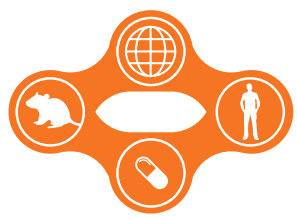


IMPC

International Mouse Phenotyping Consortium

The Function of Human Genetic Variation

IMPC Strategy 2021-2030

IMPC Drafting Group

Steve Brown | Director, MRC Harwell (Chair)
Robert E. Braun | Senior Scientific Advisor to the President and Professor, JAX
Mary E. Dickinson | Professor, Baylor College of Medicine
Colin Fletcher | Programme Director, NIH
Yann Herault | Director, Institut Clinique Souris, Strasbourg
Martin Hrabe de Angelis | Director, Institute for Experimental genetics, Helmholtz Munchen
Kent Lloyd | Director, Mouse Biology Program, UC Davis
Ann-Marie Mallon | Head of BioComputing, MRC Harwell
Colin McKerlie | Director, The Centre for Phenomics, Toronto
Terry Meehan | Project Leader, Mouse Informatics, EBI
Steve Murray | Associate Professor and Director KOMP Model Development, JAX
Lauryl Nutter | Associate Director, The Centre for Phenomics, Toronto
John Seavitt | Assistant Professor, Baylor College of Medicine
Damian Smedley | Director of Genomic Interpretation, Genomics England
Sara Wells | Director, MRC Harwell

Design by Gareth Clarke.

Contents

Executive Summary

The International Mouse Phenotyping Consortium's (IMPC) 10-year strategy presents a bold ambition for the future functional annotation of the genome and deciphering genetic variation in the context of gene function and disease. 2020 will mark the 10th anniversary of our consortium, and as we look forward to the next 10 years we present a strategy that encompasses completion of Phase 2, and execution of Phases 3 and 4. The key goal of IMPC for 2021–2030 is to deliver broad-based genomic insights into the function of human genetic variation in both coding and non-coding sequences that underpins future developments in precision medicine[1], rare disease, clinical genetics, and healthy living. This strategy document includes:

A statement on the underpinning rationale for how mouse genetics and mouse models inform gene function, human disease, and clinical genetics. In particular, we provide a description of the challenges and complexities of interpreting human genetic variation in the context of gene dysfunction (disease), and the critical role of the mouse; 1) as an informative and comparable mammalian model system for the analysis of human genetic variation, and 2) to provide robust models for hypothesis-driven discovery research, mechanistic insight, purpose-driven translational studies, preclinical safety and effect programmes, and molecular target testing platforms.

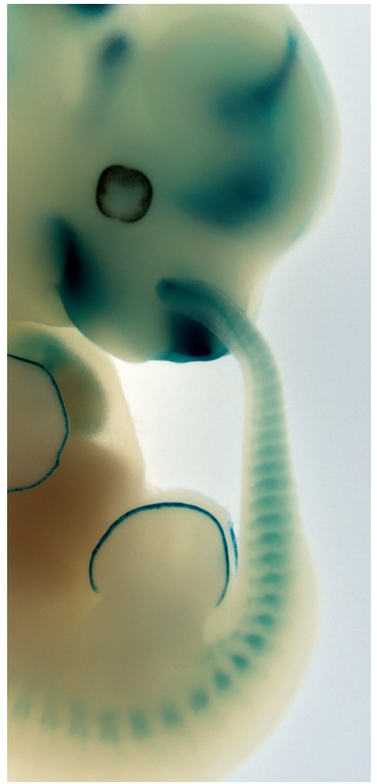
A statement and details of the IMPC's goals for the next ten years:

Goal 1

Complete the generation of a null mouse mutant resource for the coding genome that delivers a comprehensive catalogue of mutant strains available to study mammalian gene function.

Goal 2

Design and produce a genome-wide mouse strain resource of human disease-associated coding variants associated with rare disease that can be used for validation of putative functional variants and insight into disease mechanism(s).



Goal 3

Design and generate mouse strains that model genetic variation in the non-coding genome for the IMPC and the global research community to use in a collective effort to assess functionality and mechanism(s) in health and disease.

Goal 4

Phenotype mouse strains with null mutations, human-disease coding variants, and variants in the non-coding genome to provide a comprehensive catalogue of baseline information on mammalian gene function, validate putative functional variants, examine the effects of mutations in the non-coding genome, and provide data-driven insight into disease mechanism(s) and other phenotypic outcomes.

Goal 5

Explore genetic context to realise the wider potential for mouse functional studies to provide models and mechanistic insights for therapeutic development, both conventional therapeutics and for precision medicine.

Goal 6

Develop data integration, analysis, and visualisation approaches to translate mouse functional genomics studies to the human gene and disease knowledge base and vice versa, underpinning synergies in the fundamental understanding of genetic and disease systems – enabled and strengthened by dynamic interactions and networks with human genome centres, clinical genetics consortia, and data-curated biobanks that we serve.

"The key goal is to deliver broad-based genomic insights into the function of human genetic variation"

