




# Pleiotropy data resource as a primer for investigating co-morbidities/multi-morbidities and their role in disease

Violeta Muñoz-Fuentes<sup>1</sup> · Hamed Haselimashhadi<sup>1</sup> · Luis Santos<sup>2</sup> · Henrik Westerberg<sup>2</sup> · Helen Parkinson<sup>1</sup> · Jeremy Mason<sup>1</sup> 

Received: 6 July 2021 / Accepted: 6 September 2021 / Published online: 15 September 2021  
© The Author(s) 2021

## Abstract

Most current biomedical and protein research focuses only on a small proportion of genes, which results in a lost opportunity to identify new gene-disease associations and explore new opportunities for therapeutic intervention. The International Mouse Phenotyping Consortium (IMPC) focuses on elucidating gene function at scale for poorly characterized and/or under-studied genes. A key component of the IMPC initiative is the implementation of a broad phenotyping pipeline, which is facilitating the discovery of pleiotropy. Characterizing pleiotropy is essential to identify gene-disease associations, and it is of particular importance when elucidating the genetic causes of syndromic disorders. Here we show how the IMPC is effectively uncovering pleiotropy and how the new mouse models and gene function hypotheses generated by the IMPC are increasing our understanding of the mammalian genome, forming the basis of new research and identifying new gene-disease associations.

## Introduction

We still do not have a deep understanding of how mammalian genomes function. Most current genetic research focuses only on a small proportion of genes, many of which have already been studied extensively. This leaves a large amount of the genome under-studied, which results in a lost opportunity to identify disease-associated genes and explore new opportunities for therapeutic intervention (Edwards et al. 2011; Oprea et al. 2018; Stoeger et al. 2018).

The International Mouse Phenotyping Consortium (IMPC) is one of a few programmes focusing on obscure or otherwise poorly characterized and/or under-studied genes to uncover novel aspects of gene function. In turn, these newly available mouse models and gene function hypotheses stemming from the IMPC are forming the basis of new research. Notably, the IMPC broad-based phenotyping pipeline is facilitating the discovery of pleiotropy. Pleiotropy can be

defined as the effect of a gene or a variant on multiple traits and plays a central role in biology (Stearns 2010; Wang et al. 2010; Paaby and Rockman 2013) with far-reaching implications for evolution, development, canalization, ageing and disease (Wagner and Zhang 2011). Despite pleiotropy's central importance in biology, empirical datasets have not been available until recently. The analysis of these datasets is contributing to the understanding of pleiotropy, such as its genomic pattern, evolutionary implications, the evaluation of mathematical models and their theoretical predictions and the verification of hypotheses (Wang et al. 2010, Wagner and Zhang 2011, Saltz et al. 2017; Archambeault et al. 2020; Geiler-Samerotte et al. 2020; Shikov et al. 2020), as well as to identify new gene-disease associations.

In this Commentary, we focus on the relevance of pleiotropy for understanding gene function and, in particular, its critical importance to identify gene-disease associations. Here, we define the term trait as the phenotypes that are deviant in the knockout lines produced by the IMPC in comparison to wild type mice (a significant parameter in IMPC terminology), which allow us to link gene to function. We show how pleiotropy is distributed in the IMPC knockout lines and present compelling examples illustrating how this data is instrumental in making disease associations.

✉ Violeta Muñoz-Fuentes  
vmunoz@ebi.ac.uk

✉ Jeremy Mason  
jmason@ebi.ac.uk

<sup>1</sup> European Molecular Biology Laboratory-European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK

