NIH Knockout Mouse Phenotyping Program (KOMP2) and IMPC: Database to Discover New Roles of Genes in Cardiovascular Physiology and Disease

Young S. Oh*, Maarten Leerkes, Tiffany Cao, and Zorina S. Galis

Author Affiliations:

Vascular Biology and Hypertension Branch (YS, TC, and ZSG), Heart Development and Structural Diseases Branch (ML), Division of Cardiovascular Sciences, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda MD 20892

*Corresponding author:
Young S. Oh, Ph.D.
Vascular Biology and Hypertension Branch
Division of Cardiovascular Sciences
National Heart, Lung, and Blood Institute,
6701 Rockledge Drive, Room 8106
Bethesda, MD 20892
yoh@mail.nih.gov
Phone: (301) 435-0560
Fax: (301) 480-2858

Word count: 3392
Figures: 3
Online Tables: 4

Running title: NIH KOMP2 Program

Key words: Genes; Knockout Mouse; Physiology; Cardiovascular; Animal Models; Human Disease

ABSTRACT

The collaborative goal of NIH Knockout Mouse Phenotyping Program (KOMP2) and International Knockout Mouse Phenotyping Consortium (IMPC) is to discover functional insight for every gene in the mouse genome by 2021, by generating and systematically phenotyping approximately 20,000 unique knockout (KO) mouse strains. The purpose of the present study is to introduce the KOMP2/IMPC program and its publicly-accessible gene-phenotype database to the research community and to specifically illustrate its utility for the identification of novel gene candidates in cardiovascular (CV) disease. In this report, we have focused on single gene deletions found associated with CV phenotype as part of the KOMP2/IMPC phenotyping of broad physiological domains in more than 5,500 single gene KO mice. Among the 694 single genes found to result in a CV phenotype, we identified about one third (36%, n=248) that had not been previously associated with the CV system. We also searched the remainder of more than 5,500 genes that had not been previously associated with the CV system. We used the results of gene-disease relationship analysis from Medline sentences with an algorithm called Ensemble Biclustering for Classification (EBC), which uncovers relationships between biomedical entities. In addition, we studied their interactions in protein-protein interaction networks. These genes may present opportunities for new CV research and may provide new targets for therapeutic intervention for various CV diseases.

INTRODUCTION

The Knockout Mouse Project (KOMP, www.komp.org), a trans-NIH effort, was launched in 2006 and completed in 2011, successfully creating more than 17,000 single gene deletions in mouse embryonic stem cells in collaboration with the International Mouse Phenotyping Consortium (IMPC). This public repository of embryonic stem cells, vectors, and generated single-gene deletion knockout (KO) mice is available for use by any investigator worldwide. The renewal of this initiative, the Knockout Mouse Phenotyping Program (KOMP2), is currently underway, and its goal is to generate and phenotype systematically ~20,000 KO mouse strains by 2021 in continued collaboration with IMPC. The IMPC generates high-quality, high-throughput phenotype data, including 16 mandatory tests, obtained by 11 phenotyping centers located in Europe, North America, and Asia (4). As of April 2019, more than 5,500 KO mice have been generated and broadly phenotyped in multiple physiological domains, including the cardiovascular (CV) system. The catalogue of these KO mouse lines and their phenotyping data is publicly available via the IMPC web portal (http://www.mousephenotype.org).

The IMPC phenotype pipeline is expected to cover all major human diseases and help in the creation and identification of new mouse models of many human diseases. Recent IMPC data analysis of 3,328 genes has identified new mouse models for 360 rare human diseases, including the first models for type C Bernard–Soulier, Bardet–Biedl, and Gordon Holmes syndromes (19). This analysis also provided functional evidence for 1,092 genes and candidates in genetically uncharacterized diseases, such

as arrhythmogenic right ventricular dysplasia. In the present study, we have focused on the gene deletions that produced a CV-related phenotype.

METHODS

In the first part of the analysis, we accessed the IMPC web portal on April, 2019 (version 10.0) and downloaded the list of genes that showed abnormal CV system phenotypes when knocked out in a mouse. Phenotyping protocols of KO mice are available at the IMPC web site (http://www.mousephenotype.org) under "IMPReSS (International Mouse Phenotyping Resource of Standardised Screens)," and the methods for detailed statistical analysis under "Documentation." All mutants are generated on the C57BL/6N background, and the target number of animals to be generated per KO strain is fourteen (seven per sex). Published reports related CV-associated genes were searched via PubMed. The search terms included, among other terms: "gene name AND heart," "gene name AND cardi*," "gene name AND athero*," "gene name AND arter*" or "gene name AND vascular." The publications identified by these search terms were accessed and reviewed online through Medline. When ambiguous results were obtained, additional searches via Google, GeneCards (www.genecards.org), Monarch Initiative (https://monarchinitiative.org), and/or Pharos (https://pharos.nih.gov/index) were performed to determine an initial categorization. The gene was categorized as "known" if we were able to identify any publication to show biological function of a gene product associated with the CV system (e.g., vascular cells, blood vessel, heart structure or morphology, or CV-related diseases). Further, "known" CV genes also included genes that have a report(s) at the level of DNA

