IMPC

The International Mouse Phenotyping Consortium

ANNUAL HIGHLIGHTS

2020

Contents

Introduction & Year Review

2020 has brought change, trials and a level of uncertainty not seen in decades. In the face of a global pandemic, the scientific community has banded together to combat this new health challenge and continue essential research in other areas.

We assisted COVID-19 research this year by collating our COVID-19 related resources. You can find COVID-19-related IMPC alleles on our resource page. We're also part of the Global Mouse Models for COVID-19 Consortium (GMMCC), a global partnership of genetic and genomic centres and repositories. GMMCC aims to underpin global research into SARS-CoV-2 by the efficient delivery of relevant mouse strains and mouse genetics expertise.

The International Mouse Phenotyping Consortium, IMPC, comprises 21 of the leading mouse genetics centres worldwide. We have been developing a complete functional catalogue of the mouse genome, linking each gene to disease, enabling a better understanding of how genetic variation in the human population causes disease and identifying new targets for therapeutic intervention.

Our mission is to generate a knockout (KO) mouse model for all protein-coding genes in the mouse genome for which there is an orthologue, or equivalent gene, in the human genome. Each KO mouse line undergoes phenotyping via our comprehensive and standardised pipelines to identify significant changes in phenotype. The data generated is processed and released on our open-access web portal.

We have generated KOs for nearly 10,000 genes, over half the orthologous genome, and have already phenotyped over 7,400 KOs – a staggering achievement, which we discuss in more detail below.

Mouse welfare is a top priority; our animal technicians continued to conduct 24/7 mouse care. Our mouse centres have continued phenotyping throughout the year and our team at the IMPC Data Coordination Centre, DCC, has continued to load and analyse data, also providing data support to all our users.

Like all other disciplines, research in the health and medical sector has been affected by the pandemic. We believe it is important to celebrate the scientific successes and achievements made in such a difficult climate.

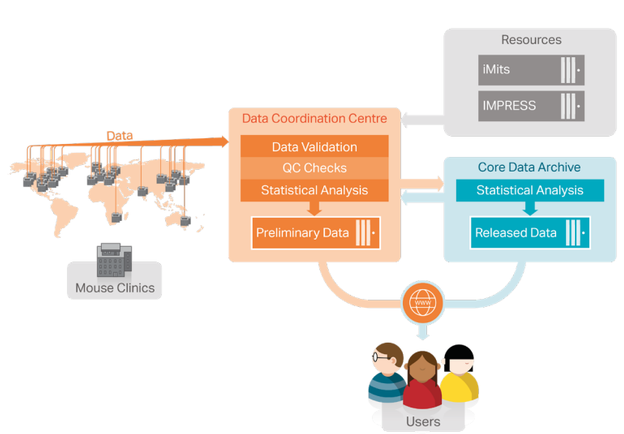
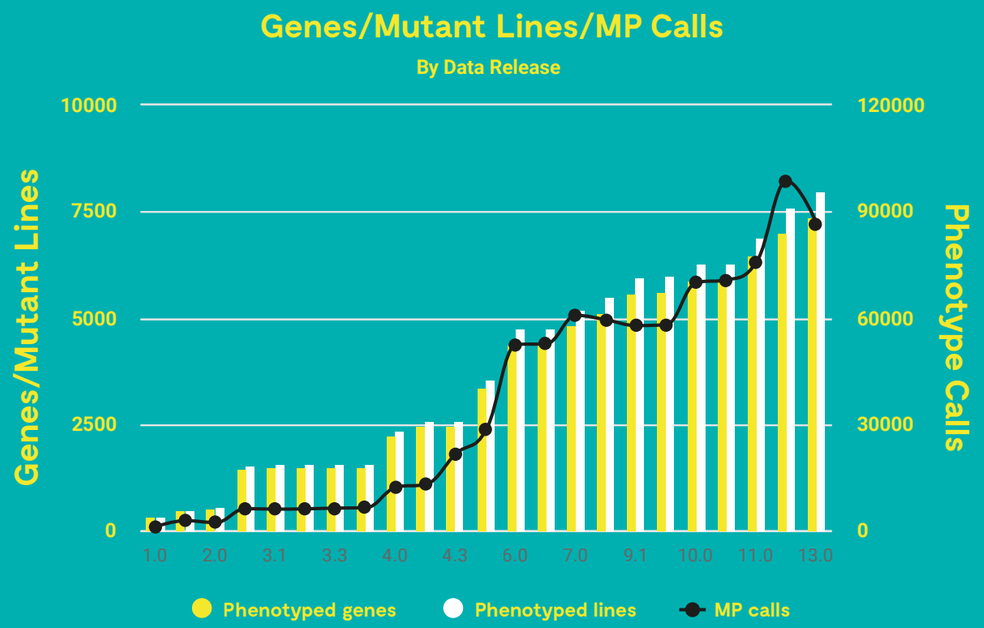


Image: IMPC data flow

Join us as we highlight achievements across our consortium, celebrating scientific success in **2020** >>>>>

2020 New Dimensions for IMPC

The IMPC database grows larger every year. In 2020, we reached a new milestone, generating KOs for over 9,700 genes. These gold-standard mouse lines represent over half the orthologous genome between mouse and human. Moreover, over 7,400 of these KOs have been phenotyped, generating 98.8 million data points and over 572,000 images!



