Mapping the aetiological foundations of the heart failure spectrum using human genetics

Word count: 3889 (excluding methods)

Supplementary files

- **Supplementary Figures 1-20**
- **Supplementary Tables 1-23**
- **Supplementary Note**
- **Supplementary Materials** (QC summary and locus characterisation)

Authors

Albert Henry[1,2], Xiaodong Mo[3], Chris Finan[1], Mark D. Chaffin[4], Doug Speed[5], Hanane Issa[2], Spiros Denaxas[6,2,7,8], James S. Ware[9,10,11,12,4], Sean L. Zheng[9,11], Anders Malarstig[13,14], Jasmine Gratton[1], Isabelle Bond[1], Carolina Roselli[15,4], David Miller[16], Sandesh Chopade[1], A. Floriaan Schmidt[1,17,18], Erik Abner[19], Lance Adams[20], Charlotte Andersson[21,22], Krishna G. Aragam[23,24,4], Johan Ärnlöv[25,26], Geraldine Asselin[27], Anna Axelsson Raja[28], Joshua D. Backman[29], Traci M. Bartz[30], Kiran J. Biddinger[23,4], Mary L. Biggs[30,31], Heather L. Bloom[32], Eric Boersma[33], Jeffrey Brandimarto[34], Michael R. Brown[35], Søren Brunak[36], Mie Topholm Bruun[37], Leonard Buckbinder[14], Henning Bundgaard[28], David J. Carey[38], Daniel I. Chasman[39,40], Xing Chen[14], James P. Cook[41], Tomasz Czuba[42], Simon de Denus[27,43], Abbas Dehghan[44], Graciela E. Delgado[45], Alexander S. Doney[46], Marcus Dörr[47,48], Joseph Dowsett[49], Samuel C. Dudley[50], Gunnar Engström[51], Christian Erikstrup[52,53], Tõnu Esko[19,4], Eric H. Farber-Eger[54], Stephan B. Felix[47,48], Sarah Finer[55], Ian Ford[56], Mohsen Ghanbari[57], Sahar Ghasemi[58,48,59], Jonas Ghouse[60], Vilmantas Giedraitis[61], Franco Giulianini[39], John S. Gottdiener[62], Stefan Gross[47,48], Daníel F. Guðbjartsson[63,64], Hongsheng Gui[65], Rebecca Gutmann[66], Sara Hägg[67], Christopher M. Haggerty[38], Åsa K. Hedman[13], Anna Helgadottir[63], Harry Hemingway[2,6], Hans Hillege[15], Craig L. Hyde[14], Bitten Aagaard Jensen[68], J. Wouter Jukema[69,70], Isabella Kardys[33], Ravi Karra[71,72], Maryam Kavousi[57], Jorge R. Kizer[73], Marcus E. Kleber[45], Lars Køber[74], Andrea Koekemoer[75], Karoline Kuchenbaecker[76,77], Yi-Pin Lai[14], David Lanfear[65,78], Claudia Langenberg[79,80,81], Honghuang Lin[22,82], Lars Lind[83], Cecilia M. Lindgren[84,4,85], Peter P. Liu[86,87,88], Barry London[89], Brandon D. Lowery[54], Jian'an Luan[81], Steven A. Lubitz[90,4], Patrik Magnusson[67], Kenneth B. Margulies[34], Nicholas A. Marston[91], Hilary Martin[92], Winfried März[93,94,45], Olle Melander[95], Ify R. Mordi[46], Michael P. Morley[34], Andrew P. Morris[41,85], Alanna C. Morrison[35], Lori Morton[96], Michael W. Nagle[14], Christopher P. Nelson[75], Alexander Niessner[97], Teemu Niiranen[98,99], Raymond Noordam[100], Christoph Nowak[25], Michelle L. O'Donoghue[91], Sisse Rye Ostrowski[49,101], Anjali T. Owens[34], Colin N.