(GWAS, Genome-Wide Association Studies, or sequence variants) or RNA expression associated with the CV system; but no functional studies. When we could not identify any report linking an individual gene with the CV system, we classified the gene as "unknown." The lists of "known" and "unknown" genes were validated and cross-checked by searching the Open Targets Platform (www.targetvalidation.org), which is a comprehensive data integration portal to link genes, drug targets, and human diseases. In addition, in-silico validation was done by verifying references for CV-associated literature citations against all citations in PubMed that were identified by a text-mining approach. To do this, an open access research data repository containing results of Ensemble Biclustering for Classification (EBC) was used. These results of EBC text mining of over 23 million articles is available in zenodo as a general-purpose open-access repository (https://zenodo.org/record/1243969#.W67DWpNKjUI) and contains labeled, weighted networks of gene-disease relationships which we interrogated with custom in-house scripts (20). This zenodo repository is the result of biomedical relationships derived from text by the EBC algorithm (20) that applies strategies to uncover relationships between biomedical entities, such as drugs, genes and phenotypes. We focused on three relationship types for extraction and characterization from unstructured biomedical text. We classified IMPC genes through statistical dependency parsing to extract descriptions of gene to CV disease relationships from Medline sentences. By applying EBC to recognize when two gene-disease pairs, as well as gene-gene and gene-drug relationships share a similar relationship, we did an in-house functional re-annotation of cardiovascular function.

Information about the existence of a human ortholog as well as the viability of homozygous KO mice was obtained from the IMPC database. The viability of KO mice was categorized into three groups: viable (> 16 weeks survival), subviable (3-4 weeks), or embryonically lethal. Since KO mice production and phenotype characterization are dynamic processes involving 11 phenotyping centers, many genes are still under characterization (4). When the life stage of KO mice was not finalized, they were included in a group “under characterization.”

In the second part of the analysis, after analyzing the downloaded IMPC phenotyped list of genes that showed abnormal CV system phenotypes when knocked out in a mouse, we performed a reverse query on the remainder of IMPC KO genes to search for novel candidates with CV annotations. We searched for genes that were linked with the CV system identified in either the National Center for Biotechnology Information (NCBI) database or in a zenodo data repository (obtained via multiple CV key word searches from an EBC methodology). By selecting from four overlapping categories such as gene-disease, gene-chemical, gene-gene interaction and overlapping NCBI CV genes, a list of genes with at least two overlapping CV categories was identified, which produced no abnormal CV phenotype in the KO mice.

Protein-protein interaction networks were created by NetworkAnalyst, (http://www.networkanalyst.ca/faces/home.xhtml), a comprehensive web-based tool for biological network analysis and visualization. Nodes in the networks can be interpreted

as important players in CV disease. The edges can be interpreted as molecular relationships between the disease-associated cellular components.

RESULTS AND DISCUSSION

Using data deposited in the IMPC database (version 10.0) containing phenotypic information of 5,684 genes tested on the CV system, we found that a total of 694 genes were associated with a CV system phenotype. Online Table 1 summarizes all the CV system phenotypes characterized by IMPC up to the date of our inquiry: based on their protocols the two most common CV phenotypes were enlarged heart and abnormal heart morphology. Four hundred forty six genes were classified as “known” CV-associated genes (64%) because these genes are reported in at least one research publication linked with the CV system (Figure 1). The remaining 248 genes (36%) were grouped as “unknown” CV-associated genes, and these could be considered as novel CV-associated gene candidates. The great majority (97%) of these genes have a human ortholog.

The deletion of either “known” or “unknown” CV phenotype-associated genes resulted in viable, subviable, and embryonically lethal KO animals. Dickinson et al. (7) have recently analyzed 1,751 unique gene KO mice from IMPC production colonies and have identified that about 35% of these genes are essential genes for survival: 24% lethal and 11% subviable. The most common phenotype observed during embryo development was growth and developmental delay, followed by abnormal CV

development. In the present analysis, within the “known” CV-associated genes, 28% of genes were lethal and 9% were subviable (Figure 1). Within the “unknown” CV-associated genes, 18% of genes were lethal and 10% were subviable. Some genes are still under characterization for their viability (13% of “known” and 15% of “unknown” genes), and it is likely that more genes will be classified into a viable group eventually.

