



The International Mouse Phenotyping Consortium (IMPC): a functional catalogue of the mammalian genome that informs conservation

Violeta Muñoz-Fuentes¹ · Pilar Cacheiro² · Terrence F. Meehan¹ · Juan Antonio Aguilar-Pimentel³ · Steve D. M. Brown⁴ · Ann M. Flenniken^{5,6} · Paul Flicek¹ · Antonella Galli⁷ · Hamed Haseli Mashhadi¹ · Martin Hrabě de Angelis^{3,8,9} · Jong Kyoung Kim¹⁰ · K. C. Kent Lloyd¹¹ · Colin McKerlie^{5,6,12} · Hugh Morgan⁴ · Stephen A. Murray¹³ · Lauryl M. J. Nutter^{5,12} · Patrick T. Reilly¹⁴ · John R. Seavitt¹⁵ · Je Kyung Seong¹⁶ · Michelle Simon⁴ · Hannah Wardle-Jones⁷ · Ann-Marie Mallon⁴ · Damian Smedley² · Helen E. Parkinson¹ · the IMPC consortium

Received: 5 December 2017 / Accepted: 3 May 2018 / Published online: 19 May 2018
© The Author(s) 2018

Abstract

The International Mouse Phenotyping Consortium (IMPC) is building a catalogue of mammalian gene function by producing and phenotyping a knockout mouse line for every protein-coding gene. To date, the IMPC has generated and characterised 5186 mutant lines. One-third of the lines have been found to be non-viable and over 300 new mouse models of human disease have been identified thus far. While current bioinformatics efforts are focused on translating results to better understand human disease processes, IMPC data also aids understanding genetic function and processes in other species. Here we show, using gorilla genomic data, how genes essential to development in mice can be used to help assess the potentially deleterious impact of gene variants in other species. This type of analyses could be used to select optimal breeders in endangered species to maintain or increase fitness and avoid variants associated to impaired-health phenotypes or loss-of-function mutations in genes of critical importance. We also show, using selected examples from various mammal species, how IMPC data can aid in the identification of candidate genes for studying a condition of interest, deliver information about the mechanisms involved, or support predictions for the function of genes that may play a role in adaptation. With genotyping costs decreasing and the continued improvements of bioinformatics tools, the analyses we demonstrate can be routinely applied.

Keywords Cheetah · Endangered species · Loss-of-function · Non-model species · Panda · Polar bear · Phenotype · Wolf · Essential genes · IMPC · Knockout · Mouse

The IMPC: a functional catalogue of the mammalian genome

The goal of the International Mouse Phenotyping Consortium (IMPC, <http://www.mousephenotype.org>) is to generate a functional catalogue of the mammalian genome by producing a knockout mouse line for every protein-coding

gene. This is achieved by characterising the phenotypes of mutants and controls, which increases our understanding of development and gene function, and identifies models for disease. Knockout mouse lines are produced on a uniform genetic background using either gene targeted embryonic stem cells (Skarnes et al. 2011) or, increasingly, nuclease-mediated genome editing with CRISPR/Cas9-based methods (Singh et al. 2015; Mianne et al. 2017). A uniform genetic background across controls and mutant lines is necessary to allow for reproducible and comparable results. Some phenotypes will be strongly influenced by the genetic background and, therefore, this is an important consideration to take into account, particularly when translating mouse findings (inbred) to other species (outbred, or mostly outbred; see Discussion). Mice are characterised across a dozen research centres in a standardized phenotyping pipeline (IMPRESS,

Violeta Muñoz-Fuentes and Pilar Cacheiro shared first authorship.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10592-018-1072-9>) contains supplementary material, which is available to authorized users.

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

Extended author information available on the last page of the article

the International Mouse Phenotyping Resource of Standardised Screens) that includes strict data quality standards and requires the minimum number of animals necessary to achieve statistical significance for each test (Hrabe de Angelis et al. 2015). The IMPC data is integrated and reviewed, and statistically significant outlier phenotypes for individual lines are annotated using PhenStat (Kurbatova et al. 2015) and the Mammalian Phenotype Ontology (MPO) (Beck et al. 2009; Smith and Eppig 2015), which is actively developed to capture phenotypes of mutant mouse lines by Mouse Genome Informatics (MGI) based at Jackson Laboratory. All raw data, results of statistical pipelines and curated phenotype data are made publicly available through the IMPC website. The data are further integrated with other resources, including OMIM, MGI and Ensembl. The IMPC database is searchable by gene name, phenotype and disease, allows batch queries and the download of all data, dedicated reports, graphs and images.

To date, 5186 mutant lines have been phenotyped (data release 7.0), with an average of 163 parameters measured on any given mouse, represented by over 128,000 knockouts and 35,000 wildtype or control mice. In addition, embryonic lethal mouse lines are analysed in a specialized embryonic development pipeline that utilizes high-resolution 3D imaging to understand structural changes (Dickinson et al. 2016). These data allow the IMPC to identify the physiological systems that are disrupted when a gene is disabled and make new gene-phenotype associations. Evolutionary conservation of fundamental processes governing development and support of metazoan life allows functional knowledge gained in one species to be translated to others (Kirschner and Gerhart 2006; Liao et al. 2006; Saenko et al. 2008; Bellen et al. 2010; Greek and Rice 2012). The IMPC uses its new gene-phenotype associations to identify models for human disease based on phenotypic similarity scores using PhenoDigm (Smedley et al. 2013), which establishes a link between IMPC mouse phenotypes mapped to the Mammalian Phenotype Ontology and the clinical descriptions of human diseases, as featured in OMIM and Orphanet, mapped to terms of the Human Phenotype Ontology (Kohler et al. 2017). Based on data from 3,328 genes, 360 new disease models have so far been identified by the IMPC, allowing researchers to investigate molecular mechanisms underpinning human genetic diseases, and explore new routes of therapeutic intervention (Meehan et al. 2017). While the IMPC has focused on translating knowledge from mouse to human, the translation to other species, including wild and endangered, is relevant as well.

Wild species may benefit from functional knowledge accumulated in the laboratory mouse

Endangered species typically suffer dramatic declines before remedial measures are put into place. During a species decline, genetic erosion results in the loss of genetic variation that limits a species' ability to adapt to changes in the environment and increases the chances for the accumulation of deleterious mutations that affect reproduction and fitness. Fertility-related disorders have been documented in the African cheetah, *Acinonyx jubatus* (Wildt et al. 1983; Crosier et al. 2007), the Florida panther, *Puma concolor coryi* (Roelke et al. 1993; Johnson et al. 2010) and the Iberian lynx, *Lynx pardinus* (Ruiz-Lopez et al. 2012). Similarly, bone and dental anomalies have been observed in inbred wolf (*Canis lupus*) populations in Isle Royale in North America and Scandinavia (Raikonen et al. 2009, 2013). In an attempt to reverse these situations and decrease inbreeding, breeding with closely related species has been implemented in the case of the Florida panther and the puma *Puma concolor stanleyana* (Johnson et al. 2010). These genetic rescue approaches need to be carefully considered, as they may cause increased inbreeding as well as loss of species-specific adaptations (Hedrick and Fredrickson 2010), and even forfeiture of legal protected status, e.g., Endangered Species Act (Haig and Allendorf 2006). Clearly, identifying the critical genes associated with disorders as well as species-specific adaptations is important from a conservation perspective to maximise conservation of adaptive potential and, if needed, preserve genetic fitness through selective breeding.

The genomes of many mammals have been sequenced in the last 15 years. We selected a number of mammalian species for which functional adaptations have been explored and illustrate how knockout mouse phenotype information can support or complement predictions for the function of genes that may play a role in adaptation, provide a panel of genes for studying a phenotype of interest, or aid deciphering the mechanisms involved in underlying certain conditions.