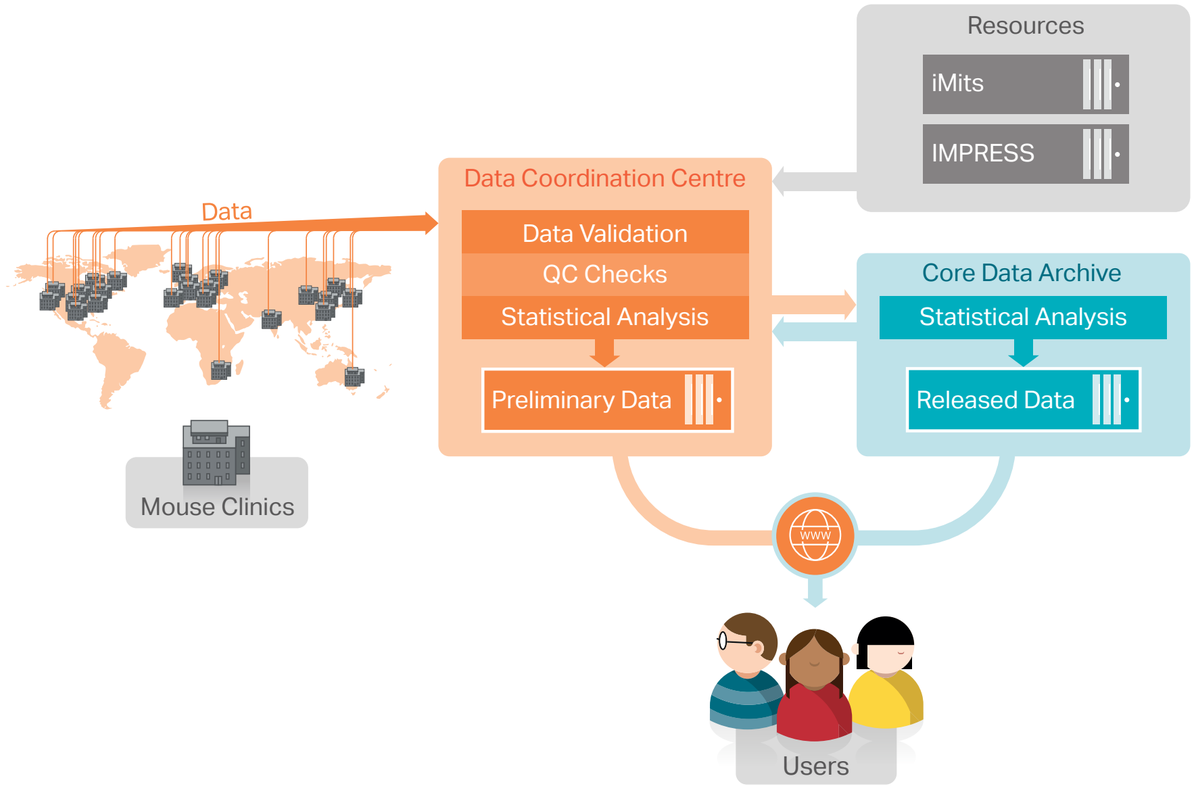


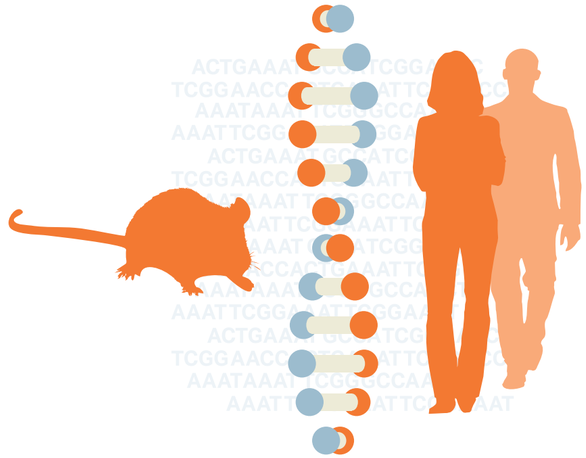
Fig 1. Flow of data through the IMPC.

iMits *International Micro-Injection Tracking System.*
IMPRESS *International Mouse Phenotyping Resource of Standardised Screens*

The Mission

The mission of the IMPC is to generate a comprehensive catalogue of mammalian function and provide the foundations for the functional analysis of human genetic variation.

We aim not only to develop insight into the function of every gene by the creation and analysis of null mutations, but also to further explore the relationship between genome and phenome by the generation of putative human pathogenic variants in both coding and non-coding sequences. These models will enable us to confirm pathogenicity, understand disease mechanisms, explore the impact of genomic context on the expressivity of disease phenotypes, and work with like-minded partners to design and undertake preclinical studies. Our aim is to provide transformative insights into the genetic bases of disease that will impact on clinical diagnosis and management, the exploitation of mouse models for therapeutic development, and ultimately support the funders' goals to prevent, detect, diagnose, and treat disease.



"Our aim is to provide transformative insights into the genetic bases of disease"

The Challenge

Defining the functional role of human genetic variation

There are around 10^{12} variants across the human population and a large number of these will be associated with disease susceptibility[2]. A very large worldwide programme encompassing human and clinical genetics, GWAS, Biobank, and sequencing studies has explored the relationship between disease and genetic variation in human populations and identified many putative loci and variants awaiting analysis to differentiate disease-causing from disease-associated variants.

Coding sequences account for less than 2% of the genome while non-coding elements responsible for genetic regulation encompass a very much larger fraction of mammalian genomes. Thus, it is not surprising that while some of the identified variants lie within coding sequences, the majority of putative variants lie within the non-coding genome in elements that are responsible for gene expression and functional regulation[3].

The challenge is to rapidly and definitively establish causality of both coding and non-coding variants and determine the functional role of these variants in normal and disease-related development and physiology.

There are profound difficulties in determining the functional role of human genetic variation, which are compounded by a lack of progress and a dearth of knowledge in several critical areas:

- *Much of the mammalian coding genome has very limited information on biological function in humans, mice, or other species – that part of the genome which is referred to as the 'dark' genome*[4].
- *The functional consequences of human genetic variation, and its impact on both common and Mendelian disorders, are poorly understood.*
- *The functional analysis of the various elements of the non-coding genome has been limited, and will require targeted mutagenesis of individual elements along with analysis of the impact of cis- and trans- effects on gene regulation.*
- *The number of variants of unknown significance is increasing dramatically and is not being met by a commensurate analysis of variant function.*

In summary, it is currently very difficult to fully establish the pathogenic and/or causal nature of human genetic variation using patient data alone. The gap between our knowledge of genome variation and our ability to survey the genome-phenome landscape is increasing exponentially.

Future developments in biomedical sciences, particularly precision medicine, depends upon a comprehensive understanding of mammalian genome function and the impact of human genetic variation on disease causation, susceptibility and risk. Progress in biomedical research, translational studies, clinical research, and ultimately therapeutic development and healthcare delivery are all at risk if we fail to tackle this challenge.

Fig 2. Gene Serpinb12, Oral Epithlium.

"There are profound difficulties in determining the functional role of human genetic variation"

Mouse Genetics and the IMPC

Meeting the challenge

The mouse is an essential model system to dissect out and understand the importance of genetic variation. In many instances, the analysis of genetic variants in the mouse from null to targeted alleles has provided profound insights for the biomedical and clinical sciences through hypothesis-driven discovery research, mechanistic insight, purpose-driven translational studies, and preclinical safety and effect programmes.[5,6,7,8]

Importantly, the design and generation of a variety of mouse mutant alleles will have a significant impact across all disease classes from rare Mendelian disorders to common disease.