We analyzed all the published reports of “known” CV genes and further categorized these genes based on evidence of a biological function association (Online Table 2A, n=335) or association only at the level of DNA (GWAS or sequence variants) or RNA expression (Online Table 2B, n=111). The availability of a list of previously “unknown” to CV research genes could spur new CV discovery (Online Table 3, n=248). A frequently used reason for prioritizing further research of unknown genes is discovering an association with a phenotype of interest. For instance, we closely examined the list of unknown genes in relation to an abnormal CV phenotype determined by electrocardiogram (n=32). In this regard, there is evidence for significant conservation of gene expression clusters and network-based prediction of gene function between human and mouse (2). Among genes in this category Vsig8 (V-set and immunoglobulin domain containing 8) stood out based on its “clean” CV phenotype in the KO and significantly altered electrocardiogram phenotype. Vsig8 KO mouse displayed a CV-specific phenotype (i.e., decreased heart rate, shortened ST segment, and shortened RR interval) without affecting any of the other physiological domains investigated by IMPC in this KO (i.e., mortality, reproduction, growth, bone, skeletal muscle, hearing, listening, etc.). Vsig8, the hair shaft protein, was previously

192 investigated for its role in epithelial differentiation and function in the upper alimentary
193 tract (21). By phenotypic similarity analysis provided by IMPC, Vsig8 KO mouse
194 displayed a phenotype close to that of the Brugada syndrome (11). The Brugada
195 syndrome is a rare arrhythmia disorder with complex inheritance, associated with high
196 risk of sudden cardiac death in the young adult. So far, only one gene, the cardiac
197 sodium channel (SCN5A), has been found associated with the syndrome, and this
198 association accounts for only about 20-25% of cases. Thus, Vsig8 might be a gene
199 candidate associated with the Brugada syndrome. However, precautions should be
200 taken when translating the results of mice model to human because the physiological
201 mechanisms underlying the electrocardiogram of mice and human are somewhat
202 different (3).

203

204 We also tested another possible, more agnostic, bioinformatics-driven approach
205 to prioritize the list of previous "unknown" CV-associated genes. Using NetworkAnalyst,
206 a protein-protein interaction network analysis was performed on all "known" and
207 "unknown" CV-associated genes (Figure 2). Several "unknown" genes, including Arid4b,
208 Htr1f, and Nmur1, were identified via zero-degree network analysis, indicating direct
209 protein-protein interactions with "known" CV-associated genes. These genes likely play
210 a role in direct or indirect functional roles in the CV system. Upon close examination, the
211 following hypothesis may be pursued: in the network of Figure 2A, a "known" CV-
212 associated gene product, HDAC1 (histone deacetylase 1) may interact directly with
213 ARID4B (AT rich interactive domain 4B), an "unknown" CV-associated gene product.
214 HDAC1 is well known for its a key role in the regulation of eukaryotic gene expression,

including cardiac gene expression (17). ARID4B is a chromatin remodeling protein, a component of the HDAC1/SIN3A chromatin remodeling complex (9), which functions in diverse cellular processes including proliferation, differentiation, apoptosis, and cell fate determination (10, 22). Arid4b acts as a transcriptional repressor (10). Arid4b KO mouse shows prolonged QRS complex duration, abnormal locomotor activity and preweaning lethality. It will be interesting to examine a potential interaction between HDAC1 and ARID4B during heart development. Figure 2A also shows another interesting gene (Pax5, paired box 5), which may interact with Myb (myeloblastosis oncogene) and Lef1 (lymphoid enhancer binding factor 1). A recent GWAS has reported an association of Pax5 with coronary artery disease (5), but the functional role of Pax5 in the CV system has not been demonstrated. Pax5 is shown to play a role in early development, especially in B cell development (18). Since both Myb and Lef1 genes are involved in vasculature development (8, 23), Pax5 may also play a role in vasculature development.

Figure 2B shows another interesting network involving four “known” CV genes and two “unknown” CV genes. An “unknown” CV gene, Htr1f (5-Hydroxytryptamine Receptor 1F, a G-protein coupled receptor for serotonin), seems to interact with two “known” CV genes: Pmch (Pro-Melanin Concentrating Hormone) and S1pr3 (Sphingosine-1-Phosphate Receptor 3). PMCH is a preproprotein that is proteolytically processed in the brain to generate multiple protein products, including melanin-concentrating hormone (MCH), neuropeptide-glutamic acid-isoleucine, and neuropeptide-glycine-glutamic acid. Among these, MCH has been implicated in a

238 variety of functions, including food intake, sleep, and the CV system (13). In this
239 network, PMCH (via its processed peptides) interacts with two “known” CV-associated
240 G-protein coupled receptors, such as GPR65 (G Protein-Coupled Receptor 65) and
241 KISS1R (KISS1 Receptor or G Protein-Coupled Receptor 54). Figure 2B shows that
242 PMCH may interact with another “unknown” CV gene product, NMUR1 (Neuromedin U
243 Receptor 1), which is a receptor for neuromedin U, a multifunctional neuropeptide
244 involved in blood pressure regulation, energy balance, immune system, and cancer
245 (16). Investigation of these potentially interacting proteins may not only reveal new
246 biological roles for these proteins and correlations to CV diseases, but may also identify
247 novel targets for therapeutic interventions, especially needed for many medical
248 conditions still lacking effective and safe therapies.

