IMPC Strategy 2021–2030

The function of human genetic variation

A focus on the next 5 years

The International Mouse Phenotyping Consortium (IMPC) released its 10-year strategy in April 2019. This supplementary document builds on that plan to hone our strategic aims for 2021-2030, provide focus on the next five years (2021-2026), and reflect on achievements and impact of the past decade.

Our Vision

Produce a genome-wide resource of mutant mouse lines with gene function data which is distributed and used globally to model human biology and disease and enables therapeutic discovery and genomic medicine.

Our Mission

The IMPC is a global consortium of academic biomedical research centres and allied funding agencies across five continents working together to:

1. Design, produce, and systematically phenotype a null mutation mouse line across the mouse<->human orthologous genome to provide insights into gene function, disease genetics, pleiotropy, and co-morbidity.

2. Prioritize mouse<->human orthologous genes and their proteins that are poorly understood with no or minimal knowledge on biological function and limited tools for their analysis (the “dark genome”).

3. Partner with the human genetics and clinical community to produce bespoke mouse mutants with human disease-associated coding variants to confirm pathogenicity, susceptibility or resistance to infection, and accelerate research and discovery.

4. Develop and distribute technology and resources that can enhance understanding of the functional relationship between human genetic variation (coding and non-coding), environment, disease, and explore *in vivo* the mechanisms of genetic variation in health and disease.

Our Strategy 2021–2030

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

No null mutation exists for 3,999 mouse<->human orthologues (see **Box 1.** *pg. 4*). Of these genes, 1,043 other mutant allele types are available with some phenotype information. We consider these 1,043 genes to be low priority but still essential to produce as a null mutation with systematic phenotype data to complete genome-wide coverage. If we exclude the large olfactory gene receptor family, there are 2,753 genes that remain with no null mutant. This gene set overlaps significantly with the dark genome (Oprea et al. 2018; and **Figure 1.** *pg. 5*) and is a high priority for IMPC production and phenotyping.

Click to go to Box 1.

Click to go to Fig 1.

Over the next five years we expect IMPC global capacity for mutant generation and phenotyping to exceed 3,000 genes. A key strategic goal for our Consortium is to complete the production, cryopreservation, and systematic phenotype analysis of the 2,753 high priority genes.

2. Significantly increase output of mouse mutants with human disease-associated coding variants to enhance studies of rare disease, accelerate validation of candidate pathogenic variants, and identify disease mechanisms (including susceptibility and resistance to infectious disease).

The IMPC has used CRISPR to produce 348 mutant mouse lines carrying human disease-associated coding variants, many of which have been distributed to researchers and allied consortia. This is an important and emerging activity for the IMPC. Since the inception of CRISPR technology, the IMPC has produced nearly 4,000 CRISPR mutant lines. By comparison ~800 CRISPR mutant lines have been produced worldwide by other public sector labs and centres. The IMPC is a global powerhouse for production of mutant mice and a key strategic aim is to leverage this know-how and capacity for functional analysis of human genetic variation.

Over the next five years we will use the IMPC's global CRISPR design and production capacity to engineer an estimated 1,000 mutant mouse lines with human disease-associated coding variants, and where possible carry out systematic or purpose-driven phenotyping in collaboration with the clinical science community and patient-facing clinical experts.

3. Undertake a pilot programme to design and generate mouse strains that model genetic variation in the non-coding genome.

The role of non-coding variation in human disease is of increasing interest. It will be critical to explore the functionality and mechanisms of non-coding variation in health and disease *in vivo*. The IMPC's experience and capacity for technology development will be used to initiate a programme to generate targeted mutations in non-coding sequences, focused initially on conserved non-coding elements (CNEs).

Over the next five years, the IMPC will undertake a pilot programme to generate a significant number of mouse strains with non-coding mutations and partner with the global research community to assess functionality, pathogenicity, and mechanisms of action.

4. Explore genetic and environmental context to realise the potential of functional studies in the mouse to translate findings of clinical genetics and disease mechanisms.

The IMPC Centres possess an enormous breadth and depth of expertise on the utility of mouse genetics to explore genetic and environmental context and how they impinge on functional and disease outcomes. Many IMPC Centres utilise and have access to mouse diversity resources which is complementary to the IMPC's approach, providing parallel insights into gene<->gene and gene<->environment interactions.