Indeed, it is clear from studies in the human population that there are thousands of associations between simple and complex disease that is manifest in a phenotype code linking Mendelian loci to complex disorders[9]. As such, human variants associated with common diseases are enriched in genes that encapsulate this Mendelian code, underlining the utility of exploring a range of mouse alleles.

The IMPC has established a global network of mouse genetic centres with the expertise and facilities to undertake genome-wide genome-phenome studies on a large-scale and, as discussed below, the consortium has already made remarkable progress in establishing a comprehensive catalogue of mammalian gene function.

Manipulation of the mouse genome is increasingly facile and is being applied by the IMPC to study the basic function of mammalian genes and investigate the 'dark' genome. Design, production, and phenotyping programmes are underpinned by standardised procedures and appropriate statistical power delivering robust, reproducible findings.

In addition, mouse genetics can avail itself of extraordinarily diverse and well characterised genetic resources. Importantly, the vast majority of genes in the mouse and human genomes are orthologous, and much of the critical variation in the human genome lies in sequences that are conserved between mouse and human.

Consequently, the IMPC is in a position of strength to build on its current success, and to deploy its 2021–2030 strategy to meet the challenge of understanding the functional role of human genetic variation.

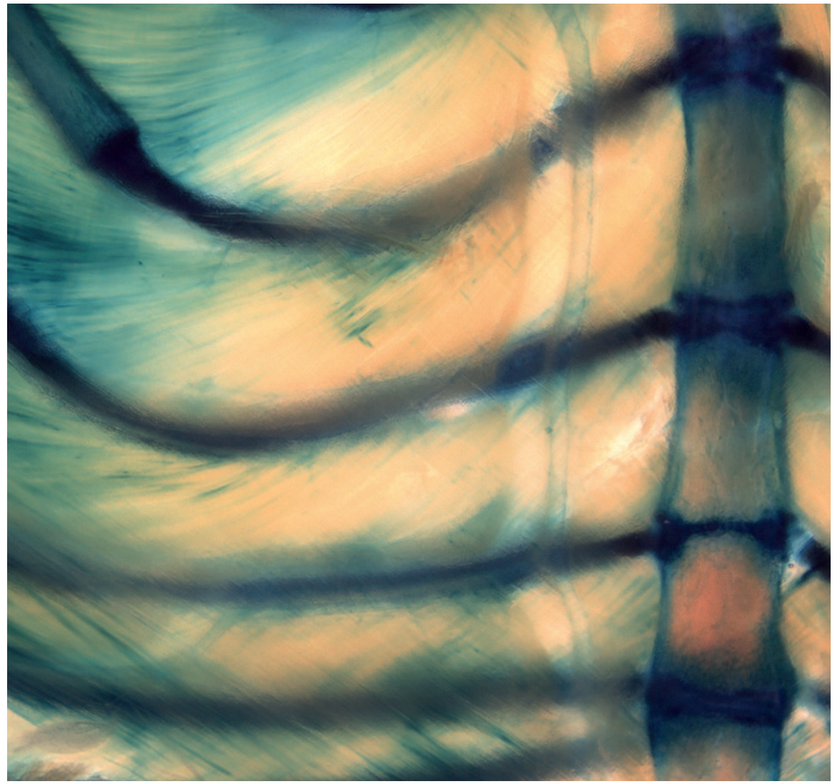


Fig 3. Gene Gfpt1, Bone, cartilage, and skeletal muscle.

"Mouse mutant alleles will have a significant impact across all disease classes"

“The mouse is an essential model system”

Achievements to Date

The IMPC has created a distributed infrastructure for the analysis of mouse gene function, allying all the major mouse genetics centres worldwide to leverage an extraordinary breadth of expertise in mouse genetics from genome editing and phenotyping to data analysis. The consortium began work in 2011 with its first and most immediate goal to generate a null mutant and undertake broad based phenotyping for every gene in the mouse genome. As of 2018, we have completed 8,000 genes – more than a third of the mouse coding genome.

The IMPC for the first time has created a highly robust, quality controlled, comprehensive and open dataset of genome-phenome information that surveys a significant proportion of the genome landscape.

8173 genes

409k images

6508 lines phenotyped

74 million data points

Data Release (9.2) for 6006 lines

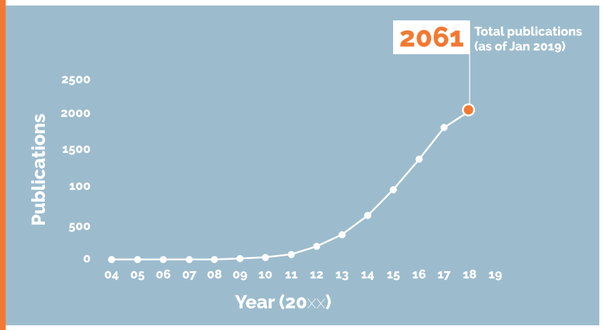


Fig 4. Yearly IKMC/IMPC related publications.

An IMPC Case Study:

Microglial Ramification, Surveillance, and Interleukin-1 Release Are Regulated by the Two-Pore Domain K+ Channel THIK-1

Neuron, vol. 97. Issue 2. January 2018

Microglia exhibit two modes of motility: they constantly extend and retract their processes to survey the brain, but they also send out targeted processes to envelop sites of tissue damage. Using IMPC Kcnk13 null mice, Madry et al demonstrate how mutant mice showing decreased tissue surveillance but normal motility response to tissue damage. This identifies Kcnk13 as an important target to modulate microglial activities with implications in protecting from neurological diseases such as Alzhemiers and autism.

The impact of this work is reflected in a number of seminal and high-profile publications that have appeared over the last 18 months describing work that is transforming biomedical sciences:

High-throughput discovery of novel developmental phenotypes
Nature, 2016
Analysis of 410 embryo lethal genes from the first 1,750 mutant strains, revealing extraordinary variable expressivity and large numbers of subviable strains, along with a high degree of enrichment for human disease genes in the mouse essential gene dataset.

Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium
Nature Genetics, 2017
Analysis of the first 3,328 null mutants analysed for abnormal phenotype. 90% of gene-phenotype annotations had not been reported before, and hitherto there had not been a mouse mutant available for 1,830 out of the 3,328 genes analysed. Moreover, for 1,092 genes the first functional knowledge was provided.

A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction
Nature Communications, 2017
Analysis of 3,006 IMPC mouse knockouts identified 67 hearing loss genes, of which the vast majority, 52, were novel and had not been hitherto associated with deafness in any species. The analysis uncovers a large and unexplored genetic landscape involved with auditory function.

