gene knockout materials produced by the three programs: KOMP, EUCOMM, NorCOMM. Duplication will be avoided wherever possible through cooperation of the three organizations. The KOMP-related data will be collected and disseminated through the Data Coordination Center (DCC) that is being established at the Jackson Laboratory. The KOMP ES cell collection will be available to the entire research community. It is envisioned that the investigators will order ES cells from repositories with the option of receiving cells or of having the repository generate mice for an additional fee. A competition for a repository for the KOMP materials is underway. EUCOMM data on gene targeting vectors, mutant ES cells, mutant mouse lines, Cre-recombinase expressing transgenic mouse lines, mutation identification tool kits, and Standard Operating Procedures will be presented at the EUCOMM website (http://www.eucomm.org). Vectors and ES cells will be distributed by the German Resource Center for Genome Research (RZPD; http://www.rzpd.de). All transgenic mouse lines will be archived as frozen embryos and disseminated by the European Mouse Mutant Archive (EMMA; http://www.emmanet.org). The NorCOMM ES cell resource will be integrated with the Canadian Mouse Mutant Repository (CMMR) in cooperation with FIMRe (http://www.fimre.org). At their discretion, on a case-by-case basis, KOMP, EUCOMM, and NorCOMM may require investigators to send mice generated from the ES cells to a repository for cryo-preservation.

[a]The above numbers reflect estimates of unique genes trapped by the respective programs, but it is anticipated that there will be a 30%–40% overlap, resulting in ~14,000 unique gene traps from the combined efforts.

vectors, 2000 targeted conditional mutant ES cell lines, and a toolbox of vectors and associated molecular reagents under development by the Samuel Lunenfeld Research Institute, Mount Sinai Hospital, to enhance the re-engineering of mutant lines.

All three efforts rely on high-quality annotation of genomic sequence for the identification of genes to ablate, for the design of targeting constructs, and for the interpretation of gene-trapping results. Toward that end, both EUCOMM and KOMP are supporting vertebrate genome annotation (VEGA; http://vega.sanger.ac.uk/) that will be performed largely at the Sanger Institute.

The resources generated by these programs will be available from three distribution centers: the German Resource Center for Genome Research (RZPD; http://www.rzpd.de), the KOMP repository (to be announced later, not yet funded), and the Canadian Mouse Consortium (CMC; http://www.mousecanada.ca/index.htm). Investigators can find reagents at these sites or at the International Mouse Strain Resource (IMSR; http://www.informatics.jax.org/imsr/index.jsp).

Mouse Strain Background

Historically, the vast majority of mouse knockouts have been generated using ES cells derived from the 129 mouse strain. Unfortunately, the 129 mouse strain has not been widely used in other fields of experimentation and poses particular problems in immunology, neuroscience, and physiology. Although ES cells have been made from C57BL/6, DBA, and BALB/c mouse strains, they have not yet demonstrated the robustness needed to make a large-scale gene knockout project feasible. A workshop held at the NIH in March 2005 recognized the expediency of using 129 mouse ES cells, but participants strongly endorsed the development and use of C57BL/6 ES cells. Thus, three groups (University of Pennsylvania;

Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto; and Regeneron Pharmaceuticals) are moving forward to further characterize and develop robust C57BL/6 ES cells for high-throughput targeting efforts and implementation into the KOMP program.