Data is coming faster, with continuing improvements in quality

IMPC publishes regular data releases (DRs), continuing to improve the quality and robustness of data. This year, we've published four data releases, DR 10, 11, 12 and 13. There have been some important changes in our data releases, including updating our statistical program from PhenStat to OpenStats and our first integration of data from our late adult pipeline (where mice are aged before they are phenotyped.) Our next release is coming up, so keep an eye on our web portal or follow us on @impc for the next update.

Phenotypes exhibited by aged individuals are important for understanding gene function. Selected mouse lines enter our ageing pipeline, in which mice are aged before they are phenotyped at 52 weeks or later. This data allows researchers to observe the effects of gene knockouts later in life. This is vital for understanding the underlying genetics of late-onset disease. Many human diseases are age-related, such as dementia, osteoporosis, diabetes, hearing loss and eye issues. Late adult data will help researchers understand these types of diseases, find relevant candidate genes and identify mouse models that accurately represent the late-onset phenotypes.

We will continue to release late adult data in 2021, supporting the ongoing research into these areas.

Know your Genes - Late Adult Data

Adcy1

Adenylate cyclase 1 (***Adcy1***) is found in the nucleus, plasma membrane and synapses. It has several protein binding functions and both adenylate cyclase and phosphorus-oxygen lyase activity. Evidence shows ***Adcy1*** is involved in various key biological processes, such as circadian rhythms, brain development and intracellular signal transduction. ***Adcy1*** is also an orthologue for the human ***ADCY1*** gene, which is associated with deafness. (See all disease models associated with ***Adcy1*** by annotation and orthology or phenotypic similarity.)

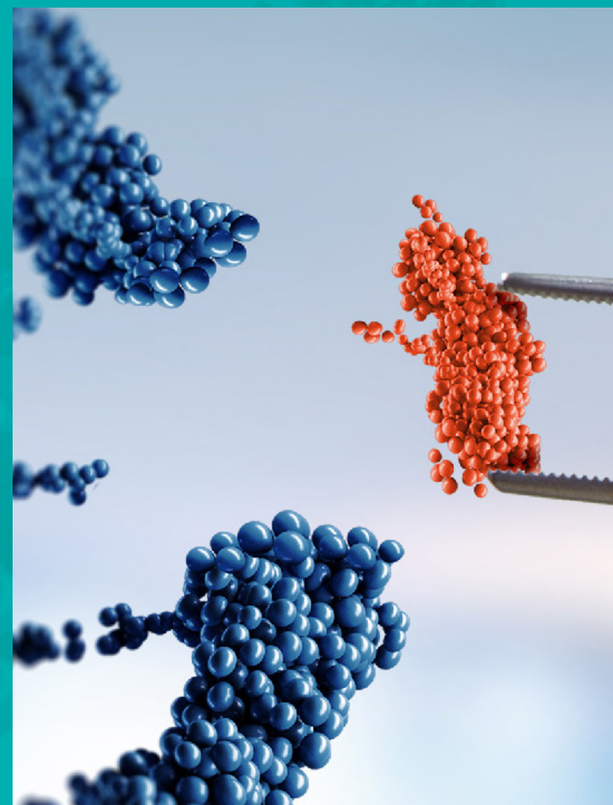
- IMPC phenotyping, found that a viable phenotype was present with both copies of the gene were knocked out (Adcy1 KO mice homozygous viable.)
- Early adults exhibited one significant phenotype affecting the nervous system.
- Middle-aged adults exhibited one significant phenotype affecting pigmentation.
- Aged mice exhibited six significant phenotypes affecting the nervous system, pigmentation, metabolism and hematopoietic/immune system.

IMPC expression data found adult mutant expression in 24 tissue types and embryo expression in 17 tissues types. NCBI mouse ENCODE transcriptome data shows ***Adcy1*** is highly expressed in the cerebellum and cortex. See available IMPC models and ES cells for this gene.

CRISPR has allowed us to knock out genes and phenotype them much more quickly than our traditional ES cell method

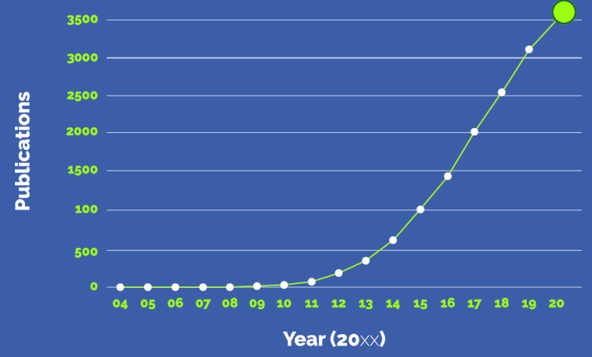
GenTaR, our new tracking platform for phenotyping, is another data change this year – read our news post for more information.

In 2021, we will still strive for excellence with our data – continuing to improve and update our resources.



Publication Highlights

We reached a record of over **3500** IMPC-related publications this year, even with COVID-19 affecting research around the world.



Here are some highlights:

Highlight 1

The Deep Genome Project Could Change our Understanding of Disease and Healthcare

Our work across the IMPC aims to identify every protein-coding gene in the mouse genome, which makes up 3-5% of mouse DNA. This is a vital first step in understanding gene function.