Prevalence of sexual dimorphism in mammalian phenotypic traits
Nature Communications, 2017
The IMPC was able to undertake a seminal and comprehensive, genome-wide view of the effects of sexual dimorphism, particularly the interaction between genotype effects and sex. Extraordinarily, we found that for continuous (numeric) parameters around one-sixth of phenotypes in mutants were affected by sex. Sexual dimorphism is pervasive across the mammalian genome.

Identification of novel genetic elements in metabolism by high-throughput mouse phenotyping
Nature Communications, 2018
From an analysis of more than 2,000 genes, the IMPC have identified novel candidate metabolic genes associated with newly uncovered regulatory gene networks and metabolic traits in human GWAS. Of 974 genes with strong metabolic phenotypes, 429 were novel metabolic genes of which 51 had not been functionally unannotated previously.

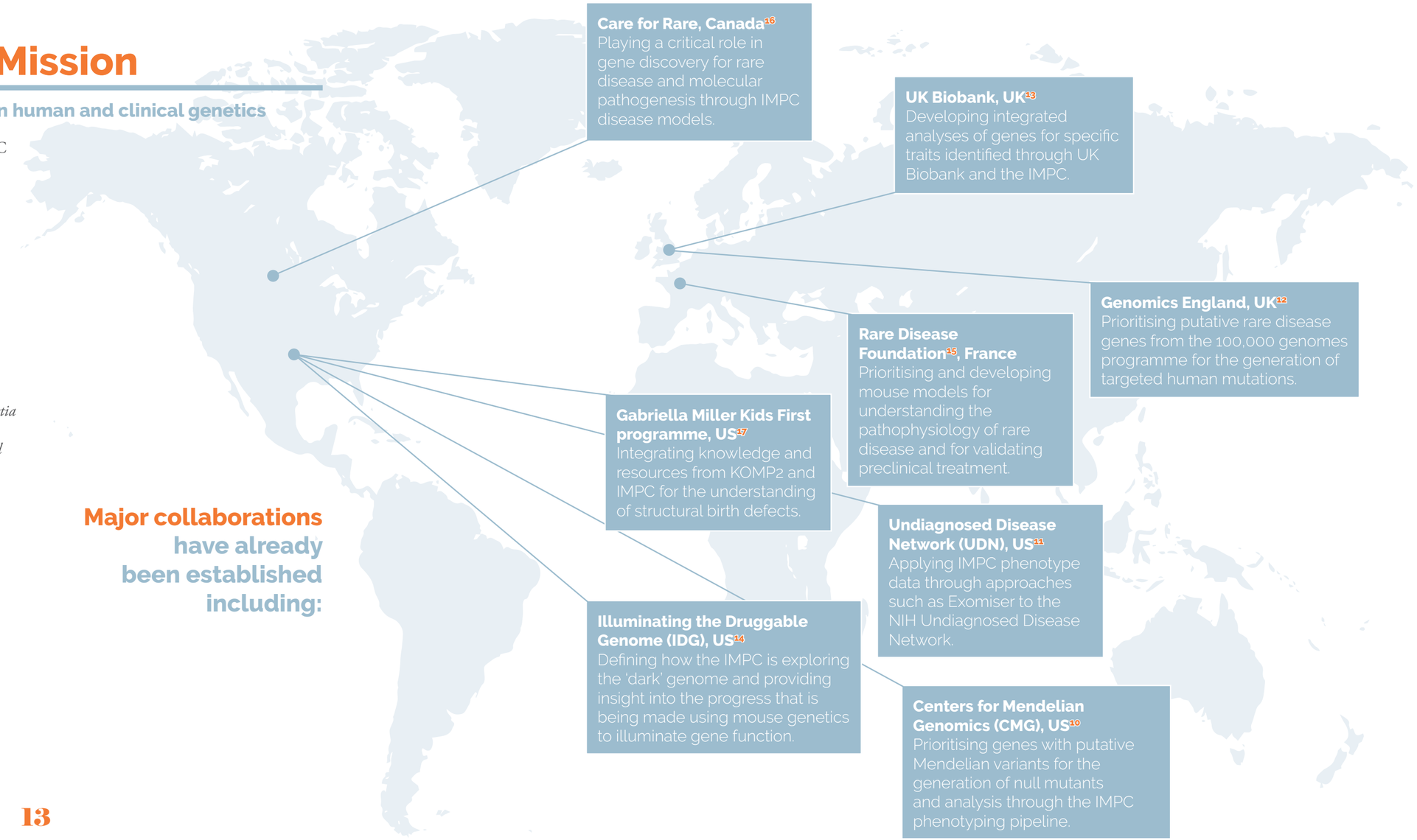
Delivering the Mission

Partnership with consortia in human and clinical genetics

To deliver our mission, the IMPC has begun to establish strategic and active partnerships with human genetics initiatives and consortia across the globe that aim to:

- *Inform the IMPC's strategy so that it best meets the needs of human genetics in defining and validating the functional role of human genetic variation*
- *Prioritise the generation of relevant mutants and the delivery of genome-phenome information to vested consortia*
- *Develop and integrate links with, and cross-species analysis of human and mouse multi-dimensional datasets.*

Major collaborations have already been established including:



Care for Rare, Canada[16]
Playing a critical role in gene discovery for rare disease and molecular pathogenesis through IMPC disease models.

UK Biobank, UK[13]
Developing integrated analyses of genes for specific traits identified through UK Biobank and the IMPC.

Genomics England, UK[12]
Prioritising putative rare disease genes from the 100,000 genomes programme for the generation of targeted human mutations.

Rare Disease Foundation[15]**, France**
Prioritising and developing mouse models for understanding the pathophysiology of rare disease and for validating preclinical treatment.

Gabriella Miller Kids First programme, US[17]
Integrating knowledge and resources from KOMP2 and IMPC for the understanding of structural birth defects.

Undiagnosed Disease Network (UDN), US[11]
Applying IMPC phenotype data through approaches such as Exomiser to the NIH Undiagnosed Disease Network.

Illuminating the Druggable Genome (IDG), US[14]
Defining how the IMPC is exploring the 'dark' genome and providing insight into the progress that is being made using mouse genetics to illuminate gene function.

Centers for Mendelian Genomics (CMG), US[10]
Prioritising genes with putative Mendelian variants for the generation of null mutants and analysis through the IMPC phenotyping pipeline.