<sup>2</sup> MRC Harwell Institute, Harwell OX11 0RD, UK

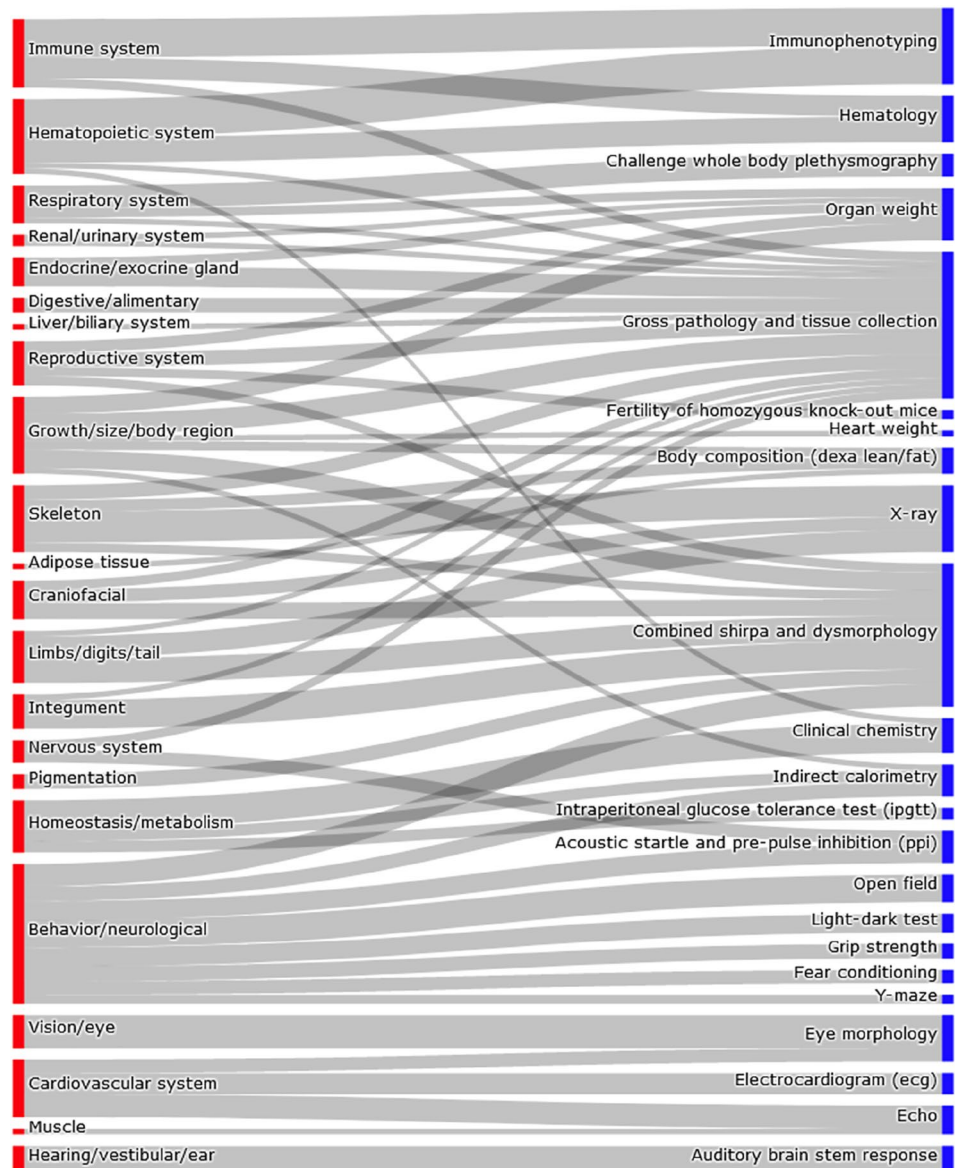
## Hypothesis-free, broad phenotyping pipeline enabling systematic analysis of pleiotropy

The IMPC is applying hypothesis-free systematic phenotyping screens to characterise single-gene knockout mice. The IMPC phenotyping pipeline is broad, encompassing many physiological systems and parameters (Fig. 1). Phenotypes in the knockout mouse line differing from the controls (wild type) identify the physiological systems that are disrupted when a gene is disabled, thus generating large, multidimensional gene-phenotype associations that are unveiling gene function. Effectively, the null allele can be associated with biochemical, physiological and developmental phenotypic changes. By investigating the impact of a modified gene on

multiple organs and physiological systems simultaneously, the IMPC pipeline is well suited to uncover pleiotropy (Brown et al. 2018; Brown and Lad 2019).

The IMPC is ultimately aiming to discover gene-disease associations. Indeed, the phenotypic tests applied by the IMPC have been selected and extended to enable the discovery of phenotypes that will facilitate modelling human disorders (Brown et al. 2018). In disease, pleiotropy manifests as multi-morbidities. In some cases, variants in one gene can lead to apparently unrelated phenotypes (Cerrone et al. 2019). Thus, characterizing pleiotropy is important in the identification of the genetic causes of any disease and, in particular, in the case of syndromic disorders. Some examples using IMPC data are provided below (see section “Associating Genes with Disease” for more detail on this).

**Fig. 1** Correspondence between physiological systems and IMPC procedures. The thickness of the lines corresponds to the number of IMPC parameters associated with the physiological system (note that the same parameter can be counted more than once when it is associated with more than one physiological system; see Supp Table 1 for details)



In addition to the characterization of adult young mice, the IMPC has two additional pipelines, conducted at the embryonic developmental stage and on aged mice. The assessment of morphology at embryonic stages allows us to characterize early-in-life lethal mutations, with 3D imaging and micro-CT data revealing phenotypes that would have been otherwise missed by gross inspection (Dickinson et al. 2016). Thus, the IMPC pipeline can detect malformations such as spina bifida, abnormal curvature in the spinal column, discoloration, kidney or spleen defects, hypoplastic lungs or thin myocardium, among others (Dickinson et al. 2016). The ageing pipeline investigates phenotypes appearing late in life. The IMPC phenotyping strategy has so far been instrumental in identifying phenotypes for genes of unknown function and provided the first hints for extensive pleiotropy (White et al. 2013; Hrabe de Angelis et al. 2015). It also provides the means to reveal the genetic networks intervening across diverse systems (Brown and Lad 2019).