Essential genes in mice and humans: mining wildlife genomes for LoF gene variants to identify basis of reduced fitness—a pilot study

A previous analysis of IMPC's high-throughput mammalian embryonic phenotype data for 1751 knockout mouse lines resulted in 24, 11 and 65% of the lines being associated to a lethal, subviable and viable phenotype, respectively; this led to the conclusion that, in mice, approximately 35% of the genes are essential for organism viability (Dickinson

et al. 2016). We hypothesize that these genes are essential in other mammalian species, and variants causing loss of function in these critical genes might, therefore, be undesirable. To test this hypothesis, we first compared genes identified as essential in IMPC mice with those identified in humans based on cell viability. We then used these essential genes to gain further insight into loss-of-function (LoF) variants (protein-coding genes containing substitutions that introduce a stop codon, frameshift indels, or modifications of essential splice sites) identified in inbred populations of gorillas.

The IMPC viability screen identifies genes essential for organism viability by identifying mouse lines which are lethal (absence of homozygote pups for the knockout allele or homozygote null pups), subviable (the frequency of homozygote null pups is less than 12.5%, or less than 50% of the 25% predicted in a heterozygote × heterozygote crossing) or viable (all others). We conducted an updated analysis on viability data for 4237 genes currently available in IMPC DR7.0, which included the 1751 previously analysed in Dickinson et al. (2016) (see Supplementary Methods). We found that 25, 9, and 66% of the lines resulted in a lethal, subviable and viable phenotype, respectively (Supplementary Datafile S1), nearly identical proportions to those reported in Dickinson et al. (2016) (see above). These results support the conclusion that about one-third of the genes are essential for life, as described in an earlier publication surveying the knockout mouse literature (Adams et al. 2013).

Screens of knockout human cells have identified ~2000 genes essential for cell viability in studies of 11 cell lines

Table 1 Overlap of mouse IMPC lethal and viable genes (DR7.0) and human cell essential and non-essential genes

Overlaps		Number of genes
Mouse	Human cell lines ^a	
Lethal	Essential	353 (35.9%)
Lethal	Non-essential	631 (64.1%)
Viable	Essential	9 (0.4%)
Viable	Non-essential	2499 (99.6%)

^aEssential: genes essential for cell viability in > 50% of the cell lines and studied in > 50% of the cell lines (that is, equivalent to ≥ 6 cell lines)

(Blomen et al. 2015; Hart et al. 2015; Wang et al. 2015). Combining these data sets, 18,862 genes were unequivocally mapped to their HUGO Gene Nomenclature Committee (HGNC) identifiers, of which 17,675 were studied in > 50% of the cell lines (at least 6 cell lines). We defined a set of core essential genes comprising 1568 genes (9%) which were essential for viability in over 50% of the cell lines where the gene was studied. To understand how gene essentiality compares between human cells and mice, we inferred mouse-to-human orthologues and looked at their distribution in the IMPC and human-viability categories. We obtained a dataset containing 4115 IMPC mouse-to-human orthologues (see Supplementary Methods), of which 4026 were included in the human cell studies (Supplementary Datafile S2). We found that 36% of the mouse genes identified as embryonic lethal (i.e., essential) corresponded to genes identified as essential in the human cell lines, while 64% corresponded to genes that are non-essential in cells. In the case of genes identified as embryonic viable in mice (i.e., non-essential), almost all (99.6%) were associated with non-essentiality in the human cell lines (Table 1). These results indicate a strong correspondence between non-essential genes and that about two-thirds as many genes are essential for organismal than for cell viability.

We then investigated the critical importance of LoF variants in gorillas *Gorilla gorilla*, western Africa, and *G. beringei*, eastern Africa; Xue et al. (2015). Notably, homozygous LoF alleles were found in 241 genes in apparently healthy individuals, and we determined which of these genes are identified as essential in mice or humans. We inferred gorilla-to-mouse orthologues (Supplementary Methods) and obtained a mouse orthologue for 169 out of the 241 gorilla genes, resulting in 192 mouse genes (due to one-to-many conversions, Datafile S3). Western lowland gorillas (*G. g. gorilla*) had 136 homozygous LoF orthologues, eastern lowland gorillas (*G. b. graueri*) had 81, and mountain gorillas (*G. b. beringei*) had 84. Overlap with the viability data obtained by the IMPC (reported above) indicated a distribution of the LoF alleles in the three viability categories similar to that obtained for any protein-coding gene in the IMPC catalogue (Table 2). The percentage of lethal genes in the gorilla populations was lower than in the IMPC

Table 2 Mouse orthologues with homozygous LoF alleles as identified by Xue et al. (2015) and their association to a lethal, subviable or viable phenotype based on viability data collected by the IMPC (DR7.0)

	Sample size	Lethal (<i>n</i> = 1052)	Subviable (<i>n</i> = 383)	Viable (<i>n</i> = 2802)
Mountain gorillas (Gbb)	7	5 (24%)	3 (14%)	13 (62%)
Eastern lowland gorillas (Gbg)	9	4 (19%)	3 (14%)	14 (67%)
Western lowland gorillas (Ggg)	27	5 (14%)	6 (16%)	26 (70%)
Total (unique)	43	9 (18%)	6 (12%)	36 (70%)

Sample size refers to the number of gorillas as indicated in the original publication

viability data (14–24% vs 25%), but the difference was not significant ($P=0.731$, $P=0.659$ and $P=0.130$ for mountain, eastern lowland and western lowland gorillas, respectively, Table S1).

We then proceeded to gain a better understanding of the potential phenotypic impact of the LoF mutations in gorillas. First, we obtained gorilla-to-human orthologues (168 human genes, Datafile S4) and assessed their essentiality using the data from the human cell studies (Fig. 1a). We found that between 5–7% of the genes were essential for cell survival, lower than what would be expected for any gene selected at random (9%), but the difference was not significant ($P=0.818$, $P=0.369$ and $P=0.736$ for mountain, eastern lowland and western lowland gorillas, respectively, Table S2). When bringing in IMPC and MGI phenotypes, the data become increasingly complex and more difficult to interpret. For the 191 mouse orthologues, there was phenotype information for 62% of the genes, and 34% of the genes were associated with an embryonic lethal phenotype (Fig. 1b). It is important to note that any given gene may not be associated with a lethal phenotype in all mouse lines, but be linked to a variety of health-impaired phenotypes that do not cause lethality in additional lines (Datafile S3). For example, homozygotes of *Chd2* investigated in three genetic backgrounds resulted in postnatal lethality in two of them and in viable individuals in the other, but with health-impaired phenotypes associated to growth, the skeleton and the hematopoietic system. These effects can be due to

potential differences in genomic modifiers between different strains used to generate the knockouts. Further, most knockouts, including the IMPC ones, are on inbred backgrounds, while wild species will be outbred, or at least more so than laboratory mouse strains. Based on these results, it is therefore possible that gorillas carrying these variants may present clinical complications that may impact their fitness and thus it will be desirable to reduce the prevalence of these alleles, particularly in a recovery population. Alternatively, these truncated variants identified in gorillas may not be affecting the functional exon of the protein or may correspond to genes redundant in function. Although viable lines are more likely to have a paralogue than lethal lines, there are nevertheless some essential genes with paralogues (White et al. 2013; Dickinson et al. 2016) and it is possible that these provide functional compensation for the effect of LoF variants in the inactivated genes. Further research is needed to clarify this situation, including advances to detect pseudogenes e.g. Claes and De Leeneer (2014). Humans carry LoF variants (MacArthur et al. 2012) at about ~100 putative variants per individual (The Genomes Project 2012) and the identification of both deleterious and beneficial variants has fuelled significant interest in these regions (Balasubramanian et al. 2017).

