

The International Mouse Phenotyping Consortium

Concise Business Plan

November 15, 2010

Dr Mark Moore and the IMPC Steering Committee

Table of Contents

SUMMARY OF THE IMPC

TABLE OF CONTENTS	2
BACKGROUND.....	3
RECOMMENDATIONS	3
VALUE TO THE FUNDING ORGANIZATIONS	4
1.1 THE MOUSE AS A MODEL ORGANISM	6
1.2 RATIONALE FOR A SINGLE-GENE KO APPROACH	8
1.3 THE INTERNATIONAL KNOCKOUT MOUSE CONSORTIUM (IKMC).....	8
1.4 SYSTEMATIC PHENOTYPING – AN EMERGING CONSORTIUM	9
1.5 COMMUNITY CONSULTATION	10
1.6 PROPOSED IMPC GOALS	13
1.7 DELIVERABLES FOR THE IMPC	14
1.8 KEY REQUIREMENTS FOR DELIVERY OF IMPC GOALS	15
2.1 REQUIREMENTS FOR CONTRIBUTING CENTRES	17
2.2 INFRASTRUCTURE REQUIREMENTS	18
2.3 MOUSE PRODUCTION	18
2.4 ARCHIVING	19
2.5 HOUSING AND ANIMAL CARE	20
3.1 GOVERNANCE.....	22
3.2 IMPC PROJECT MANAGEMENT PLANS	23
3.3 MONITORING AND EVALUATION OF THE PIPELINES	24
3.4 INTERACTIONS WITH OTHER ORGANIZATIONS	24
3.5 FUNDING MODELS.....	25
3.6 BUDGET REQUIREMENTS	25

Executive Summary

Background

The International Mouse Phenotyping Consortium (IMPC) Steering Committee¹ was formed from a group of funding organizations and major mouse research centres to consider the opportunities presented by the large number of mouse mutants generated by global efforts under the auspices of the International Mouse Knockout Consortium (IKMC) and to maximise the biomedical value of these mutants through large-scale phenotyping. Specifically the Committee's goal was to develop a plan to harness the major world-wide mouse research programmes and infrastructures in a strategic and coordinated effort to undertake a broad-based, systematic genome-wide phenotyping project of knockout mice in order to provide the wider research community with a long lasting resource of mammalian gene function information.

In order to formulate this plan and to engage the broader scientific community for input as to the need, scope, scale, and composition of the constituent programmes, the IMPC Steering Committee engaged a Project Manager. This business plan has been produced by the Project Manager and provides details of the goals, deliverables and requirements for an IMPC based upon diligent research and significant discussion among the IMPC Steering Committee members along with end-user community engagement and discussions. The plan outlined here will take advantage of the enormous strides in mouse mutant generation, the development of experimental tools and technology platforms for phenotyping, and the operational capacities available around the world.

Recommendations

In order to accomplish the IMPC vision of producing an encyclopedia of mammalian gene function, the IMPC would need to:

- Establish a world-wide consortium of 6-10 large mouse centres with capacity and expertise for large-scale primary phenotyping;
- Establish a world-wide consortium of mouse production centres to generate germ line transmission of targeted knockout mutations in embryonic stem cells for all known and predicted mouse genes;
- Test each mutant mouse line (4,000 mouse lines in the first 5 years, and ultimately up to 20,000) through a broad based primary phenotyping pipeline in all the major adult organ systems and most areas of major human disease;
- Through this activity and employing data annotation tools, systematically aim to discover and ascribe biological function to each gene, driving new ideas and underpinning future research into biological systems;
- Establish collaborative “networks” with specialist phenotyping consortia or laboratories, providing standardized secondary phenotyping that enriches the primary

1 Michael Dunn(Wellcome Trust), Nathan Richardson (MRC), Allan Bradley (WT Sanger Institute), Steve Brown (MRC Harwell), European Commission, Martin Hrabe de Angelis (Infrafrontier/German Mouse Clinic), Colin McKerlie (Toronto Centre for Phenogenomics), Eric Green and Jim Battey (NIH), Yann Herauld (ICS Strasbourg), Chris Goodnow and Adrienne McKenzie (Australian Phenomics Network).

dataset, and end-user, project specific tertiary level phenotyping that adds value to the mammalian gene functional annotation and fosters hypothesis driven research;

- Enhance the availability and range of mouse disease models that can be used to study human disease pathology and treatment;
- Cryopreservation, or arrange to cryopreserve with a third party, a genome-wide archive of mouse mutant stocks and create a distribution portal to the end-user research community to facilitate further discovery while promoting animal welfare;
- Provision of a centralized data centre and portal for free, unrestricted access to primary and secondary data by the scientific community, promoting sharing of data, genotype-phenotype annotation, and the development of open source data analysis tools;
- Establish quality control, quality assurance, and programme management systems to guarantee the highest quality scientific and operational output;
- Act as a coordinating body to recruit additional major mouse phenotyping initiatives, for example in emerging EU countries, to complement the international effort;
- Champion the ethical use of mouse genetics and phenotyping, promoting the refinement, reduction and replacement of procedures where at all possible.

It is anticipated that the discoveries which will emerge from the IMPC programme will produce a paradigm shift in our understanding of basic molecular, cellular, and systems biology, as well as feed the biopharmaceutical discovery pipeline by enhancing our understanding of the genetic bases for disease. The study of novel genes through the phenotyping of mutant mouse lines, and particularly knockout (KO) mice, provides the opportunity to unlock new areas of biology and will lead to novel drug target discovery. Large-scale biology projects offer economies of scale allied to capacity that shifts the burden of risk and funding demands from individual labs, freeing these experts to focus on their core competencies of in-depth research discovery. In summary, the proposed endeavor provides a unique opportunity for breakthrough discoveries, as well as laying the foundation for a profound understanding of mammalian gene and physiological systems in the coming decades.

Value to the Funding Organizations

1. Immediate and Free Access to Mouse Resources of Lasting Biological and Medical Value

The IMPC plan is to harness and co-ordinate capacity on a world-wide level, with open and unfettered access for everyone. Participating organizations will be able to influence the selection of genes entering the pipeline and how they are tested. Scientists not involved directly in the consortium will also be able to access models for their own phenotyping studies. Mice will be distributed to specialist screeners who will undertake more sophisticated and higher content analysis not suited to a high-throughput environment. The network model takes full advantage of capacity for centralized infrastructure and actively engages with specialized expertise dispersed throughout the community.

2. Enhanced delivery and involvement of institutions in the identification and characterisation of disease models

Investing in IMPC presents an unprecedented opportunity to increase the scale of output and dramatically increase our understanding of disease mechanism. Organizations involved in the IMPC will help steer development of the IMPC and its programmes to

ensure the receipt of appropriate benefits and resources for their research communities. It will also improve access to these resources and platforms by the proposed increases in activities in mutant generation and phenotyping.

3. Retain or develop competitive position internationally in phenotyping approaches and disease model discovery, with consequent wide-ranging benefits to biomedical science

Involvement with the IMPC will afford organizations the opportunity to develop or maintain a competitive edge by benefiting from the expected technological advances in phenotyping platforms, new instrumentation and the application of miniaturization and automation, as well as wider developments in mouse genetics. A key component is likely to be the development of new physiological and imaging approaches.

4. Strengthen ability to develop networks of laboratories involved in phenotyping and mouse mutant characterisation, enhancing mouse genetics across ones country

The IMPC would be pivotal to continue to develop the extensive network of laboratories, including funded programmes that avail themselves of local expertise and facilities. Several IMPC members have successfully developed pre-clinical research networks in metabolism, immunology, neurosciences, bone, liver, deafness, and cardiovascular disease amongst others that have had an impact both on global phenotyping advances as well as the promulgation of individual laboratory studies. The IMPC plans to further strengthen these and other networks that will develop. The networks would be sustained and enriched through IMPC leadership and, would benefit from the competitive position that would be realized by involvement in this global project. The networks can take advantage of access to the primary phenotyping platforms and that component of the pipeline that would be devoted to investigator initiated phenotyping. Moreover, these networks would be able to bring considerable input to bear upon the research and development of the IMPC pipelines, and importantly, research and development into data analysis pipelines. Overall, individual researchers would be enormously strengthened by direct involvement in the IMPC.