The IMPC as part of its ongoing programme is continuing to provide expertise, mouse diversity resources, and phenotyping capabilities including ageing pipelines, to assist in identifying the multifactorial nature of disease states. The IMPC will also proactively engage and collaborate with public and private sector researchers on purpose-driven preclinical studies and treatment effect or toxicity evaluation of potential diagnostic and therapeutic test articles.

IMPC Achievements

2021 marks the 10th anniversary of the IMPC. Over the last 10 years the Consortium has produced resources, data, and new knowledge that have had profound impact on biomedical science and its capacity to identify dark gene function, essential genes, sexual dimorphism, genotype comorbidities (pleiotropy), and mechanisms of disease.

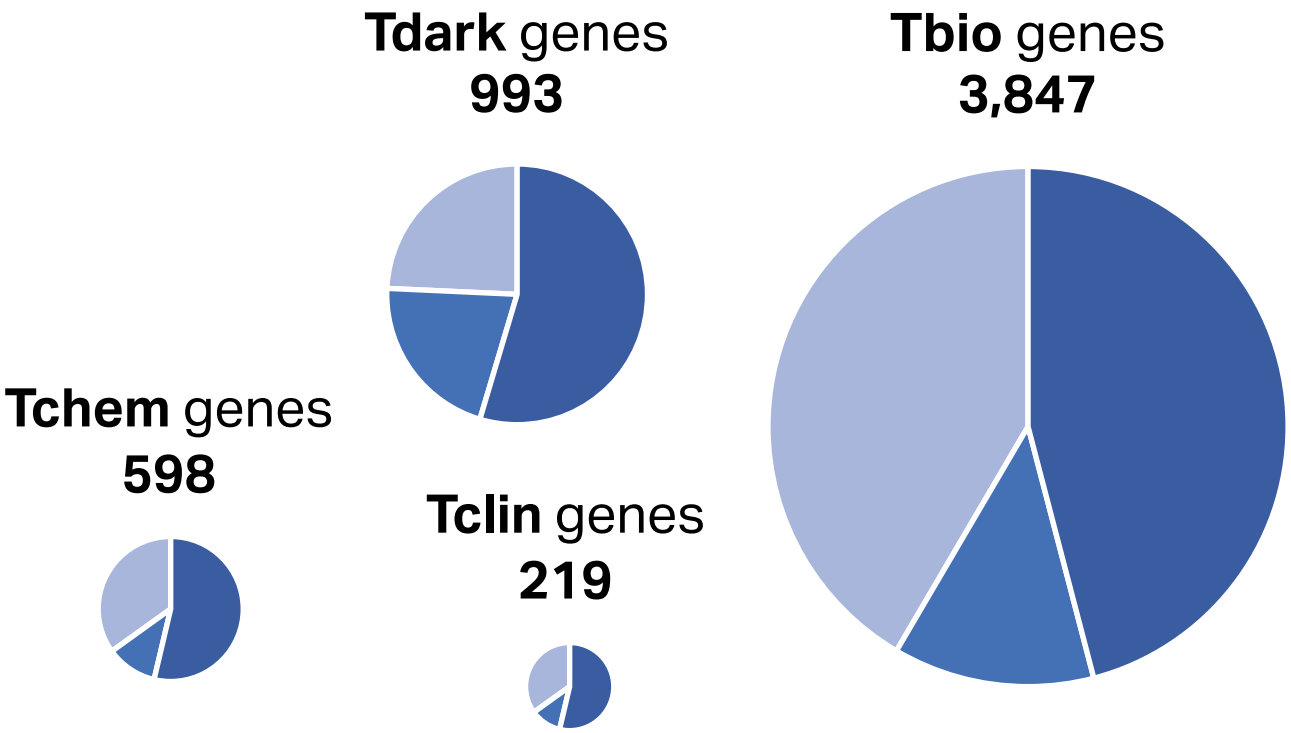
The IMPC has produced and cryopreserved null mutations for 9,719 mouse<->human orthologous genes, representing over half the orthologous genome (mousephenotype.org). 7,455 of these have been phenotyped through a systematic and standardised phenotyping pipeline. This large multidimensional dataset (>98M datapoints) has enabled an unprecedented view of the mammalian genome landscape, particularly in revealing novel loci, many of them thus far unstudied, associated with various disease states (Meehan et al. 2017; Cacheiro et al. 2019). It is noteworthy that of 13,657 total genes for which null mutations have been generated in the mouse, 5,646 genes are exclusive to the IMPC (see **Box 1.**).