Here, we investigate the pleiotropic effects of the genes for which the IMPC has produced and phenotyped knockout lines. Generally, the IMPC phenotypes 7 males and 7 females to characterize a mouse line (a cohort of mice deriving from the same gene modification event). An average of 153 parameters were collected per individual mouse. Parameters that show a significant deviation in the mutant cohort from the wildtype mice inform us about the physiological systems that are disrupted when the gene is disabled (Fig. 1). For example, the parameters in the IMPC Echocardiogram procedure (IMPC\_ECHO; described in detail here <https://www.mousephenotype.org/impress/ProcedureInfo?procID=654>) assess the functionality of the heart, thus informing us about the cardiovascular system (Fig. 1). Some IMPC parameters are associated with more than one physiological system (e.g., tibia length is a parameter of the X-ray procedure that is used to assess the limbs as well as the skeleton) and more than one parameter may inform about (sometimes different aspects of) the same physiological system (e.g. the ECHO and ECG electrocardiogram procedures about the cardiovascular system). Supp Table 1 shows a complete list of parameters and the physiological systems they inform.

The IMPC utilises a set of statistical methods (Haselimashhadi et al. 2020) as well as manual curation to determine the deviation of the knockouts from the wildtypes. Significant deviations (at the level of 0.0001) are annotated using Mammalian Phenotype Ontology (MPO) terms (Smith and Eppig 2015). All tests and procedures are described in detail in IMPReSS, the International Mouse Phenotyping Resource of Standardised Screens (<https://www.mousephenotype.org/impress>). Registering significant results to ontology terms has two main advantages. First, the systematic annotation of well-characterized phenotype descriptions ensures no loss of information and enables us to treat the results informatically (Deans et al. 2015; Thessen et al. 2020). And,

secondly, ontologies enable the possibility of systematically implementing cross-species comparisons. To elucidate gene-disease associations, mouse phenotypes expressed using MPO terms are mapped to those of patients whose clinical descriptions are annotated using the Human Phenotype Ontology (HPO; Robinson et al. 2008). More detail on this is provided below.

## Detecting pleiotropy

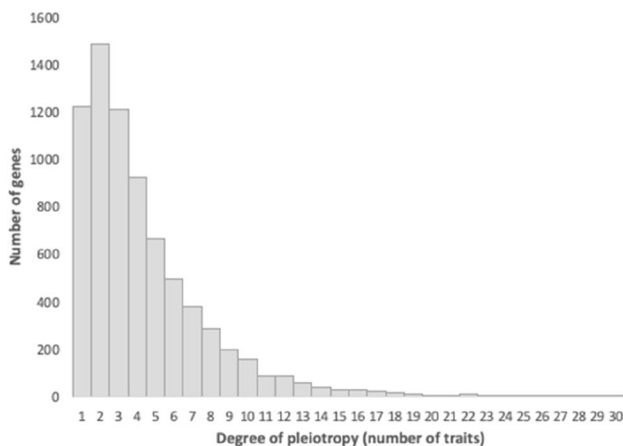
Mouse lines for a total of 7590 genes (Supp Table 2) have undergone or are currently undergoing phenotyping by the IMPC (DR 14.0, 7 May 2021, [www.mousephenotype.org](http://www.mousephenotype.org)). The IMPC focuses on phenotyping null homozygotes, that is, both alleles have been knocked out or made inactive. Thus, the default strategy is to apply the main IMPC phenotyping pipeline to null homozygotes. However, when the null homozygotes result in a lethal phenotype, the main phenotyping pipeline is applied to the heterozygotes and the null homozygotes are further investigated using the embryo pipeline as described above. Here we report pleiotropy in homozygous and heterozygous mouse lines.

Among the 7590 genes, we distinguish two special cases of genes, those for which the null homozygotes result in a lethal phenotype (lack of homozygous viable individuals) and those genes for which the individuals cannot be associated with any trait (no detectable effect in the knockout line). Essential genes can be argued to be maximally pleiotropic from a developmental or selectional point of view, even if molecularly they only affect one or a few gene's activities (Paaby and Rockman 2013). In the IMPC data, 24% of the genes present a lethal phenotype (absence of viable homozygotes or complete penetrance of preweaning lethality), while 9% present a variable degree of subviability (incomplete penetrance of preweaning lethality in null homozygotes; it is worth noting that heterozygotes are almost exclusively viable with only two lines presenting a degree of subviability). The IMPC has focused papers dealing with essential genes (Dickinson et al. 2016; Muñoz-Fuentes et al. 2018, Cacheiro et al. 2020), and we refer the interested reader to those publications. In the case of the other special gene class, identifying genes affecting zero traits requires considering a minimum number of phenotypic screens conducted on the mutant line in order to be able to conclude that the effect of the null allele is likely to be negligible (albeit in this particular background and experimental setting). We established this minimum threshold at 13 IMPC procedures. In the IMPC dataset, there are 5146 genes that fulfil this criterion and, out of them, 457 genes have no trait association. Genes with no trait association have been suggested to be an interesting class with undefined pleiotropy (Paaby and Rockman 2013).