5. Ensuring that researchers can continue to build upon current strengths in phenotype databases and disease ontologies, enhancing our understanding of disease states

The IMPC plan is to deliver an extensive database of primary and meta-data which will transform mammalian biology. Several current groups have developed key informatics strengths in functional genomics, which have contributed to developments in phenotype studies. MRC Harwell has led the development of the EuroPhenome database, which is seen as a potential model for a future IMPC database. Meeting the future informatics challenges in mouse genetics and phenotyping will require advances in informatics – the development of software tools and ontologies for data mining and analysis. Importantly, involvement in this area is critical to ensure that data will be integrated with human clinical data, and further developing methods to map mouse phenotypes to human disease states.

Part 1 – Scientific and International Context

1.1 The Mouse as a Model Organism

The biomedical community is posed with enormous challenges and opportunities in the next decade. The advent of large-scale sequencing has propelled the field of genetics and its relationship to human disease to an unprecedented pace. While next generation sequencing technologies now make Genome Wide Association Studies (GWAS), and even the sequencing of the genomes of individual patients practical, the need for basic understanding of gene function is as great as ever. Despite these advances, the function of a least half of the mammalian genome is poorly understood and nearly one third has no functional annotation whatsoever.

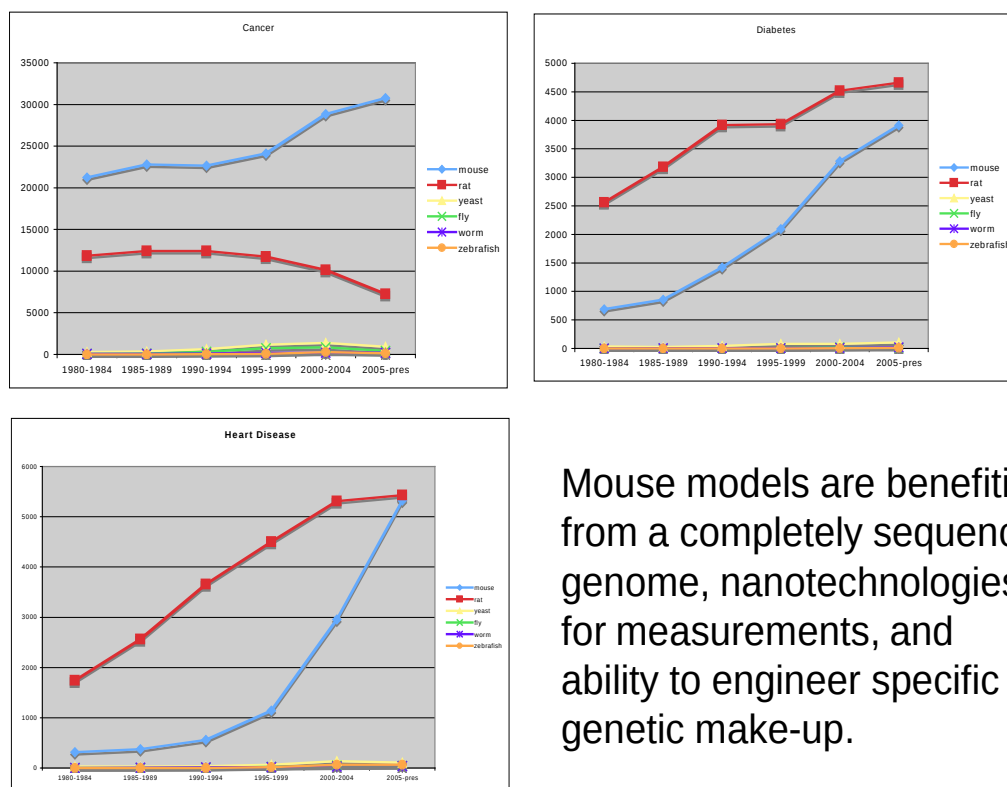
Throughout medicine, it is seldom that the underlying genetic cause of a disease is treated, but interventions in the pathway or systems related to the disease are often the entry point (drug target) utilized to relieve symptoms and suffering. The prevailing view is that many diseases are of multi-genic origin, and strongly influenced by environmental factors, in which the combination of various gene products lead to disease. This paradigm shift in our understanding of disease mechanisms underlines the necessity to understand the basic biology of pathways and systems in order to treat the symptoms and downstream deleterious effects of these diseases. The corollary is that a targeted, bottom-up approach to studying mammalian gene function, making *a priori* assumptions about the role of individual genes in disease (perhaps based upon the identification of a gene association in human disease) can only provide a limited view of gene function at most. Rather, there is an opportunity to undertake a hypothesis-generating, comprehensive and systematic programme for the phenotypic study of the function of all genes, which will help usher in a new era of genomic medicine.

One of the most important tools at our scientific disposal in understanding mammalian gene function is the laboratory mouse. The scientific community has taken advantage of its fundamental similarity to humans at the genetic level (>95% at the gene level), similar physiology and anatomy, its relative low cost compared to other mammals, and nearly 100 years of genetic study. There is an extensive toolkit for the manipulation of the mouse genome and the generation of new disease models. The development of mouse embryonic stem (mES) cells, and the ability to manipulate them *in vitro* to mutate any gene at will and transmit that mutation into the mouse germline led to a revolution in basic medical research over the past 20 years, and was recognized by the awarding of the Nobel prize in Physiology and Medicine 2007 to Drs. Capecchi, Evans, and Smithies. At the same time, large-scale chemical (ENU) mutagenesis programmes have provided hundreds of new disease models and highlighted many examples of novel gene function that could not have been predicted on the basis of existing knowledge of the gene product.

The importance of the mouse as a model organism was also recognized by the inclusion of a goal for the construction of genetic and physical maps of the mouse genome in the initial plan for the Human Genome Project (HGP). In fact, a group of centres involved with the HGP was able to complete a high quality, finished sequence of the mouse genome (strain C57BL/6) in 2005. Mouse models represent one of the best methodologies for validating and

understanding the roles that genes play in disease susceptibility, progression and response to new treatments.

Mouse mutants with phenotypes that mimic human traits have served as critical research tools in understanding the genetics underlying mammalian biology, and have been crucial in determining gene function and pathway interactions for decades. As genomics moves from gene identification to mammalian functional genomics the mouse will continue to play a critical role, particularly by utilizing gene knockouts. Janan Eppig of The Jackson Laboratory recently presented the results of a survey of the Mouse Genome Database (MGD) of published model organisms used in studies of human disease². The study demonstrates the rise of the mouse as the most common experimental organism to study human disease and gene function, and this is closely correlated in time with the development of mES cell knockout technology and the ability to build better models. In fact most of the current increases are due to the development of mouse knockout models, humanization of the mouse, and point mutations. The example in Figure 1 below demonstrates for three major disease areas the community uptake of mouse models in Diabetes, Cancer, and Cardiovascular research in the past 20 years.



Mouse models are benefiting from a completely sequenced genome, nanotechnologies for measurements, and ability to engineer specific genetic make-up.

Figure 1. Model Organism Publications. MGI search results by year demonstrating the number of publications utilizing the indicated model organism by disease area.