Box 1 Completing the null resource

- The IMPC has generated KO mutations for 9,719 genes; over half of known mouse<->human orthologs.
- The entire global biomedical research community has generated KO mutations for 13,657 genes, of which 5,646 are exclusive to the IMPC.
- There remain 3,999 orthologs without mouse null mutations.
- Of these remaining 3,999 orthologs, there are 1,043 phenotypic alleles available.
- Excluding the olfactory gene receptor family leaves 2,753 genes which are an immediate priority for generation of a null mutant line on a defined inbred background strain and comprehensive phenotyping.

Many genes analysed by the IMPC were hitherto part of the dark genome, for which functional information was not available. Detailed analyses of dark gene function confirm the IMPC has been at the forefront of illuminating the dark genome, providing freely available mutant mouse lines and substantive functional data on unexplored genes to help identify those that contribute to or Mendelian disorders and complex disease. Knockout (KO) mouse lines produced and phenotyped by the IMPC with at least one abnormal phenotype mapped to 5,657 human genes. The vast majority (85%) of these IMPC strains map to **Tdark** (993 understudied proteins or "ignorome"

genes) and **Tbio** (3,847 genes with studied biological roles) (Oprea 2019; and **Figure 1.**) and 3,123 of the 5,657 (55%) KO lines with at least one abnormal phenotype mapped to human genes that are under-studied, i.e., PubMed /PubTator score <=50 (IMPC Data Release 13.0 accessed 24 Feb 2021). 960 are **Tdark** (993 understudied proteins or "ignorome" genes) and 1,998 are **Tbio** (genes with studied biological roles).



Abbreviations:

Tdark - understudied proteins that do not meet criteria for the other 3 categories

Tbio - proteins with well-studied biology, having a fractional publication count above 5

Tchem - proteins known to bind small molecules with high potency

Tclin - proteins via which approved drugs act (i.e., mode-of-action drug targets)

Figure 1. IMPC mouse null strains with at least one abnormal phenotype are mapped onto "Target Development Levels" demonstrating significant overlap with the dark and under-studied genome (TDLs, druggablegenome.net/ProteinFam). TDLs reflect the degree to which a gene has been studied and are shown as separate plots. We illustrate knockout mouse lines produced and phenotyped by the IMPC with significant phenotypes mapped to 5,657 human genes. The completeness of mouse phenotyping is indicated by color:

- light blue < 6 procedures
- medium blue = 7—12 procedures
- dark blue = > 13 procedures

The vast majority (85%) of IMPC lines map to **Tdark** (993 "ignorome" genes) and **Tbio** (3,847 genes with studied biological roles). The remaining 817 genes map to **Tchem** or **Tclin** (associated with drugs or chemical probes). 3,123 of these genes, or 55%, of which 960 are **Tdark** and 1,998 are **Tbio**, are under-studied, i.e., PubMed /PubTator score <=50. (IMPC Data Release 13.0 accessed 24 Feb 2021).

The embryo phenotyping pipeline has identified 1,718 homozygous lethal and 620 homozygous subviable genes, significantly enriching human Mendelian disease genes among this class of essential loci (Dickinson et al. 2016; Cacheiro et al. 2020). Female and male lines with viable mutations enter an early adult pipeline with sequential tests that assess most organ and physiological systems. This has enabled a better understanding of the pervasive and wide-ranging sexual dimorphism of phenotypic traits in both wild-type and mutant mice (Karp et al. 2017).

The generation of an unbiased dataset linking genes to comprehensive phenotype information has enlightened the nature and extent of pleiotropy on a genome-wide scale. Of the 7,360 genes with complete phenotyping data in the IMPC web portal (Data Release 13.0 published 17 Dec 2020) ~65% have >1 abnormal phenotype annotation (see **Box 2.**). 90% of these gene-phenotype annotations have not been previously reported (Meehan et al., 2017 and unpublished data). The IMPC's mutant lines with associated phenotype data will allow us to further dissect the nature and interplay across multiple loci and genetic pathways of multi-morbidity in disease states.