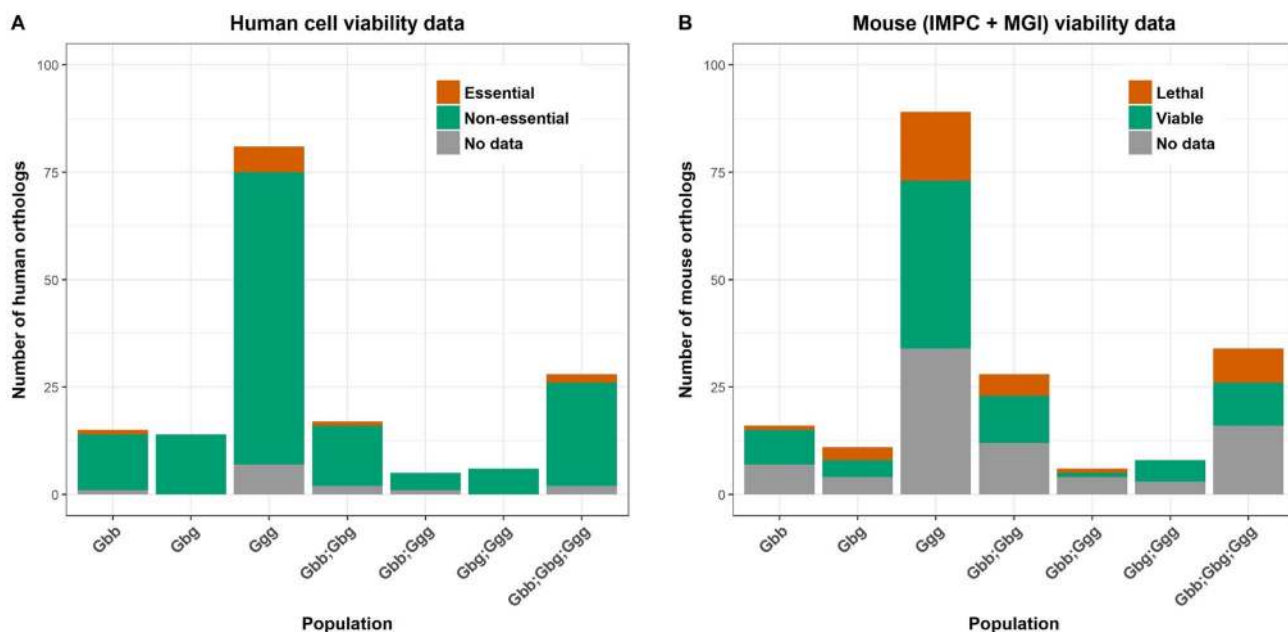


Fig. 1 Human (a) and mouse orthologues (b) of gorilla genes with homozygous LoF alleles and their association to essentiality based on human cell studies (a) or IMPC and MGI data (b). (Data in Supple-

mentary Table S3). Gorilla populations, from larger to smaller size in the wild: mountain gorillas (Gbb), eastern lowland gorillas (Gbg) and western lowland gorillas (Ggg)

IMPC aids the functional annotation of regions putatively targeted by positive selection to understand the genomic basis of adaptation

A number of recent studies in which mammalian gene function and adaptations are evaluated allow us to illustrate additional ways in which the IMPC may constitute a useful resource for mammals other than humans. The African cheetah (*Acinonyx jubatus*), a species with remarkably low levels of genome diversity relative to other mammals, exhibits signs of inbreeding depression in captive and free ranging populations, including low fecundity and malformed

spermatozoa (Dobrynin et al. 2015). An initial panel of 964 human genes with gene ontology (GO) terms associated to reproduction, yielded a set of 18 genes with accelerated rate of non-synonymous to synonymous substitution (dN/dS) accumulation in the cheetah lineage and damaging mutations previously associated to reproductive impairment (Dobrynin et al. 2015). We found a mouse orthologue for all genes, 5 with an IMPC significant phenotype (DR6.0, Datafile S5). Two had phenotypes associated with reproduction. One of them, *Rspo1*, was characterized with abnormal morphology in seminal vesicles and testes, small testes, *lacZ* expression in the *vas deferens* and the epididymis. In addition, this gene is associated with at least one infertility-related disease,

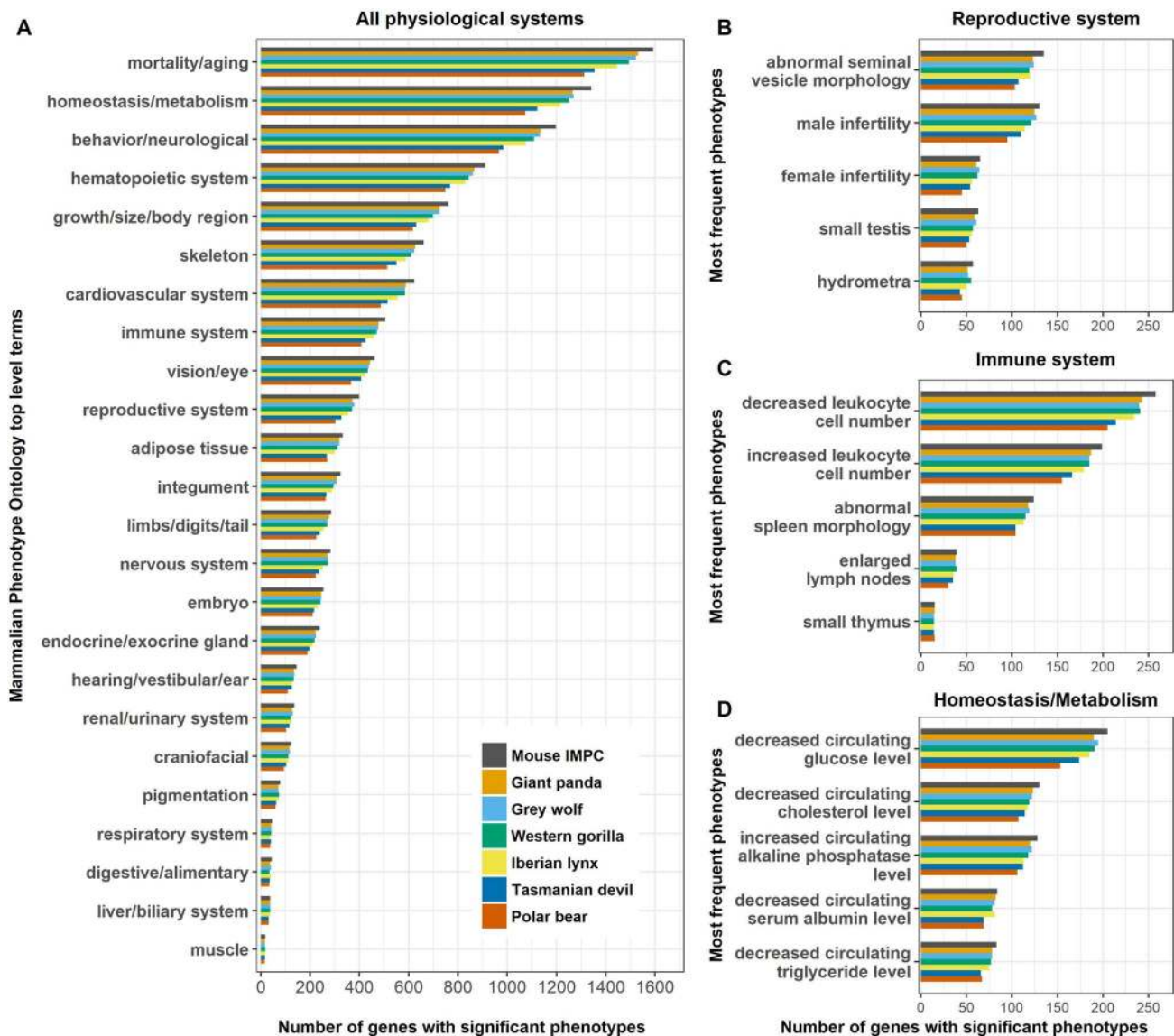


Fig. 2 Number of IMPC significant phenotypes for selected mammalian species. A mouse orthologue was found for 71–91% of the genes of each species, of which 24–25% had IMPC phenotype information

(DR6.0, Supplementary Table S4). **a** Phenotypes classified according to the top levels of the Mammalian Phenotype Ontology. **b–d** Phenotypes can be classified for more granular ontology terms

progesterone resistance (affecting females). MGI (<http://www.informatics.jax.org/>, accessed 17 November 2017) had phenotype information for 14 genes, of which 12 were related to the reproductive system. With only about ~25% of the protein-coding genes in the mouse genome explored, the IMPC currently contains around 400 mouse genes with phenotypes associated with the reproductive system (Fig. 2b), a potential useful resource to inform future studies on the genetic contributors to low fecundity.

