

# Mapping the future of mouse functional genomics: driving the translational engine

The mouse is a pivotal organism for studying mammalian gene function and modelling human disease. It offers an unparalleled toolkit for modifying genes and studying the phenotypic outcomes. Recognising these advantages and opportunities, several groups<sup>1-4</sup> have expounded the vision and implementation of creating a mouse knockout for each gene in the mouse genome in order to facilitate the next goal of genomics - to develop a functional annotation of the mammalian genome. A meeting of groups working in this area was recently organized by Genome Canada, the US National Institutes of Health and the European Commission and held in Brussels to discuss current and future approaches and to plan for the future utilization of the resources. We will summarize here the outcome of the discussions and recommendations put forth at this meeting. They offer a long-term international strategy for mouse functional genomics, with potential wide and diverse impacts upon biomedical and translational research.

## Overview

Several large scale gene knockout efforts have been funded recently by the EC (EUCOMM)(<http://www.eucomm.org>), NIH (KOMP)(<http://komp.org>), Genome Canada (NorCOMM)(<http://www.norcomm.phenogenomics.ca>), and the Texas Institute for Genomic Medicine (TIGM) (<http://www.tigm-knockouts.org>) and are well on their way to developing a complete ES cell gene knockout library<sup>3</sup>. These groups have formed the International Knockout Mouse Consortium (IKMC) to facilitate communication, share resources, reduce redundant effort and aid in distribution of materials. The next critical step in this process is to disseminate and harness appropriate mouse mutant resources, to ensure that these programs and future programs are integrated into a worldwide research network and repository. The overarching goal of these efforts is to deliver cost-effective programmes that contribute to facilitating translational research across the spectrum from basic biomedical sciences to experimental medicine. This will require the dissemination of data and materials produced to the wider research community by fostering access to mouse models and thereby enhancing functional studies. In addition to existing programs, it is critical that large scale efforts be launched to initiate the phenotyping of the mouse mutant resource to begin to map out the relationship between gene and phenotype. It will however be necessary to move beyond basic phenotyping to fully determine gene function. It will be imperative to integrate mouse mutant resources and extensive phenotyping studies with QTL approaches and molecular studies (e.g. proteomics) to elaborate the complexity of networks underpinning the relationship between genes, proteins, phenotypes and gene function. The production, integration and analysis of these datasets should lead to a better understanding of the similarities and differences between mouse and human biological systems and foster the continued development of new models for studying the underlying bases of human disease. Moreover these discoveries will illuminate where in certain circumstances humanization of mouse genes or systems are needed to create improved models. All of the ensuing resources provide enormous opportunities for discovery research with the ultimate goal of improving human health.

## Mutagenesis: Current and Future

The projects of the International Knockout Mouse Consortium will deliver a resource of conditional, deletional and trapped mutants (a mix of targeted and gene trapped ES cells) covering around 22,000 mouse genes. Details on the knockout approaches are summarized in Collins et al. (ref. 3) and full descriptions with additional information are available.<sup>5</sup> This first phase resource will be created in C57BL/6N ES cells, but extant vectors provide an opportunity to generate second phase parallel resources in other inbred strains in order to develop specialized models and test gene function on different genetic backgrounds. It will be important to consider the development of ES cells for a wider collection of inbred strains as backcrossing is expensive

and time consuming. The null allele approach will also provide a template for more subtle mutations, as one can extend the resource by targeting to create specific isoforms or perform genome wide screens over the null background to identify enhancer or suppressor mutations.