1.2 Rationale for a single-gene KO approach

While many human diseases have a multigenic basis, there is often little knowledge of the nature of disease initiating events, or the overall contributions of each genetic variant in the onset and severity of a disease. It seems reasonable that the best way to approach the complexity of such diseases is to understand the underlying function of each gene, its involvement in pathways, and function in organ systems in order to better understand its role in biological systems. This will require the study of single gene knockouts or other single mutants as a fundamental step in dissecting the biology of the disease system. Furthermore, when it comes to assessing and designing therapeutic interventions, single gene mutations will be key to understanding the effects of blocking or knocking out the functionality of that gene in the whole organism as a way to gauge not only efficacy but also potential side-effects of a treatment. If a disease involves multiple genes and hence their protein products and targets for therapeutics, it will be necessary to understand which individual targets in the process may provide the maximum clinical benefit. It is well recognized that the determination of the effects of mutations in individual genes is a powerful tool for dissecting the genetic basis of disease pathways.

In this context, knockout mouse mutants and ENU mutants will be a first step in assessing candidates revealed by GWAS. While GWAS are illuminating the potential loci involved in complex diseases in human, the mouse will be the critical vehicle for assessing the role of the genes identified in the disease processes. For example, the identification of the FTO gene in GWAS for obesity has been explored and validated by the use of a variety of mouse mutants from knock-outs to ENU mutations to over-expressing transgenics. The analysis of each of these mutant mouse lines has depended in several cases on the skills and expertise of the mouse phenotyping centres at the institutes involved.

1.3 The International Knockout Mouse Consortium (IKMC)

The increasing use of KO mouse lines as a critical primary research tool, and the evidence that many mouse lines had been created redundantly by multiple labs led to the concept of a large-scale knockout project. The proposition to use KO mice as part of a large-scale effort in functional genomics was the subject addressed by the attendees of an international meeting convened in the fall of 2003 at the Banbury Centre, Cold Spring Harbor, NY. The participants strongly supported the establishment of a focused, large-scale international effort to produce a publicly available, comprehensive collection of knockout mouse lines containing a null mutation in every protein-coding gene in the mouse genome (Austin, C.P., *et al.* Nature Genetics 36, 921-924, 2004). The meeting attendees also endorsed a plan of a phased production approach, beginning with the generation of a resource of mES cells comprising a comprehensive collection of null, ideally conditional alleles, to be followed by the production of mice from the mES cells and then phenotyping the mouse lines in a step-wise fashion with an increasingly sophisticated set of tests. The European community of scientists has also called for large-scale mutagenesis and phenotyping of every gene in the mouse genome (Auwerx, J., *et al.* Nature Genetics 36, 925-927, 2004).

Several international agencies have funded efforts to create conditional knockout mutations in mES cells; in 2005 the European Conditional Mouse Mutagenesis Programme (EUCOMM) received funding to generate 20,000 conditional mutations (12,000 conditional gene-trap mutations and 8,000 conditional targeted mutations) in mES cells, and a complementary effort in Canada, the North American Conditional Mouse Mutagenesis Programme (NorCOMM), whose goal is to produce 500 targeted mutations in mES cells was also funded

in 2005. Subsequently in 2006, the National Institutes of Health (NIH) in the US funded the Knockout Mouse Project (KOMP) to generate 8,500 knockout mES cell lines. To achieve coordination of these efforts, as well as that of the Texas A&M Institute of Genomic Medicine (TIGM; which has a large collection of gene trap mES cells), these groups have formed the International Knockout Mouse Consortium (IKMC). The IKMC has created several working groups to facilitate coordination and to provide solutions and best practices for all programmes. The efforts of these programmes have now produced >8,000 targeted KO in C57BL/6NTac mES cells, mostly (~5,800) conditional ready, accounting for one-third of the mammalian genome, and over 10,000 gene trap clones available from TIGM. Projections for KOMP, NorCOMM, and EUCOMM are that the majority (>95%) of the mouse genome will have a knockout allele by 2012, thus fulfilling the first recommendation from the Banbury meeting. Continuing the analysis of the mouse and human genomes using various and developing predictors for gene identification has the potential to increase the current gene count, and that mutagenesis efforts should branch into non-protein coding genes, the generation of point mutations linked to human disease, and humanization of the mouse.

1.4 Systematic Phenotyping – an Emerging Consortium

The success of the IKMC and emerging world-wide phenotyping efforts including those of the Wellcome Trust Sanger Institute Mouse Genetics Programme (WTSI MGP) (<http://www.sanger.ac.uk/mouseportal/>), and the European Mouse Disease Clinic (EUMODIC) programme (<http://www.eumodic.org/>), the first internationally coordinated large-scale phenotyping effort funded by the European Commission (EC), has led to much discussion regarding the possibility of a coordinated international programme in mouse phenotyping. The EUMODIC programme is comprised of four mouse phenotyping centres (MRC Harwell, WTSI, ICS Strasbourg, and the German Mouse Clinic (GMC) Munich), whose goals are to produce mutant mouse lines and to provide phenotype information on 500 IKMC knockout mouse lines by 2011. One of the early phenotyping protocols utilized in EUMODIC was termed EMPReSSslim (<http://empress.har.mrc.ac.uk/>) and was developed by another EC funded programme (the EUMORPHIA Programme). The protocol has now developed further with input from all partners to the current version of the EUMODIC pipeline (<http://www.eumodic.org/>). Ultimately, the phenotypic data is made accessible through an EC funded database, EuroPhenome. (<http://www.europhenome.org/>), and the Sanger Mouse Portal (<http://www.sanger.ac.uk/mouseportal/>). The major advances in the EUMORPHIA and EUMODIC (<http://www.eumodic.org/>) programmes have been the harmonization of phenotyping experimental platforms, through standardized protocols, based on the sharing of know-how and experience; these efforts were the first to address the enormous technological and logistical problems associated with large-scale internationally-coordinated phenotyping. This has been an important stepping-stone towards an international programme of mouse mutant phenotyping on a genome-wide scale.

Several meetings have taken place in the past 3 years, (Rome in 2007, Bar Harbor and Toronto in 2008, and a large UK focused meeting in November 2008 hosted by the Wellcome Trust and MRC). The aim of these meetings was to establish a vision for a co-ordinated global effort in mouse phenotyping and these have provided an inclusive forum to discuss international, coordinated phenotyping efforts. The consensus conclusion of these meetings was that there is an exciting opportunity for the research community to engage and develop a coordinated mouse phenotyping programme. The discussions and recommendations from these meetings have driven the structure and content of the plans for an International Mouse Phenotyping Consortium (IMPC) programme. If the IMPC is to be a community driven

project, it will require the engagement of a broad sector of the scientific and clinician-scientist communities to ensure uptake of the project deliverables, translation of phenotyping outputs to the clinical research community, and ultimately project success.

The vision of the IMPC is to provide phenotypic information on a knockout mouse line from every gene in the mouse genome and effectively create an Encyclopedia of Mammalian Gene Function. Initial planning efforts have been primarily focused on maximising utility of the IKMC resource, but the developing infrastructure and approaches will increasingly be utilized by the other major mouse mutagenesis pipelines, including ENU point mutations, transposon insertions, chromosomal translocations, deficiencies, and duplications. In developing plans to deliver this vision it has been important to consider the merits of various models for providing broad based phenotype information on the large collection of mutant mES cells being generated by the IKMC.

Similar to the Human Genome Project that was led by major genome centres with a single focus, it is considered vital to the success of the IMPC programme that large-scale centralized production centres and phenotyping centres be utilized. Community input to the centralized phenotyping centres is needed to overcome technological difficulties and ensure that the phenotyping will be biologically significant and useful to the broader scientific community.