The IMPC

Future strategy and the functional role of human genetic variation

The IMPC has established a formidable network and operation for the large-scale analysis of mammalian gene function and the dissection of the functional role of human genetic variation. It is timely now to assess the wider opportunities to bring our expertise and infrastructure to the challenges facing biomedical sciences, in particular the opportunity to provide essential underpinnings to precision medicine and the need to interpret human genetic variation and its impact on disease.

2020 will mark the 10th anniversary of the IMPC, and as we look forward to the next 10 years we present a strategy that encompasses completion of Phase 2, and execution of Phases 3 and 4.

Over the next few pages we elaborate our future vision and strategy for 2021 to 2030 that sets out bold and far-reaching goals that will provide a key foundation for human genetics and precision medicine initiatives worldwide.

Goal One

Complete a null mouse mutant resource for the coding genome that delivers a comprehensive catalogue of mutant strains available to study mammalian gene function.

Since 2011, the IMPC has been generating a gold-standard resource of null allele mutations for every gene in the mouse genome. Already, over a third of the mouse genome has been covered and the mutant alleles are distributed widely on a continuous basis. The null alleles provide basic insights into the function of each gene, importantly providing a resource for the interpretation of Mendelian and loss-of-function disorders. As discussed above, a fundamental understanding of gene function through the analysis of null alleles also provides underpinning information for the interpretation of genetic variation in the human genome that impacts on common diseases. The IMPC will extend the range of mouse alleles produced to enable a more precise modelling and interpretation of human genetic variation by the production and analysis of firstly, targeted human mutations in coding sequences, and, secondly, the generation and analysis of variants in conserved enhancer elements.

The IMPC will complete the production of a comprehensive null mutant resource for the mouse genome during Phase 3 (2021–2025). Completing the null resource will enable us to define the complete catalogue of essential genes in the mammalian genome as well as continue to elaborate and expand our view of the genome landscape for multiple disease areas. In addition to the >1,000 essential genes we have identified to date, production has and will continue to identify genes that are under extreme constraint and whose absence compromises fertility or viability. This production data provides an entry point for further functional characterisation.

By the end of Phase 2 in 2021, we expect to have generated at least 10,000 null mutations – 5,500 in the US, 3,500 from the EU, 500 across Asia, and 150 from Canada.

During Phase 3, the IMPC will continue its effort to complete the genome but we recognise that we need to assess carefully the finish line and define precisely our goals for the remaining effort. We will focus on the human orthologous genome, that is ~18,000 protein-coding genes.

Goal One The IMPC will generate 8,000 new null alleles to complete the null resource. All mouse strains will continue to be made available and accessible through open access repositories to the global biomedical research community.

Goal Two

Design and produce a genome-wide mouse strain resource of human disease-associated coding variants associated with rare disease that can be used for validation of putative functional variants and insight into disease mechanism(s).

The IMPC will initiate a global programme to generate a genome-wide resource of human disease-associated coding variants. Our aim is two-fold – to generate disease models for mechanistic and preclinical studies, but also to further enhance the discovery of the phenotype landscape of the targeted gene important in diagnosis, clinical management, and ultimately therapeutic development. The availability of a coding variant allele provides long-term value in providing insights into areas such as phenotype expansion and variable clinical presentation. Together, the combination of targeted human-disease associated coding variants and null alleles is a powerful resource for fundamental understanding of disease mechanisms and the necessary models to develop new therapeutic approaches.

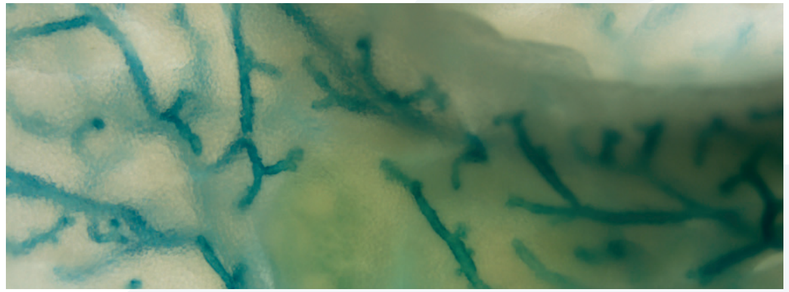


Fig 5. Gene Nxn, Mammary gland.

The IMPC, in delivering a genome-wide mouse strain resource of human disease-associated coding variants, will:

- *Establish partnership agreements with research consortia, clinical research consortia, clinical geneticists, rare disease foundations, and individuals which aim to adhere to the IMPC principles of openness and access to resources and data; and*
- *Engage these partners for curation and prioritisation of variants to deliver functional information about gene and variant function in disease and development.*

The IMPC has already successfully developed this model – for example, with Genomics England – whereby the IMPC centre at the Harwell Institute has established a pipeline, GEMM, for the generation of human disease-associated coding variants for the rare diseases under study in the UK 100,000 genomes project. Similar partnerships with other groups committed to clinical genomics, such as the UDN, CMG, Kid's First have also been established and allele variant nominations are being solicited.

Goal Two By 2025, the IMPC aims to be generating ~1,000 human coding disease-variant strains per year.

Goal Three

Design and generate mouse strains that model genetic variation in the non-coding genome for the IMPC and the global research community to use in a collective effort to assess functionality and mechanism(s).

The IMPC has considered carefully how it can bring its resources in mutagenesis and phenotyping to the analysis of human functional variation in the non-coding genome. It is recognised that functional variants that contribute to common, non-Mendelian diseases mainly lie in the non-coding genome. These non-coding pathogenic variants largely impact regulation of genes or gene networks and are highly enriched in conserved non-coding elements (CNEs).

The IMPC plans to devote a significant proportion of its pipeline capacity to the deletion of CNEs to examine their impact on biology and disease in the mammalian genome. Our strategy to ensure prioritisation and deliver immediate value is:

- *Engage and partner with diverse human genetics consortia, e.g. DDD, to prioritise relevant CNEs associated with high priority loci in the relevant disease area;*
- *Introduce targeted deletion mutations of relevant CNEs for phenotyping and transcriptomics. This leverages our experience with generating both deletion and precise disease-associated mutations in CNEs;*[18] *and*
- *Design and produce precise disease-associated variants in enhancers after validation of a CNE by targeted deletion to assist in further detailed dissection of enhancer function in relation to transcriptional networks and phenotype outcomes.*

Overall, the IMPC will engage in a large-scale assessment of critical elements of the non-coding genome that bear upon a key aspect of functional variation and its impact on disease. The consortium will generate an important resource for human genetics and clinical consortia that will enable the identification and validation of loci involved with pathways associated with the relevant disease area.