249

250 In the second part of the analysis, we have performed a reverse query to search
251 for novel CV annotations to all of the remaining KO strains considered in this report.
252 For this reverse search the starting point was all 6,002 KO genes in IMPC database
253 (April, 2019) except the previously identified 694 CV-associated genes by IMPC’s
254 phenotyping pipelines. This starting set of 5,308 genes was searched against 3 EBC
255 relationship data sets for evidence in CV phenotypes and pathways. Through this
256 approach we found there were ~1,000 unique genes that are linked with the CV system
257 identified in either an NCBI database search or in the zenodo data repository (obtained
258 via multiple CV key word searches). By selecting from overlapping categories such as
259 gene-disease, gene-chemical, gene-gene interaction and overlapping NCBI CV genes,
260 a total of 251 genes was identified (Online Table 4), which have at least two overlapping

CV categories; many have hits in 3 or 4 categories as depicted in the Venn diagram (Figure 3A). Figure 3B shows an example of a combined network of CV genes annotated by the reverse query and “unknown” CV genes.

Interestingly, two “unknown” CV genes that were identified and discussed in Figure 2B (Htr1f and Nmur1) were also identified in Figure 3B, by their interaction with multiple CV genes annotated by the reverse query approach. For example, Htr1f shows strong direct protein-protein interactions with six CV-annotated genes, such as Pomc, Cxcl12, Htr1a, Apln, Gal and Sst. Another “unknown” CV gene, Nmur1 also interacts with seven CV-annotated genes: Hcrt1, Pomc, Cxcl12, Plcb2, Apln, Gal, and Sst. These potential CV genes likely interact with each other, which is how they can form subnetworks or even disease modules. With this dual approach of text mining and network analysis, an identification of CV disease modules could be possible, which can subsequently help us to identify CV disease pathways. Interesting follow-up research could be, for example, to functionally validate these unknown players in CV disease by using sets of inter-connected proteins by overlaying additional omics datasets such as RNAseq. In this possible approach, we could define a disease state as multiple interacting perturbed pathways that are part of modules and sub-networks. Finally, intervention in a module or sub-network can alter the disease state, where any protein (node) in a network could be a potential drug target (1, 6, 12).

The current analysis has several potential limitations. First, the gene designation used in this report as “known” versus “unknown” CV-associated gene is subjective, and

this classification can change based on new studies as to the functional characterization of some of these genes. For example, genes that have only GWAS or RNA profiling study in CV-linked publications (Online Table 2B, n=111) were considered as “known” in this report; however, their functional role needs to be tested in the future. Second, the list of CV-associated genes obtained from the IMPC database is incomplete and still a work in progress because of the nature of the IMPC database, which is in dynamic status with continuous updates and optimization in statistical analysis as more KO mice are characterized. For example, we have found that 31 genes of the CV-associated genes identified in version 9.1 IMPC database (November, 2018) that became statistically non-significant CV genes in the current version 10.0 (April, 2019), which represents about 5% of CV genes in version 9.1. It is possible that differences in animal handling and housing facility among different KO mouse phenotyping centers may contribute to data variation as well. For example, the mouse gut microbiota can vary greatly due to environmental factors and may impact phenotypes (14). Third, important CV-associated genes might have not been identified though the current CV phenotyping process because of functional redundancy by paralogue(s) and/or physiological compensatory mechanisms. In this regard, we have noted that many previously identified and studied CV-associated genes did not produce any CV phenotype in the current database (see Online Table 4). Some examples include: Agtr1a (angiotensin II receptor, type 1a), Agtr2 (angiotensin II receptor, type 2), Apoe (apolipoprotein E), Fgf2 (fibroblast growth factor 2), Kcnj11 (potassium inwardly rectifying channel, subfamily J, member 11), Lepr (leptin receptor), Nox1 (NADPH oxidase 1), and Tgfb3 (transforming growth factor, beta 3). Fourth, in the present study, we have not attempted to analyze

sexual dimorphism associated with the CV phenotype although a recent report suggests the prevalence of sexual dimorphism in many mammalian phenotypic traits (15). This will be an important topic in the future to examine when more detailed data are available.