A. Palmer[102], Guillaume Paré[103,104,105], Ole Birger Pedersen[106,101], Markus Perola[99], Marie Pigeyre[107,105], Bruce M. Psaty[108,109], Kenneth M. Rice[30], Paul M. Ridker[39,40], Simon P. R. Romaine[75], Jerome I. Rotter[110,111,112], Christian T. Ruff[91], Mark S. Sabatine[91], Neneh Sallah[2], Veikko Salomaa[99], Naveed Sattar[113], Alaa A. Shalaby[114], Akshay Shekhar[96], Diane T. Smelser[38], Nicholas L. Smith[115,109,116], Erik Sørensen[117], Sundararajan Srinivasan[102], Kari Stefansson[63,118], Garðar Sveinbjörnsson[63], Per Svensson[119,120], Mari-Liis Tammesoo[19], Jean-Claude Tardif[27,121], Maris Teder-Laving[19], Alexander Teumer[48,122,59], Guðmundur Thorgeirsson[63,123,118], Unnur Thorsteinsdottir[63,118], Christian Torp-Pedersen[124], Vinicius Tragante[63], Stella Trompet[69,100], Andre G. Uitterlinden[57,125], Henrik Ullum[126], Pim van der Harst[17,15], David van Heel[127], Jessica van Setten[17], Marion van Vugt[17], Abirami Veluchamy[46,102], Monique Verschuuren[128,129], Niek Verweij[15], Christoffer Rasmus Vissing[28], Uwe Völker[48,130], Adriaan A. Voors[15], Lars Wallentin[131], Yunzhang Wang[67], Peter E. Weeke[60], Kerri L. Wiggins[31], L. Keoki Williams[65], Yifan Yang[34], Bing Yu[35], Faiez Zannad[132], Chaoqun Zheng[28], Genes & Health Research Team, DBDS Genomic Consortium, Folkert W. Asselbergs[2,8,18], Thomas P. Cappola[34], Marie-Pierre Dubé[27,121], Michael E. Dunn[96], Chim C. Lang[46], Nilesh J. Samani[75], Svati Shah[71,133,134], Ramachandran S. Vasan[22,135,141], J. Gustav Smith[136,42,137,141], Hilma Holm[63,141], Sonia Shah[3,141], Patrick T. Ellinor[90,4,138,139,141], Aroon D. Hingorani[1,141], Quinn Wells[140,141], R. Thomas Lumbers[8,6,2,141], HERMES Consortium

Corresponding author: R. Thomas Lumbers (t.lumbers@ucl.ac.uk)

Affiliations

1. Institute of Cardiovascular Science, University College London, London, England, UK
2. Institute of Health Informatics, University College London, London, England, UK
3. Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia
4. Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA
5. Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark
6. Health Data Research UK, London, England, UK
7. British Heart Foundation Data Science Centre, London, England, UK
8. The National Institute for Health Research University College London Hospitals Biomedical Research Centre, University College London, London, England, UK
9. National Heart & Lung Institute, Imperial College London, London, England, UK
10. MRC London Institute of Medical Sciences, Imperial College London, London, England, UK
11. Royal Brompton & Harefield Hospitals, Guy's and St. Thomas' NHS Foundation Trust, London, England, UK
12. Hammersmith Hospital, Imperial College Hospitals NHS Trust, London, England, UK
13. Cardiovascular Medicine unit, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden
14. Pfizer Worldwide Research & Development, 1 Portland St, Cambridge, MA, USA
15. Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

16. Division of Biosciences, University College London, London
17. Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands
18. Department of Cardiology, Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centers, Amsterdam, The Netherlands
19. Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia
20. Geisinger Health System, Danville, PA, USA
21. Department of Cardiology, Herlev Gentofte Hospital, Herlev Ringvej 57, Herlev 2650, Denmark
22. National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, Massachusetts, USA,
23. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
24. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
25. Department of Neurobiology, Care Sciences and Society/ Section of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden
26. School of Health and Social Sciences, Dalarna University, Falun, Sweden
27. Montreal Heart Institute, Montreal Heart Institute, Montreal, Canada
28. Department of Cardiology, The Heart Centre, Copenhagen University Hospital, Rigshospitalet, Ingen Lehmanns vej 7, Copenhagen 2100, Denmark
29. Analytical Genetics, Regeneron Genetics Center, Tarrytown, NY, USA
30. Department of Biostatistics, University of Washington, Seattle, WA, USA
31. Department of Medicine, University of Washington, Seattle, WA, USA
32. Department of Medicine, Division of Cardiology, Emory University Medical Center, Atlanta, GA, USA
33. Department of Cardiology, Erasmus University Medical Center, Rotterdam, The Netherlands
34. Penn Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
35. Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas School of Public Health, Houston, TX 77030, USA
36. Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
37. Department of Clinical Immunology, Odense University Hospital, Odense, Denmark
38. Department of Molecular and Functional Genomics, Geisinger, Danville, PA, USA
39. Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA
40. Harvard Medical School, Boston, MA, USA
41. Department of Biostatistics, University of Liverpool, Liverpool, England, UK
42. Department of Molecular and Clinical Medicine, Institute of Medicine, Gothenburg University and Sahlgrenska University Hospital, Gothenburg, Sweden
43. Faculty of Pharmacy, Université de Montréal, Montreal, Canada
44. MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London, England, UK
45. Vth Department of Medicine (Nephrology, Hypertensiology, Endocrinology, Diabetology, Rheumatology), Medical Faculty of Mannheim, University of Heidelberg, Heidelberg, Germany
46. Division of Molecular & Clinical Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, Scotland, UK
47. Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany
48. DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany
49. Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