1.5 Community Consultation

The IMPC Project Manager, with support from the Wellcome Trust and MRC, formulated an online survey with a list of 7 key questions for consideration for the role of the IMPC, and distributed it to over 100 UK researchers, to gain additional information and help recruit workshop participants (please see Appendix 1 for the survey results). In order to develop plans and garner community support, focused workshops were held at the Wellcome Trust on 19, 21, and 23 October 2009 to engage end users, specialist phenotypers, clinicians, and pharmaceutical experts to gain their perspective on what the resource should deliver, and what would facilitate their research. The workshops were held for 3 hour sessions attended by investigators and members from the IMPC steering committee, the Wellcome Trust, the MRC, the WTSI, and MRC Harwell, and divided into the following areas of biology based on the survey responses:

- Immunology/infectious disease/respiratory/skin
- Cardio/metabolism/endocrinology
- Neurobiology/vision/hearing
- Cancer/stem cells/development/fertility/bone/muscle

As a starting point the participants were given the EUMODIC pipeline procedures (EMPreSSslim) (Figure 2) and the WTSI MGP phenotyping pipelines (Figure 3) for evaluation and comment.

EMPreSSslim Primary Phenotyping Pipelines

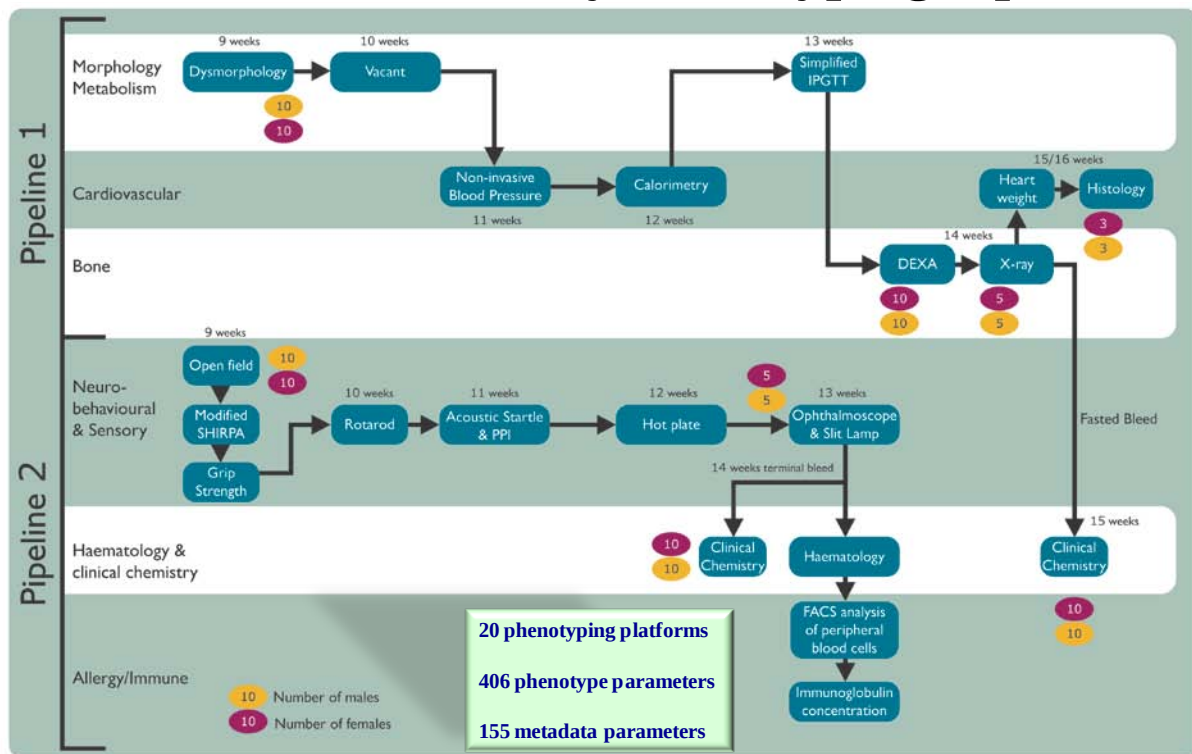


Figure 2. The EMPReSSslim and EUMODIC Phenotyping Pipeline. The above figure outlines the current phenotypic pipelines, disease areas and tests currently in use in EUMODIC.

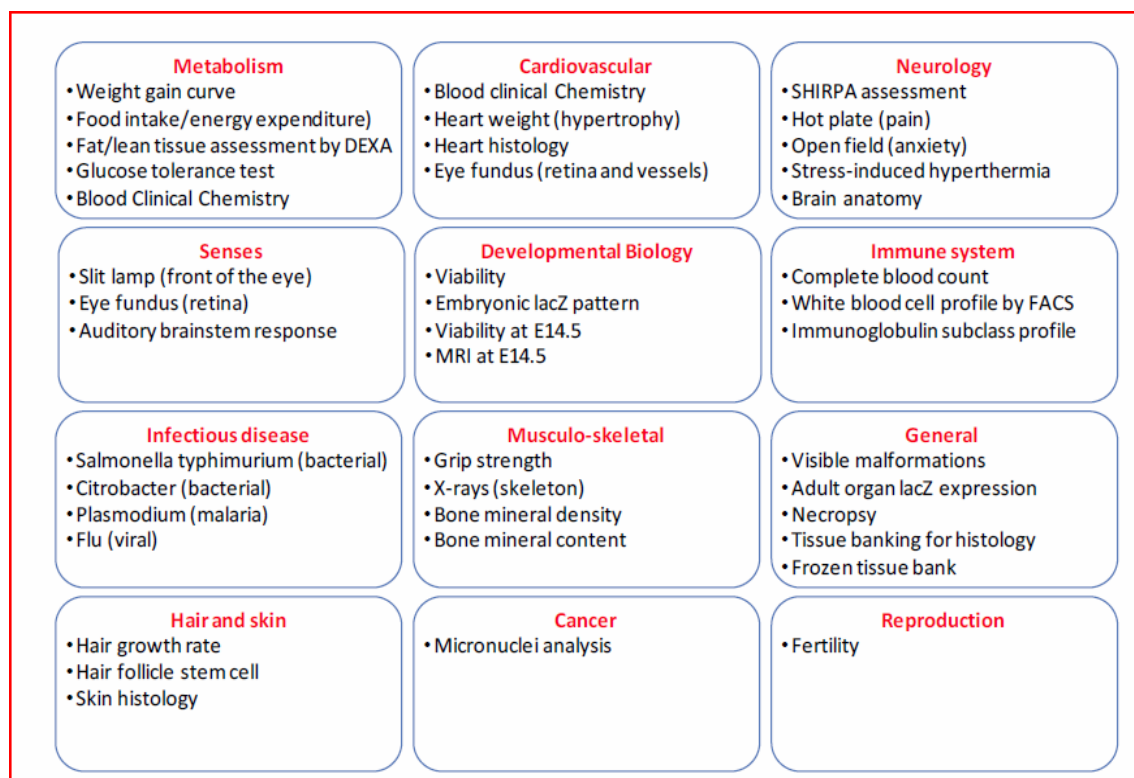


Figure 3. MGP Phenotyping Platform. The figure indicates areas of biology and disease focus that are under study in the MGP phenotyping platform. A list of each test is provided under each area of study.

In coordination with the UK surveys and workshops, the NIH issued a Request for Information (RFI) and Dr. Kent Lloyd, UC Davis and PI of the KOMP repository programme, conducted a survey of over 2,000 customers of the Mutant Mouse and mES cell Repository, using the same set of questions developed by the IMPC steering committee. The results were quite remarkable, in that there was nearly universal enthusiasm about the project in all surveys and workshops, and two important conclusions emerged:

1. Strong recommendation for an IMPC pipeline similar to a combination of the WTSI MGP and the current EUMODIC pipelines of phenotyping, as a basis for the development of future pipelines.
2. A very strong desire for broad-based phenotyping, incorporating more tests and challenge models.

Therefore, it has become increasingly clear that an IMPC needs to develop more phenotyping tests and challenge tests at the high throughput level to uncover clinically relevant and biologically significant phenotypes, in order to meet the research communities' needs. It was suggested strongly by a large majority that all of the broad based screening should be done in centralized phenotyping groups for quality control and to save on mouse production costs.