The wealth of IMPC procedures applied to single-gene knockouts for 7590 genes resulted in 1491 genes associated with one trait and 4862 genes (77%) associated with two or more traits. These results indicate that pleiotropy is widespread, but most genes in the genome are not highly pleiotropic and a few affect a great number of traits (Fig. 2).

For the genes affecting two or more traits, 4157 genes showed effects spanning across at least two physiological systems (as defined in the IMPC pipeline, Fig. 1, in turn dictated by the Mammal Phenotype Ontology, Smith and Eppig 2015); it is worth noting that some of these systems are highly related, such as the hematopoietic and the immune systems and certain traits may be mapping to more than one system. Indeed, for the genes affecting one trait, 139 genes had effects that spanned across two or three systems, due to the mapping of those traits to more than one physiological system (Fig. 1).

We explored this relationship further using a normalized co-occurrence matrix which shows the degree of association of any two physiological systems based on the count of genes that affect traits associated with those two systems. We normalized using the geometric mean to give an idea of the strength of the association based on the underlying data. We show this for early adults (homozygote and heterozygote lines) and embryos (homozygotes not surviving to preweaning stage; Fig. 3a–c). Generally, the homozygotes show higher degree of pleiotropy, a pattern that can be expected as the complete inactivation of the gene normally leads to more abnormal phenotypes than a partial inactivation. Perhaps not so surprisingly, genes affecting the hematopoietic system also affect the immune system (Fig. 3a, b), while other, potentially more interesting relationships may include genes affecting the skeleton also having an effect on adipose



**Fig. 2** Frequency distribution of the degree of gene pleiotropy showing extensive pleiotropy for 7590 genes studied by the IMPC; 4862 (77%) of the genes affect two or more phenotypes or traits. Degree of pleiotropy here is equivalent to the number of traits

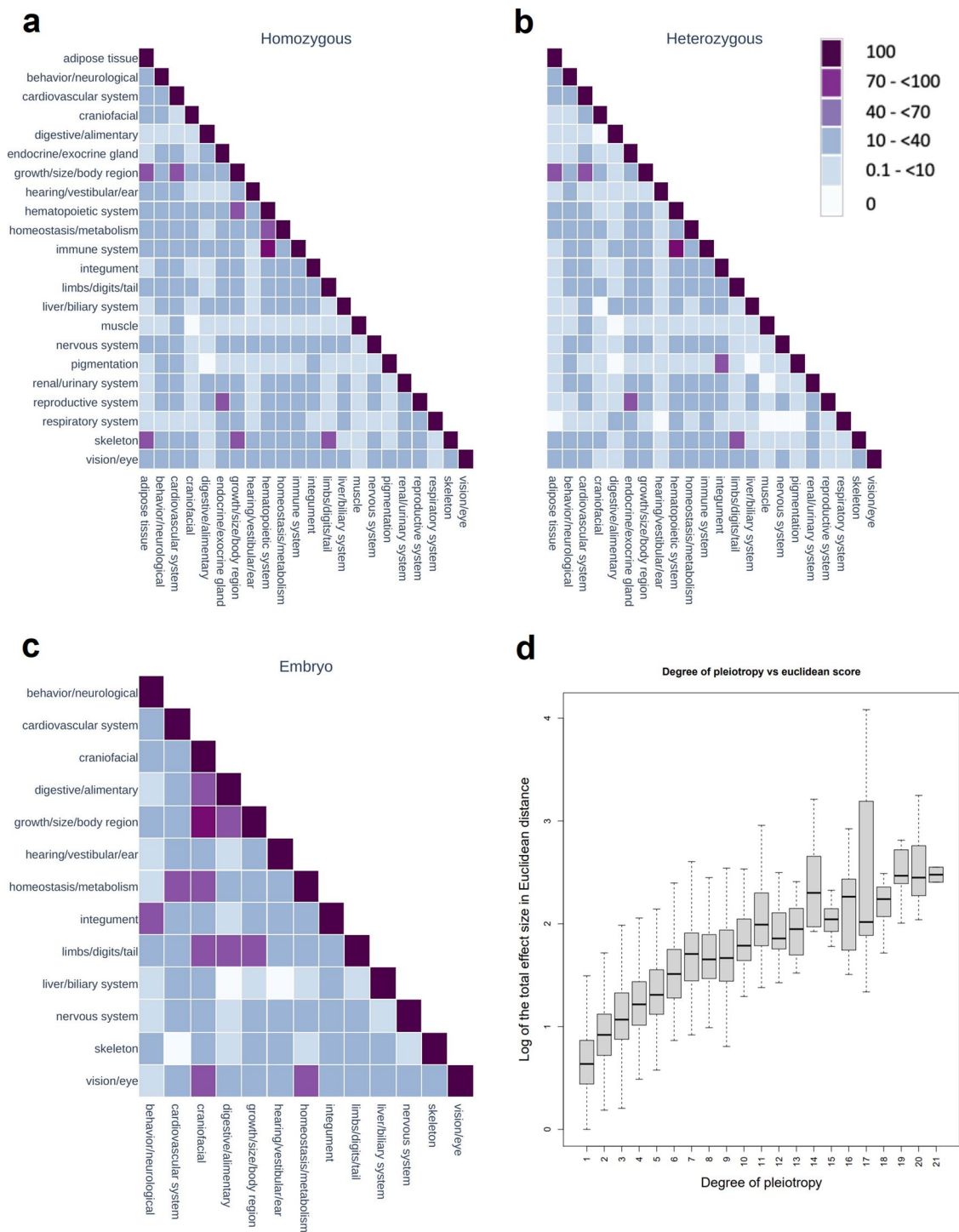
tissue, or genes affecting the digestive system not normally having an effect on pigmentation phenotypes (Fig. 3a, b).