50. Department of Medicine, Cardiovascular Division, University of Minnesota, Minneapolis, MN, USA
51. Department of Clinical Sciences, Lund University, Malmö, Sweden
52. Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark
53. Deparment of Clinical Medicine, Health, Aarhus University, Aarhus, Denmark
54. Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA
55. Centre for Primary Care and Public Health, Wolfson Institute of Population Health, Queen Mary University of London, London E1 4NS, England, UK
56. Robertson Center for Biostatistics, Institute of Health and Wellbeing, University of Glasgow, Glasgow, Scotland, UK
57. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands
58. Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
59. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany
60. Department of Cardiology, Rigshospitalet, Copenhagen University Hospital, , Denmark
61. Department of Public Health and Caring Sciences, Geriatrics, Uppsala, Sweden
62. Department of Medicine, Division of Cardiology, University of Maryland School of Medicine, Baltimore, MD, USA
63. deCODE genetics/Amgen Inc., Sturlugata 8, Reykjavik 101, Iceland
64. School of Engineering and Natural Sciences, University of Iceland, Sæmundargata 2, Reykjavik 101, Iceland
65. Center for Individualized and Genomic Medicine Research, Department of Internal Medicine, Henry Ford Hospital, Detroit, MI, USA
66. Division of Cardiovascular Medicine, University of Iowa Carver College of Medicine, Iowa City, IA, USA
67. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
68. Department of Clinical Immunology, Aalborg Univeristy Hospital, Aalborg
69. Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands
70. Einthoven Laboratory for Experimental Vascular Medicine, LUMC, Leiden, The Netherlands
71. Department of Medicine, Division of Cardiology, Duke University Medical Center, Durham, NC 27710, USA
72. Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA
73. Cardiology Section, San Francisco Veterans Affairs Health System, and Departments of Medicine, Epidemiology and Biostatistics, University of California San Francisco, CA, USA
74. Department of Cardiology, Nordsjaellands Hospital, Hilleroed, Copenhagen, Denmark
75. Department of Cardiovascular Sciences, University of Leicester and NIHR Leicester Biomedical Research Centre, Glenfield Hospital, Leicester, England, UK
76. Division of Psychiatry, University College London, London W1T 7NF, England, UK
77. UCL Genetics Institute, University College London, London WC1E 6BT, England, UK
78. Heart and Vascular Institute, Henry Ford Hospital, Detroit, MI, USA
79. Precision Healthcare University Research Institute, Queen Mary University of London, London, England, UK
80. Computational Medicine, Berlin Institute of Health (BIH) at Charité - Universitätsmedizin Berlin, Berlin, Germany
81. MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge CB2 0QQ, England, UK
82. Department of Medicine, University of Massachusetts Chan Medical School, Worcester, Massachusetts, USA,

83. Department of Medical Sciences, Uppsala University, Uppsala, Sweden
84. Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, England, UK
85. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, England, UK
86. University of Ottawa Heart Institute, Ottawa, ON, Canada
87. Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada
88. Department of Medicine, University of Ottawa, Ottawa, ON, Canada
89. Division of Cardiovascular Medicine and Abboud Cardiovascular Research Center, University of Iowa, Iowa City, IA, USA
90. Cardiac Arrhythmia Service and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
91. TIMI Study Group, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Cambridge, MA, USA
92. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1RQ, England, UK
93. Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria
94. Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany
95. Department of Internal Medicine, Clinical Sciences, Lund University and Skåne University Hospital, Malmö, Sweden
96. Cardiovascular Research, Regeneron Pharmaceuticals, Tarrytown, NY, USA
97. Department of Internal Medicine II, Division of Cardiology, Medical University of Vienna, Vienna, Austria
98. Department of Medicine, Turku University Hospital and University of Turku, Turku, Finland
99. National Institute for Health and Welfare, Helsinki, Finland
100. Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands
101. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
102. Division of Population Health and Genomics, University of Dundee, Ninewells Hospital and Medical School, Dundee DD1 9SY, Scotland, UK
103. Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada
104. Thrombosis and Atherosclerosis Research Institute, Hamilton, ON, Canada
105. Population Health Research Institute, Hamilton, ON, Canada
106. Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark
107. Department of Medicine, McMaster University, Hamilton, ON, Canada
108. Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA
109. Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, WA, USA
110. The Institute for Translational Genomics and Population Sciences, Harbor-UCLA Medical Center, Torrance, CA, USA
111. Departments of Pediatrics and Medicine, Harbor-UCLA Medical Center, Torrance, CA, USA
112. Los Angeles Biomedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA, USA
113. BHF Cardiovascular Research Centre, University of Glasgow, Glasgow, Scotland, UK
114. Department of Medicine, Division of Cardiology, University of Pittsburgh Medical Center and VA Pittsburgh HCS, Pittsburgh, PA, USA
115. Department of Epidemiology, University of Washington, Seattle, WA, USA
116. Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research & Development, Seattle, WA, USA
117. Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, Copenhagen, Denmark