In individuals belonging to the three subspecies of gorillas (*G. g. gorilla*, *G. b. graueri* and *G. b. beringei*) in the above-mentioned study, polymorphisms were identified in 23 genes corresponding to human disease-causing variants, significantly enriched for blood coagulation phenotypes, with 3 (*TNNT2*, *KCNE1*, *PKP2*) associated to cardiomyopathies (Xue et al. 2015). Indeed, cardiovascular disease is an important cause of death for gorillas in captivity (McManamon and Lowenstine 2012). Mouse orthologues were obtained for all except one gene, of which 5 have IMPC phenotype information (*F13A1*, *HNF4A*, *KLKB1*, *NPC1*, *SHH*), two associated to a cardiovascular phenotype, one to a skeleton phenotype and two to pre-weaning lethality in homozygotes, respectively (Datafile S6). The MGI database included 20 of these genes, of which 11 were associated to cardiovascular phenotypes and 5 to the hematopoietic system. In the IMPC, there are predictions for the association of 585 mouse orthologues of Western gorilla genes to cardiovascular phenotypes, and 844 to hematopoietic system phenotypes (Fig. 2a), potentially constituting an important resource for understanding cardiomyopathies in gorillas.

A study on speciation and adaptation in polar bears (*Ursus maritimus*) identified 20 genes as strong candidates to have been positively selected in polar bears, in what is a prime example of speciation through adaptation to an extreme environment (Liu et al. 2014). The authors reported disease associations in humans and other mammalian model organisms, including mice, suggesting a function for 11 of these genes associated to adipose tissue development and fatty acid metabolism (*APOB*), cardiovascular function (*APOB*, *ABCC6*, *ALPK3*, *ARID5B*, *CUL7*, *EHD3*, *TTN*, *VCL*, *XIRP1*) or white fur pigmentation (*LYST*, *AIM1*), which may be advantageous in the Arctic. Information derived from the IMPC and MGI databases (Datafile S7) supported these predictions and provided evidence for new roles. An exception was *AIM1* (currently *CRYBG1*). IMPC data indicates no association with coat colour or pigmentation. Homozygotes for *AIM1* were not viable, and the heterozygotes presented phenotypes associated with vision and the nervous system, but the fur was normal. In addition, there were phenotypic associations for 3 genes out of the 9 for which no function was reported in the paper. Knockouts for *COL5A3*, *LAMC3* and *SH3PXD28* presented a variety of phenotypes, including associations with adipose tissue,

the cardiovascular and the immune systems and homeostasis, which are functions that the authors indicated might be relevant in adapting to the Arctic.

A recent study on grey wolves (*Canis lupus*) from North America aimed to identify candidate genes under selection and environmentally driven functional variation (Schweizer et al. 2016a, b). In this study, nonsynonymous mutations were significantly correlated with environmental variables in genes associated to lipid metabolism (*APOB*, *LIPG*), immunity (*DLA-DQA*, *DLA-DRB1*), olfaction (*OR4S2*, *OR5B17*, *OR6B1*), vision and hearing (*PCDH15*, *USH2A*), and pigmentation (*TYR*, *TYRP1*), where 4 genes had variants with predicted deleterious impact *LIPG*, *OR4S2*, *OR5B17*, *USH2A* (Schweizer et al. 2016a). Information derived from the IMPC database for 6 of these genes and the MGI database for 23 of them reproduces previous findings and provides evidence for new roles (Datafile S8). *LIPG* is reported to be associated with the metabolic and cardiovascular systems, *USH2A* with the nervous system but also with vision and hearing, and no information is available for genes *OR4S2* and *OR5B17*, potentially related to olfaction.

We identified at least one study focusing on wild species, giant and red pandas *Ailuropoda melanoleuca* and *Ailuurus fulgens* (Hu et al. 2017) and two on cattle (Kadri et al. 2014; Biase et al. 2016) that have used the mouse knockout database information to further characterize genes or processes. The panda species are predicted to have independently acquired adaptations in 70 genes to a bamboo-rich diet, including a pseudthumb (limb development genes *DYNC2H1* and *PCNT*) and features related to digestion and nutrient utilization (in particular genes *GIF*, *CYP4F2*, *ADH1C* and *CYP3A5*). The IMPC database complements data collected by MGI by providing information for 5 additional genes (Datafile S9). In the two cattle studies, mouse knockout data informed about processes related to infertility (Kadri et al. 2014; Biase et al. 2016).

Outlook

Here we show how viability data collected by the IMPC are defining a set of essential genes that are likely also relevant in other species, particularly mammals. Identifying deleterious mutations is important for the design of captive breeding strategies (Bosse et al. 2015), and we encourage exploring the potential of the analyses presented here to identify critical functional variants. An assessment of human and mouse genes orthologous to gorilla genes containing homozygous LoF variants indicates that a number of them are strong candidates to compromise fitness and, therefore, further investigation of the phenotypes of gorillas with these variants will be required. The phenotypic effects of LoF or any other variants will manifest under certain genetic conditions

(genetic background) or environmental conditions. While a number of phenotypes in the mouse have shown to correlate directly with humans (e.g., Brophy et al. 2017; Santiago-Sim et al. 2017), these findings should, in general, be taken as indicative of directions for further investigations. Currently, mouse outbred stocks that are genetically heterogeneous and diverse, and thus more appropriately mimicking human or wild animal populations, are being used for mapping genes and quantitative trait loci (QTLs) (Winter et al. 2017). Additionally, the prediction of LoF variants is not straightforward and improved methods are under development.

We have shown that the IMPC, by elucidating mammalian gene function, provides experimental evidence to support novel or previously hypothesised relationships between gene function and processes, and aids in characterising hereditary diseases in mammalian species other than human. The IMPC is focusing on characterizing many of the poorly understood genes (the *ignorome*). It is also making relevant contributions to our understanding of mammalian gene function, in terms of sexual dimorphism (Karp et al. 2017; Rozman et al. 2018), pleiotropy (Brown et al. 2018) and disease (Bowl et al. 2017; Meehan et al. 2017; Perez-Garcia et al. 2018; Rozman et al. 2018). Formidable challenges remain ahead, including understanding, for example, incomplete penetrance, co-regulation of promoters or gene networks and the function of non-coding sequences, especially ultra-conserved non-coding regions that are more highly conserved across species than most protein-coding genes. Recognizing the effects of processes such as epistasis and hitchhiking of variants closely linked to selected genes in the wild species genomes pose a challenge. Another obvious challenge will be to determine gene function of wildlife phenotypes not present in human and mouse (e.g. aquatic phenotypes) and the co-opting of gene function for other biological processes via gene duplication.