These associations require careful consideration of the underlying traits. Evaluating these co-occurrence matrices in the context of the complete list of parameters to associated phenotypic systems (Supp Table 1) is necessary to get an overall picture of how pleiotropy is represented in the IMPC dataset. By design, a single term in the MPO may inherit semantic meaning from multiple parents at each level of the ontology, thus providing the potential for multiple paths from a single highly-specific term to multiple broadly descriptive terms, e.g., a specific term, “abnormal leukocyte physiology,” relates to both “hematopoietic system phenotype” and “immune system phenotype”. For example, many genes associated with the cardiovascular system are also growth/size phenotypes due to annotation with “enlarged heart” and “increased heart weight” phenotypes (Fig. 3a, b), and the overlap of genes associated with craniofacial and digestive systems in embryos is driven by the identification of a “cleft palate” phenotype (Fig. 3c). In contrast, the association between the skeleton and adipose tissue (Fig. 3a) is not driven by phenotypes that annotate to both systems (there are none, Supp Table 1; e.g. *Prdm14*, *Ube4a*, *Arl10*, *Dnd1*). While phenotypes of the respiratory systems, muscles or digestive system have few associations with phenotypes in other systems, this could be a consequence of the way we measure e.g., number of traits investigated to obtain a full characterization of that system (Supp Table 1). Thus, further investigation is needed in this space.

A widely cited relationship has been previously described in which a larger per-trait effect size is observed for genes affecting more traits, which has implications for our understanding of the evolvability and the adaptation of organisms (Wagner et al. 2008; Wang et al. 2010). We also investigated this relationship between the degree of pleiotropy and the effect size for each gene (Fig. 3d). To do this, we used a subset of the genes for which we had continuous measurements (Supp Table 3) to calculate the Euclidean distance from the individual gene-trait associations. The effect size is calculated using the Cohen method (Haselimashadi et al. 2020). Our results based on the extensive IMPC data set are consistent with previous observations.

## Associating genes with disease: mouse strains that model the patients’ phenotypes

Detecting co-morbidities certainly plays a role in the association of genes with disease in humans and the IMPC broad-phenotyping pipeline attempts to capture the collection of phenotypes that may be manifested when a gene is disabled. Key to this endeavour is matching the comprehensive collection of mouse phenotypes associated with



**Fig. 3** Pleiotropy across physiological systems and in relation to the effect size of the gene on the trait. Co-occurrence matrix of genes by physiological system for viable homozygotes (a) and heterozygotes (b), assessed with the early adult phenotyping pipeline, and lethal

homozygotes (c), assessed with the embryo pipeline. Relationship between the total effect size of gene-trait associations in Euclidean distance versus the degree of pleiotropy (d)

a gene with the patients' phenotypes. To identify models of human diseases, including rare diseases, the IMPC applies a translational pipeline which uses the PhenoDigm

algorithm, developed by the Monarch Initiative (Smedley et al. 2013), the human phenotype ontology (HPO) annotations associated with a disease, maintained by the Monarch