Box 2 Pleiotropy across the mouse genome

- Data in the IMPC web portal (Data Release 13.0 published 17 Dec 2020) has complete phenotyping data on 7,360 genes.
- For 846 genes which failed to show a phenotype, a minimum of 13 procedures was not carried out, and they are excluded from the analysis.
- Number (%) of genes:

with no phenotype	**452 (6.9%)**
with >zero phenotypes	**6,062 (93.1%)**
with one phenotype	**1245 (19.1%)**
with ≥two phenotypes	**4817 (74%)**

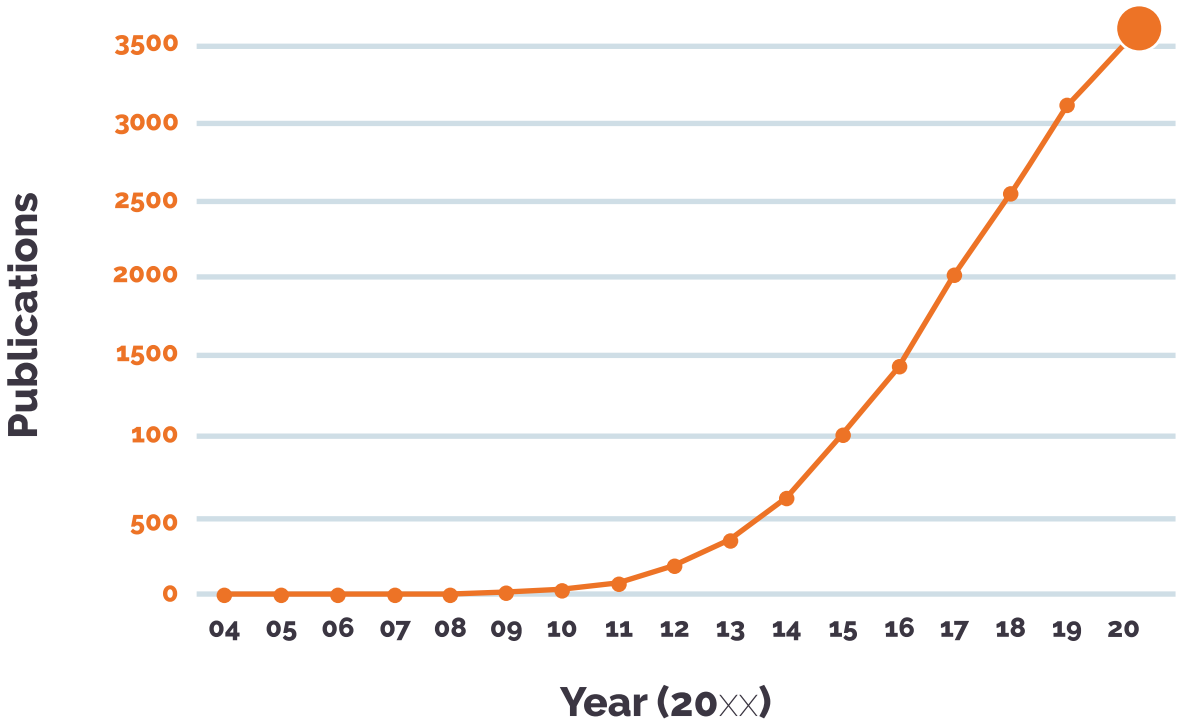
Since 2016, the IMPC has published several high impact seminal papers delivering pan-genomic insight into the relationship between genes and function (see IMPC reference list, **Annex 1** *pg. 12*). All of the papers demonstrate and underline a key notable outcome; *identification across multiple disease areas of an extensive, unexplored landscape of novel genes and potential mechanisms in areas as diverse as development* (Dickinson et al. 2016), metabolism (Rozman et al. 2018), deafness (Bowl et al. 2017), vision (Moore et al. 2018), and bone pathologies (Swan et al 2020).

Click to go to **Annex 1**

IMPC Impact

The IMPC has had transformative impact across biomedical discovery research, medical genetics and assisting rare disease diagnosis, enabling preclinical studies, and freely providing data and biological resources that are being used in drug target discovery (Oprea et al. 2018, Oprea 2019; and **Figure 1.** *pg. 5*) and the development of therapeutics and clinical trials that will bring target treatments to market.

Click to go to Fig 1.



1. By the end of 2020, there were >3,600 publications by the global research community that used IMPC mutant lines, technology or technical resources, or data (Phenotyping Consortium, 2020).