Other mutagenesis approaches, such as ENU, RNAi and transposon mutagenesis, will continue to make a significant impact upon functional annotation. Notably, there is an important threshold in the cost effectiveness of re-sequencing approaches whereby comprehensive characterisation of the mutations in ENU archives would become feasible providing a powerful resource of multiple mutant alleles for every gene in the mouse genome. Furthermore, RNAi strategies enable the simultaneous downregulation of multiple genes without extensive mouse breeding and the generation of inducible-reversible mutations. Nearly all of the ES cells created by the various approaches in the IKMC will afford investigators the opportunity to remove selectable markers, re-activate gene expression, knock-in other genes and create inducible knockouts in selective tissues, dependent on the specific program that produced each ES cell. In parallel to the mutant ES cell projects, efforts to generate the complementary cre/flip/φ31 expressing “driver” mouse lines needed for further manipulation of mutated genes in ES cells or mice requires additional investment to fully utilize the resource. For example, while it is estimated that over 200 CRE expressing mouse lines have been produced on a variety of genetic backgrounds, full systematic characterization of these lines has not been performed leaving in question which lines are most useful in specific biological systems.<sup>6</sup> Modest investment is being made in the generation and characterisation of new CRE lines (for example, EUComm is investing in the generation of some 20 new lines) and further activity in the characterisation of extant lines and the generation of new lines is merited. The availability of the targeting constructs generated by the IKMC, as well as the placement of loxP and FLP sites in genetic loci of interest, following both targeting and trapping, provides excellent materials for further genetic manipulations using these recombinase systems. Furthermore, with the availability of a comprehensive mutant resource we envisage that individual investigators will increasingly shift their efforts to include areas such as exploring the feasibility of reversible mutations. Genetic analysis of a number of inbred mouse strains has documented numerous duplications and deletions of regions of the mouse genome, thus knockouts and other subtle mutagenic approaches will provide the ability to assess the effects of copy number variation and their epistatic interactions with other genomic alterations in population analysis of complex traits. One of the next big challenges in genome research is to determine the function of non-coding conserved regions/regulatory sequences systematically (ENCODE project, Nature 2007, 447, 801) since many of these regions are transcribed and SNPs in these regions are often associated with human disease conditions.

### **Repositories**

The availability of comprehensive ES cell mutant libraries is likely to generate significant demand for materials to generate live mice for experimentation. While some laboratories will be able to receive and utilise ES cell lines, many will require frozen embryos, frozen sperm or live animals. Repositories (both existing and planned) will need to adopt a mixed distributive approach able to disseminate a variety of resources reflecting the capability of the customer. Production centres (likely allied with or embedded within repositories) will need to be established with capacity for high throughput mouse production, generating mice from ES cells in response to customer requests, re-archiving frozen embryos and frozen sperm from live mice, and working with repositories to distribute materials as appropriate. Careful consideration needs to be given to the expansion of existing facilities as future demands are likely to outstrip current capacity. Financial support to develop these facilities needs to be considered in line with a business model for how they will become self-supporting in the long-term.

The delivery and utilisation of frozen embryos (or frozen sperm) should be encouraged to reduce cost associated with maintaining a mouse resource, improve efficiency of distribution and solve many potential issues with animal pathogen status. Increased investment in training courses for re-derivation from frozen material is urgently required with the aim of all major mouse facilities worldwide becoming proficient in the long term. Consideration needs to be given to the long-term

goal of phasing out the delivery of live animals from mutant ES or knockout mouse repositories worldwide. Repositories and production centres worldwide should establish common operating procedures and quality control standards, which would underpin performance and benchmark indicators. It will be important that production centres and repositories share developments in technology and promulgate these advances to the wider user community. For example, it is vital that centres and repositories are active in training and dissemination of procedures for the utilisation of C57BL/6N ES lines.

FIMRE, (International Federation of Mouse Repositories, <http://www.fimre.org/>), for example is well placed to lead and monitor the development of common standards with representation from Asia, Australia, Europe and North America<sup>7</sup>. Building links between databases at Production Centres, Repositories and Phenotyping Centres will allow users to efficiently identify the location of resources and relate the available mouse lines to phenotype datasets. The IMSR (International Mouse Strain Resource, <http://www.informatics.jax.org/imsr/index.jsp>) is a prototype model for a comprehensive portal that would provide a one-stop shop for searching the worldwide repositories. Consortia such as CASIMIR (Coordination and Sustainability of International Mouse Informatics Resources, <http://www.casimir.org.uk/>) are also well placed to develop a consolidated approach to the informatics solutions required for the efficient dissemination and utilisation of mouse mutant resources. In anticipation of the expansion of the frozen embryo archives at individual repositories, discussions are underway to implement the sharing of frozen material between repositories, facilitating the delivery of lines to the customer irrespective of location thus avoiding tedious delays and reducing cost due to international shipping. Additional planning is needed to develop cost recovery programs to allow the repositories to reach a level of self-sufficiency or at least to operate with minimal subsidies from external funding. A revenue sharing plan for the participating repositories is required to encourage the sharing of materials and maintain level pricing irrespective of who is distributing the materials.