The alternative suggestion to make mouse lines and distribute these to groups that would perform a specific set of tests for phenotyping but not the entire pipeline, for example behavior, and to send mouse lines to other research groups for a different type of analysis, such as metabolic screening has been considered. This model was discussed above and should be rejected on cost grounds. Moreover, there are also practical impediments. Such a system would be impractical because it is not easy or accepted general practice to ship the

same cohort of mice multiple times, as the shipping itself causes changes in certain phenotypic tests and there are substantive technical problems in that most modern day mouse facilities employ strict quarantine procedures and would not accept such mice. The alternative of breeding multiple cohorts of the same mouse line to ship to multiple groups would be cost prohibitive as described above and would increase costs by 2-4 fold per mouse line. A centralized broad-based phenotyping of mice, which is performed on a single cohort at each mouse phenotyping centre, is the most practical way to proceed, and was strongly supported by the research community in the surveys and workshops. The distributive model for secondary and tertiary phenotyping is an excellent way to move forward to place mouse lines in the hands of specialist who do not or cannot perform large-scale mouse production.

In summary, the overall recommendations from the community are as follows:

- Build upon the pioneering work of EUMODIC and the WTSI MGP.
- Adopt a pipeline similar to the standard phenotyping tests, analyses, and examinations used in the EUMODIC and the WTSI MGP pipelines.
- Future additional primary tests should be selected based on their reliability, power, and likelihood to reveal disease-relatedness and human clinical relevance. The most important identified was imaging. Opportunities exist to bring in actively developing atlases of mouse anatomy, combined with current high throughput imaging technologies, such as MRI and Micro-CT, which could benefit the IMPC.
- Selection of targets for secondary specialized phenotyping should be based on results from primary phenotyping screens conducted at centralized sites on as many genes as possible.
- Phenotyping data and information should be made immediately available without restriction through a centralized web portal that is user friendly and well advertised. It was strongly recommended to study novel genes, with no known function, as a priority. It was highlighted that it is difficult to obtain funding for the study of novel genes through individual grants but this holds high promise for breakthrough

1.6 Proposed IMPC Goals

The vision for an IMPC requires a long-term (10 year) commitment and concerted effort from many partners. This project is a huge challenge and it is recommended that it be approached in two phases:

- Phase I (2011-2016): Complete and assess existing programmes (e.g. EUMODIC), build capacity in existing or add additional production and phenotyping centres, create a central database, data centre and portal, standardized production and phenotyping of 4,000 targeted genes.
- Phase II (2016-2021): Assess performance of Phase I, adjust pipelines and operations as necessary, and scale operations to complete the genome i.e. standardized production and phenotyping of 16,000 targeted genes.

The goals set out for Phase I of the IMPC, have been planned based on existing experiences and current technical expertise. Phase I would build on the already robust efforts at the WTSI MGP and EUMODIC programmes and ensure that this phase of IMPC would add significantly to our understanding of the function of up to 4,000 genes. In order to achieve the aggressive goals of Phase I, capacity issues need to be carefully considered, and it will be

necessary for most current large production and phenotyping centres to contribute to the project, and for current centres to expand or enhance their operations in order to be in position to scale-up for Phase II. During a ramp-up period in Phase I, there will be an opportunity to explore new technologies and additional phenotyping tests to enhance and optimize the programmes in preparation for the fully operational Consortium.

The following overarching goals are recommended for Phase I (2011-2016) of the IMPC programme:

- Generate broad based phenotype data for at least 4,000 knockout mouse lines with a minimum of at least 100 genes per year for each phenotyping centre and incorporating a proportion of investigator driven requests for gene prioritization;
- Refine the phenotyping pipeline, bringing technological improvements to bear and providing significant cost savings;
- Develop a centralized database for IMPC and disseminate phenotype data to the wider community through a common web portal;
- Deposit all of the mutant mouse lines into biorepositories and promote their distribution and sharing throughout the community;
- Develop new data visualization and analysis tools for phenotype data mining;
- Enhance interactions with the wider biomedical sciences community, fostering networks of interaction with centres performing secondary and tertiary phenotyping.

There has been substantial ground work that supports the potential of the IMPC to realize these goals. Several centres have the capability to phenotype 100 lines per year (already exceeded by WTSI) and with the recent commitment of funding by the NIH to fund the production and phenotyping of 500 lines per year for the next 5 years, these goals are achievable.

1.7 Deliverables for the IMPC

The IMPC aims to ultimately deliver an encyclopedia of mammalian gene function and associated materials, but will need to meet milestones and other deliverables to accomplish this goal. For Phase I of the IMPC the following deliverables are required:

- Microinjection or aggregation of 6,000 mES cell lines
- Selection of IMPC core phenotypic tests and platforms
- Release of the database
- Population of the database with information on at least 4,000 KO mouse lines
- Cryopreservation of 4,000 mouse lines
- Demonstration of successful distribution of biological materials

Progress in the programme will be evaluated by the funding agencies to ensure that the resource is being delivered efficiently, has maximum utility and that the overall value of the knockout mouse phenotyping effort warrants continued funding. If successful, decisions will be required as to whether or not to proceed with the programme and, if so, how to ramp-up to complete the functional analysis of the mammalian genome.

1.8 Key Requirements for Delivery of IMPC Goals

The IMPC needs to complete certain required elements in order to create thousands of knockout mouse lines from mES cells, phenotype these mouse lines, and provide a high quality data set that is of maximum utility to the scientific community. Evidence from interviews of all of the current EUMODIC clinics and many other large-scale mouse producers and users clearly indicates key areas that will be required for the IMPC.

1. Mouse Production. This is the first step in the project and by far the greatest expense and potential bottleneck. The IMPC needs mES cell microinjection and aggregation expertise and capacity to ensure the production of high quality germline chimeras to produce phenotyping-ready cohorts of knockout mouse lines. Any improvements in this process would ensure the delivery of the product, increase the speed and stability of the resource, and could greatly reduce costs. Indeed, as was the case in the Human Genome Project, EUComm, NorComm, and KOMP, great efficiencies were realized as part of running large operations, and led to technical and process improvements, lowered costs and made the projects successful. It is likely that the efficiencies of mouse production will improve as the IMPC project proceeds. Recent technology developments in the IKMC's mES cell programmes have already significantly addressed the efficiencies of generating germline chimeras from C57BL/6NTac mES cells.

2. Standardized high value phenotyping. One of the strengths of the IMPC plan is to utilize a standardized pipeline of broad-based phenotypic assays on a defined and standardized single mouse strain background, as well as a battery of approved optional tests that will increase the likelihood of discovery of a function for all genes. It is critical that the IMPC stay abreast of current technologies, the output of existing programmes, and monitor the quality of the data. As the data is more valuable as a complete collection, it is important that members of the IMPC phenotyping centres maintain the same set of core phenotypic tests. The IMPC must maintain this approach and implement changes if necessary in common. Each phenotyping centre may add any additional tests that are proven to be valid, but should not alter the agreed upon core tests unless all groups agree to the changes.

3. IMPC Database. The informatics efforts of this project are just as valuable as the phenotyping data itself, as the web based interface is what the scientific community will see and use to access the data. The IMPC needs a custom high quality and industrial-strength relational database to collect, store, manipulate and mine the data produced from the IMPC. The initial and substantive efforts by both the EuroPhenome project and the WTSI MGP can serve as a springboard for the IMPC database. However, due to the complexity of the number of different tests, multiple international partners and a customer base of tens of thousands of users, the IMPC will require a substantial investment to develop and maintain this database.