Much of the coding genome still remains 'dark'. There are many genes for which we have no functional information at all and, as a consequence, we are ignorant of their role in disease. We know even less about the role of the non-coding portion of the genome (which makes up 95%) in disease. Researchers are only just beginning to understand how non-coding DNA can regulate gene function and contribute to disease.

Forty-four leading scientists, clinicians, and academics authored an editorial in *Genome Biology* to raise awareness of how limited our understanding is of the function of the majority of our genes and their products. In addition to IMPC members, the authors of this editorial include numerous human geneticists and clinical researchers. Together, they laid out a four-step plan on how to illuminate the dark genome.

1. Study the protein-coding section of the mouse genome (3-5%.) The IMPC is working towards completing this step. This includes introducing human coding variants into mice to explore their function and providing a critical pathway for understanding the role of human genetic variation in disease.
2. Target the non-coding section of the genome (95%.) Non-coding DNA has vital roles in gene regulation and can have a significant effect on gene function.
3. Translate this genetic information and data into clinical knowledge. Doctors, Specialists, Clinician-Scientists, and Researchers anywhere in the world will be able to use this information to study the role of genes in health and disease and find new targets for therapies.
4. Ensure fast and easy access to this data for integration into the clinical decision-making process. By streamlining the production and analysis of these mouse models, clinicians could diagnose patients more easily. Rapid administration of targeted therapies would increase the chances of treatments being effective earlier in a patient's course of disease.

They firmly believe that generating knockout models for coding and non-coding genes will transform biology, medicine, and global health. The Deep Genome Project will develop and deliver a better understanding of the genome and improve the current resources that researchers and clinicians use for the study of disease, efficient diagnoses, development of treatments, and improved patient care.

Full IMPC article:
mousephenotype.org/news/impc-editorial-how-the-deep-genome-project-could-change-our-understanding-of-disease-and-transform-healthcare

Journal Paper:
Lloyd, K.C.K., Adams, D.J., Baynam, G. et al. The Deep Genome Project. *Genome Biol* **21**, 18 (2020). doi.org/10.1186/s13059-020-1931-9

Highlight 2

IMPC Experts are Analysing Gene Essentiality to Find New Possible Causes of Disease

IMPC researchers published a paper on gene essentiality this year, introducing a new gene classification system, Full Spectrum of Intolerance to Loss-of-function (FUSIL), which identifies genes associated with disease.

FUSIL classifies genes based on organismal viability and cellular viability. Loss of gene function is often referred to as a binary concept; lethal or viable. This study shows that gene essentiality is more of a spectrum ranging from cellular lethal, developmental lethal, subviable, viable with a significant phenotype, and viable without a significant phenotype.

The study found several interesting associations, such as how essential genes were almost always either cellular lethal or developmentally lethal. Developmentally lethal genes were also more likely to be associated with early-onset diseases and/or diseases affecting multiple bodily systems.

Developmentally lethal genes, therefore, make great targets for further analysis. The researchers tested the FUSIL system on three large rare disease sequencing programs for unsolved diagnostic cases which flagged two genes, VPS4A and TMEM63B.

Two unsolved cases associated with VPS4A showed patients with intellectual disability, developmental delay, delayed motor development and eye abnormalities. IMPC knockout mice for VPS4A had smaller embryo, abnormal brain development and spine curvature phenotypes.

Know your Genes - Gene Essentiality

Vps4a

Vacuolar protein sorting 4A (***Vps4a***) is found in the cytosol, endosome, nucleus and plasma membrane. It has a few protein binding functions and ATPase activity. Evidence suggests ***Vps4a*** is involved in processes such as the cell cycle, cell division, vesicle budding, and endosomal transport. ***Vps4a*** is also an orthologue of the human gene ***VPS4A***. See disease models associated with ***Vps4a*** by annotation and orthology or phenotypic similarity.

- IMPC phenotyping found that a lethal phenotype was present with both copies of the gene that were knocked out (***Vps4a*** KO mice homozygous lethal.)
- Early adults exhibited six significant phenotypes affecting the nervous system, integument, immune and hematopoietic systems and vision.
- Two embryonic phenotypes were found at E18.5, affecting growth and size. See embryo viewer.

IMPC expression data found adult mutant expression in 26 tissue types, with a further three tissues having ambiguous expression. Embryo expression was found in all 62 tissue types tested, suggesting ***Vps4a*** expression is widespread during development. NCBI mouse ENCODE transcriptome data shows ***Vps4a*** is most highly expressed in the testis and cerebellum in adulthood and the central nervous system (CNS) during development. See available IMPC models and ES cells for this gene.

TMEM63B was mutated in one clinical case, with intellectual disability, abnormal movement and brain morphology symptoms. IMPC knockout mice for this gene had phenotypes like abnormal behaviour and hyperactivity.

The majority of rare disease patients remain undiagnosed due to a lack of detection or due to the difficulty in interpreting a variant in an uncharacterized gene. Undiagnosed patients also often suffer from physical, social and financial costs. FUSIL, as an open-access resource, can be used to identify candidate genes linked to rare disease. Clinicians and researchers can use FUSIL to more easily prioritise candidate genes for studies, making it easier to find the causes of undiagnosed cases and de novo genetic disorders.