### **Phenotyping and Informatics**

The ultimate purpose of the ES/mouse knockout projects is to reduce the cost of production and expedite the availability of materials for phenotypic analysis of genetic alterations to elucidate gene function. There would be significant economies and synergies to be gained by linking production centres/repositories with phenotyping centres. Customers would be able to request not only mice or frozen materials but, allied to the generation of mice, primary phenotyping could be undertaken and offered as an additional service to the customer and the information provided to the research community in a database format. This would reduce redundant efforts in primary screens that would be needed by most researchers regardless of their major areas of interest, and primary phenotypes may elicit interest from researchers who heretofore had not studied a particular gene. Areas of phenotyping under consideration extends from the molecular (the transcriptome and proteome) to cellular and organ systems. There is currently considerable focus on the development and application of observational, physiological and biochemical screens studying diverse features including amongst others dysmorphology, neurological, sensory, behaviour, metabolism, immunology and cardiovascular.

There has already been considerable focus and investment on the development and standardisation of phenotyping approaches<sup>8</sup>, particularly broad based primary screens, by a number of consortia (for example, the EUMORPHIA (<http://www.eumorphia.org/>) and EUMODIC (<http://www.eumodic.org/>) consortia)<sup>9,10</sup>. Standardisation of the phenotype procedures and detailed protocols is critical if we are to generate comparable datasets of phenotypes. Phenotyping centres with the breadth of expertise to carry out broad based phenotyping will be required if we are to undertake large-scale systematic phenotyping of the mouse mutant catalogue. It would be highly advantageous to establish an international mouse phenotyping consortium (IMPC), where the constituent phenotyping centres will work together to develop common standards and adopt common SOPs. Representatives of phenotyping centres worldwide met in Rome in December 2007 to consider the formation of an IMPC and to outline both early and longer term goals for a global phenotyping effort. The current phenotyping centres perform a

dual role: firstly, undertaking systematic phenotyping of the mouse mutant catalogue; secondly, operating as clinics and providing services (at cost) to the wider biomedical sciences community in characterising mouse mutants. Phenotyping centres, while able to offer phenotyping facilities of considerable range and sophistication, will nevertheless need to be networked within the wider community in order to provide access to the most comprehensive range of advanced phenotyping platforms and provide data and mice to more specialised phenotyping laboratories and individual investigators. In addition to the study of mutants, primary phenotyping of the 40 inbred strains that are the constituent lines of the mouse phenome project (<http://phenome.jax.org/pub/cgi/phenome/mpdcgi>) will not only provide baseline data for systematic phenotyping of mouse mutant resources, but also identify phenotypes that may more accurately reflect human disease, effectively identifying the best “humanised” background for both mutant and QTL studies (see below).

The IMPC can work together to introduce technology improvements and drive down costs both in phenotyping and mouse production (for example, homozygosing mutant ES cells and adopting techniques for the generation of 100% chimeras) in order to increase the speed and number of genes that can be studied. There are a number of key informatics issues that need to be resolved as the international phenotyping consortium emerges<sup>11</sup>. A common approach to phenotype ontologies that underpins phenome database structures needs to be developed. Critically, the phenotyping protocols need to be incorporated as part of the ontological structure. Adoption of common operating procedures and output parameters incorporated within an agreed upon mouse phenotype ontology will be a major step forward in fostering phenotype data exchange which is pivotal to the success of the enterprise.

In order to drive these phenotyping efforts, we propose that the IMPC consider the goal of completing broad based primary phenotyping of 5000 mouse lines by 2013. The IMPC should present an operational plan including indicative costs to the constituent funding agencies in the next year to allow consideration of these efforts. The expectation will be that as technology improves and costs decreases, the opportunity to phenotype increasing numbers of lines will improve.