In summary, KOMP2/IMPC KO mice are broadly phenotyped in multiple physiological domains (i.e., CV, eye, muscle, bone, behavior, immune, blood, development, etc.), thus investigators in different research fields may be able to find new information related to the physiology and function of their genes of interests by searching the IMPC database. In many physiological research domains, the KOMP2/IMPC database is likely to serve as a great resource to study gene-gene interactions and to identify novel gene candidates. The current study illustrates the opportunity to use the KOMP2/IMPC database to identify novel gene candidates associated with the CV system. Follow-up studies with these uncharacterized genes may reveal novel insights into targeting CV genes that have not been highlighted in the CV context as of yet.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Marc Charette for reading the manuscript and colleagues in Vascular Biology and Hypertension Branch for their support and advice during this project.

DISCLOSURES

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

SUPPLEMENTARY TABLES

Supplementary tables are available from the DOI: 10.6084/m9.figshare.9170492 (https://figshare.com/articles/KOMP2_Supplemental_Tables/9170492).

FIGURE LEGEND

Figure 1. Summary of known and unknown cardiovascular (CV)-associated genes identified from the IMPC database (5,684 genes tested).

Figure 2. Network analysis of CV-associated genes (n=694) using NetworkAnalyst (http://www.networkanalyst.ca/faces/home.xhtml). The network analysis was done on the STRING Interactome database, which integrates protein-protein interaction data (both direct and indirect interactions) from many sources with confidence score cutoff of 900. Two subnetworks of the zero-order network are shown, which represent a network of strong (direct) protein-protein interactions. **A:** A network involving many "known" CV genes such as Ehmt2 (euchromatic histone lysine N-methyltransferase 2) and Hdac1 (histone deacetylase 1). An "unknown" CV gene, Arid4b (AT-Rich Interactive Domain-Containing Protein 4B), may interact with Hdac1. In this network, a GWAS-identified CV-associated gene is also present: Pax 5 (paired box 5). It is interesting to note that

Pax5 may interact with two “known” CV genes: Myb (myeloblastosis oncogene) and Lef1 (lymphoid enhancer binding factor 1). **B:** A network involving 4 “known” and 2 “unknown” CV genes. Potential protein interactions between “known” and “unknown” CV genes in these two networks are discussed in more detail in the text. Future studies are needed to explore the function of these CV candidate genes. Red circle: known CV-associated genes. Blue circle: unknown CV-associated genes. Yellow circle: GWAS-identified CV gene.

Figure 3. Reverse query to search for novel CV-annotated genes identified in the zenodo data repository or NCBI CV genes (obtained via multiple CV key word searches). **A:** Venn diagram showing results of CV genes annotated from three EBC relationship data sets (gene-disease, gene-chemical, and gene-gene interaction) in the zenodo data repository and NCBI CV genes. A total of 251 genes (Online Table 4) which have at least two overlapping CV categories were identified; KO of some of these genes did not produce any CV phenotype in the mice. See text for discussion.

B: Network analysis of the reverse query CV-annotated genes (n=251) combined with "unknown" CV genes (n=248, blue circles). The network analysis was done using the NetworkAnalyst on the STRING Interactome database with confidence score cutoff of 900. One particular subnetwork with the first-degree order is shown, highlighted in a circle with dotted lines which represent a network of direct protein-protein interactions of Htrf1 and Nmur1, which are highly interconnected with multiple CV-annotated genes.

Htr1f (blue circle) shows 6 direct protein-protein interactions with Pomc, Cxcl12, Htr1a, Apln, Gal, and Sst, while Nmur1 (blue circle) shows 7 protein-protein interactions, forming a dense network of highly connected CV edges with Hcrt1, Pomc, Cxcl12, Plcb2, Apln, Gal, and Sst.