The identification of critical functional variants can be of particular importance for endangered species or bottlenecked populations to aid attempts to reduce the incidence of genetically-determined traits that decrease fitness, or limit recovery. However, the benefits of a genetic rescue approach would require that the conditions that led to the accumulation of deleterious alleles are removed from the population. In the case of endangered species with low effective population sizes, the effect of genetic drift will be much greater than that of natural selection. Hence, even with an optimized breeding program in place, the potential gains of selecting critical functional variants in the breeders might be offset by the stochastic effects of genetic drift. When attempting to develop strategies to preserve adaptive variation, a design where breeding can be managed closely might be desired. For example, the establishment of a captive insurance meta-population for the Tasmanian devil aims at maximising genetic diversity and keeping a healthy stock of individuals

that can be used as a source population for re-wilding and genetic rescue (Gooley et al. 2017).

Advances in genotyping and whole-genome sequencing are resulting in an increase in the number of available genomes and transcriptomes, as well as improved methods to analyse these data, infer orthologous relationships and generate cross-species knowledge. In addition, integration of phenotype data is expected to become prominent in evolutionary studies. In order to produce databases that are computationally tractable and that allow for cross-species integrations, as well as to avoid loss of information, adhering to standards and persistent genetic identifiers (e.g., Ensembl, HGNC or MGI identifiers), as well as applying purpose-oriented ontologies, will be critical. In evolutionary biology and phylogenetic systematics, efforts to computationally integrate genetic, phenotypic and anatomical data include the ‘Phenotype And Trait Ontology’ (PATO; Mabee et al. 2007) and the Phenoscope project (Dahdul et al. 2010) but improvements in this area will certainly be needed (McMurry et al. 2017).

Animal models have proved useful to develop assisted reproductive technologies for endangered species, including lessons learned from oocyte and embryo culture in domestic animals and humans, and oncofertility techniques applied to humans (Comizzoli et al. 2010). Recently, cryopreservation of gametes was used to recover past genetic diversity in the black-footed ferret (*Mustela nigripes*; Wildt et al. 2016) and in vitro fertilization of frozen oocytes and spermatozoa is now the only way in which the northern white rhino (*Ceratotherium simum cottoni*) may be rescued (Saragusty et al. 2016). Studies on domestic mammals provide molecular markers that can be transferred for use in non-model species to inform about molecular processes with potentially phenotypic implications (Munoz-Fuentes et al. 2015). Moreover, understanding consequences of gene variants in other species may be of importance for human health and disease; for example, polar bears have evolved adaptations to deal with extremely fat-rich diets (see above), which are a major concern in human health. Currently, methods based on genomic data are being put forward to improve breeding strategies of wild species to attempt to minimize the impact of undesirable genetic variants while maintaining acceptable levels of genetic diversity (Bosse et al. 2015; Irizarry et al. 2016) and rapid advances in CRISPR/Cas9 technology in animal models to reduce the risk of off-target mutagenesis opens up opportunities to eliminate deleterious mutations in zygotes. In the case of wild species, such methods would allow the persistence of fitness-linked alleles and the avoidance of deleterious mutations without the risks associated with inbreeding or breeding between two similar species. The combined accumulation of gene function annotation by the IMPC and their advances in the use of CRISPR/Cas9 technology will be able to assist in future conservation efforts.

Acknowledgements This work was supported by the United States National Institutes of Health (NIH) Grants U54 HG006370, U42 OD011185, U54 HG006332, U54 HG006348, U54 HG006364, U42 OD011175 and UM1 OD023221, Government of Canada through Genome Canada and Ontario Genomics NorComm2 project (OGI-051), Korea Mouse Phenotyping Project (2017M3A9D5A01052447) of the Ministry of Science and ICT through the Korea National Research Foundation, and by the German Federal Ministry of Education and Research (INFRAFRONTIER grant 01KX1012).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.





References

- Adams D, Baldock R, Bhattacharya S, Copp AJ, Dickinson M, Greene NDE, Henkelman M, Justice M, Mohun T, Murray SA, Pauws E, Raess M, Rossant J, Weaver T, West D (2013) Bloomsbury report on mouse embryo phenotyping: recommendations from the IMPC workshop on embryonic lethal screening. *Dis Models Mech* 6:571
- Balasubramanian S, Fu Y, Pawashe M, McGillivray P, Jin M, Liu J, Karczewski KJ, MacArthur DG, Gerstein M (2017) Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nat Commun* 8:382
- Beck T, Morgan H, Blake A, Wells S, Hancock JM, Mallon AM (2009) Practical application of ontologies to annotate and analyse large scale raw mouse phenotype data. *BMC Bioinform* 10(Suppl 5):S2
- Bellen HJ, Tong C, Tsuda H (2010) 100 years of *Drosophila* research and its impact on vertebrate neuroscience: a history lesson for the future. *Nat Rev Neurosci* 11:514–522
- Biase FH, Rabel C, Guillomot M, Hue I, Andropolis K, Olmstead CA, Oliveira R, Wallace R, Le Bourhis D, Richard C, Champion E, Chaulot-Talmon A, Giraud-Delville C, Taghouti G, Jammes H, Renard JP, Sandra O, Lewin HA (2016) Massive dysregulation of genes involved in cell signaling and placental development in cloned cattle conceptus and maternal endometrium. *Proc Natl Acad Sci USA* 113:14492–14501
- Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Starling J, Sacco R, van Diemen FR, Olk N, Stukalov A, Marceau C, Janssen H, Carette JE, Bennett KL, Colinge J, Superti-Furga G, Brummelkamp TR (2015) Gene essentiality and synthetic lethality in haploid human cells. *Science* 350:1092–1096
- Bosse M, Megens HJ, Madsen O, Crooijmans RPMA., Ryder OA, Austerlitz F, Groenen MAM, de Cara MAR (2015) Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome Res* 25:970–981
- Bowl MR, Simon MM, Ingham NJ, Greenaway S, Santos L, Cater H, Taylor S, Mason J, Kurbatova N, Pearson S, Bower LR, Clary DA, Meziante H, Reilly P, Minowa O, Kelsey L, International Mouse Phenotyping C, Tocchini-Valentini GP, Gao X, Bradley A, Skarnes WC, Moore M, Beaudet AL, Justice MJ, Seavitt J, Dickinson ME, Wurst W, de Angelis MH, Herault Y, Wakana S, Nutter LMJ, Flenniken AM, McKelrie C, Murray SA, Svenson KL, Braun RE, West DB, Lloyd KCK, Adams DJ, White J, Karp N, Flicek P, Smedley D, Meehan TF, Parkinson HE, Teboul LM, Wells S, Steel KP, Mallon AM, Brown SDM (2017) A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. *Nat Commun* 8:886
- Brophy PD, Rasmussen M, Parida M, Bonde G, Darbro BW, Hong X, Clarke JC, Peterson KA, Denegre J, Schneider M, Sussman CR, Sunde L, Lildballe DL, Hertz JM, Cornell RA, Murray SA, Manak JR (2017) A gene implicated in activation of retinoic acid receptor targets is a novel renal agenesis gene in humans. *Genetics* 207:215
- Brown SDM, Holmes CC, Mallon AM, Meehan TF, Smedley D, Wells S (2018) High-throughput mouse phenomics for characterizing mammalian gene function. *Nat Rev Genet*. <https://doi.org/10.1038/s41576-018-0005-2>
- Claes KB, De Leeneer K (2014) Dealing with pseudogenes in molecular diagnostics in the next-generation sequencing era. *Methods Mol Biol* 1167:303–315
- Comizzoli P, Songsasen N, Wildt DE (2010) Protecting and extending fertility for females of wild and endangered mammals. *Cancer Treat Res* 156:87–100
- Crosier AE, Marker L, Howard J, Pukazhenthil BS, Henghali JN, Wildt DE (2007) Ejaculate traits in the Namibian cheetah (*Acinonyx jubatus*): influence of age, season and captivity. *Reprod Fertil Dev* 19:370–382
- Dahdul WM, Balhoff JP, Engeman J, Grande T, Hilton EJ, Kothari C, Lapp H, Lundberg JG, Midford PE, Vision TJ, Westerfield M, Mabee PM (2010) Evolutionary characters, phenotypes and ontologies: curating data from the systematic biology literature. *PLoS ONE* 5:e10708
- Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, Meehan TF, Weninger WJ, Westerberg H, Adissu H, Baker CN, Bower L, Brown JM, Caddle LB, Chiani F, Clary D, Cleak J, Daly MJ, Denegre JM, Doe B, Dolan ME, Edie SM, Fuchs H, Gailus-Durner V, Galli A, Gambadoro A, Gallegos J, Guo S, Horner NR, Hsu CW, Johnson SJ, Kalaga S, Keith LC, Lanoue L, Lawson TN, Lek M, Mark M, Marschall S, Mason J, McElwee ML, Newbigging S, Nutter LM, Peterson KA, Ramirez-Solis R, Rowland DJ, Ryder E, Samocha KE, Seavitt JR, Selloum M, Szoke-Kovacs Z, Tamura M, Trainor AG, Tudose I, Wakana S, Warren J, Wendling O, West DB, Wong L, Yoshiki A, International Mouse Phenotyping C, Jackson L, Infrastructure Nationale Phenomin ICdIS, Charles River L, Harwell MRC, Toronto Centre for P, Wellcome Trust Sanger I, MacArthur DG, Tocchini-Valentini GP, Gao X, Flicek P, Bradley A, Skarnes WC, Justice MJ, Parkinson HE, Moore M, Wells S, Braun RE, Svenson KL, de Angelis MH, Herault Y, Mohun T, Mallon AM, Henkelman RM, Brown SD, Adams DJ, Lloyd KC, McKelrie C, Beaudet AL, Bucan M, Murray SA (2016) High-throughput discovery of novel developmental phenotypes. *Nature* 537:508–514
- Dobrynin P, Liu S, Tamazian G, Xiong Z, Yurchenko AA, Krasheninnikova K, Kliver S, Schmidt-Kuntzel A, Koepfli KP, Johnson W, Kuderna LF, Garcia-Perez R, Manuel M, Godinez R, Komissarov A, Makunin A, Brukhin V, Qiu W, Zhou L, Li F, Yi J, Driscoll C, Antunes A, Oleksyk TK, Eizirik E, Perelman P, Roelke M, Wildt D, Diekhans M, Marques-Bonet T, Marker L, Bhak J, Wang J, Zhang G, O'Brien SJ (2015) Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol* 16:277
- Gooley R, Hogg CJ, Belov K, Grueber CE (2017) No evidence of inbreeding depression in a Tasmanian devil insurance population despite significant variation in inbreeding. *Sci Rep* 7(1):1830
- Greek R, Rice MJ (2012) Animal models and conserved processes. *Theor Biol Med Model* 9:40
- Haig SM, Allendorf FW (2006) Hybrids and policy. In: Scott JM, Goble DD, Davis FW (eds) *Conserving biodiversity in human dominated landscapes*. Island Press, Washington, pp 150–163
- Hart T, Chandrashekar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland OS, Durocher D, Angers S, Moffat J (2015) High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163:1515–1526
- Hedrick PW, Fredrickson R (2010) Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conserv Genet* 11:615–626