118. Department of Medicine, University of Iceland, Sæmundargata 2, Reykjavik 101, Iceland
119. Department of Cardiology, Söderjukhuset, Stockholm, Sweden
120. Department of Clinical Science and Education-Södersjukhuset, Karolinska Institutet, Stockholm, Sweden
121. Faculty of Medicine, Université de Montréal, Montreal, Canada
122. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
123. Department of Internal Medicine, Division of Cardiology, National University Hospital of Iceland, Eiríksgata 5, Reykjavik 101, Iceland
124. Department of Epidemiology and Biostatistics, Aalborg University Hospital, Aalborg, Denmark
125. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands
126. Statens Serum Institut, Artillerivej 5, Copenhagen 2300, Denmark
127. Centre for Genomics and Child Health, Blizard Institute, Queen Mary University of London, London E1 4AT, England, UK
128. Department Life Course and Health, Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
129. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
130. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
131. Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden
132. Université de Lorraine, CHU de Nancy, Inserm and INI-CRCT (F-CRIN), Institut Lorrain du Coeur et des Vaisseaux, Vandoeuvre Lès Nancy 54500, France
133. Duke Clinical Research Institute, Durham, NC 27710, USA
134. Duke Molecular Physiology Institute, Durham, NC 27710, USA
135. Sections of Cardiology, Preventive Medicine and Epidemiology, Department of Medicine, Boston University Schools of Medicine and Public Health, Boston, Massachusetts, USA,
136. Department of Cardiology, Clinical Sciences, Lund University and Skåne University Hospital, Lund, Sweden
137. Wallenberg Center for Molecular Medicine and Lund University Diabetes Center, Lund University, Lund, Sweden
138. Cardiac Arrhythmia Service and Cardiovascular Research Center, Massachusetts General Hospital, Cambridge, MA, USA
139. Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA, The Broad Institute of MIT and Harvard, Cambridge, MA, USA
140. Division of Cardiovascular Medicine, Vanderbilt University, Nashville, TN, USA
141. These authors contributed jointly

Summary paragraph

Heart failure (HF), a syndrome of symptomatic fluid overload due to cardiac dysfunction, is the most rapidly growing cardiovascular disorder. Despite recent advances, mortality and morbidity remain high and treatment innovation is challenged by limited understanding of aetiology in relation to disease subtypes. Here we harness the de-confounding properties of genetic variation to map causal biology underlying the HF phenotypic spectrum, to inform the development of more effective treatments. We report a genetic association analysis in 1.9 million ancestrally diverse individuals, including 153,174 cases of HF; 44,012 of non-ischaemic HF; 5,406 cases of non-ischaemic HF with reduced ejection fraction (HFrEF); and 3,841 cases of non-ischaemic HF with preserved ejection fraction (HFpEF). We identify 66 genetic susceptibility loci across HF subtypes, 37 of which have not previously been reported. We map the aetiologic contribution of risk factor traits and diseases as well as newly identified effector genes for HF, demonstrating differential risk factor effects on disease subtypes. Our findings highlight the importance of extra-cardiac tissues in HF, particularly the kidney and the vasculature in HFpEF. Pathways of cellular senescence and proteostasis are notably uncovered, including *IGFBP7* as an effector gene for HFpEF. Using population approaches causally anchored in human genetics, we provide fundamental new insights into the aetiology of heart failure subtypes that may inform new approaches to prevention and treatment.

Main

Genome-wide association studies (GWAS) of heart failure (HF) and cardiac traits have begun to yield important translational insights, however, existing studies are limited by phenotypic heterogeneity and a lack of stratification by major aetiologies[1–3]. To address these limitations, we performed a GWAS meta-analysis of 153,174 HF cases in 1.9 million ancestrally diverse individuals, including analysis of 4 HF subtypes defined by aetiology and left ventricular ejection fraction (**Supplementary Figure 1**). Multimodal definitions were used to harmonise the ascertainment of heart failure (hereafter referred to as HF), non-ischaemic HF (ni-HF), and non-ischaemic HF stratified by left ventricular ejection fraction (LVEF) below or above 50% (ni-HFrEF and ni-HFpEF respectively), together with corresponding control populations. We characterise the common genetic architecture of HF subtypes and use this information to identify the key effector tissues, pathways and genes underlying aetiology. We map the genetic pleiotropy of HF genomic risk loci to identify aetiological phenotype clusters and use Mendelian randomisation to estimate subtype-specific effects of cardiometabolic traits and diseases. These findings provide new insights into the aetiology of HF subtypes and provide a basis for the development of new therapeutic approaches.