REFERENCES

1. **Barabasi AL, Gulbahce N, and Loscalzo J**. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12: 56-68, 2011.
2. **Berg J, and Lassig M**. Cross-species analysis of biological networks by Bayesian alignment. *Proc Natl Acad Sci U S A* 103: 10967-10972, 2006.
3. **Boukens BJ, Rivaud MR, Rentschler S, and Coronel R**. Misinterpretation of the mouse ECG: 'musing the waves of Mus musculus'. *J Physiol* 592: 4613-4626, 2014.
4. **Brown SDM, Holmes CC, Mallon AM, Meehan TF, Smedley D, and Wells S**. High-throughput mouse phenomics for characterizing mammalian gene function. *Nat Rev Genet* 19: 357-370, 2018.
5. **Charmet R, Duffy S, Keshavarzi S, Gyorgy B, Marre M, Rossing P, McKnight AJ, Maxwell AP, Ahluwalia TVS, Paterson AD, Tregouet DA, and Hadjadj S**. Novel risk genes identified in a genome-wide association study for coronary artery disease in patients with type 1 diabetes. *Cardiovasc Diabetol* 17: 61, 2018.
6. **Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabasi AL, and Loscalzo J**. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun* 9: 2691, 2018.
7. **Dickinson ME, Flenniken AM, Ji X, Teboul L, et al**. High-throughput discovery of novel developmental phenotypes. *Nature* 537: 508-514, 2016.
8. **Farrell KA, Withers SB, and Holt CM**. C-Myb function in the vessel wall. *Front Biosci (Elite Ed)* 3: 968-977, 2011.
9. **Fleischer TC, Yun UJ, and Ayer DE**. Identification and characterization of three new components of the mSin3A corepressor complex. *Mol Cell Biol* 23: 3456-3467, 2003.
10. **Goldberger N, Walker RC, Kim CH, Winter S, and Hunter KW**. Inherited variation in miR-290 expression suppresses breast cancer progression by targeting the metastasis susceptibility gene Arid4b. *Cancer Res* 73: 2671-2681, 2013.
11. **Gourraud JB, Barc J, Thollet A, Le Scouarnec S, Le Marec H, Schott JJ, Redon R, and Probst V**. The Brugada Syndrome: A Rare Arrhythmia Disorder with Complex Inheritance. *Front Cardiovasc Med* 3: 9, 2016.
12. **Guney E, Menche J, Vidal M, and Barabasi AL**. Network-based in silico drug efficacy screening. *Nat Commun* 7: 10331, 2016.
13. **Hayward LF, Hampton EE, Ferreira LF, Christou DD, Yoo JK, Hernandez ME, and Martin EJ**. Chronic heart failure alters orexin and melanin concentrating hormone but not corticotrophin releasing hormone-related gene expression in the brain of male Lewis rats. *Neuropeptides* 52: 67-72, 2015.
14. **Hufeldt MR, Nielsen DS, Vogensen FK, Midtvedt T, and Hansen AK**. Variation in the gut microbiota of laboratory mice is related to both genetic and environmental factors. *Comp Med* 60: 336-347, 2010.
15. **Karp NA, Mason J, Beaudet AL, Benjamini Y, et al**. Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat Commun* 8: 15475, 2017.

16. **Martinez VG, and O'Driscoll L**. Neuromedin U: a multifunctional neuropeptide with pleiotropic roles. *Clin Chem* 61: 471-482, 2015.
17. **Mathiyalagan P, Keating ST, Du XJ, and El-Osta A**. Chromatin modifications remodel cardiac gene expression. *Cardiovasc Res* 103: 7-16, 2014.
18. **Medvedovic J, Ebert A, Tagoh H, and Busslinger M**. Pax5: a master regulator of B cell development and leukemogenesis. *Adv Immunol* 111: 179-206, 2011.
19. **Meehan TF, Conte N, West DB, Jacobsen JO, et al**. Disease model discovery from 3,328 gene knockouts by The International Mouse Phenotyping Consortium. *Nat Genet* 49: 1231-1238, 2017.
20. **Percha B, and Altman RB**. Learning the Structure of Biomedical Relationships from Unstructured Text. *PLoS Comput Biol* 11: e1004216, 2015.
21. **Rice RH, Phillips MA, and Sundberg JP**. Localization of hair shaft protein VSIG8 in the hair follicle, nail unit, and oral cavity. *J Invest Dermatol* 131: 1936-1938, 2011.
22. **Wu RC, Zeng Y, Pan IW, and Wu MY**. Androgen Receptor Coactivator ARID4B Is Required for the Function of Sertoli Cells in Spermatogenesis. *Mol Endocrinol* 29: 1334-1346, 2015.
23. **Zhang XF, Yang Y, Yang XY, and Tong Q**. LEF-1 gene silencing inhibits pulmonary vascular remodeling and occurrence of pulmonary arterial hypertension through the beta-catenin signaling pathway. *Biomed Pharmacother* 108: 817-827, 2018.