Find out how we're using FUSIL on our Essential Genes data focus page. Our researchers are working on an IMPC haploessential screen to support the discovery of essential genes associated with disease. Our FUSIL and haploessential data will be incorporated into our upcoming Essential Genes Data Portal where researchers can submit gene lists to find human and mouse essential scores. Through this, the IMPC will continue to support the discovery of novel candidate genes linked to rare disease.

Full IMPC article:
mousephenotype.org/news/how-impc-experts-are-categorically-analysing-gene-essentiality-to-find-new-possible-genetic-causes-of-disease

Journal Paper:
CACHEIRO, P., et al. (2020). Human and mouse essentiality screens as a resource for disease gene discovery. *Nature Communications*. Published online 31 01; DOI: 10.1038/s41467-020-14284-2

Highlight 3

Researchers Invent New Algorithm to Find Genes Linked to Circadian Misalignment

Researchers from our consortium member CAM-SU Genomic Resource Center, Soochow University, have developed a machine-learning algorithm to identify candidate genes linked to disruption of circadian behaviour, eating behaviour and other cyclical biological processes.

Circadian misalignment can be caused by lifestyle or genetics. It can promote conditions such as sleep disorders, depression, seasonal affective disorder, bipolar disorder, and metabolic disorders. Research into circadian misalignment is crucial for our understanding of complex health issues.

Indirect calorimetry (IC) data was collected from five IMPC centres: the Riken BioResource Center, The Centre for Phenogenomics, the Institut Clinique de la Souris, the Wellcome Trust Sanger Institute and Helmholtz Zentrum Munich. With this data, Zhang *et al.* compared 750 knockout mouse models to 2000 wild-type mice, examining differences between food intake and activity peaks over time.

After they tested the algorithm against genes known to disrupt circadian rhythms (which were flagged successfully) they tested it against the 750 mutant mouse models. The algorithm flagged 88 genes, which were further narrowed down by the researchers to a group of five: ***Slc7a11***$^{tm1b/tm1b}$, ***Rhbdl1***$^{+/tm1.1}$, ***Spop***$^{+/tm1b}$, ***Oxtr***$^{tm1.1/tm1.1}$ and ***Ctcl***$^{+/tm1b}$. Mice in these mutant lines shared similar phenotypes, had statistically significant differences from wild-type mice and had an effect size greater than 1.2.

The researchers further investigated *Slc7a11*. The *Slc7a11*$^{tm1b/tm1b}$ mutant line expressed early activity onset and potentially impaired metabolism due to altered glucose tolerance. They generated an IMPC knockout model for this gene and found that the mice had disrupted light/dark cycles. They also found that *Slc7a11* plays a role in suprachiasmatic nucleus (SCN) neuron synchronisation and messaging between cells.

Know your Genes - Circadian Rhythms

Slc7a11

Solute carrier family 7 member 11 (***Slc7a11***) is found in the plasma membrane, cytoskeleton and endoplasmic reticulum and has transmembrane transporter activity. Evidence suggests ***Slc7a11*** is involved in processes such as cell death, cell proliferation, homeostasis, visual learning and synapse organisation regulation, among many others. ***Slc7a11*** is also an orthologue of the human gene ***SLC7A11***. See disease models associated with ***Slc7a11*** by annotation and orthology or phenotypic similarity.

- IMPC phenotyping, found that a viable phenotype was present with both copies of the gene were knocked out (***Slc7a11*** KO mice homozygous viable.)
- Early adults exhibited five significant phenotypes, affecting the metabolism, vision, hearing and size/growth.

IMPC expression data is not currently available for this gene. NCBI mouse ENCODE transcriptome data shows ***Slc7a11*** is highly expressed in the thymus and subcutaneous and genital fat pads. See available IMPC models and ES cells for this gene.

The SCN is our internal master clock that regulates our innate circadian rhythms. *Slc7a11*'s apparent role in entrainment to light/dark cycles and the synchronisation of SCN neuron activity suggests it could play a role in circadian misalignment if disrupted by a mutation. Zhang *et al.* concluded that the algorithm could successfully find novel candidate genes for this condition.

This algorithm can be used for future dataset analysis, opening new research possibilities into circadian misalignment for metabolism, sleep and behaviour. Future research into the genetic background of circadian misalignment is vital for our understanding of an ever-growing list of conditions.

Full IMPC article:
mousephenotype.org/news/researchers-invent-new-algorithm-to-find-genes-linked-to-circadian-misalignment

Journal Paper:
Zhang T, Xie P, Dong Y, Liu Z, Zhou F, Pan D, *et al.* (2020) High-throughput discovery of genetic determinants of circadian misalignment. PLoS Genet 16(1): e1008577. doi.org/10.1371/journal.pgen.1008577

Highlight 4

Researchers Investigate the Cause of Fam151b-related Retinal Degeneration

Researchers from the MRC Human Genetics Unit (Institute for Genetics and Molecular Medicine, University of Edinburgh), in collaboration with the Mary Lyon Centre (MRC Harwell Institute) and the Roslin Institute (University of Edinburgh), recently used IMPC data to inform their research into ***Fam151b***.

We previously flagged *Fam151b* as a gene linked to eye morphology and retinal degeneration, specifically 'abnormal retina morphology' at a significance of 2.55×10^{-38}. Research into this gene could improve our understanding of the retina and retinal diseases, such as retinitis pigmentosa (RP) and other conditions that lead to blindness.