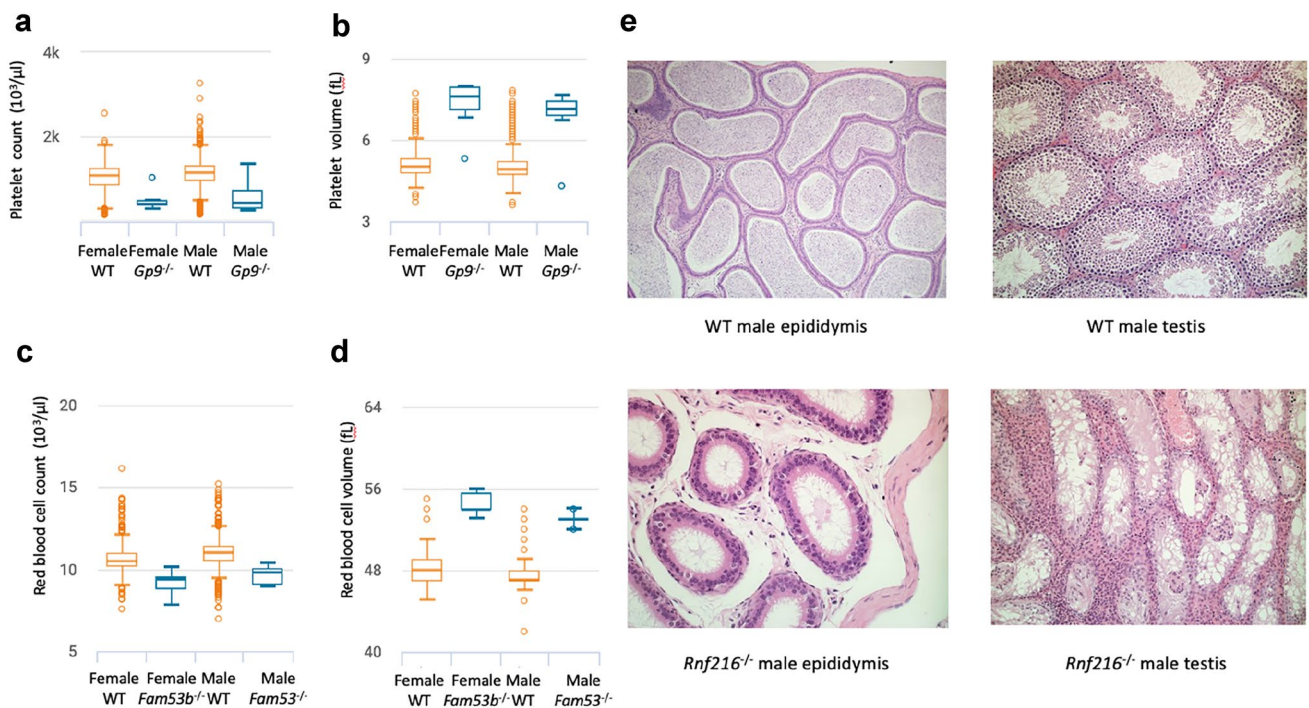
Initiative (Köhler et al. 2021; Mungall et al. 2017) and the Mammalian Phenotype Ontology (MPO) annotations from both the Mouse Genome Database (Bult et al. 2019) and the IMPC. A score, between 0 and 100, provides a quantitative measure of the phenotypic similarity between a particular mouse strain and the clinical description of a disease. The results of the pipeline are integrated in the IMPC website. Mouse strains, by modelling the patients' phenotypes, help to establish gene-disease associations.

This strategy was exemplified in a study focusing on the phenotypes of IMPC knockouts for 3328 genes (Meehan et al. 2017). Some examples from this study included the null homozygotes for *Gp9*, which recapitulated key features of the Bernard-Soulier syndrome, a bleeding disorder, or *Rnf216* homozygous-null mice which, characterized by hypogonadism and cerebellar ataxia, were associated with the Gordon-Holmes syndrome; likewise, *Fam53b* was proposed as a new model for anemia (Fig. 4). In another study, a gene essentiality classification using IMPC (mouse) and human screens was presented as a criterion to prioritise gene candidates for developmental disorders (Cacheiro et al. 2020). The authors presented evidence for the association of *VPS4A* and *TMEM63B* with two unsolved cases associated with intellectual disabilities in humans. The observed phenotypes were relevant for

establishing the link between the clinical manifestation of the disease and these genes. The mouse knockouts of the orthologous *Vps4a* gene presented lens opacity and brain abnormalities, and the mouse knockouts of the orthologous *Tmem63b* gene abnormal behaviour, hyperactivity and limb-grasping phenotypes, all consistent with the human patients' phenotypes.

## Conclusion

The IMPC, with its broad-phenotyping pipeline, is building a pleiotropy-data resource. Data deriving from mouse models suggests that pleiotropy is potentially more widespread than initially anticipated, and that the majority of genes affect only a few traits, and a few affect many traits. We show that the frequency distribution of pleiotropy or its relationship with gene effect sizes based on the wealth of data collected by the IMPC is consistent with previous findings, based on data from mice and other organisms (Wagner et al. 2008; Albert et al. 2008; Wang et al. 2010; White et al. 2013; Hrabe de Angelis et al. 2015; Cerrone et al. 2019). Thus, IMPC data is well suited to contribute to the understanding of the impact of pleiotropy on adaptive evolution and the evolvability of complex organisms, the understanding of the



**Fig. 4** Selected traits of mouse models for Mendelian disease identified using the IMPC translational pipeline. *Gp9*<sup>tm1.1(KOMP)Vlcs</sup> homozygotes have decreased platelet number (a) and increased platelet volume (b). *Fam53b*<sup>tm1b(EUCOMM)Hmgu</sup> homozygotes have decreased red blood cell number (c) and increased red blood cell volume (d).