Results

Multi-ancestry genetic association analysis highlights novel heart failure loci

We performed a meta-analysis of case-control GWAS across 42 studies to investigate the association of up to 10,199,961 common genetic variants (minor allele frequency

(MAF) >1%) with risk of four HF phenotypes. The study population comprised 1,946,349 individuals representing five major ancestry groups: European (86.6%), African (0.8%), South Asian (1.5%), East Asian (10.9%), and Hispanic (0.1%). 153,174 HF cases were identified, including 44,012 cases of non-ischaemic HF; 5,406 cases of non-ischaemic HFrEF; and 3,841 cases of non-ischaemic HFpEF (**Supplementary Figure 2**). We identified 59 conditionally independent (sentinel) genetic variants at 56 non-overlapping genomic loci (distance > 500 kilobase pairs) associated with HF at a genome-wide significance ($P < 5 \times 10^{-8}$) (**Figure 1**, **Supplementary Table 1**). Sentinel variants showed consistent allelic effects across ancestry groups, except one variant at the *LPA* locus ($P_{het\text{-}ancestry} < 0.05 / 59$) (**Supplementary Figure 3, Supplementary Table 2**). GWAS of ni-HF, ni-HFrEF, and ni-HFpEF subtypes highlighted 10 additional sentinel variants that were not identified in the HF GWAS. In total, 66 independent genomic risk loci were identified across HF phenotypes[1,4,5] among which 37 have not previously been reported in GWAS of HF[1–5]. Of the 66 loci, 46 (70%) were associated with at least one of the non-ischaemic HF phenotypes at $P < 0.05 / 66$. These loci were then classified as non-ischaemic HF loci, and the remaining 20 (30%) HF loci were classified as secondary HF (s-HF) loci. Amongst the non-ischaemic loci, 17 were associated with ni-HFrEF and 3 were associated with ni-HFpEF at $P < 0.05 / 66$. In addition, we replicated 76 / 87 (87%) previously reported HF sentinel variants that were available in our study ($P < 0.05 / 87$) (**Supplementary Figure 4, Supplementary Table 3)**.

Genetic architecture and heritability heart failure subtypes

The genetic architecture of HF was found to be highly polygenic, evidenced by an elevated genomic inflation factor (λ_{GC} = 1.22) in the absence of population stratification (linkage disequilibrium score [LDSC] regression intercept = 1.01) **(Supplementary Figures 5-6)**, and there was an exponential relationship between allele frequency and effect size for associated variants (**Supplementary Figure 7**). The estimated proportion of variance in disease liability explained by common genetic variants, i.e. SNP-based heritability (h^2_g), was 5.4 ± 0.2% for HF; 6.1 ± 0.5% for non-ischaemic HF; 11.8 ± 2.6% for non-ischaemic HFrEF, and 1.8 ± 1.3% for non-ischaemic HFpEF (**Supplementary Figure 8**)[6]. All HF phenotype pairs demonstrated positive genetic correlations, with estimates ranging from 0.42 (standard error, SE = 0.18) between ni-HFrEF and ni-HFpEF to 0.93 (SE = 0.15) between ni-HF and ni-HFpEF (**Supplementary Figure 9**). The genetic correlation between ni-HFrEF and both HF (0.66, SE = 0.06) and ni-HF (0.74, SE = 0.06) was lower than for ni-HFpEF (0.85, SE = 0.14 for HF and 0.93, SE = 0.15 for ni-HF), suggesting heritable components that are specific to ni-HFrEF.

To explore the potential utility of SNP-based heritability for prediction, we derived a polygenic score (PGS_{HF}) from the HF GWAS excluding UK Biobank (UKB) samples and evaluated the association with HF in UKB. Among 346,667 UKB European participants (13,824 HF cases), the PGS_{HF} was associated with HF (OR per PGS SD 1.37 [95% CI 1.27 to 1.46], $P<2x10^{-16}$) after adjusting for sex, age and first 10 ten genetic principal components (PC)s. Individuals in the top PGS_{HF} decile had a 1.76-fold higher odds of developing heart failure compared to those in the fifth decile (OR = 1.76, 95% CI = 1.64 to 1.89) and 2.89-fold compared to those in the bottom decile (OR = 2.89, 95% CI = 2.66 to 3.14) (**Supplementary Figure 10**).

352

353 **Prioritisation of effector genes for heart failure**

354 To identify HF effector genes, we characterised the functional properties of variants
355 and genes within each identified GWAS locus using a range of orthogonal approaches.
356 Functionally-informed fine-mapping[7] identified 70 credible sets containing 547 likely
357 causal variants (cumulative posterior inclusion probability > 0.95) at 47/66 HF loci.
358 These credible sets included 11 single nucleotide variants with predicted
359 deleteriousness ranked amongst the top 1% in the reference human genome (CADD
360 Phred score >20)[8], including exonic variants in established dilated cardiomyopathy
361 (DCM) genes such as *FLNC*, *BAG3* and *HSPB7*[9–11] (**Supplementary Figure 11,**
362 **Supplementary Table 4**).