Findlay *et al.* first determined whether *Fam151b* had similar functions to its C. elegans homologue, *menorin* (*mnr-1.*) They ran the gene through several databases, including HHPred which matched the Fam151 profile with the Phospholipase D domain of a spider (*Sicarius terrosus*) toxin. Other statistically significant matches included members of the 'Phospholipase D and glycerophosphodiester phosphodiesterase' (GDPD) families. These GDPD members, like other PLC-like phosphodiesterase enzymes, have a TIM barrel fold that is also found in menorin and Fam151b proteins.

Fam151b is expressed at low levels in most tissues, including the retinal pigmented epithelium (RPE), retina, iris, ciliary body, lens and cornea.

The Mary Lyon Centre (MRC Harwell Institute), one of our consortium members, generated a *Fam151b* knockout mouse model (*Fam151b*$^{KO/KO}$). Findlay *et al.* examined this model across several weeks and found "a patchwork pattern from fundal imaging indicative of retinal degeneration" at week 11. Histological analysis found a "severe reduction in the length of the outer segments of the photoreceptors [and reduction] in the number of nuclei…found in the outer nuclear layer (ONL.)"

Degeneration began occurring around day 15 at eye-opening, after which it gradually progressed. They examined several possible causes for this degeneration, ruling out retinal pigmented epithelium damage, light toxicity and nerve cell patterning issues. The exact cause of the degeneration remained unclear.

This study creates a platform for future research into *Fam151b's* gene function and how it may relate to human disease, as well as research into the Fam151 family on a larger scale. An understanding of the underlying mechanisms involved in photoreceptor loss is vital for the production of treatments. The mouse model in this study could be used to find effective therapies for photoreceptor loss, which is a leading cause of blindness.

Full IMPC article:
mousephenotype.org/blog/2020/10/12/researchers-investigate-the-cause-of-fam151b-related-retinal-degeneration

Journal Paper:
Findlay, A.S., McKie, L., Keighren, M. *et al.* *Fam151b*, the mouse homologue of C.elegans menorin gene, is essential for retinal function. Sci Rep **10**, 437 (2020). doi.org/10.1038/s41598-019-57398-4

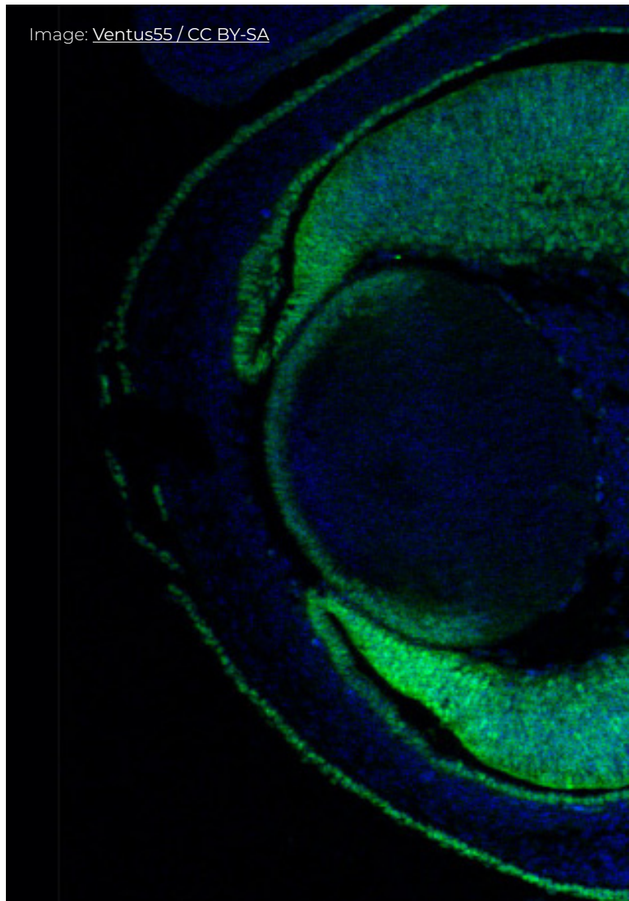


Image: Ventus55 / CC BY-SA

Highlight 5

Arf1-KO Induces Anti-Tumour Response with Protective Vaccination

IMPC *Arf1* embryonic stem cells were used in a study investigating the *Arf1*-mediated lipid metabolism pathway's role in sustaining Cancer Stem Cells (CSCs) and how disrupting this pathway would affect tumour growth and number.

Wang *et al.* knocked out *Arf1* from a known model for CSCs (Lgr5-CreERT2/Apc$^{f/f}$). The generated Lgr5/Arf1/Apc mice had a dramatic reduction in stem cell tumour numbers and a significant increase in lifespan.

Wang *et al.* found that "the *Arf1*-regulated lipolysis pathway selectively sustains stem cells, progenitors and cells enriched with CSCs in mice and that disrupting the pathway in these cells results in lipid droplet accumulation, mitochondrial defects and cell necrosis."

Arf1 depleted mice showed evidence of a T-cell immune response, with dendritic cells (DCs) and inflammatory DCs also significantly increased. "Knocking out *Arf1* in Lgr5+ [intestinal] stem cells triggers T-cell infiltration and activation, leading to CSC death and prolonged survival." Similar effects were seen on liver CSCs, potentially via downregulation of PD-L1.