*Rnf216*<sup>tm1b(EUCOMM)Wtsi</sup> homozygotes are infertile; histopathology images show epididymal aspermia and testicular degeneration and atrophy and lack of spermatogenesis in the mutants, while the wild types are unaffected (e)

genotype–phenotype maps and the validation of mathematical models and their predictions (Wagner et al. 2008, Wang et al. 2010, Wagner and Zhang 2011).

Notably, the IMPC constitutes a data resource for investigating co-expressed abnormal phenotypes, including those of poorly characterized genes, which makes this functional catalogue indispensable in detecting co-morbidities and identifying genes associated with disease. By combining these data with other resources, such as the Mouse Genome Informatics (MGI) database (<https://www.informatics.jax.org>) or human essentiality screens, as shown above, the power to identify genes associated with disease can be augmented. Furthermore, the embryo pipeline allows investigation of developmental malformations early in life, which is indispensable in order to understand the effects of lethal genes. Finally, the IMPC, with its specific pipeline applied to aged mice, will facilitate the study of the link between gene and phenotypes appearing later in life, and the implications for disease associations.

Such a resource does not come without limitations. Because of its high-throughput nature, the IMPC pipeline focuses on phenotyping the minimum number of animals necessary to detect strong phenotypic effects (Hrabe de Angelis et al. 2015) and characterizes only protein coding single-gene knockout effects. The phenotypic screens applied by the IMPC have been selected to facilitate modelling human disorders; however, some key biological functions are not fully covered or not investigated (e.g. respiration, inflammation), in some cases only in centre-specific pipelines or pilot studies (e.g. immunophenotyping, <https://www.mousephenotype.org/impress/PipelineInfo?id=15>; pain pilot study, Wotton et al. 2020; <https://www.mousephenotype.org/understand/data-collections/pain/>). The goal of the IMPC is to ascribe altered physiological function in the presence of a null mutation, hence additional challenges are intentionally minimised to increase the signal of the mutated genotype. Detecting pleiotropy in the context of gene by environmental interaction would be highly insightful, but falls outside the scope of the IMPC. By working together with the medical and scientific communities this resource may be expanded in the future to produce and characterize more complex genetic modifications of potentially clinical relevance.

The IMPC, through its website [www.mousephenotype.org](http://www.mousephenotype.org), facilitates access to data and mouse models as they become available (3–4 data releases a year). The IMPC portal not only displays phenotype data and available mouse lines, but also provides access to sophisticated tools to visualize phenotypes encompassing quantitative, categorical and image data. IMPC mice, produced using either embryonic-stem-cell or CRISPR-Cas9 technology, are available for future research. Thus, the IMPC paves the way to increased knowledge about under-studied or poorly characterized

genes, as well as to new experiments aiming at increasing our understanding of gene function, disease mechanisms or new therapies.

The IMPC, through its website [www.mousephenotype.org](http://www.mousephenotype.org), facilitates access to data and mouse models as they become available (3–4 data releases a year). The IMPC portal not only displays phenotype data and available mouse lines, but also provides access to sophisticated tools to visualize phenotypes encompassing quantitative, categorical and image data. IMPC mice, produced using either embryonic-stem-cell or CRISPR-Cas9 technology, are available for future research. Thus, the IMPC paves the way to increased knowledge about under-studied or poorly characterized genes, as well as to new experiments aiming at increasing our understanding of gene function, disease mechanisms or new therapies.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00335-021-09917-w>.

**Acknowledgements** Research reported in this publication was supported by the European Molecular Biology Laboratory core funding and the National Human Genome Research Institute of the National Institutes of Health under Award Number UM1HG006370. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Albert AY, Sawaya S, Vines TH, Knecht AK, Miller CT, Summers BR, Balabhadra S, Kingsley DM, Schluter D (2008) The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution* 62:76–85
- Archambeault SL, Bärtschi LR, Merminod AD, Peichel CL (2020) Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback *Eda* locus. *Evol Lett.* <https://doi.org/10.1002/evl3.175>