- Hrabe de Angelis M, Nicholson G, Selloum M, White JK, Morgan H, Ramirez-Solis R, Sorg T, Wells S, Fuchs H, Fray M, Adams DJ, Adams NC, Adler T, Aguilar-Pimentel A, Ali-Hadji D, Amann G, Andre P, Atkins S, Auburtin A, Ayadi A, Becker J, Becker L, Bedu E, Bekeredjian R, Birling M-C, Blake A, Bottomley J, Bowl MR, Brault V, Busch DH, Bussell JN, Calzada-Wack J, Cater H, Champy M-F, Charles P, Chevalier C, Chiani F, Codner GF, Combe R, Cox R, Dalloneau E, Dierich A, Di Fenza A, Doe B, Duchon A, Eickelberg O, Esapa CT, Fertak LE, Feigel T, Emelyanova I, Estabel J, Favor J, Flenniken A, Gambadoro A, Garrett L, Gates H, Gerdin A-K, Gkoutos G, Greenaway S, Glasl L, Goetz P, Da Cruz IG, Gotz A, Graw J, Guimond A, Hans W, Hicks G, Holter SM, Hofler H, Hancock JM, Hoehndorf R, Hough T, Houghton R, Hurt A, Ivandic B, Jacobs H, Jacquot S, Jones N, Karp NA, Katus HA, Kitchen S, Klein-Rodewald T, Klingenspor M, Klopstock T, Lalanne V, Leblanc S, Lengger C, le Marchand E, Ludwig T, Lux A, McKerlie C, Maier H, Mandel J-L, Marschall S, Mark M, Melvin DG, Meziane H, Micklich K, Mittelhauser C, Monassier L, Moulart D, Muller S, Naton B, Neff F, Nolan PM, Nutter LMJ, Ollert M, Pavlovic G, Pellegata NS, Peter E, Petit-Demouliere B, Pickard A, Podrini C, Potter P, Pouilly L, Puk O, Richardson D, Rousseau S, Quintanilla-Fend L, Quwailid MM, Racz I, Rathkolb B, Riet F, Rossant J, Roux M, Rozman J, Ryder E, Salisbury J, Santos L, Schable K-H, Schiller E, Schrewe A, Schulz H, Steinkamp R, Simon M, Stewart M, Stoger C, Stoger T, Sun M, Sunter D, Teboul L, Tilly I, Tocchini-Valentini GP, Tost M, Treise I, Vasseur L, Velot E, Vogt-Weisenhorn D, Wagner C, Walling A, Wattenhofer-Donze M, Weber B, Wendling O, Westerberg H, Willershauser M, Wolf E, Wolter A, Wood J, Wurst W, Yildirim AO, Zeh R, Zimmer A, Zimprich A, Consortium E, Holmes C, Steel KP, Herault Y, Gailus-Durner V, Mallon A-M, Brown SDM (2015) Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. *Nat Genet* 47:969–978
- Hu Y, Wu Q, Ma S, Ma T, Shan L, Wang X, Nie Y, Ning Z, Yan L, Xiu Y, Wei F (2017) Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proc Natl Acad Sci USA* 114:1081–1086
- Irizarry KJ, Bryant D, Kalish J, Eng C, Schmidt PL, Barrett G, Barr MC (2016) Integrating genomic data sets for knowledge discovery: an informed approach to management of captive endangered species. *Int J Genom*. <https://doi.org/10.1155/2016/2374610>
- Johnson WE, Onorato DP, Roelke ME, Land ED, Cunningham M, Belden RC, McBride R, Jansen D, Lotz M, Shindle D, Howard J, Wildt DE, Penfold LM, Hostetler JA, Oli MK, O'Brien SJ (2010) Genetic restoration of the Florida panther. *Science* 329:1641–1645
- Kadri NK, Sahana G, Charlier C, Iso-Touru T, Gulbrandsen B, Karim L, Nielsen US, Panitz F, Aamand GP, Schulman N, Georges M, Vilkki J, Lund MS, Druet T (2014) A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in nordic red cattle: additional evidence for the common occurrence of balancing selection in livestock. *Plos Genet* 10:e1004049
- Karp NA, Mason J, Beaudet AL, Benjamini Y, Bower L, Braun RE, Brown SDM, Chesler EJ, Dickinson ME, Flenniken AM, Fuchs H, Angelis MH, Gao X, Guo S, Greenaway S, Heller R, Herault Y, Justice MJ, Kurbatova N, Lelliott CJ, Lloyd KCK, Mallon AM, Mank JE, Masuya H, McKerlie C, Meehan TF, Mott RF, Murray SA, Parkinson H, Ramirez-Solis R, Santos L, Seavitt JR, Smedley D, Sorg T, Speak AO, Steel KP, Svenson KL, International Mouse Phenotyping C, Wakana S, West D, Wells S, Westerberg H, Yaacoby S, White JK (2017) Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat Commun* 8:15475
- Kirschner MW, Gerhart JC (2006) *The plausibility of life*. Yale University Press, New Haven
- Kohler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Ayme S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, Brudno M, Buske OJ, Chinnery PF, Cipriani V, Connell LE, Dawkins HJ, DeMare LE, Devereau AD, de Vries BB, Firth HV, Freson K, Greene D, Hamosh A, Helbig I, Hum C, Jahn JA, James R, Krause R, SJ FL, Lochmuller H, Lyon GJ, Ogishima S, Olry A, Ouwehand WH, Pontikos N, Rath A, Schaefer F, Scott RH, Segal M, Sergouniotis PI, Sever R, Smith CL, Straub V, Thompson R, Turner C, Turro E, Veltman MW, Vulliamy T, Yu J, von Ziegenweid J, Zankl A, Zuchner S, Zemojtel T, Jacobsen JO, Groza T, Smedley D, Mungall CJ, Haendel M, Robinson PN (2017) The human phenotype ontology in 2017. *Nucleic Acids Res* 45:D865–D876
- Kurbatova N, Mason JC, Morgan H, Meehan TF, Karp NA (2015) PhenStat: a tool kit for standardized analysis of high throughput phenotypic data. *PLOS ONE* 10:e0131274
- Liao TS, Call GB, Guptan P, Cespedes A, Marshall J, Yackle K, Owusu-Ansah E, Mandal S, Fang QA, Goodstein GL, Kim W, Banerjee U (2006) An efficient genetic screen in *Drosophila* to identify nuclear-encoded genes with mitochondrial function. *Genetics* 174:525–533
- Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C, Wray G, Li J, He W, Wang Z, Fu W, Xiang X, Morgan CC, Doherty A, O'Connell MJ, McInerney JO, Born EW, Dalen L, Dietz R, Orlando L, Sonne C, Zhang G, Nielsen R, Willerslev E, Wang J (2014) Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157:785–794
- Mabee PM, Ashburner M, Cronk Q, Gkoutos GV, Haendel M, Segerdell E, Mungall C, Westerfield M (2007) Phenotype ontologies: the bridge between genomics and evolution. *Trends Ecol Evol* 22:345–350
- MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, Albers CA, Zhang ZDD, Conrad DF, Lunter G, Zheng HC, Ayub Q, DePristo MA, Banks E, Hu M, Handsaker RE, Rosenfeld JA, Fromer M, Jin M, Mu XJ, Khurana E, Ye K, Kay M, Saunders GI, Suner MM, Hunt T, Barnes IHA, Amid C, Carvalho-Silva DR, Bignell AH, Snow C, Yngvadottir B, Bumpstead S, Cooper DN, Xue YL, Romero IG, Wang J, Li YR, Gibbs RA, McCarroll SA, Dermitzakis ET, Pritchard JK, Barrett JC, Harrow J, Hurles ME, Gerstein MB, Tyler-Smith C, Consortium GP (2012) A Systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335:823–828
- McManamon R, Lowenstein LJ (2012) Cardiovascular disease in great apes. *Zoo Wild Anim Med Curr Ther* 7:408–415
- McMurry JA, Juty N, Blomberg N, Burdett T, Conlin T, Conte N, Courtot M, Deck J, Dumontier M, Fellows DK, Gonzalez-Beltran A, Gormanns P, Grethe J, Hastings J, Heriche JK, Hermjakob H, Ison JC, Jimenez RC, Jupp S, Kunze J, Laibe C, Le Novere N, Malone J, Martin MJ, McEntyre JR, Morris C, Muilu J, Muller W, Rocca-Serra P, Sansone SA, Sariyar M, Snoep JL, Soiland-Reyes S, Stanford NJ, Swainston N, Washington N, Williams AR, Wimalaratne SM, Winfree LM, Wolstencroft K, Goble C, Mungall CJ, Haendel MA, Parkinson H (2017) Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biol* 15:e2001414
- Meehan TF, Conte N, West DB, Jacobsen JO, Mason J, Warren J, Chen CK, Tudose I, Relac M, Matthews P, Karp N, Santos L, Fiegel T, Ring N, Westerberg H, Greenaway S, Sneddon D, Morgan H, Codner GF, Stewart ME, Brown J, Horner N, International Mouse Phenotyping C, Haendel M, Washington N, Mungall CJ, Reynolds CL, Gallegos J, Gailus-Durner V, Sorg T, Pavlovic G, Bower LR, Moore M, Morse I, Gao X, Tocchini-Valentini GP, Obata Y, Cho SY, Seong JK, Seavitt J, Beaudet AL, Dickinson