Knocking out *Arf1* also affected other aspects of the immune system. They found *Arf1* inhibition may induce inflammasome-mediated cell necrosis/pyroptosis. *Arf1* inhibition or ablation also induced key DAMPS, as well as ER-stress markers. They concluded that *Arf1* inhibition triggers ER stress and, as a result, induced DAMPS and DC infiltration, enhancing the T-cell infiltration.
Arf1 knockdown was not directly cytotoxic to tumour cells and, after further tests, it was determined that the anti-tumour effect was via DCs-ATP-IFN-γ-mediated T-cell immunity. Knockdown also prevented tumour metastasis.

Vaccination with knockout *Arf1* cells also proved successful in protecting animals from developing tumours. The vaccination was effective against melanomas and a variety of histological types like colon cancer and breast carcinomas. *Arf1* inhibition triggered a localised T-cell immune response that then attacked cancer throughout the body.

If results like the above are reproducible in humans, DAMP-mediated anti-tumour immune therapy could be a breakthrough for the treatment of reoccurring tumours, particularly those with high *Arf1* expression. When used in conjunction with other already effective treatments or with the addition of successful checkpoint blockades, this method could prove even more effective.

Full IMPC article:
mousephenotype.org/blog/2020/04/27/arf1-ko-induces-anti-tumour-response-with-protective-vaccination

Journal Paper:
Wang, G., Xu, J., Zhao, J. *et al.* *Arf1*-mediated lipid metabolism sustains cancer cells and its ablation induces anti-tumor immune responses in mice. Nat Commun11, 220 (2020).
doi.org/10.1038/s41467-019-14046-9

"If the results are reproducible in humans, DAMP-mediated anti-tumour immune therapy could be a breakthrough for the treatment of reoccurring tumours"

Consortium Members, Funders and Collaborations

2020 has been an important year for the growth of the IMPC consortium and the further development of critical research programmes with our collaborators.

Firstly, we'd like to welcome our new consortium member, the Institute of Laboratory Animal Sciences (ILAS) at the Chinese Academy of Medical Sciences, Beijing. ILAS is a comprehensive national research centre for laboratory animals and comparative medicine. They integrate the conservation, breeding, and supply of laboratory animals and animal disease models, combining the comparative study of medical technology with technical training.

We want to thank all of our consortium members for their exceedingly hard work in 2020. The IMPC, as an international collaborative effort, integrates data from many mouse centres across the globe. The expertise and capabilities of our 21 consortium members are the backbone of the IMPC.

We'd also like to thank all our funders from 2020. Without their support, we wouldn't have made the substantial progress we have made this year.

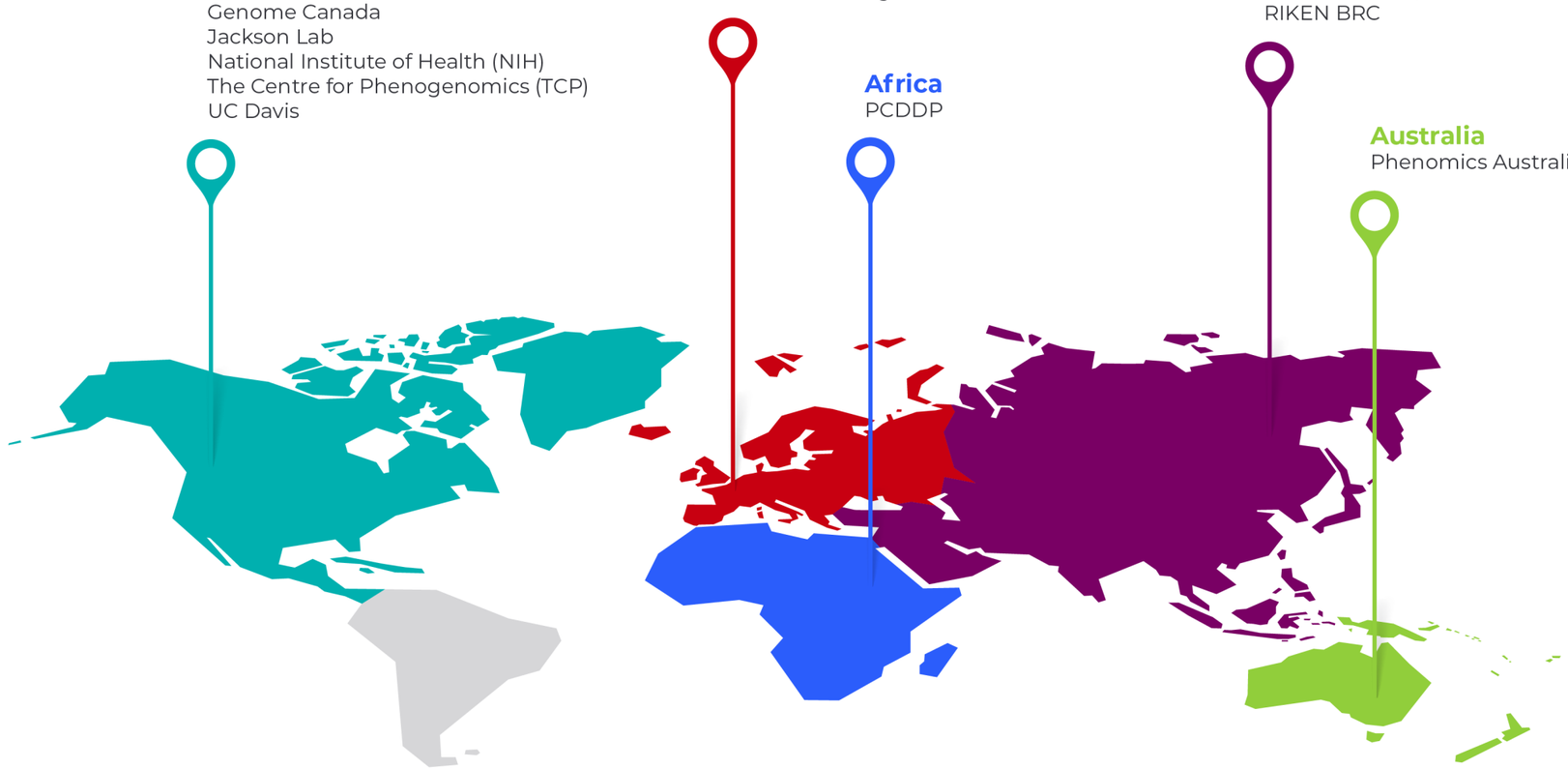
North America
Baylor College of Medicine
CIHR
Genome Canada
Jackson Lab
National Institute of Health (NIH)
The Centre for Phenogenomics (TCP)
UC Davis