Goal Three By 2025, the IMPC aims to be generating around 500 targeted deletions of CNEs per year.

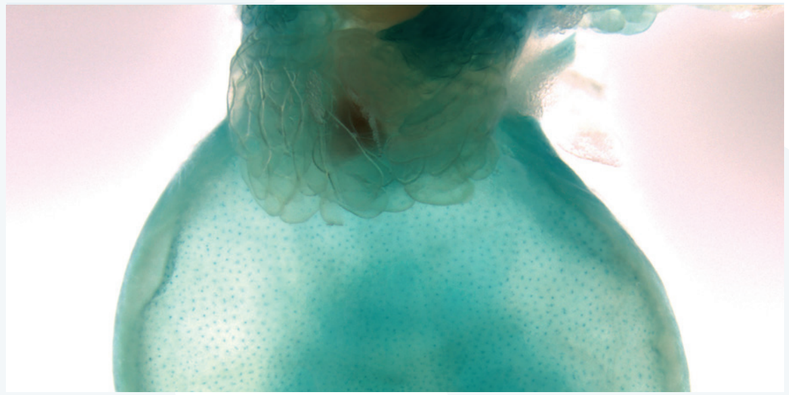


Fig 6. Gene Leprotl1, Bladder and prostate gland.

Goal Four

Phenotype mouse strains with null mutations, human-disease coding variants, and variants in the non-coding genome to provide a comprehensive catalogue of baseline information on mammalian gene function, validate putative functional variants, examine the effects of mutations in the non-coding genome, and provide data-driven insight into disease mechanism(s) and other phenotypic outcomes.

The IMPC will need to be a driver of development of large-scale phenotyping technology and be an active collaborator and early adopter of new or refined approaches being developed by the biomedical sciences and clinical communities. It is essential that the IMPC continues to deliver high quality and robust phenotyping data that adds significant value to research aimed at discovery and validation of functional human genetic variation, and disease mechanisms, and provides a high-value resource of well-characterised models ready for reproducible and predictive development of new prevention, diagnostic and treatment strategies.

Since its inception, the IMPC has been at the forefront of developments in mouse phenotyping combining innovation, globally-distributed validation, and standardisation to deliver comprehensive yet robust and reproducible phenotyping tests in many disease areas. The consortium has a track record of success in revealing novel features of the phenome landscape, including the wide extent of pleiotropy in the mammalian genome, sexual dimorphism, and variable expressivity of embryo phenotypes. The IMPC has set ambitious ongoing targets for future phenotyping developments. Our 2021–2030 strategy includes cutting-edge standardised pipelines across the life course with a special emphasis on humanised phenotyping that is applied to the wide range of mutants generated.

There are four major areas where the IMPC will continue to be a leader and make major advances in phenome studies.

1 Automate, update, and develop

Our experience, combined with recent advances in technology, present unparalleled opportunities to envision and deploy next generation phenotyping pipelines that have as their hallmarks cutting edge experimental platforms, a high degree of automation and machine data capture, and the ability to acquire an even broader range of phenotypes relevant to human clinical diagnostic observations. New test areas in renal function and visual acuity phenotyping are two of many examples. Moreover, the IMPC will continue to develop an increasing focus on molecular phenotyping, including metabolomics and the microbiome, where pilot studies are underway. Automation will be core to future developments, enabling 24-hour monitoring within the home-cage environment and providing a step change in our ability to observe and dissect phenotyping information.

2 Changing protocols and husbandry to humanise testing

We will continue to focus on the correlation of mouse phenotype tests and their outputs with key phenotypes and indicators of human clinical disease. Our continual evolution and selection of protocols for the core IMPC pipeline is increasingly driven by the need to ensure that we provide clinically relevant phenotype information for the validation and analysis of human genetic variation. Critical to this strategic objective, we will adapt and refine our husbandry and phenotyping protocols to humanise testing – performing our work in a manner designed to maximise the relevance of biological observations in mice to human pathophysiology – and includes diverse approaches from red light phenotyping to metabolic assessment under conditions of thermoneutrality.

3 Cross-centre validation and standardisation – setting standards for data reproducibility

Reaffirming our commitment of resources to achieve the highest standards of data reproducibility is an essential component of this strategy document. The demonstrated value of cross-centre verification and validation of experimental results from standardised protocols in ensuring reproducibility is the cornerstone on which the consortium was founded, and determines a key mandate for the IMPC for the next 10 years. New phenotype tests must continue to be assessed for reproducibility across the IMPC before widespread introduction.

4 Develop new tools and resources for mouse to human phenotype comparisons – phenotype to molecular mechanism matching

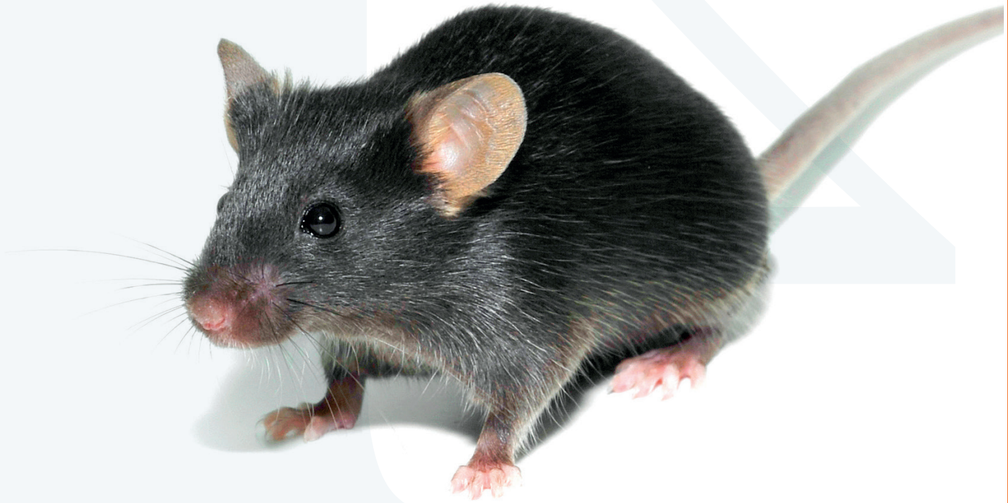
With our colleagues from MPI2 the IMPC will develop enhanced informatics approaches to deliver better interrogation of mouse to human phenotype comparisons, incorporating data from phenotype to molecular mechanism. This will be increasingly valuable as we expand the range of molecular phenotyping in the IMPC to enable improved correlation between human disease states and mouse phenotype data.