2. Medical genetics' pursuit of the genetic basis of disease increased dramatically over the last 10 years owing to the sharply reduced costs of genome sequencing. But thousands of variants are identified. Which one causes the disease? IMPC mutant lines have been used to point to the answer by providing candidate gene mutants whose function has or can be tested and used to unravel the molecular mechanisms of genetic disease. In the case of rare disease, when a novel rare disease gene is first discovered, there is minimal insight into its biological function, the pathogenic mechanisms of how disease-causing variants cause a complex phenotype, and how therapy might be approached (see for example **Figure 2.** *pg. 8*). An example is IMPC mutant lines that have helped confirm variant pathogenicity for the NIH funded Centres for Mendelian Genomics and Undiagnosed Diseases Network.

Click to go to Fig 2.

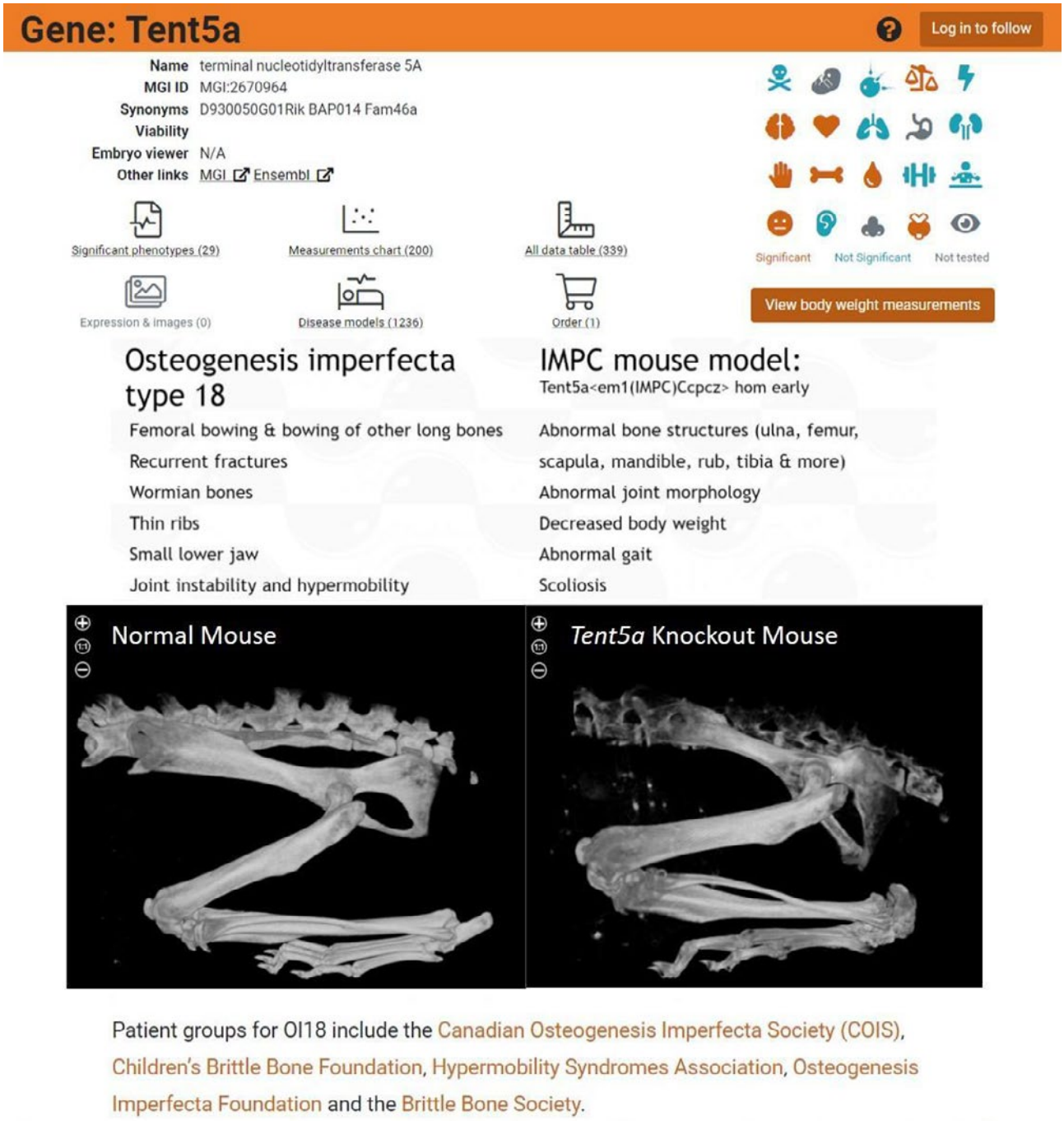


Figure 2. Osteogenesis Imperfecta Type XVIII (OI18) is caused by autosomal recessive mutations in human *TENT5A*. The number of individuals with OI Type 18 in the US or worldwide is unknown. The overall prevalence of OI (types I-XIX) is estimated at 0.5 per 10,000 individuals. Approximately 20,000-50,000 individuals in the United States have OI. Osteogenesis imperfecta type XVIII (OI18) patients have congenital bowing of the long bones, wormian bones, blue sclerae, vertebral collapse, and multiple fractures in the first years of life. The *Tent5a*$^{em1(IMPC)Ccpcz}$ knockout produced and phenotyped by the IMPC has multiple phenotypes including abnormal bone structure, joint morphology, scoliosis (curved spine) and abnormal gait (mobility), recapitulating many of the phenotypes in OI patients.

3. Enabling preclinical studies

Disease gene discovery and diagnosis is just the beginning of the quest to improve the lives of those living with genetic diseases. Mouse models can be essential in the preclinical studies required to validate potential therapeutic effect and safety. An example is the novel *Kctd13* IMPC mutant. Data from the IMPC's pipeline identified a cognitive phenotype associated with humans with autism spectrum disorder and intellectual disability. Additional purpose-driven research by experts in this disease area identified the Rho-kinase inhibitor and vasodilator Fasudil hydrochloride (HA1077), approved in 1995 for treating cerebral blood clots and to improve cognitive decline in stroke patients, and demonstrated that it restored learning and memory loss in *Kctd13* mice (Lorenzo et al. 2021). A repurposing patent has been filed, and these IMPC mice are being used in the preclinical studies en route to a potential new treatment for autism spectrum disorder and intellectual disability.