- ME, Herault Y, Wurst W, de Angelis MH, Lloyd KCK, Flenniken AM, Nutter LMJ, Newbigging S, McKerlie C, Justice MJ, Murray SA, Svenson KL, Braun RE, White JK, Bradley A, Flicek P, Wells S, Skarnes WC, Adams DJ, Parkinson H, Mallon AM, Brown SDM, Smedley D (2017) Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nat Genet* 49:1231–1238
- Mianne J, Codner GF, Caulder A, Fell R, Hutchison M, King R, Stewart ME, Wells S, Teboul L (2017) Analysing the outcome of CRISPR-aided genome editing in embryos: Screening, genotyping and quality control. *Methods* 121:68–76
- Munoz-Fuentes V, Marcet-Ortega M, Alkorta-Aranburu G, Linde Forsberg C, Morrell JM, Manzano-Piedras E, Soderberg A, Daniel K, Villalba A, Toth A, Di Rienzo A, Roig I, Vila C (2015) Strong artificial selection in domestic mammals did not result in an increased recombination rate. *Mol Biol Evol* 32:510–523
- Perez-Garcia V, Fineberg E, Wilson R, Murray A, Mazzeo CI, Tudor C, Sienerth A, White JK, Tuck E, Ryder EJ, Gleeson D, Siragher E, Wardle-Jones H, Staudt N, Wali N, Collins J, Geyer S, Busch-Nentwich EM, Galli A, Smith JC, Robertson E, Adams DJ, Weninger WJ, Mohun T, Hemberger M (2018) Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature* 555:463
- Raikkonen J, Vucetich JA, Peterson RO, Nelson MP (2009) Congenital bone deformities and the inbred wolves (*Canis lupus*) of Isle Royale. *Biol Conserv* 142:1025–1031
- Raikkonen J, Vucetich JA, Vucetich LM, Peterson RO, Nelson MP (2013) What the Inbred Scandinavian wolf population tells us about the nature of conservation. *PLoS One* 8:e67218
- Roelke ME, Martenson JS, O'Brien SJ (1993) The consequences of demographic reduction and genetic depletion in the endangered Florida panther. *Curr Biol* 3:340–350
- Rozman J, Rathkolb B, Oestereicher MA, Schutt C, Ravindranath AC, Leuchtenberger S, Sharma S, Kistler M, Willershauser M, Brommage R, Meehan TF, Mason J, Haselimashhadi H, Consortium I, Hough T, Mallon AM, Wells S, Santos L, Lelliott CJ, White JK, Sorg T, Champy MF, Bower LR, Reynolds CL, Flenniken AM, Murray SA, Nutter LMJ, Svenson KL, West D, Tocchini-Valentini GP, Beaudet AL, Bosch F, Braun RB, Dobbie MS, Gao X, Herault Y, Moshiri A, Moore BA, Kent Lloyd KC, McKerlie C, Masuya H, Tanaka N, Flicek P, Parkinson HE, Sedlacek R, Seong JK, Wang CL, Moore M, Brown SD, Tschop MH, Wurst W, Klingenspor M, Wolf E, Beckers J, Machicao F, Peter A, Staiger H, Haring HU, Grallert H, Campillos M, Maier H, Fuchs H, Gailus-Durner V, Werner T, Hrabe de Angelis M (2018) Identification of genetic elements in metabolism by high-throughput mouse phenotyping. *Nat Commun* 9:288
- Ruiz-Lopez MJ, Ganan N, Godoy JA, Del Olmo A, Garde J, Espeso G, Vargas A, Martinez F, Roldan ER, Gomendio M (2012) Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. *Conserv Biol* 26:1121–1129
- Saenko SV, French V, Brakefield PM, Beldade P (2008) Conserved developmental processes and the formation of evolutionary novelties: examples from butterfly wings. *Philos Trans R Soc Lond B* 363:1549–1555
- Santiago-Sim T, Burrage LC, Ebstein F, Tokita MJ, Miller M, Bi W, Braxton AA, Rosenfeld JA, Shahrouh M, Lehmann A, Cogné B, Küry S, Besnard T, Isidor B, Béziau S, Hazart I, Nagakura H, Immken LL, Littlejohn RO, Roeder E, Afawi Z, Balling R, Barisic N, Baulac S, Craiu D, De Jonghe P, Guerrero-Lopez R, Guerrini R, Helbig I, Hjalgrim H, Jähn J, Klein KM, Leguern E, Lerche H, Marini C, Muhle H, Rosenow F, Serratos J, Sterbová K, Suls A, Moller RS, Striano P, Weber Y, Zara F, Kara B, Hardies K, Weckhuysen S, May P, Lemke JR, Elpeleg O, Abu-Libdeh B, James KN, Silhavy JL, Issa MY, Zaki MS, Gleeson JG, Seavitt JR, Dickinson ME, Ljungberg MC, Wells S, Johnson SJ, Teboul L, Eng CM, Yang Y, Kloetzel P-M, Heaney JD, Walkiewicz MA (2017) Biallelic variants in *OTUD6B* cause an intellectual disability syndrome associated with seizures and dysmorphic features. *Am J Hum Genet* 100:676–688
- Saragusty J, Diecke S, Drukker M, Durrant B, Ben-Nun IF, Galli C, Goritz F, Hayashi K, Hermes R, Holtze S, Johnson S, Lazzari G, Loi P, Loring JF, Okita K, Renfree MB, Seet S, Voracek T, Stejskal J, Ryder OA, Hildebrandt TB (2016) Rewinding the process of mammalian extinction. *Zoo Biol* 35:280–292
- Schweizer RM, Robinson J, Harrigan R, Silva P, Galverni M, Musiani M, Green RE, Novembre J, Wayne RK (2016a) Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Mol Ecol* 25:357–379
- Schweizer RM, vonHoldt BM, Harrigan R, Knowles JC, Musiani M, Coltman D, Novembre J, Wayne RK (2016b) Genetic subdivision and candidate genes under selection in North American grey wolves. *Mol Ecol* 25:380–402
- Singh P, Schimenti JC, Bolcun-Filas E (2015) A mouse geneticist's practical guide to CRISPR applications. *Genetics* 199:1–15
- Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A (2011) A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* 474:337–342
- Smedley D, Oellrich A, Kohler S, Ruef B, Sanger Mouse Genetics P, Westerfield M, Robinson P, Lewis S, Mungall C (2013) PhenDigm: analyzing curated annotations to associate animal models with human diseases. *Database (Oxford)* 2013: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 Semantics* 6:11
- The Genomes Project C (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56
- Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM (2015) Identification and characterization of essential genes in the human genome. *Science* 350:1096–1101
- White JK, Gerdin AK, Karp NA, Ryder E, Buljan M, Bussell JN, Salisbury J, Clare S, Ingham NJ, Podrini C, Houghton R, Estabel J, Bottomley JR, Melvin DG, Sunter D, Adams NC, Tannahill D, Logan DW, MacArthur DG, Flint J, Mahajan VB, Tsang SH, Smyth I, Watt FM, Skarnes WC, Dougan G, Adams DJ, Ramirez-Solis R, Bradley A, Steel KP, Project SIMG. (2013) Genome-wide Generation and Systematic Phenotyping of Knockout Mice Reveals New Roles for Many Genes. *Cell* 154:452–464
- Wildt DE, Bush M, Howard JG, O'Brien SJ, Meltzer D, Van Dyk A, Ebedes H, Brand DJ (1983) Unique seminal quality in the South African cheetah and a comparative evaluation in the domestic cat. *Biol Reprod* 29:1019–1025
- Wildt DE, Lynch C, Santymire RM, Marinari PE (2016) Recovery of gene diversity using long-term cryopreserved spermatozoa and artificial insemination in the endangered black-footed ferret: response to commentaries. *Anim Conserv* 19:118–119
- Winter JM, Gildea DE, Andreas JP, Gatti DM, Williams KA, Lee M, Hu Y, Zhang S, Mullikin JC, Wolfsberg TG, McDonnell SK, Fogarty ZC, Larson MC, French AJ, Schaid DJ, Thibodeau SN, Churchill GA, Crawford NPS (2017) Mapping complex traits in a diversity outbred F1 Mouse population identifies germline modifiers of metastasis in human prostate cancer. *Cell Syst* 4:31–45
- Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, de Manuel M, Hernandez-Rodriguez J, Lobon I, Siegmund HR, Pagani L, Quail MA, Hvilson C, Mudakikwa A, Eichler EE, Cranfield MR, Marques-Bonet T, Tyler-Smith C, Scally A (2015) Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* 348:242–245