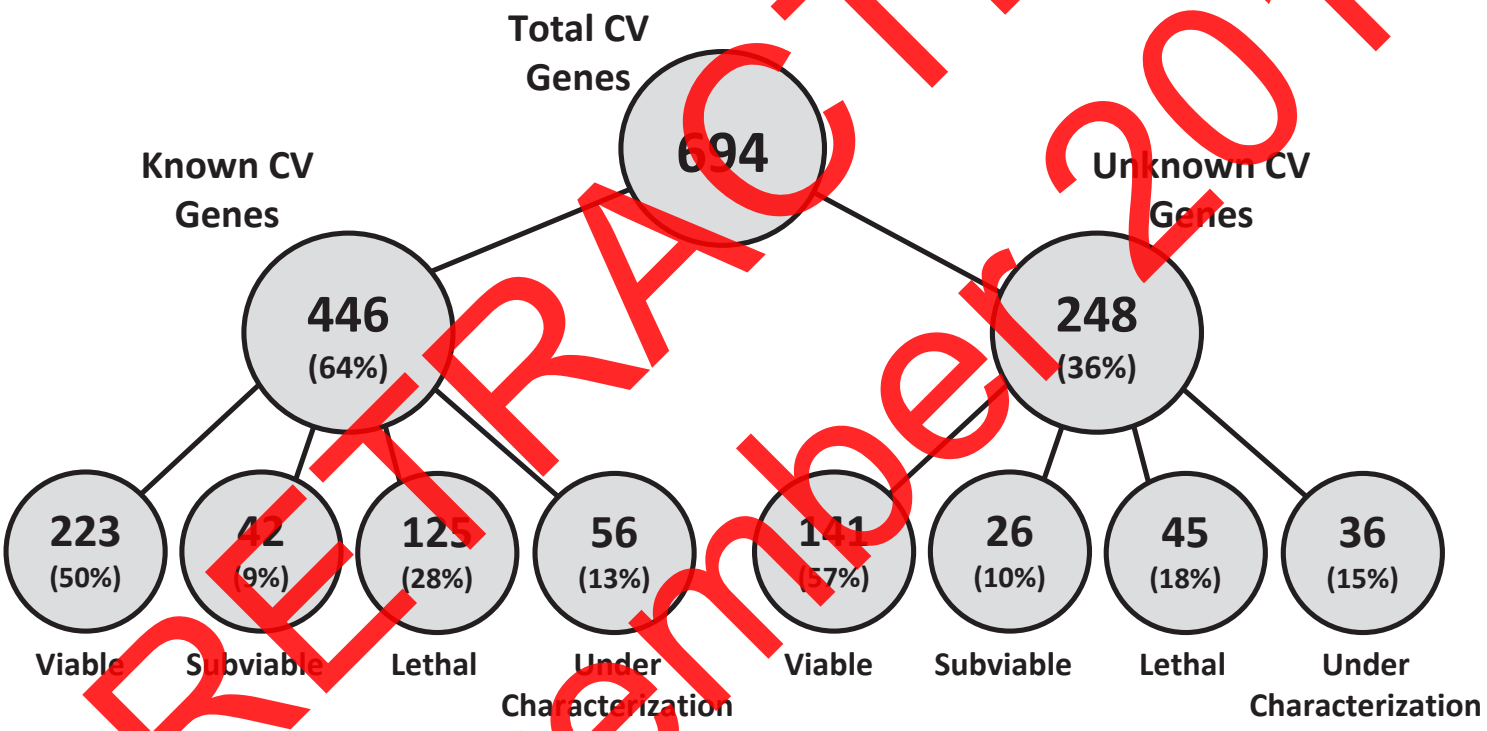

Fig 2A

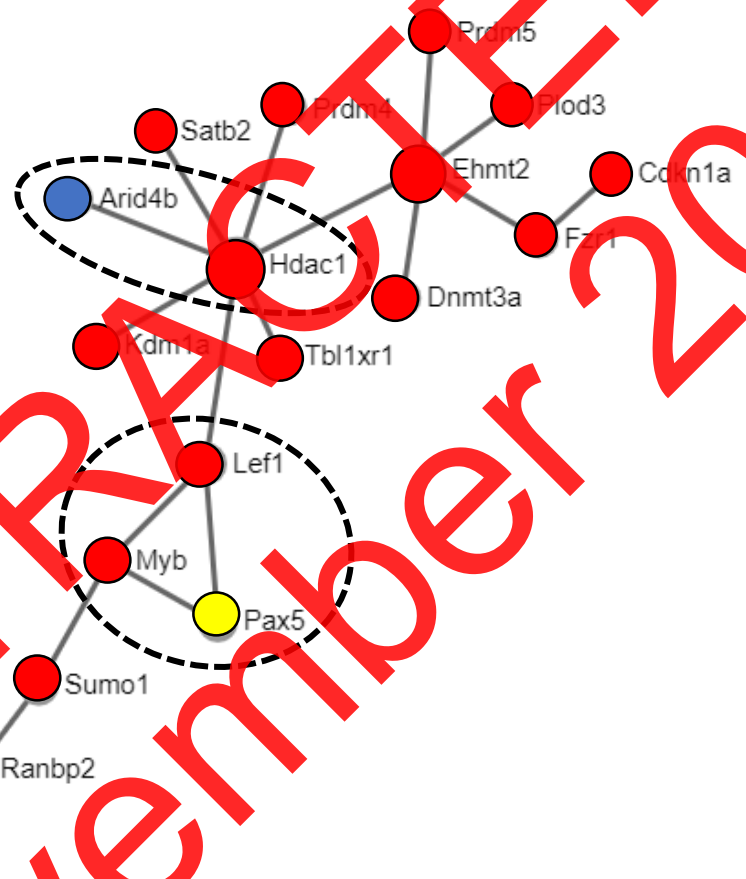

Fig 2B