363 We then prioritised putative causal genes among 758 protein-coding genes
364 overlapping HF loci by triangulating evidence from three predictors of gene relevance:
365 1) variant-to-gene (V2G) score derived from functional properties of fine-mapped and
366 sentinel variants[12], 2) polygenic priority score (PoPS) calculated from enrichment of
367 gene features[13], and 3) predicted gene expression levels across tissues derived from
368 multi-tissue transcriptome-wide association study (TWAS)[14]. 142 genes were ranked
369 by at least one predictor (labelled as *candidate genes*), of which 71 were further
370 prioritised based on: ranking highest on a combined predictor-score; highest ranking
371 with at least two predictors; colocalisation with gene transcript expression in a relevant
372 tissue[15]; or association with a phenotypically relevant Mendelian disorder (labelled as
373 *prioritised genes*)[16] (**Figure 2a-c**, **Supplementary Table 5**).

The prioritised genes at secondary HF loci included a high proportion with established evidence linking them to atherosclerotic diseases, for example via regulatory effects on cholesterol and lipoproteins metabolism *LDLR*[17], *LPL*[18], *ABCG5*[19,20], *LPA*[21], and *CDKN2A/CDKN2B*[22] (commonly known as 9p21.3 locus). Prioritised genes at the non-ischaemic HF loci included established cardiomyopathy genes *BAG3*, *FLNC*, *ACTN2*, and *HSPB7,* [9–11,23], *CAMKD2* (linked to cardiac hypertrophy[24]), *STRN* (linked to canine DCM[25]), as well as *PITX2*, *KLF12*, and *ATP1B1* which have been associated with cardiac arrhythmia[26–28]. Other notable findings include *CAND2,* a muscle-specific gene of which upregulation has been linked to pathological cardiac remodelling[29] and *NKX2-5*, a core cardiac transcription factor associated with congenital heart disease and cardiomyopathy[30–32].

To facilitate genomic appraisal at each identified locus, we constructed an online dashboard visualisation containing regional genetic association, gene prioritisation, cross-trait association, and study-level estimate (**Supplementary Materials**).

Biological pathways associated with heart failure subphenotypes

To identify convergent molecular disease mechanisms, we performed a gene set (candidate and prioritised) enrichment analysis. We identified 91 enriched biological pathways and cellular phenotypes for HF and 11 for non-ischaemic HF (adjusted $P < 0.05$) (**Figure 3a**, **Supplementary Table 6**). Several pathways involved in adult tissue homeostasis were identified, including regulation of proteostasis, cellular senescence and extracellular matrix remodelling. Putative effector genes of HF and ni-HF were enriched for proteoglycan pathways, including *ERBB2* the target of cancer therapies

that is associated with impaired cardiac function[33,34]. Genes relating to the formation of aggresomes, cellular organelles store misfolded proteins for subsequent disposal throughautophagy[35], were enriched, including *HSBP7* and *BAG3*. BAG3-mediated sarcomeric protein turnover is implicated as a mechanism underlying heart failure[36]. Pathways regulating cell division and senescence were identified, implicating tissue homeostasis and renewal in predisposition to age-associated organ dysfunction, Notably, these included the senescence-associated secretory phenotype, a distinct phenotype with a range of cell-autonomous and non-cell-autonomous effects that re-inforce senescence. *IGFBP7,* a circulating anti-angiogenic factor produced by the kidney and the vasculature, was prioritised within a ni-HFpEF locus. IGFBP7 was recently reported to stimulate cardiomyocyte senescence and cardiac remodelling[37]. Pathways related to cardiac arrhythmia, including regulation of cardiac muscle contraction by calcium ion sensing and cell communication by electrical coupling were specifically enriched for ni-HF. The glucocorticoid-related pathways were enriched for all HF subtypes; glucocorticoids are primary stress hormones that are implicated in heart failure through both glucocorticoid and mineralocorticoid receptor-mediated effects[38]. Genes that were identified in the HF but not the non-ischaemic subtypes were enriched in pathways related to dyslipidaemia.

A map of the organs, tissue and cells mediating heart failure risk

To identify tissues and cell types contributing to HF aetiology, we performed a heritability enrichment analysis across 206 tissues and cell types, representing 12 organs and systems, using sets of uniquely expressed genes (**Figure 3b**,

Supplementary Table 7). 46 unique tissues and cell types were enriched (one-sided $P < 0.05$), highlighting the breadth of tissues and organ systems contributing to HF aetiology. Enrichment of cardiac tissues was the most frequent (15 enrichments), followed by musculoskeletal / connective tissues (12 enrichments), and nervous system tissues (8 enrichments). The most highly enriched tissues for HF, ni-HF, and ni-HFrEF, were cardiac whilst the kidney and pancreas were the most highly enriched for ni-HFpEF, highlighting the relative importance of these organs for specific HF phenotypes. Given the importance of cardiac tissues for ni-HFrEF, we extended the resolution of our enrichment analysis to individual cardiac cell populations, using single nucleus gene expression data from 16 non-failing human hearts[39] (**Supplementary Figure 12a**, **Supplementary Table 8**). Cardiomyocytes were enriched for non-ischaemic HF and HFrEF subtypes, consistent with the pattern observed in the organ-level analysis.