4. Mouse Repositories. The mice produced will be of huge value for future research programmes throughout the world and it is absolutely essential for the IMPC to adequately plan for the long-term sustainability and open access to these resources (sperm, embryos, tissue, mice). Biorepositories like Harwell in the UK and the MMRRRC and the IMR at The Jackson Laboratory in the US exist, as do distribution networks like EMMA in the EU. The IMPC should take full advantage of these existing infrastructures,

indeed most partners are already directly involved in managing biorepositories. However, the scale of the IMPC presents a challenge to the capacity and sustainability of existing efforts so an important aspect of the IMPC will be to raise sufficient operational funding and invest in capital infrastructure to maintain these programmes over the 10 year time frame.

Part 2 – Operational & Infrastructure Requirements

2.1 Requirements for Contributing Centres

The criterion of requiring phenotyping of a minimum of 100 mouse lines/year in order to be an IMPC partner and whether a lower number could be regarded as a useful and participating contribution has been seriously considered and debated. With the goal of covering the entire genome in the next 10 years, there is a strong need for phenotyping of a large number of mouse lines to be performed in each phenotyping centre; otherwise there is the less manageable alternative of a very large number of centres doing fewer lines. This could pose serious logistical issues; for example IT issues and data quality issues multiply. Delivery of robust phenotyping standards is underpinned in high-throughput (HTP) biology programmes, providing better data assurance and quality. Moreover, a standard IT model is to deliver more data from fewer centres ensuring quality assurance and data management. The number of 100 mouse lines is recommended as a starting point as all of the current EUMODIC partners and the Toronto Centre for Phenogenomics have experience with numbers in this range and have made predictions on the cost of the programme at this level. It would not seem to be practical to undertake this programme with 40 partners working on 50 mouse lines per year to achieve the goal of 20,000 genes functionally annotated in 10 years. Overall, one might expect that it will be much more efficient, ensure quality data standards, and lower overall cost to engage centres that can analyze 100 mouse lines per year level in Phase I - and subsequently develop plans to increase this further to over 200 mouse lines per centre in Phase II. These goals appear achievable as one centre, the WTSI, is already operating at the level of 200 mouse lines per year. Some groups will make different types of contributions to the IMPC. Given current funding realities, the European Infrafrontier programme (<http://www.infrafrontier.eu/>) has proposed to work collectively to reach the level of 150 knockouts per year. The Australian Phenomics Network (APN) has expertise in a wide variety of phenotypic assays and plan to contribute with downstream secondary analysis. The APN also propose to create an archived resource of mouse strains with defined deleterious single nucleotide variants (SNVs) changing the protein sequence of all mammalian genes together with accompanying phenotypic data from challenge screens. The APN propose to conduct a pilot project to be run on 200 pedigrees of ENU mutant mice, providing a searchable, requestable resource of phenotype and SNV genotype data and archived sperm for **~9400 protein-changing SNVs** in C57BL/6 mice.

Phenotyping centres involved in the primary IMPC programme are likely to:

- **Perform agreed upon core broad-based phenotyping.**
- **Analyse a minimum of 50-100 mouse lines per year and preferably demonstrate the capacity to ramp up to at least 100 mouse lines after 3-5 years (or comparable accepted contribution).**
- **Deliver data in the appropriate formats and in a timely fashion to a centralized database.**
- **Assist in the analysis and curation of their data.**
- **Be flexible to add or drop tests in conjunction with the guidance of the IMPC.**

- **Develop collaborations or internal programmes to advance phenotypic hits to higher level phenotyping such as secondary and tertiary analysis.**
- **Cooperate, share, and distribute know-how and experience to enhance the IMPC programme and assist new centres to develop.**

As the IMPC grows it will be important to maintain high standards but to also encourage the active participation of other groups that can make substantial contributions such as Secondary Phenotyping Centres, technology groups, and specialty areas.

2.2 Infrastructure Requirements

It is anticipated that all mouse production centres involved in the IMPC will participate in some if not all of the work activities. The main resources required for a production centre fall into the access to adequate capital facilities (e.g. high health status vivarium, transgenic labs, etc.), capital equipment (mainly phenotyping platforms), trained staff, and direct operational expenditure for salaries, running costs etc. In addition, budget forecasts for most institutions now require “full economic costing” which includes allocated expenditure for site and corporate overheads (e.g. capital depreciation, HR, security, finance, etc.). Each production centre will work to its own cost models.

All mouse production centre managers agreed that the production of mice is the single greatest expense in this process (70-80% of the total cost to generate a mouse line from mES cell to complete phenotyped dataset) and that any means to decrease this, increases the number of mouse lines and number of tests that can be performed. Another way to view this is that the addition of an additional cohort of mice only adds 10-20% to the cost, while doubling the number of data points or types of assays. A major issue is that currently only WTSI and the TCP Transgenic Core in Toronto can take hundreds (250) of mouse lines per year from mES cells to mice. Other centres will need to ramp up this expensive part of the programme or take mice or embryos from other sources. The other major obstacle is that very few mouse facilities (such as Helmholtz) will take in mice from the outside but must re-derive mice from embryos (Harwell and WTSI).

The alternative to performing the highly technical and expensive process of microinjection or aggregation of mES cells to obtain the mice from another centre or contract with a commercial vendor. This strategy has been employed successfully by the German Mouse Clinic for the majority of its efforts. An alternative strategy is presented here and highlights the process and effort needed to produce cohorts of mice for phenotyping at the MLC Harwell.

2.3 Mouse Production

This necessity to building up a breeding colony to produce the KO mouse line for phenotyping is a large part of the cost of the Programme. This process would have to be repeated at multiple production centres if targets are tested in different systems or shared for confirmation of phenotypes. If large production facilities could take mES cells and make germline transmitting mice, then produce large numbers of vendor health status mice or embryos, or simply provide frozen sperm from germline mice, that could then be shipped to phenotyping centres, the cost savings could be very significant. If large numbers of embryos are produced and shipped it would solve the animal health importation issue, but the cost factor must be considered carefully. Centralized mouse production would also open up large scale specialty phenotyping to groups that have neither the resources nor the desire to do

large scale mouse production. The workflow below (Figure 4) provides a two option scheme for consideration in the IMPC phenotyping programmes. As taking mES cells through to the first cohort of mice is ~80% of the cost of the programmes, it is imperative that groups “be creative” and not continue business as usual but look for innovations in this area. Cost savings in this areas (and space requirements) could fund the expansion of the number of lines studied in Phase II of the programme, as well as fund the expansion of other pipelines for additional challenge models that are needed by the research community.

The IMPC should work to explore this and other mechanisms to increase the availability of mice and technologies to decrease the cost of production of mice from mES cells, but it will ultimately be the responsibility of each centre to obtain the funding needed to generate the mice and the phenotypic information for their portion of the programme.

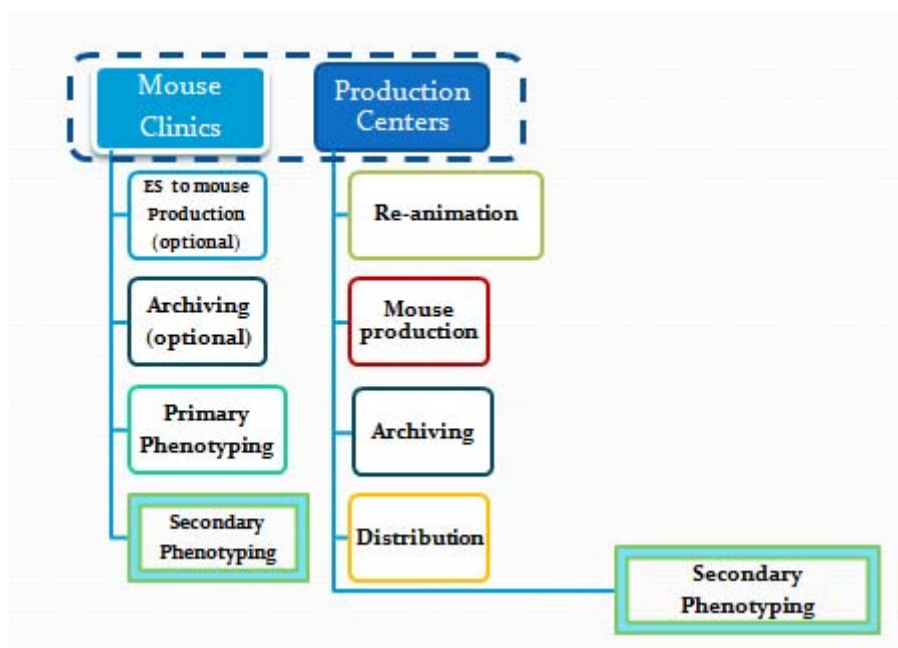


Figure 4. IMPC Mouse Production and Phenotyping Structures. While alternatives are possible, the figure depicts two models for mouse production and phenotyping; one in which all activities take place under one integrated facility and the other in a more distributive model with separate production and phenotyping centres.