Affiliations

Violeta Muñoz-Fuentes¹  · Pilar Cacheiro²  · Terrence F. Meehan¹  · Juan Antonio Aguilar-Pimentel³  · Steve D. M. Brown⁴ · Ann M. Flenniken^{5,6} · Paul Flicek¹ · Antonella Galli⁷ · Hamed Haseli Mashhadi¹ · Martin Hrabě de Angelis^{3,8,9} · Jong Kyoung Kim¹⁰ · K. C. Kent Lloyd¹¹ · Colin McKerlie^{5,6,12} · Hugh Morgan⁴ · Stephen A. Murray¹³ · Lauryl M. J. Nutter^{5,12} · Patrick T. Reilly¹⁴ · John R. Seavitt¹⁵ · Je Kyung Seong¹⁶ · Michelle Simon⁴ · Hannah Wardle-Jones⁷ · Ann-Marie Mallon⁴ · Damian Smedley² · Helen E. Parkinson¹ · the IMPC consortium

¹ European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

² Clinical Pharmacology, William Harvey Research Institute, School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK

³ German Mouse Clinic, Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

⁴ Medical Research Council Harwell Institute (Mammalian Genetics Unit and Mary Lyon Centre), Harwell, Oxfordshire OX11 0RD, UK

⁵ The Centre for Phenogenomics, Toronto, ON M5T 3H7, Canada

⁶ Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada

⁷ Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK

⁸ German Center for Diabetes Research (DZD), Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

⁹ School of Life Science Weihenstephan, Technische Universität München, Alte Akademie 8, 85354 Freising, Germany

¹⁰ Department of New Biology, DGIST, Daegu 42988, Republic of Korea

¹¹ Mouse Biology Program, University of California, Davis, CA 95618, USA

¹² The Hospital for Sick Children, Toronto, ON M5G 1X84, Canada

¹³ The Jackson Laboratory, Bar Harbor, ME 04609, USA

¹⁴ PHENOMIN-iCS, 1 Rue Laurent Fries, 67404 Illkirch Cedex, Alsace, France

¹⁵ Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

¹⁶ Laboratory of Developmental Biology and Genomics, College of Veterinary Medicine, Interdisciplinary Program for Bioinformatics and Program for Cancer Biology, Seoul National University, Seoul, Republic of Korea