To determine whether HF susceptibility genes were differentially regulated in failing versus non-failing hearts, we investigated cell-type specific expression profiles of the identified HF genes. 30/95 (30%) of ni-HF genes were differentially regulated in failing hearts, predominantly within fibroblasts and cardiomyocytes (**Supplementary Figure 12b-d**, **Supplementary Table 9**). Nuclear Receptor Subfamily 3 Group C Member 1 (*NR3C1*) which encodes glucocorticoid receptors was upregulated in 12 out of 14 cardiac cell types from failing hearts, consistent with the glucocorticoid pathway enrichment among the identified HF genes. These findings support the reported cardioprotective effect of glucocorticoid signalling via inhibition of pathological mineralocorticoid receptor signalling[40] across human cardiac cell types, which may partially corroborate the unexplained beneficial effect of mineralocorticoid receptor

antagonist drugs in HF[41]. Another prosurvival HF gene, *KLF12*, was upregulated in cardiomyocytes and cardiac fibroblasts. KLF12 is a transcription factor that is upstream of p53 and p21 (*CDKN1A*); methylation of the KFL12 transcription factor binding motif has been identified as a mechanism for *LMNA* cardiomyopathy [42]. In contrast to prior reports, we did not observe differential expression of *IGFBP7* in cardiomyocytes from failing hearts [37].

Graph modelling of phenome-wide pleiotropy identifies aetiologic clusters

To identify aetiological clusters underlying HF based on shared genetic aetiology, we tested the associations between sentinel variants at the 66 HF susceptibility loci and 294 diseases in UK Biobank, and modelled the associations using a pleiotropy network graph[43]. We found 207 pleiotropic associations at FDR <1% (corresponding to $P < 0.001$) among 46 / 66 (70%) HF loci and 79 / 294 (27%) unique phenotypes (**Supplementary Figures 13-14, Supplementary Table 10**). We constructed a HF pleiotropy graph network by representing loci and diseases as nodes and locus-phenotype associations as edges (**Supplementary Figure 15a**). Community detection analysis on the network revealed distinct 18 locus-disease aetiological clusters (**Figure 4, Supplementary Figure 15b**, **Supplementary Table 11**). The largest cluster, cluster 1, contained traits and diseases related to atherosclerosis and genes that were predominantly related to secondary HF loci. Cluster 2 comprised cardiac arrhythmias and dilated cardiomyopathy, and genes mostly associated with non-ischaemic HF. Cluster 4 centred on hypertension but contained genes for infection

and cancer, while Cluster 5 comprised genes associated with non-ischaemic HF and ni-HFpEF, and disease including adiposity, diabetes, and carpal tunnel syndrome.

Given the extensive pleiotropic effects of HF susceptibility loci, we further investigated the effects of sentinel variants in each locus on 24 GWAS traits representing major organ systems and image-based cardiac endophenotypes. We identified 270 genotype-phenotype associations at FDR <1%, of which 95 (35%) reached genome-wide significance in the original GWAS (**Supplementary Figure 16**, **Supplementary Table 12**). All tested traits were associated with at least one HF locus at FDR <1%, with coronary artery disease (CAD) and systolic blood pressure (SBP) showing the highest number (25 out of 66 loci / 38%). All lead variants in HF loci that were associated with DCM (6 loci) showed a concordant allelic effect (i.e. additional HF risk-increasing allele was associated with an increased disease risk), whereas all shared associations with HCM (14 loci) showed a discordant allelic effect, consistent with the opposite genetic effects on both traits [44]. Colocalisation analysis revealed evidence of 105 common causal variants across 22 of the 24 traits tested (posterior probability > 0.8) at 42 out of 66 loci (**Figure 5a-c**, **Supplementary Table 14**), adding further evidence for pleiotropic associations and unveiling novel associations. For example, we found colocalisation of the *BACH1* ni-HFrEF locus with both coronary artery disease and dilated cardiomyopathy. *BACH1* is a transcription factor expressed in cardiac fibroblast and endothelial cells in response to mechanical stress and regulates YAP[45]. The BACH1-YAP transcriptional network exerts essential roles in both atherosclerosis and cardiac regeneration through the regulation of proliferative programs in cardiomyocytes[46]. Notably, six HF genes were found to colocalise with glomerular filtration rate (eGFR) or chronic kidney disease (CKD), including prefoldin

490 subunit 1 (*PFDN1*) an important molecular chaperone for protein folding associated
491 with mortality and cardiovascular phenotypes in mouse knockouts[47,48].