Europe
CNR Monterotondo (IMC)
Czech Centre for Phenogenomics (IMG)
EBI
Helmholtz Zentrum Munich (GMC)
MRC Harwell Institute
PHENOMIN
Universitat Autonoma de Barcelona (UAB)
Wellcome Sanger Institute

Africa
PCDDP

Asia
CAM-SU GRC, Suzhou
IISER
ILAS, Chinese Academy of Medical Sciences
Korea Mouse Phenotyping Center
MARC, Nanjing
NLAC NARlabs, Taiwan
RIKEN BRC

Australia
Phenomics Australia



Lastly, we would like to thank our collaborators, whose partnerships enable us to have a continued and increasing global impact with our data and resources. Here are some of our collaborator highlights from **2020** >>>>>

Gabriella Miller Kids First Pediatric Research Program

Kids First aims to uncover new biological insights into childhood cancer and structural birth defects. Children with birth defects have an increased risk of developing childhood cancer, suggesting that these disorders may share genetic pathways. Kids First is using whole genome sequencing and the Gabriella Miller Kids First Data Resource, a large-scale database of clinical and genetic data, to discover new genetic pathways between these disorders.

Kids First has now over 5,000 families and 14,000 participants in 23 studies. In collaboration with IMPC mouse centres, they're investigating genes that have been implicated in facial weakness, diaphragmatic hernias (CDH), adolescent idiopathic scoliosis and much more.

The IMPC and Kids First are collaborating on a pilot project to create mouse strains to study, phenotype, and validate coding and non-coding genetic variants (e.g. missense, structural variants, copy number variants, INDELS, frameshifts) identified from Kids First datasets.

The IMPC and Kids First will continue to collaborate, investigating more genes to find links between birth defects and childhood cancer.

Illuminating the Druggable Genome IDG

Around 3,000 genes are part of the druggable genome. However, less than 10% of druggable proteins are targeted by FDA-approved drugs. IDG is a program to identify and collect information on genes within druggable protein families, specifically genes that are not well-studied. These protein families include kinases, G-protein coupled receptors and ion channels. IDG is integrating this data into a single resource, Pharos, and is also helping support technology development for the study of druggable proteins.

The IMPC (KOMP2) is generating knockout mouse strains for the IDG project. Human genes that IDG are studying can be mapped to mouse orthologues. The IMPC can then generate and phenotype knockout mouse models for these orthologues. The IMPC currently has data for 41% of IDG's target genes across 25 groups of phenotypes, including ageing, reproductive, immunity, craniofacial, nervous, renal, cardiovascular and many more.

Information sharing between the KOMP2 and IDG projects allows researchers to prioritize understudied genes and increase the pace and scientific depth of phenotyping studies. The combination of IDG's human protein/gene data with the IMPC's phenotype and mouse model data helps researchers to find candidate genes for drug discovery. It also accelerates our ability to discover new biological insights and apply findings to health and disease.

Centers for Mendelian Genomics CMG

The CMG use genome-wide sequencing and other methods to discover the genetic basis underlying Mendelian traits. They accelerate discoveries by sharing their data, knowledge and effective approaches via reaching out to individual investigators and coordinating with other rare disease programs worldwide.

Discovery and understanding of genetic variants that underlie human Mendelian disorders will assist in the faster diagnosis of patients and lead to new therapeutic approaches.

We collaborate with the CMG by prioritising genes with putative Mendelian variants for the generation and analysis of null mutants via the IMPC pipeline. Our mouse models support the functional investigation of candidate genes, mechanistic evaluation of candidate variants and the identification of Mendelian conditions. We have KO models and data for over 2,000 of the genes CMG are investigating.

Over eight and a half years, CMG has made over 2,600 novel discoveries. This includes variants associated with congenital diaphragmatic hernias, Robinow syndrome, Mayer-Rokitansky-Kuster-Hauser syndrome and TBX6-associated congenital scoliosis (TACS.)