Converging Scientific Strategies

In contemplating the range of strategies that might be used to generate mutants of a large number of mouse genes at high efficiency and low cost, much discussion has centered around which technical approaches to use—for example, gene targeting or trapping, null or conditional alleles, and whether to include reporter genes (e.g., β-galactosidase). Further discussions that have taken into consideration the complexities of gene structure and expression have yielded a synthesis of approaches, while continuing evolution of genetic resources and high-throughput technologies have also influenced these plans. For example, one advantage of trapping is that only a single (or small number) of vectors needs to be constructed to trap thousands of genes (Skarnes et al., 2004). In contrast, with targeting, a vector has to be made for each gene that is to be mutated. Now, however, end-sequenced BAC libraries and recombineering methods (Copeland et al., 2001; Zhang et al., 1998) with the implementation of robotics allow the rapid construction of hundreds of targeting vectors (Valenzuela et al., 2003), which can in principle be used to knock out any gene in ES cells. Moreover, development and implementation of high-throughput screens for ES cells that are correctly targeted using such vectors has enabled the efficient utilization of large numbers of targeting vectors (Valenzuela et al., 2003). As a second example, the adoption of promoterless selection cassettes into targeting vectors has greatly improved the efficiency in targeting expressed genes (Friedel et al., 2005; Testa et al., 2004), decreasing the advantage of trapping in large-scale efforts. On the other hand, an advantage of targeting, that it allows the design of conditional alleles, is no longer as compelling because conditional gene traps are now being made (Schnütgen et al., 2005). Finally, the use of long homology arms in the constructs used for recombination may obviate the need for DNA constructs from the same strain (isogenic) (Valenzuela et al., 2003), eliminating a disconnect between genome reagents, that is, the standard C57BL/6 BAC library that was used to sequence the mouse genome, and ES cells that are derived from 129 or other mouse strains. Thus, a combination of methods allows various approaches to be used in the construction of universal alleles that can be re-engineered as nulls, hypomorphs, or conditional knockouts. The development of technologies, the recognition of medical need for experimental models, and the common interests of strained funding agencies have resulted in, from the perspective of the mouse genome, a grand conjunction in the field of genetics.

Prioritization

Given the scale of the mouse knockout project, it should be possible to rapidly provide a collection of gene knockouts of immediate interest. EUCOMM has already selected a list of high-priority genes that is available at http://www.eucomm.org. For the KOMP, a working group will make recommendations to the NIH, based on community input (http://www.knockoutmouse.org) regarding prioritization of which genes to knockout. Presumably, not all genes will be nominated, so that after ordering of nominated genes the working group will complete recommendations for prioritization of the remaining genes based on completion of known pathways, gene families, disease-related genes, expression profiles, and completeness of the gene annotation. KOMP, EUCOMM, and NorCOMM have agreed to share gene lists and data to help coordinate these international efforts.

Future Directions

The participants of the 2003 Banbury Conference envisioned that the completion of the ES cell knockout resource for the entire mouse genome would provide a foundation of maximum utility to the research community and serve as a springboard for developing additional high-throughput cost-saving efforts in functional genomics. For example:

- Proven constructs with the presence of loxP and FRT (Flp recombinase target) sites will afford the opportunity to use recombinases to exchange reporter genes or introduce expression cassettes with human genes (or any transgene or siRNA) into the vectors or directly in the targeted ES cell lines. These reagents will help to facilitate the future re-engineering of allelic mutations as our knowledge of particular gene functions and human disease alleles grows. The specific introduction of disease-related mutations should significantly enhance modeling of human diseases in mice.
- While all the above-mentioned efforts have incorporated the highly versatile and broadly accepted lacZ reporter, there are many instances where redesign of the reporter cassette to provide live-cell imaging, molecular tags, or other markers could be highly advantageous. Having the reagents available from the original targeting experiment will greatly facilitate this switch.
- Reporter expression profiles should also provide valuable data to allow the selection of the next generation of Cre driver strains, which can be easily developed using the construct and targeted ES cell resources.
- The BacVectors with their very long regions of homology will also enhance the rate of homologous recombination in ES cell lines from other mouse strains that, while less robust for these large-scale projects, are from mouse strains preferred in specific areas of research.
- The ES cells themselves are a valuable tool to study gene function in

vitro. The redesign of vectors with a different selection cassette could allow the production of homozygous mutant ES cells that can be studied in cell-based assays, differentiated into multiple cell types, and used as recipients for human transgenes (or knockin of a humanized locus) for drug screening or toxicology without the potential background of the mouse gene product. In fact, entire pathways potentially can be engineered.

- Homozygous ES cells can be used as the "null benchmark" for the development of other mutagenesis approaches, such as siRNA, to validate these approaches in vitro prior to their application in whole animals.