4. Provided data and biological resources being used in drug target discovery

IMPC mutant mouse lines have also been used to identify therapeutic targets for a known disease gene(s). Wolfram syndrome is an inherited rare disease typically associated with childhood-onset insulin-dependent diabetes mellitus and progressive optic atrophy. Wolfram syndrome is also considered a prototype of human endoplasmic reticulum (ER) disease. Despite identification of two causative genes, Wolfram syndrome 1 (WFS1) and Wolfram syndrome 2 (WFS2), discovery of druggable targets to prevent the death of CNS neurons and insulin-producing pancreatic β-cells remained elusive. A research team in St. Louis used the IMPC's *Cisd2* mutant to first implicate the intracellular enzyme calpain 2 in the mechanism of cell death in Wolfram syndrome and demonstrate that calpain and the pathway leading its activation provides potential therapeutic targets for Wolfram syndrome and other ER diseases (Lu et al. 2014).

5. Provided data and biological resources being used in therapeutic development

Currently, there are 18 clinical trials in the US that included an IMPC publication in their approval submission to the FDA; testing therapeutic or diagnostic candidates for cancer, metabolic disease, childhood obesity and diabetes, kidney disease, retinal disease, colitis, and tuberculosis. An example is the observational study initiated in October 2019 to assess urinary dickkopf-3 (DKK3) as a biomarker of the course of progressive chronic kidney disease, specifically which patients are likely to experience life-threatening loss of glomerular filtration rate. Essential proof-of-principle studies using two mouse models of renal fibrosis, the *Dkk3* IMPC mutant and commercially available *Rag2* mice showed that genetic as well as antibody-mediated neutralization of DKK3 led to reduced tubular atrophy and decreased interstitial matrix accumulation (Federico et al. 2016).

Summary Statement

The IMPC multidimensional production and phenotype dataset is available at mousephenotype.org and mouse mutants are freely available from IMPC centres around the world for the cost of shipping and handling. Critical to this endeavour has been the focus on quality assurance of processes and quality control of products to deliver gold-standard mouse and data resources to biomedical scientists, clinician researchers, and the biopharmaceutical industry.