2.4 Archiving

The IMPC centres that produce mice from mES cells by microinjection or aggregation will, either internally or in partnership, create or provide viable frozen sperm or embryos to IMPC approved repositories and distribution centres. Current IKMC repositories include:

- UC Davis
- Canadian Mouse Mutant Repository Toronto
- Helmholtz Munich

These existent repositories have proved to be highly efficient programmes by the provision of thousands of mES lines globally. Since 2007 the IKMC consortium has distributed >3,000 mES cells and reagents to international academic and industrial users.

It is recommended that IMPC mouse production centres, either allied with phenotyping centres or acting independently (Figure 4, section 3.3), would utilize established

infrastructure, operations, and distribution processes to act as formal distribution centres for IMPC materials, such as:

- MLC, Harwell
- European Mouse Mutant Archive
- The MMRCC
- The Jackson Laboratory
- UC Davis
- Canadian Mouse Mutant Repository Toronto
- Helmholtz Munich

All of the above have collectively and effectively archived and distributed thousands of mutant mouse lines over the past 10 years. The IMPC should require that each member that produces mice from mES cells submit a plan for archiving at one of the established repositories or one with demonstrably similar capabilities and agree to ensure that mouse lines and germplasm will be provided without any intellectual property or financial encumbrances. As previously discussed, secondary and tertiary phenotyping may be conducted at either phenotyping centres that are also performing the primary analysis or at other institutions that will rely on the Production/Distribution centres to deliver mice, embryos, or frozen sperm. The successful archiving and distribution of IMPC biological materials is a key component of the programme.

2.5 Housing and Animal Care

All mouse lines in IMPC phenotyping centres should be maintained at the highest possible health status and in appropriate high standard mouse caging and ventilation systems. It is recommended that the IMPC should adopt some of the ground breaking work of EUMODIC in developing procedures and terms for mouse welfare:

- A controlled vocabulary of mouse welfare (MW) terms (www.mousewelfareterms.org) - hierarchical structure of terms with standardised descriptors
- Agreed criteria for breeding threshold for moving to phenotyping heterozygous animals
 - 3+ homozygotes from 28 pups -> continue
 - 3+ non-healthy -> option to terminate
 - <3 -> phenotype only heterozygotes

The IMPC should also institute an IMPC Animal Welfare Committee that will report directly to the Steering Committee to ensure that the health and well being of animals within the IMPC production and phenotyping centres are in compliance with each countries animal welfare practices and the IMPC's own high standards. It is further recommended that the IMPC appoint an external, nationally or internationally recognized Chair of the IMPC Animal Welfare Committee. This Committee would continue to assist the IMPC to address each member country's concerns for animal welfare, for example the UK mandated RRR, of Reduce, Replace and Refinement. Indeed the entire IMPC programme by performing high-throughput analysis and placing the data in public databases will serve to dramatically decrease redundant work with mice, by decreasing the instances where multiple groups perform the same or similar work. The generation of mice from mES cells and making those animals available, will reduce greatly the number on animals that are used in the competitive and redundant efforts of multiple labs to create the same mouse lines from the same mES cells. Additionally, the work by EUMODIC and MGP

Sanger has already used statistical tests to reduce the size of the mouse cohorts used in phenotyping while maintaining scientific validity.

Part 3 – Organisational Requirements

3.1 Governance

It is proposed that IMPC is organized as an international confederation of mouse phenotyping programmes, comprising organizations, including international, national and regional funding agencies, production centres, scientific centres, and a central data centre that deliver the experimental goals of the programme and the acquisition and dissemination of data. The confederation should agree upon and establish terms of operation for the programme, including agreement on governance structures through a formal memorandum of understanding (MOU). This flexible but accountable structure (as opposed to setting up a separate legal entity) is favorable because it allows effective project co-ordination and delegated management, distributed funding from multiple organizations, and importantly has a proven track record for delivering the objectives of other consortia such as the IKMC, although it is anticipated that the IMPC will need to be more actively involved in coordination and Quality Control than the IKMC due to the added complexity of mouse production and phenotyping platforms. This will necessitate more structure and support from the participating members.

The current prevailing thoughts at this stage of the IMPC plan is that representatives from the funding agencies, production centres, and scientific centres will form a Steering Committee (SSC) to manage the programme and ensure coordination across the IMPC. The SSC will be responsible for delivery of the project and reporting progress to the various agencies involved. It will be chaired by a scientist of international stature agreeable to the funding agencies and the production groups, who will serve a defined term, to be decided upon by the IMPC Steering Committee. The Steering Committee from time to time and as appropriate for the development of the project will establish working groups to undertake detailed planning for the implementation of the project, or to address particular scientific, technological or organizational issues.

A working group dedicated to the IT programme is one of the most critical, and an informal team has already met on several occasions and was responsible for drafting the IT strategy incorporated in this document.

The steering Committee will be advised by an international scientific advisory board (SAB) appointed by the funding agencies involved in IMPC. The SAB will provide important oversight of the progress on deliverables and achievements of milestones, scientific and technological developments, and future project planning. The steering Committee will meet with the SAB at least once a year. It is also anticipated that the IMPC funding agencies will also meet at regular intervals to discuss and assess the progress of the programme.

It is proposed that the Steering Committee will be chaired by a scientist involved in one of the major phenotyping programmes. To reflect the international nature of the consortium, it is proposed that the chair position can rotate. The Chair and the Steering Committee should also be supported by a Secretariat, which will be responsible for assisting in delivering project management and ensuring coordination across the consortium, organizing working groups and other meetings, including outreach meetings and conferences based around

developments and achievements of IMPC. Members of the Secretariat will be defined as needed and formalized as the IMPC is formalized, but it is anticipated that the Secretariat will report to the Steering Committee.

3.2 IMPC Project Management Plans

Throughout this plan there have been recommendations that working groups be established for each critical aspect of the IMPC programme. Once completed this will be one of the most integrated, complex set of mammalian datasets ever produced, collected, stored and presented to the public. Some of these committees and other additional Working Groups have immediate action items that are needed to continue development of the IMPC Business Plan and to prepare for the full launch of the IMPC Programme (Figure 5).