### **Next Level Studies: Elaborating gene networks and Humanization**

A variety of tools and experimental approaches in conjunction with ES derived mutant mice are also available to elucidate gene function. The function of any gene involves the interaction, regulation and control of a myriad of other genes. The elaboration of genetic networks will require the study of mutants on a variety of genetic backgrounds, which if carried out systematically would be an enormous undertaking. However, we propose two “variation screens” as pilot studies which will prove instructive over the long term. These variation screens will be underpinned by primary phenotyping of the 40 phenome strains (see above) and the generation of high density (~250K) SNP maps and Copy Number Variation (CNV) analysis (~250K sites) of these and other reference strains (CSS, RI, CC). The first pilot study, stage 1 of the variation screen to be completed within 3 years, envisages gene expression and phenotype analysis of F1 mice from 10 mutants crossed to 20 inbred strains. A second pilot study, stage 2 of the variation screen and to be completed within 5 years, envisages gene expression and phenotype analysis of F2 populations generated from 5 mutants crossed to 5 strains. A corollary of these efforts will be the continued development of the Collaborative Cross (<http://lsd.ornl.gov/mgg/projects/collabcross.html>), a randomized cross of eight inbred mouse strains designed by members of the Complex Trait Consortium<sup>12</sup>. The cross features a randomized assortment of 8 inbred strains, A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, NZO, CAST/Ei, PWK/Ph, WSB/Ei. The lines are first crossed pairwise to make all 56 possible G1 parents and all 8 genomes are brought together in G2:F1, and the offspring of this cross are inbred. This study will enable high resolution QTL analysis for a large number of phenotypes. Current goals are to be able to map QTLs rapidly to about 100kb of DNA. These regions could potentially be further analysed using genome modifications of 100-200kb using BAC-exchange technology in ES cells. The creation and end sequencing of additional BAC libraries would

facilitate these efforts. This will be an important technology for allowing identification of the underlying causes of inbred strain variation.

The development of genome-wide mutant resources along with technical advances that will allow genome-wide tagging opens up new approaches to cataloguing the mammalian proteome. Research on appropriate tags is underway and, building on developments in high throughput mass spectrometry technology, rapid progress can be made in large-scale studies on the immensely complex and exciting study of the mammalian proteome. As shown by recent progress in yeast, proteomic mapping by tagging, generic affinity chromatography<sup>13-16</sup> and mass spectrometry provides a core dataset that is predictive for functional interactions<sup>13-16</sup>. Obviously mammalian proteomes are significantly more complex; however, the acquisition of core proteomic maps from a few well chosen cell types will provide extremely valuable guidance for the understanding of the functional landscape. Steps towards tackling the technical challenges involved in mapping mammalian core proteomes have been made<sup>17</sup> and it is likely that large-scale programs can be initiated in the near future. One such effort, the International Regulome Consortium, has already been initiated ([www.internationalregulomeconsortium.ca](http://www.internationalregulomeconsortium.ca)). Ideally, an integrated program will be based on international co-operation akin to the arrangements behind the EUComm/KOMP/NorComm relationships. Moreover, comprehensive analysis of the mouse epigenome will be critical in assessing the mechanisms and pathways underlying regulatory networks and the analysis of the impact of epigenetic changes will also be transformed by the availability of comprehensive mutant resources.

The phenotyping and utilisation of a wide range of inbred strains, along with the generation of ES lines from these lines, will contribute to providing a better understanding and improved tools for the use of the most appropriate genetic background for studying human disease states. Nevertheless, it will be important to develop and apply a wider toolbox for the humanisation of mice to create better disease models and direct study of therapeutic compounds in animal models. The introduction of human gene variants into the corresponding mouse mutants will provide a wider range of disease models associated with individual loci and can account for the genetic variation in the human population that often leads to loss of efficacy and in some cases increased side-effects. It will be important to prioritise the generation and dissemination of mouse mutants in key systems that will benefit from subsequent humanisation. Areas of interest such as the P450 genes that are responsible for drug clearance and metabolism and drug-drug interaction, the MAP-kinase pathway, the NF- $\kappa$ B pathway, GPCRs, vasculogenesis, insulin utilization, cell degeneration and apoptotic pathways, to name a few are areas amenable to humanization with current technology. The production of humanized mice or more fully immunocompromised mice would enhance current models for human cell transplants, including stem cell studies. Additional humanization can be implemented to create better models not only in immunology but in a wide variety of fields especially in mimicking neurological and psychiatric diseases. The combination of mouse models with human stem cells poses a very exciting and potentially fruitful area of research in nearly all fields of biology, potentially leading to more effective treatment of human disease.