We recognize that as biomedical research and therapeutic discovery and development evolves, so must the IMPC's strategy. The Consortium will revisit, re-prioritize, and adjust its strategy but always with a focus to deliver new knowledge, relevant and publicly available mouse mutants and data to ensure the highest impact across biomedical science, clinical genetics, genomic medicine, and therapeutics that will improve healthcare.

For more information on the IMPC visit

Cited References

Oprea TI, et al., Zahoránszky-Köhalmi G." Unexplored therapeutic opportunities in the human genome." *Nat Rev Drug Discov* 2018 May;17(5):317-332. doi: 10.1038/nrd.2018.14. PMID: 29472638.

Meehan TF, et al., Smedley D. "Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium." *Nat Genet.* 2017 Aug;49(8):1231-1238. doi: 10.1038/ng.3901. PMID: 8650483.

Cacheiro P, et al., Brown SDM. "New models for human disease from the international mouse phenotyping consortium." *Mamm Genome* 2019 Jun;30(5-6):143-150. doi: 10.1007/s00335-019-09804-5. PMID: 31127358.

Oprea TI. *Mamm Genome* 2019 Aug;30(7-8):192-200. doi: 10.1007/s00335-019-09809-0. PMID: 31270560.

Dickinson ME, et al., Murray SA. "High-throughput discovery of novel developmental phenotypes." *Nature* 2016 Sep 22;537(7621):508-514. doi: 10.1038/nature19356. PMID: 27626380.

Cacheiro P, et al., Smedley D; Genomics England Research Consortium; International Mouse Phenotyping Consortium. "Human and mouse essentiality screens as a resource for disease gene discovery." *Nat Commun* 2020 Jan 31;11(1):655. doi: 10.1038/s41467-020-14284-2. PMID: 32005800.

Karp NA, et al., White JK. "Prevalence of sexual dimorphism in mammalian phenotypic traits." *Nat Commun* 2017 Jun 26;8:15475. doi: 10.1038/ncomms15475. PMID: 28650954.

Rozman J, et al., Hrabe de Angelis M. "Identification of genetic elements in metabolism by high-throughput mouse phenotyping." *Nat Commun* 2018 Jan 18;9(1):288. doi: 10.1038/s41467-017-01995-2. PMID: 29348434.

Bowl MR, et al., Brown SDM. "A large-scale hearing loss screen reveals an extensive, unexplored genetic landscape for auditory dysfunction." *Nat Commun* 2017 Oct 12;8(1):886. doi: 10.1038/s41467-017-00595-4. PMID: 29026089.

Moore BA, et al., Moshiri A. "Identification of novel genes required for eye development by high-throughput screening of mouse knockouts." *Nature Commun Biol* 2018 Dec 21;1:236. doi: 10.1038/s42003-018-0226-0. eCollection 2018. PMID: 30588515.

Swan AL, et al., Hrabe de Angelis M. "Mouse mutant phenotyping at scale reveals novel genes controlling bone mineral density." *PLoS Genet* 2020 Dec 28;16(12):e1009190. doi: 10.1371/journal.pgen.1009190. eCollection 2020 Dec. PMID: 33370286.

Phenotyping Consortium. (2020, December 17). Publications with IMPC Alleles: International Mouse Phenotyping Consortium. Retrieved March 10, 2021, from mousephenotype.org/data/publications.

Lorenzo SM, et al., Herault Y. "Targeting the RHOA pathway improves learning and memory in adult Kctd13 and 16p11.2 deletion mouse models." *Mol Autism* 2021 Jan 13;12(1):1. doi: 10.1186/s13229-020-00405-7. PMID: 33436060.

Lu S, et al., Urano F. "A calcium-dependent protease as a potential therapeutic target for Wolfram syndrome." *Proc Natl Acad Sci U S A* 2014 Dec 9;111(49):E5292-301. doi: 10.1073/pnas.1421055111. PMID: 25422446.

Federico G, et al., Gröne H-J. "Tubular Dickkopf-3 promotes the development of renal atrophy and fibrosis." *JCI Insight.* 2016 Jan 21;1(1):e84916. doi: 10.1172/jci.insight.84916. PMID: 27699213.

Annex 1. Select papers and References from the IMPC

2016: DICKINSON et al. (2016) "High-throughput discovery of novel developmental phenotypes." *Nature.* 2016 Sep 22;537(7621):508-514. doi: 10.1038/nature19356. PMID: 27626380.