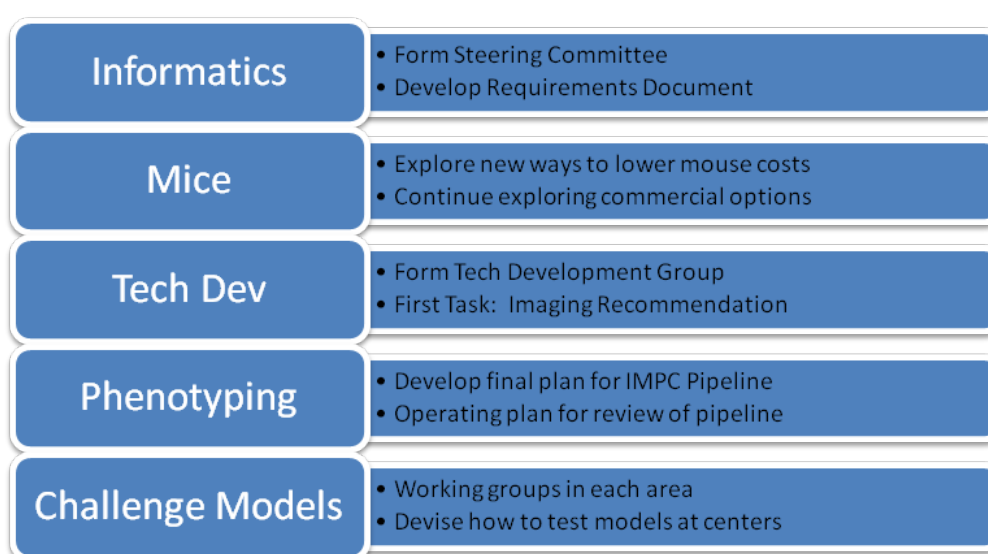


Figure 5. IMPC Working Groups. The groups indicated above are considered to be an immediate need of the IMPC and are in the process being formed.

Other key committees or working groups that are needed to manage and operate the IMPC include the following all of which will report to the SSC:

- QA and QC
- Finance
- Animal Assurance
- Outreach

Each of these committees or working groups should be formed as needed and will report to the Steering Committee. The composition of each committee should be decided by the SSC on an as need basis.

It is anticipated that the IMPC will grow to include members of technology groups, secondary phenotyping centres, potentially mouse production groups, and specialty phenotyping groups. In order to manage this organization it is recommended that a smaller Scientific Steering Committee remain in place that is composed of funding organizations, major phenotyping centres, and ad hoc members.

3.3 Monitoring and Evaluation of the Pipelines

The IMPC has the opportunity to help encourage funding, provide leadership, oversight, and assurance of data quality and standards, and manage coordination among members. The IMPC should be empowered to shape decisions and implement them, keep members on track for production, quality of data and numbers of mouse lines produced in order to meet its objectives, as well as make recommendations for investment and allocation of resource requirements that may be needed to complete the goals of the project.

It is recommended that the IMPC implement and maintain monthly teleconferences and require monthly updates to keep abreast of progress and track and deal with any issues. The complexity of the IMPC programme will require special monitoring and many processes depend on other steps that take months for completion. The total time from receipt of a mES cell line to completion of a phenotyped cohort is ~18 months and several steps, particularly microinjection or aggregation, germline transmission, and mouse breeding can cause serious downstream delays.

In addition to monthly tracking, the aforementioned Phenotyping Working Group will monitor the data in the IMPC with the assistance of special working groups assigned to QC & QA. Essential tools for quality control and quality assurance are critical to ensure the success and acceptance of the IMPC data and will require substantial support from the IMPC Data Centre. A programme of this scale and complexity will require data tracking, analysis, and management tools to be effective and these will be addressed further in the IT planning section.

In order to reach the scientific objectives of phenotyping a KO mouse line for every gene in the mouse genome, it will be necessary to take advantage of the considerable skills and resources that are currently available and wherever possible to expand these efforts. It will also be necessary to bring in new partners to achieve these ambitious goals. These are needed to obtain not only the total number of KO mouse lines studied, but enhance the type of data in the programme, such as aging which was mentioned in detail previously. New partners also have the potential to bring forth new perspectives and technologies that could further advance the IMPC programmes.

3.4 Interactions with Other Organizations

The current IMPC steering committee members and its representatives are actively engaged with groups from Japan (RIKEN), China (UK-China workshops) and continental Europe (the Infrafrontier Project <http://www.infrafrontier.eu/>) to encourage the flow of information and encourage broader participation. The Infrafrontier Project was established in 2008 to prepare Europe for an estimated increase of at least 4-fold in demand for mouse space, housing and experimentation over the next 10 years. Based on the successful European programmes EUMODIC (systemic phenotyping) and EMMA service (archiving and distribution), it brings together the leading European scientific institutions in this field as well as national ministries and funding bodies. Its main purpose is to initiate all necessary steps for the establishment of a sustainably funded pan-European research infrastructure for systemic phenotyping, archiving and distribution of mouse models of human disease.

Recent developments within Europe indicate that it now appears that the Infrafrontier Project may be the best way for most groups, including the German Mouse Clinic and the Institut

Clinique de la Souris (Strasbourg), to raise the substantial part of the funding needed for phenotyping for the IMPC project. Infrafrontier has offered to act as the lead in coordinating efforts across continental Europe to encourage more phenotyping, but the indications are that these efforts will be smaller than the recommended 100 mouse lines per year for IMPC. As such, Dr. Martin Hrabe de Angelis, the leader of Infrafrontier has proposed that Infrafrontier act to run and coordinate the potential IMPC projects with the smaller research groups, such as UAB (Barcelona), Italy (Monterotondo) and the Czech Republic, to reduce the burden on the IMPC infrastructure and to open the IMPC to more groups.

3.5 Funding Models

As discussed above, it is proposed that the IMPC acts as a confederation of funders and centres engaged in supporting and delivering the phenotyping of mutant mouse lines. In addition, there needs to be a substantive informatics effort to support the acquisition, analysis, and dissemination of phenotype data. The delivery of the IMPC programme also needs to be supported by the activities of a Secretariat, which will assist the Steering Committee in delivering the goals of the programme. Each of these three activities will require consideration of the appropriate funding model as described below.

1. Distributed Funding of the core activity of mouse production and phenotyping.

The experimental activities of IMPC (mouse generation, archiving, distribution, breeding, and phenotyping) will be supported at each of the production and phenotyping centres by funding from the relevant international, national or regional agency or entity. These core activities will depend upon a one-to-one relationship between the centre and the relevant national funding agency (e.g. Harwell and the MRC).

2. Collective Funding of Informatics, including the Data Centre and the associated Informatics Resource Development Groups (IRDGs). While funding for mouse production centres is likely to be granted by individual funding agencies, there is a need to ensure clearly coordinated funding for the centralised informatics activity.

One option for funding of the informatics activity could be that each funding agency makes a subscription to the overall costs of a centralized Data Centre independent from production and phenotyping centres and the accompanying Informatics Resource Development Groups (IRDGs) located at the production and phenotyping centres. The informatics budget will be held and distributed by the organization hosting the Data Centre, and budget dispensations will be overseen by the Steering Committee.

3. Collective Funding of the Secretariat. It is recommended that the Secretariat that will assist the Steering Committee in developing and managing the project also be supported by subscriptions from funding members of the SSC. It is proposed in this case the budget will be held by one of the partner funding agencies; this might be the agency that hosts the Secretariat.

3.6 Budget Requirements

The current projections for the IMPC estimate phenotyping costs (including mouse production) at ~£28,000 per KO mouse line (Full Economic Costing). Although different cost

models may be applied by other participating centres, depending on an institute's financial requirements, £28,000/line is a realistic estimate based on extensive analysis by the Mary Lyon Centre over the last 12 months. These numbers will likely increase slightly to £30,000 as new phenotyping tests are added as discussed in previous sections. The cost per 100 mouse lines is therefore £3.0M per year, but do not include IT support. The costs represent an enormous saving over the current practice of targeted, hypothesis-driven support of projects focusing on individual mutants, irrespective of the enormous benefits of data integration and standardisation. A formal budget for the project will only be possible when all centres, their cost structures and funding are known.

The IMPC received a significant endorsement and push forward by a \$110 million (US) initiative from the NIH (US) to commence Phase I of the KOMP² program whose goal is to produce and phenotype 2,500 KO mice from the IKMC KO ES line collection over the next 5 years in coordination with the IMPC. The IMPC is continuing to engage and recruit new members to help meet the ultimate Phase I goal of 4,000 KO lines phenotyped by 2016, in pursuit of the ultimate goal of ascribing function to every gene in the mouse genome by 2021.