The IMPC aims to continue to provide KO models for projects like CMG and the Mendelian Genomics Research Center. Through this, we will continue to support the solving of unsolved rare disease cases, increase our understanding of the molecular mechanisms of disease and find potential therapeutic targets.

Undiagnosed Disease Network UDN

The UDN is a study that aims to improve diagnosis and care for patients with undiagnosed diseases and facilitate research into the aetiology of undiagnosed disease. *De novo* and rare mutations are difficult to diagnose. In many cases, rare mutations can be found only in a single-family. The inability to discover the causes for a patient's symptoms can lead to difficulty finding affective treatments, extensive stress and financial difficulties.

Many of these mutations can also be severely physically disabling and have wide-ranging symptoms, such as craniofacial deformities, intellectual disabilities, growth issues, and immunological and haematological irregularities.

The UDN uses IMPC data and mouse models to assist diagnoses. Our phenotype data can provide evidence to support the pathogenicity of variants associated with rare and undiagnosed disease cases. Our mouse models provide resources for functional studies for pathogenicity and therapeutic approaches.

The UDN has now reached over 1400 evaluated participants and, with an additional 400 participants accepted and 333 under-review applications, they will continue to make progress throughout 2021. At the IMPC, we aim to continue to support the UDN study with our resources, supporting the UDN's aim to give valuable diagnoses to those with rare mutations.

Our collaborations are not limited to this year's highlights!

Our continued work with other collaborators such as **MMPC** | **Genomics England** | **Rare Disease Foundation** **UK Biobank** | **Care4Rare** | **INFRAFRONTIER** | **3i** | **Monarch Initiative** and more is vital for our aims. These collaborations not only support our global reach and impact but enable the vital step of translating our data into clinical knowledge and results.

Check out our collaborators page for more information.

Other Exciting Updates

Given the extraordinary circumstances of 2020, the IMPC has had to adapt and change, particularly seeking out new opportunities for virtual engagement with our users. We've been undergoing user experience (UX) testing throughout the year and have been building and improving our web portal behind the scenes. We'll be rolling out these changes in 2021 when testing is complete, offering a new, more fluid and easier to use portal.

Normally, you could expect to see us at some of the largest genetic conferences of the year. This year, we still made it to the virtual versions of ESHG and ASHG as attendees or exhibitors. It's great to see that, even as countries entered lockdowns across the world, the scientific community continued to perform and share excellent work.

To help our users, we also produced a free and virtual IMPC course and webinar this year. New and older users from various biology disciplines can learn how to fully take advantage of our resources. Check them out if you want to learn how to get started with our data, how to order mouse models or how IMPC resources are being used in studies of genetic disease, development and sexual dimorphism, amongst other topics.

We were also very active on social media in 2020! We post all our article highlights, data release announcements and conference information on our @impc Twitter account. You can also follow @geneoftheday for quick and exciting snapshots of IMPC data.

Know your Genes - IMPC Data

Lep

Leptin (***Lep***) is found in the cytosol, cytoplasm and extracellular space. It has DNA, leptin and signalling receptor binding functions and hormone activity. Evidence suggests ***Lep*** is involved in a wide range of processes, such as the regulation of transcription and phosphorylation, ovulation, placenta development, cytokine production, circadian rhythms, blood pressure, metabolism, and signal transduction. ***Lep*** is also an orthologue of the human gene ***LEP***, which is associated with leptin deficiency and obesity. See all disease models associated with ***Lep*** by annotation and orthology or phenotypic similarity.

- IMPC phenotyping, found that a viable phenotype was present with both copies of the gene were knocked out (***Lep*** KO mice homozygous viable.)
- Early adults exhibited 20 significant phenotypes, affecting the nervous system, cardiovascular system, integument, skeleton, metabolism, vision and reproductive system.

IMPC expression data found adult mutant expression in only three tissue types – brown and white adipose tissue and testis. NCBI mouse ENCODE transcriptome data shows ***Lep*** is highly expressed the subcutaneous and genital fat pads. See available IMPC models and ES cells for this gene.

Looking to the Future - 2021

IMPC's vision to generate a complete catalogue of mammalian gene function is progressing at a rapid pace, providing a critical springboard for future research into human disease and the development of new therapeutic approaches. With over half of the protein-coding portion of the genome available as mouse mutants, we are looking forward in 2021 to reach our target of 9,000 phenotyped mouse genes – providing functional information on half of the orthologous mouse and human genomes.

We're also excited to begin work on our aims to use the mouse as a key tool to analyse the function of human genetic variation and its impact on disease. First, this includes introducing human coding variants into mice to explore their function. Second, we aim to begin to explore the non-coding section of the mouse genome. This is a major challenge but will open up new vistas on the functional relationships between the non-coding genome and human disease.

In 2021, we're planning to attend three conferences: Genomics of Rare Disease, ESHG and ASHG. These and other conferences are exciting opportunities to reach our existing audience as well as new users. We'll display our resources, highlighting the best of IMPC-related research, providing support and learning and collaborating with the scientific community. We hope to see you there!

Would you like to keep track of our progress throughout the year? Follow us on social media **@impc** | **@geneoftheday** and sign-up to our seasonal **newsletter** for regular updates.