2017: KARP et al. (2017) "Prevalence of sexual dimorphism in mammalian phenotypic traits." *Nat Commun.* 2017 Jun 26;8:15475. doi: 10.1038/ncomms15475. PMID: 28650954.

2017: MEEHAN et al. (2017) "Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium." *Nat Genet.* 2017 Aug;49(8):1231-1238. doi: 10.1038/ng.3901. PMID: 28650483.

2017: BOWL et al. (2017) "A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction." *Nat Commun.* 2017 Oct 12;8(1):886. doi: 10.1038/s41467-017-00595-4. PMID: 29026089.

2018: ROZMAN et al. (2018) "Identification of novel genetic elements in metabolism by high-throughput mouse phenotyping." *Nat Commun.* 2018 Jan 18;9(1):288. doi: 10.1038/s41467-017-01995-2. PMID: 29348434.

2018: BROWN et al. (2018) "High-throughput mouse phenomics for characterising mammalian gene function." *Nat Rev Genet.* 2018 Jun;19(6):357-370. doi: 10.1038/s41576-018-0005-2. PMID: .29626206.

2018: NUTTER et al. (2018) Response to "Unexpected mutations after CRISPR-Cas9 editing in vivo". *Nat Methods.* 2018 Apr;15(4):235-236. doi: 10.1038/nmeth.4559. PMID: 29600991.

2018: MUNOZ-FUENTES et al. (2018) "The International Mouse Phenotyping Consortium (IMPC): A functional catalogue of the mammalian genome that informs conservation." *Conserv Genet.* 2018;19(4):995-1005. doi: 10.1007/s10592-018-1072-9. PMID: 30100824.

2018: MOORE et al. (2018) "Identification of genes required for eye development by high-throughput screening of mouse knockouts." *Commun Biol.* 2018 Dec 21;1:236. doi: 10.1038/s42003-018-0226-0. eCollection 2018. PMID: 30588515.

2019: BARUPAL et al. (2019) "Comprehensive Plasma Metabolomics Dataset for a Cohort of Mouse Knockouts within the International Mouse Phenotyping Consortium." *Metabolites.* 2019 May 22;9(5):101. doi: 10.3390/metabo9050101. PMID: 31121816.

2019: HASELIMASHHADI et al. (2019) "Soft windowing application to improve analysis of High-throughput phenotyping data." *Bioinformatics.* 2020 Mar 1;36(5):1492-1500. doi: 10.1093/bioinformatics/btz744. PMID: 31591642.

2019: CACHEIRO et al. (2019) "New models for human disease from the International Mouse Phenotyping Consortium." *Mamm Genome.* 2019 Jun;30(5-6):143-150. doi: 10.1007/s00335-019-09804-5. PMID: 31127358.

2020: ZHANG et al. (2020) "High-throughput discovery of genetic elements of circadian misalignment." *PLoS Genet.* 2020 Jan 13;16(1):e1008577. doi: 10.1371/journal.pgen.1008577. PMID: 31929527.

2020: CACHEIRO et al. (2020) "Human and mouse essentiality screens as a resource for disease gene discovery." *Nat Commun.* 2020 Jan 31;11(1):655. doi: 10.1038/s41467-020-14284-2. PMID: 32005800.

2020: LLOYD et al. (2020) "The Deep Genome Project." *Genome Biol.* 2020 Feb 3;21(1):18. doi: 10.1186/s13059-020-1931-9. PMID: 32008577.

2020: SWAN et al. (2020) "Mouse mutant phenotyping at scale reveals novel genes controlling bone mineral density." *PLoS Genet.* 2020 Dec 28;16(12):e1009190. doi: 10.1371/journal.pgen.1009190. PMID: 33370286.

2020: Wotton et al. (2020) "Machine learning-based automated phenotyping of inflammatory nocifensive behavior in mice." *Mol Pain.* Jan-Dec 2020;16:1744806920958596. doi: 10.1177/1744806920958596. PMID: 32955381.

2020: ZHANG et al. (2020) "Cage-lid hanging behavior as a translationally relevant measure of pain in mice." *Pain.* 2020 Oct 11. doi: 10.1097/j.pain.0000000000002127. Online ahead of print. PMID: 33230005.

2021: BIRLING et al. (2021) "A resource of targeted mutant mouse lines for 5,061 genes." *Nat Genet.* 53, 416–419 (2021). https://doi.org/10.1038/s41588-021-00825-y