### **Summary**

The development of an international mouse knockout resource provides unparalleled prospects for mammalian functional genomics and the study of the genetic bases of human disease. The creation of this resource raises a number of opportunities and challenges for the biomedical sciences community. Firstly, it will be critical to determine the function of the conserved-noncoding sequences in respect to genome organisation and gene regulation. Secondly, it will be important to consider harnessing the mutant resource to investigations of for example the mammalian proteome and to begin to decipher networks that will aid our understanding of cellular systems. Thirdly, there needs to be a commensurate development in infrastructure provision for the dissemination and archiving of these new biological tools. This should also be accompanied by a renewed focus on the development and characterisation of “driver” lines that will be critical for the analysis of conditional mutants, a key feature of the new mouse mutant resources.

Fourthly, systematic and comprehensive phenotyping of the mouse mutant resource will generate a rich dataset that will enormously enhance the discovery and characterisation of disease models that will underpin our understanding of disease processes. However, this goal is not without many formidable obstacles that will require innovative and cost-effective solutions in phenotyping development, standardization of protocols and informatics. Fifthly, it will be vital to reflect on the interplay between programmes in mutant generation and QTL analysis and to begin to undertake pilot variation screens that will underpin future programmes in the elaboration of genetic networks. Finally, careful consideration needs to be given to the humanisation of the mouse genome as we seek to develop better models that are more appropriate phenotypically in uncovering disease mechanisms and assessing compound efficacy. In conclusion, advances in mouse biology are leading to a new era of genetics in which we are poised for the first time to attempt a comprehensive functional annotation of a mammalian genome with enormous potential impacts upon basic and translational research.

## References

1. Auwerx, J. et al. (2004) The European dimension for the mouse genome mutagenesis programme. *Nature Genetics* **36**: 925-927.
2. Austin, C.P. et al. (2004) The Knockout Mouse Project. *Nature Genetics* **36**: 921-924.
3. Collins, F.S., Rossant, J., Wurst, W. (2007) A mouse for all reasons. *Cell* **128**: 9-13.
4. Collins FS, Finnell RH, Rossant J, Wurst W. A new partner for the international knockout mouse consortium. *Cell* 2007;129:235.
5. <http://www.eucomm.org>, <http://www.genome.gov/17515708>, <http://knockoutmouse.org/>, and <http://norcomm.phenogenomics.ca/index.htm>
6. [jaxmice.jax.org/models/cre\\_intro.html](http://jaxmice.jax.org/models/cre_intro.html), [www.mshri.on.ca/nagy](http://www.mshri.on.ca/nagy), <http://strains.emmanet.org/strains>, [www.taconic.com](http://www.taconic.com) and [www.criver.com](http://www.criver.com).
7. Davisson, M. (2006) FIMRe: Federation of International Mouse Resources: global networking of resource centers. *Mamm. Genome* **17**: 363–364.
8. Brown, S.D.M., Hancock, J.M., Gates, H. (2006) Understanding mammalian genetic systems: the challenge of phenotyping in the mouse. *PLoS Genetics* **2**: e118.
9. Brown S.D.M and the Eumorphia Consortium (2005) EMPRESS: standardised phenotype screens for functional annotation of the mouse genome. *Nature Genetics* **37**: 1155.
10. Gailus-Durner, V. et al. (2005) Introducing the German Mouse Clinic: open access platform for standardised phenotyping. *Nat. Methods* **2**: 403-404.
11. Hancock, J.M. et al. (2007) Integration of mouse phenome data resources. *Mammalian Genome*, in press.
12. Complex Trait Consortium (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature* **26**: 1133–1137.
13. Gavin, A.C. et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**: 631-6.
14. Krogan, N.J. et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**: 637-43.
15. Collins, S.R. et al (2007) Towards a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol Cell Proteomics* **6**: 439-450.
16. Collins, S.R. et al. (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**: 806-810.
17. Blagoev, B., Mann, (2006) Quantitative proteomics to study mitogen-activated protein kinases. *M. Methods* **40**: 243-50.