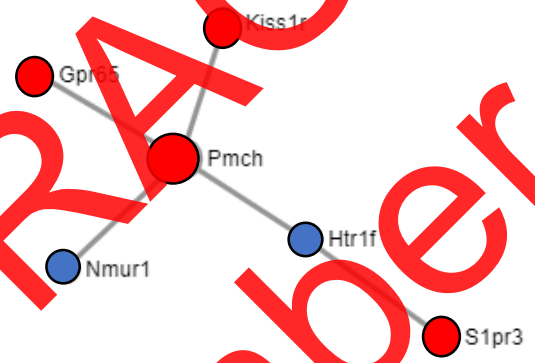

Fig 3A

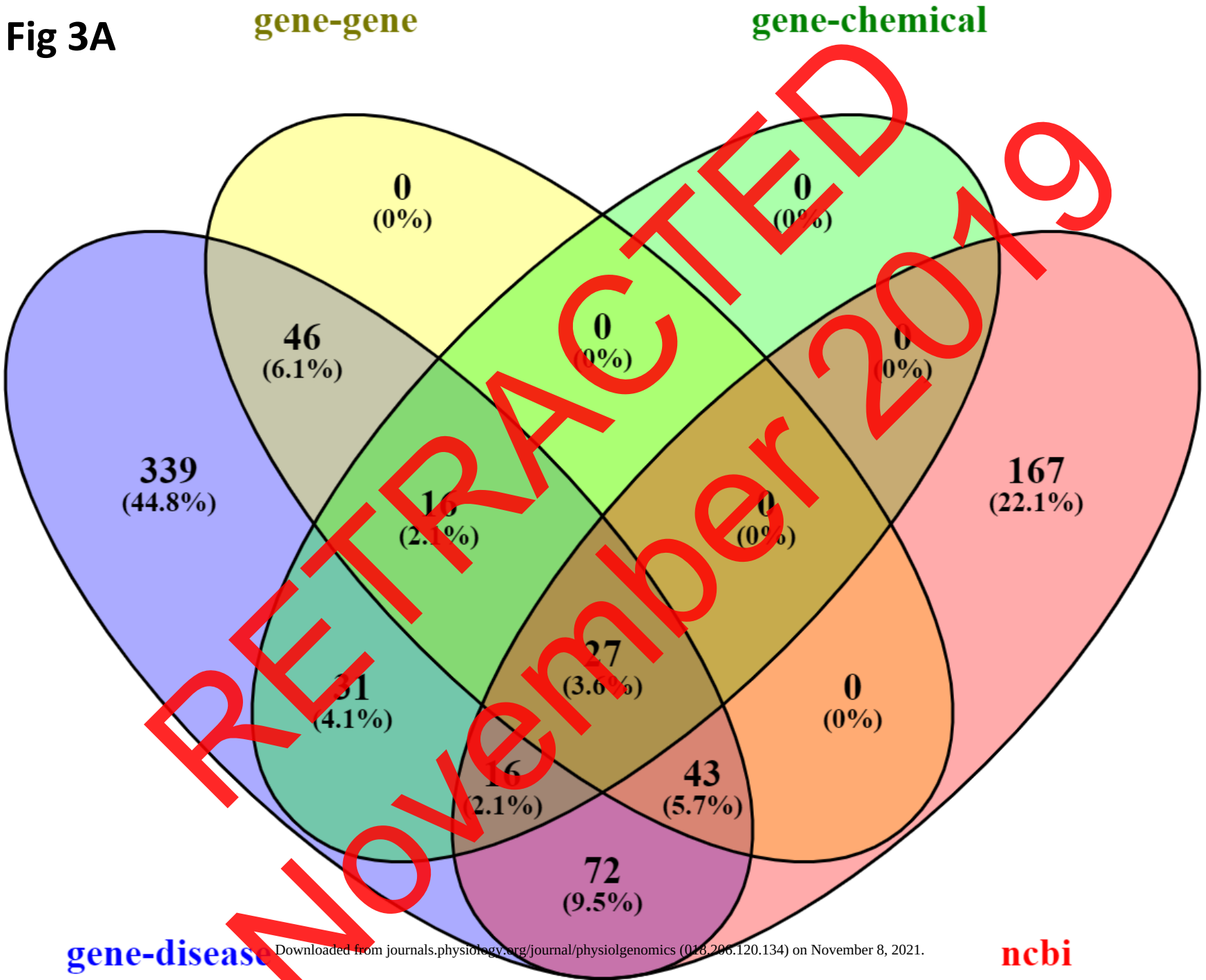

Fig 3B