Tuning mouse phenotyping to the needs of human variant analysis

We recognise the value of broad based phenotyping and the insight it brings to our understanding of pleiotropy, the multiple functions of genes and their role in disease and pathophysiology, and areas in human genetics such as phenotype expansion. However, as we complete the null resource and increasingly turn our attention to the functional validation and analysis of human variants, we will focus our capacity on a tiered strategy for phenotyping that will allow us to maximise impact on human and clinical genetics. We will develop this strategy inclusively with our consortium partners to take advantage of their expertise in multidisciplinary clinical genetics and diagnosis.

In summary, we will:

- *Work closely with our clinical partners to identify tests for each relevant disease area in which variants are being tested;*
- *Phenotype mutant strains using a standardised set of essential tests that relate to the disease phenotype observed in the human population, providing validation of the variant and initial insights into disease mechanism; and*
- *Analyse variants whenever capacity is available using the expanded comprehensive IMPC phenotyping pipeline to identify the full range of pleiotropic effects.*



Goal Five

Explore genetic context to realise the wider potential for mouse functional studies to provide models and mechanistic insights for therapeutic development, both conventional therapeutics and for precision medicine.

The provision of a genome-wide gold-standard mutation resource on a standard inbred background is a critical foundation for understanding gene function as well as exploiting the resource for mechanistic studies, including the production and analysis of oligogenic mutants. Nevertheless, **the IMPC recognises the importance of identifying and describing pleiotropic phenotype features which may only be elicited on other genetic backgrounds.** In order to understand the impact of genetic context, the IMPC will:

- *Work closely with our clinical partners to identify tests for each relevant disease area in which variants are being tested;*
- *Phenotype mutant strains using a standardised set of essential tests that relate to the disease phenotype observed in the human population, providing validation of the variant and initial insights into disease mechanism; and*
- *Analyse variants whenever capacity is available using the expanded comprehensive IMPC phenotyping pipeline to identify the full range of pleiotropic effects.*

The IMPC will provide the expertise and infrastructure to investigate how genetic context impacts upon the phenotype and disease outcomes that will be key to further elaborating the function of human genetic variation.

Goal Six

Develop data integration, analysis, and visualisation approaches that will be required to translate mouse functional genomics studies to the human gene and disease knowledge base and vice versa – enabled and strengthened by dynamic interactions and networks with human genome centres, clinical genetics consortia, and data-curated biobanks that we serve.

A critical goal of the IMPC for 2021–2030 will be the development of informatics tools and methodologies for improved data analysis and integration that will be required to translate mouse functional genomics data to the human gene and disease knowledge base and vice versa, to enable synergies in the fundamental understanding of genetic and disease systems. **The IMPC aims to continue to be a leader in mouse genetics data science, with an increasing emphasis on interactions and integration with human genome data and clinical genetics.** There are many opportunities to engage the IMPC experience and know-how in data structures, data wrangling, analytics, and methods development in both human and other species.

The IMPC has several core aims for developments in data science and our interface with the wider biomedical sciences and clinical communities that will impact upon our understanding of the function of human genetic variation:

1. Develop multivariate analysis of null mutant multi-dimensional datasets along with comparative analysis to human genetics data

We aim to develop methods for identifying new functional knowledge from IMPC datasets, with a particular emphasis on:

- *Improving multivariate analytical tools to identify robust phenotypes and latent factors;*
- *Applying machine learning to provide further insights into genetics and disease systems, including novel members of pathways and novel phenotype associations; and*
- *Establishing comparative projects with human genetic datasets to facilitate gene-variant functional discovery (e.g. UK Biobank), developing tools for the comparison of variants associated with human traits with the functional information from the IMPC to identify and validate new function.*

2. Improve comparative ontologies and assessment and computation of cross-species phenotype comparisons, particularly human and mouse

In parallel, there is an important need to improve approaches to comparative ontologies and our ability to relate disease phenotypes to phenotypes identified in the mouse or other species. To facilitate this, we will:

- *Undertake a systematic comparison of all rare disease generic HPO phenotype annotations from OMIM and Orphanet diseases with known genes;*
- *Undertake a systematic comparison between extant patient clinical data and diagnoses from programmes such as the CMG and the 100,000 Genomes Project and the phenotype parameters assessed through the IMPC phenotyping pipeline, aiming to identify gaps and weaknesses in the current phenotyping platforms.*

These comparative analyses will enable a better understanding of the impact of each phenotyping test on assessment of disease state in the mouse, including the appropriate granularity to apply, as well as critical areas of the HPO that are not covered. In particular, analysis of IMPC phenotyping calls on orthologues of human disease genes will provide an ongoing measure of rates of phenotype detection and how they vary based on mode of inheritance and human allele type. Any future development of algorithms importantly needs to account for negative phenotypes, including those that may reflect genetic context.

- *Develop automated dysmorphology mapping of mutant mouse embryos to the EMAPA datasets;*
- *Provide longitudinal representations of common disease phenotypes, taking advantage of the IMPC ageing screen and directly integrating mouse phenotype data to human epidemiological data for age-related disease.*

3. Develop informatics tools and shared space to facilitate prioritisation and delivery of human disease and trait associated variants for the IMPC's production and phenotyping

A key goal for the IMPC is to develop substantive interactions and collaborations with key networks and consortia in human genome science and clinical genetics that will underpin the selection of both coding and non-coding sequence variants for production and phenotype analysis. Systems from tracking of mutant production to annotation, visualisation and dissemination of phenotype information will be a priority for IMPC informatics development. We will extend the IMPC gene, phenotype, and disease web pages to display the available phenotypes in the context of the generated allele, null or otherwise. Importantly, we will also provide support and assistance to all end users in data interrogation and analysis.