It is important to note that generating mice from an ES cell line, by blastocyst injection and breeding, and subsequent phenotyping of such mice is significantly more expensive than the creation of the knockout ES cell line. Furthermore, many labs have specialized research interests, such that many phenotypes may be missed for lack of expertise or broad systematic experimentation. If an expanded program can be mounted in the future, the proposed areas for centralized phenotyping might include expression analysis of the reporter gene, stage of lethality, activity, appearance and general behavior of the mice, blood chemistry, skeletal structure, cardiovascular function, urinalysis, MRI, and histology of major organs and tissues. The number of targeted genes that can be studied in these programs will depend critically on funding. It is envisioned that specialized second and third tier phenotyping such as behavioral studies, immunological challenge, physiology, cardiovascular disease assessment, and cancer occurrence, to name a few, will require highly specialized assays and can only reasonably be performed on a subset of genes. Therefore, those assays that are of higher cost and that require a higher degree of specialization will most likely need to be investigator driven.

The development of the vectors, mutant ES cells, mice, and distribution networks for the efforts of KOMP, EUCOMM, and NorCOMM to produce a knockout mutation for every gene in the mouse genome will not only provide valuable research tools in the short-term, decrease overall cost to the research community, and reduce time in producing experimental data, but it will also provide a foundation for future research in a wide variety of areas for many years to come.

Note

Definition of gene trapping and gene targeting: Gene trapping relies on the random integration in the genome of a DNA construct carrying a splice acceptor sequence and selection marker. This strategy exploits the fact that when constructs integrate into introns, they create fusion transcripts through the action of the vector's splice acceptor and express the selection marker. The resulting truncation/fusion often leads to a knockout or a severe hypomorphic allele of the endogenous gene. Gene trapping has the advantage of being relatively inexpensive and can be performed using one or a few vectors. A significant limitation of this strategy is that it is only effective for genes that are expressed in ES cells, and there is also bias toward larger genes. Gene targeting is a directed approach that uses homologous recombination to mutate an endogenous gene. Specifically, this method can be used to delete a gene, remove exons, and introduce point mutations either permanently or in a conditional fashion. Gene targeting requires the creation of a specific vector for each gene of interest but can be used for any gene, regardless of transcriptional activity or gene size.

Supplemental Data

Supplemental Data include the names and affiliations of all members of the International Mouse Knockout Consortium and can be found with this article online at http://www.cell.com/cgi/content/full/128/1/9/DC1/.

ACKNOWLEDGMENTS

We would like to thank Allan Bradley, Colin Fletcher, Geoff Hicks, Mark Moore, William Skarnes, and David Valenzuela for providing information for this Commentary and for their collegial spirit in undertaking this international collaboration.

REFERENCES

Austin, C.P., Battey, J.F., Bradley, A., Bucan, M., Capecchi, M., Collins, F.S., Dove, W.F., Duyk, G., Dymecki, S., Eppig, J.T., et al. (2004). Nat. Genet. *36*, 921–924.

Auwerx, J., Avner, P., Baldock, R., Ballabio, A., Balling, R., Barbacid, M., Berns, A., Bradley, A., Brown, S., Carmeliet, P., et al. (2004). Nat. Genet. *36*, 925–927.

Copeland, N.G., Jenkins, N.A., and Court, D.L. (2001). Nat. Rev. Genet. *2*, 769–779.

Friedel, R.H., Plump, A., Lu, X., Spilker, K., Jolicoeur, C., Wong, K., Venkatesh, T.R., Yaron, A., Hynes, M., Chen, B., et al. (2005). Proc. Natl. Acad. Sci. USA *102*, 13188–13193.

Schnütgen, F., De-Zolt, S., Van Sloun, P., Hollatz, M., Floss, T., Hansen, J., Altschmied, J., Seisenberger, C., Ghyselinck, N.B., Ruiz, P., et al. (2005). Proc. Natl. Acad. Sci. USA *102*, 7221–7226.

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.

Testa, G., Schaft, J., van der Hoeven, F., Glaser, S., Anastassiadis, K., Zhang, Y., Hermann, T., Stremmel, W., and Stewart, A.F. (2004). Genesis *38*, 151–158.

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.

Zhang, Y., Buchholz, F., Muyrers, J.P., and Stewart, A.F. (1998). Nat. Genet. *20*, 123–128.