- Brown SDM, Lad HV (2019) The dark genome and pleiotropy: challenges for precision medicine. *Mamm Genome* 30:212–216
- Brown SDM, Holmes CC, Mallon A-M, Meehan TF, Smedley D, Wells S (2018) High-throughput mouse phenomics for characterizing mammalian gene function. *Nat Rev Genet* 19:357–370
- Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, The Mouse Genome Database Group (2019) Mouse genome database (MGD) 2019. *Nucleic Acids Res* 47(D1):D801–D806
- Cacheiro P, Muñoz-Fuentes V, Murray SA et al (2020) Human and mouse essentiality screens as a resource for disease gene discovery. *Nat Commun* 11:655. <https://doi.org/10.1038/s41467-020-14284-2>
- Cerrone M, Remme CA, Tadros R, Bezzina CR, Delmar M (2019) Beyond the one gene-one disease paradigm. *Circulation* 140:595–610
- Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP et al (2015) Finding our way through phenotypes. *PLoS Biol* 13(1):e1002033. <https://doi.org/10.1371/journal.pbio.1002033>
- Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK et al (2016) High-throughput discovery of novel developmental phenotypes. *Nature* 537(7621):508–514. <https://doi.org/10.1038/nature19356>
- Edwards AM, Isserlin R, Bader GD, Frye SV, Willson TM, Yu FH (2011) Too many roads not taken. *Nature* 470:163–165
- Geiler-Samerotte KA, Li S, Lazaris C, Taylor A, Ziv N, Ramjeawan C, Paaby AB, Siegal ML (2020) Extent and context dependence of pleiotropy revealed by high-throughput single-cell phenotyping. *PLoS Biol*. <https://doi.org/10.1371/journal.pbio.3000836>
- Haselimahhadi H, Mason JC, Mallon A-M, Smedley D, Meehan TF, Parkinson H. (2020). OpenStats: a robust and scalable software package for reproducible analysis of high-throughput phenotypic data. *bioRxiv*:2020.2005.2013.091157.
- Hrabe De Angelis M, Nicholson G, Selloum M et al (2015) Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. *Nat Genet* 47:969–978. <https://doi.org/10.1038/ng.3360>
- Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM et al (2021) The human phenotype ontology in 2021. *Nucleic Acids Res* 49:D1207–D1217
- Meehan TF, Conte N, West DB, Jacobsen JO, Mason J, Warren J, Chen C-K, Tudose I, Relac M, Matthews P et al (2017) Disease model discovery from 3,328 gene knockouts by the international mouse phenotyping consortium. *Nat Genet* 49:1231–1238
- Mungall CJ, McMurry JA, Köhler S, Balhoff JP, Borromeo C, Brush M, Carbon S, Conlin T, Dunn N, Engelstad M et al (2017) The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res* 45:D712–D722
- Muñoz-Fuentes V, Cacheiro P, Meehan TF, Aguilar-Pimentel JA, Brown SD, Flenniken AM et al (2018) The International Mouse Phenotyping Consortium (IMPC): a functional catalogue of the mammalian genome that informs conservation. *Conserv Genet* 19:995–1005. <https://doi.org/10.1007/s10592-018-1072-9>
- Oprea TI, Bologa CG, Brunak S et al (2018) Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov* 17:317–332
- Paaby AB, Rockman MV (2013) The many faces of pleiotropy. *Trends in Genet* 29:66–73
- Robinson PN, Kohler S, Bauer S, Seelow D, Horn D, Mundlos S (2008) The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. <https://doi.org/10.1016/j.ajhg.2008.09.017>
- Saltz JB, Hessel FC, Kelly MW (2017) Trait correlations in the genomics era. *Trends Ecol Evol* 32:279–290
- Shikov AE, Skitchenko RK, Predeus AV, Barbitoff YA (2020) Phenome-wide functional dissection of pleiotropic effects highlights key molecular pathways for human complex traits. *Sci Rep*. <https://doi.org/10.1038/s41598-020-58040-4>
- Smedley D, Oellrich A, Köhler S et al (2013) PhenoDigm: analyzing curated annotations to associate animal models with human diseases. *Database* 2013:bat025. <https://doi.org/10.1093/database/bat025>
- Smith CL, Eppig JT (2015) Expanding the mammalian phenotype ontology to support automated exchange of high throughput mouse phenotyping data generated by large-scale mouse knockout screens. *J Biomed Semant*. <https://doi.org/10.1186/s13326-015-0009-1>
- Stearns FW (2010) One hundred years of pleiotropy: a retrospective. *Genetics* 186:767
- Stoeger T, Gerlach M, Morimoto RI, Nunes Amaral LA (2018) Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biol* 16:e2006643
- Thessen AE, Walls RL, Vogt L, Singer J, Warren R, Buttigieg PL et al (2020) Transforming the study of organisms: phenomic data models and knowledge bases. *PLoS Comput Biol* 16(11):e1008376. <https://doi.org/10.1371/journal.pcbi.1008376>
- Wagner GP, Kenney-Hunt JP, Pavlicev M, Peck JR, Waxman D, Cheverud JM (2008) Pleiotropic scaling of gene effects and the ‘cost of complexity.’ *Nature* 452:470–472
- Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat Rev Genet* 12(3):204–213. <https://doi.org/10.1038/nrg2949>
- Wang Z, Liao B-Y, Zhang J (2010) Genomic patterns of pleiotropy and the evolution of complexity. *Proc Natl Acad Sci* 107(42):18034–18039. <https://doi.org/10.1073/pnas.1004666107>
- White JK, Gerdin A-K, Karp NA et al (2013) Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* 154:452–464. <https://doi.org/10.1016/j.cell.2013.06.022>
- Wotton JM, Peterson E, Anderson L, Murray SA, Braun RE, Chesler EJ, White JK, Kumar V (2020) Machine learning-based automated phenotyping of inflammatory nociceptive behavior in mice. *Mol Pain* 16:1744806920958596

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.