4. Identify and establish vibrant interactions and networks with human genome centres, clinical genetics consortia, and data-curated biobanks that we serve, to ensure rapid translation of mouse genome analyses on our understanding of human genetic variation, function, and disease mechanisms

The IMPC strategic plan and scientific goals require us to engage broadly with and establish new interactions and networks with global precision medicine initiatives, various human genetics and rare disease networks, as well as clinical consortia and biobanks. Interactions, collaborations and partnerships will be vital for ensuring rapid and significant impact of mouse genome analyses on our understanding of human genetic variation and disease, and translating that understanding into diagnostic and therapeutic outcomes. Already the IMPC has been developing relationships with a number of networks and stakeholders (e.g. the CMG, Genomics England) and it is vital for the IMPC to extend its engagement and develop strategies for outreach and establishment of new networks. Our engagement plan aims to:

- *Expand interactions with human genome centres, disease networks, clinical genetics consortia, and biobank studies.*

We aim to establish relationships with all major centres, networks, consortia and biobanks by the end of Phase 2 (2020). Some of these relationships will be deep, working relationships receiving and prioritising information on coding sequence and non-coding sequence targets and mutations, and providing mutant resources and phenotype information. We will identify and develop our engagements by invitations to IMPC meetings and conferences, active dialogue with key principals of each network and centre, and wider outreach and communication activities at symposia, workshops, and major conferences (see below).

- *Developing the IMPC communication, outreach, and marketing (COM) strategy*

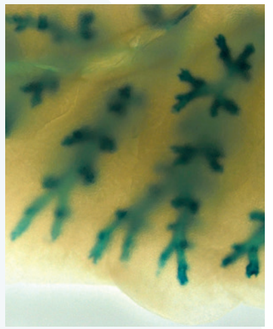


Fig 7. Gene Shh, Lung.

The IMPC will develop and deliver an improved communication programme taking account of the stakeholders identified by the strategy and goals outlined above. The communications programme will include all channels of broadcast distribution (media-voice-email-print) to make contact and share information. A significant component of the programme will be devoted to social media. The outreach programme will include events and engagement at a variety of forums to proactively find individuals, established or nascent projects and consortia, and funders who are not aware of the IMPC, but who need IMPC data or resources or advice to support their research enterprise. Examples include an interactive information booth at large multi-disciplinary meetings (SfN, ASHG, etc.). The outreach plan will include increasingly dynamic interactions with potential stakeholders, including biotech, pharma, patient groups, and the wider public. Moreover, our outreach events will be extended to include focused interactions, particularly where there are key impacts from insights emerging from the IMPC into the wider genome landscape (e.g. Developmental Biology meetings, ISMB). The IMPC will also develop and deliver a dynamic marketing programme, evolved over time to promote research products or services available on a transactional basis to maximise provision of its scientific resources to sectors of the global academic or industry community more aligned with a fee-for-service model of collaboration.

Summary

Understanding the functional consequences of variations in the human genome is an enormous challenge but is an essential prerequisite for progressing fundamental knowledge, translational research, and precision medicine. The challenge is heightened by our limited knowledge of basic gene function (coding and non-coding) and the extent of the 'dark' genome. Through the generation of null mutations the IMPC has already made a considerable impact on knowledge of gene function across many disease areas and our wider understanding of the mammalian genome landscape. Our ability to interpret the function of human genetic variation is enhanced by the significant steps that IMPC has already made in our understanding of the coding genome. We set out here an ambitious but achievable strategy to engage the expertise and facilities of the IMPC in developing large-scale analyses of both coding and non-coding sequence variation. Over ten years from 2021–2030, the IMPC, interacting with all stakeholders across the biomedical science community, will generate new knowledge on the mammalian genome, in particular the function of human genetic variation, that will have a transformative impact on clinical diagnosis and management, preventative and therapeutic development, precision medicine, and healthcare delivery.

"We set out here an ambitious but achievable strategy"

Footnotes

[1] *Precision medicine uses information to identify which therapy will be effective for which patients based on genetic, molecular, environment, and lifestyle factors.*

[2] *The 1000 Genomes Project Consortium (2015)* ***A global reference for human genetic variation.*** Nature *doi:10.1038/nature15393. Telenti et al. (2016)* ***Deep sequencing of 10,000 genomes.*** PNAS *doi/10.1073/pnas.1613365113*

[3] *Short et al. (2018)* ***De novo mutations in regulatory elements in neurodevelopmental disorders.*** Nature *doi:10.1038/nature25983; Iulio et al. (2018)* ***The human noncoding genome defined by genetic diversity.*** Nature Genetics *doi.org/10.1038/s41588-018-0062-741588-018-0062-7*

[4] *Oprea et al. (2018)* ***Unexplored therapeutic opportunities in the human genome.*** Nature Reviews Drug Discovery *doi:10.1038/nrd.2018.14; Stoeger et al. (2018)* ***Large-scale investigation of the reasons why potentially important genes are ignored.*** PLoS Biol. *doi.org/10.1371/journal.pbio.2006643*

[5] *Musunuru et al. (2018)* ***Functional assays to screen and dissect genomics hits.*** Circ Genom Precis Med. *11: e002178. DOI: 10.1161/CIRCGEN.118.002178*

[6] *Brown et al. (2018)* ***High-throughput mouse phenomics for characterizing mammalian gene function.*** Nature Reviews Genetics *https://doi.org/10.1038/s41576- 018-0005-2*

[7] *Liao et al. (2017)* ***In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation.*** Cell *doi.org/10.1016/j.cell.2017.10.025*

[8] *National Academies of Sciences, Engineering, and Medicine (2018)* ***Advancing Disease Modeling in Animal-Based Research in Support of Precision Medicine: Proceedings of a Workshop.*** Washington, DC: The National Academies Press. *doi: https://doi.org/10.17226/25002.*

[9] *Blair, D. R. et al. (2013)* ***A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk.*** Cell *155, 70–80, doi:10.1016/j.cell.2013.08.030.*

[10] ***http://mendelian.org***

[11] ***https:// undiagnosed.hms.harvard.edu***

[12] ***https://www.genomicsengland. co.uk***

[13] ***http://www.ukbiobank.ac.uk***

[14] ***https://druggablegenome.net***

[15] ***http://www.phenomin.fr/en-us/services/disease-investigation/18***

[16] ***http://care4rare.ca***

[17] ***https://commonfund.nih.gov/kidsfirst***

[18] *Preliminary data suggests that deletions uncover both substantive phenotype and transcriptional effects, while precise disease-associated mutations through modelling transcriptional effects may not reveal strong disease phenotypes. This may reflect subtle differences between mouse and human in the action of transcriptional factors at the CNE.*