492

493 **Genetic appraisal of risk factors across the heart failure spectrum**

494 Next, we sought to explore evidence for causal relationships between risk phenotypes
495 associated with HF in the pleiotropy analysis, using genetic correlation and Mendelian
496 randomisation (MR) analysis (**Figure 6, Supplementary Figure 17-19**,
497 **Supplementary Tables 14-15**). First, we compared the estimated effects of upstream
498 risk factors on HF and non-ischaemic HF. Trait associations were broadly similar
499 between HF and ni-HF, except for coronary artery disease (CAD) for which there was
500 evidence of causal effects on HF [MR odds ratio per doubling prevalence with inverse
501 variance weighted estimator / $OR_{MR\text{-}IVW}$ = 1.20, 95% confidence intervals = 1.18-1.22]
502 but not on non-ischaemic HF [$OR_{MR\text{-}IVW}$ = 1.02, 1.00-1.04]. We found evidence for the
503 causal effects of systolic blood pressure (SBP), body mass index (BMI), and liability
504 to atrial fibrillation (AF) across the HF phenotypic spectrum. SBP had the largest effect
505 on ni-HFrEF [$OR_{MR\text{-}IVW}$ = 1.92, 1.70-2.17 per standard deviation / SD] while BMI had
506 the largest effects on ni-HFpEF [$OR_{MR\text{-}IVW}$ = 1.87, 1.70-2.06 per SD]. Although type 2
507 diabetes and HF shared ~23% genetic risk [squared genetic correlation / r_g^2 = 0.23],
508 there was little evidence of a causal association between the two disorders [$OR_{MR\text{-}IVW}$
509 = 1.04, 1.03-1.05; $OR_{MR\text{-}Egger}$ = 0.98, 0.96-1.00 per doubling prevalence], suggesting
510 that observational associations between T2D and HF largely reflect common upstream
511 disease mechanisms, rather than direct causal effects of T2D on HF. Higher functional
512 lung capacity (forced vital capacity, FVC; forced expiratory volume in 1 second, FEV1)

513 was protective for HF [$OR_{MR\text{-}IVW}$ = 0.93, 0.87-0.94 per SD FVC; $OR_{MR\text{-}IVW}$ = 0.87, 0.83-
514 0.92 per SD FEV1] and ni-HFpEF [$OR_{MR\text{-}IVW}$ = 0.70, 95% CI = 0.61-0.80 per SD FVC;
515 $OR_{MR\text{-}IVW}$ = 0.86, 0.76-0.97 per SD FEV1]. We found evidence for effects of smoking
516 behaviour on HF [$OR_{MR\text{-}IVW}$ = 1.26, 95% CI = 1.19-1.33 per SD pack-years of smoking].
517 Positive associations of higher alcohol consumption with HF [$OR_{MR\text{-}IVW}$ = 1.17, 95%
518 CI = 1.06-1.29 per SD drinks / week] and ni-HFrEF [$OR_{MR\text{-}IVW}$ = 1.87, 95% CI = 1.37-
519 2.55 per SD drinks / week] were observed, consistent with clinical studies relating
520 alcohol consumption to DCM[49].

521 Although renal tissues and traits (glomerular filtration rate, eGFR; chronic kidney
522 disease, CKD) were associated with ni-HFpEF in the heritability enrichment (**Figure**
523 **4**) and colocalisation analyses (**Figure 6**), we found little or no evidence of genetic
524 correlation [r_g = 0.17, SE = 0.05 between HF and CKD; and r_g = -0.001, SE = 0.052
525 between HF and eGFR]. These findings may reflect directional heterogeneity of locus-
526 specific effects which results in a global genetic correlation. For example, there were
527 13 lead variants of HF associated with eGFR at FDR 1%, of which 8 had a concordant
528 allelic effect and 5 had a discordant allelic effect on HF risk (**Supplementary Figure**
529 **16**). Paradoxically, we found suggestive evidence of a protective effect of CKD on ni-
530 HFpEF [$OR_{MR\text{-}IVW}$ = 0.89, 0.81-0.98] which may indicate confounding by competing
531 diagnosis, whereby a CKD diagnosis reduces the likelihood of diagnosis with HF in
532 patients presenting with fluid congestion[50].

533

The relationship between cardiac structure and function and heart failure

Traits and diseases associated with reduced contractile function were found to be associated with risk of all HF phenotypes, except ni-HFpEF (**Figure 6**, **Supplementary Figure 19**). As expected, the strongest effect was observed in ni-HFrEF [$OR_{MR\text{-}IVW}$ = 0.18, 0.12-0.27 per SD LVEF; $OR_{MR\text{-}IVW}$ = 1.83, 1.64-2.03 per doubling prevalence of DCM]. Conversely, traits associated with increased contractility were protective of ni-HFrEF [$OR_{MR\text{-}IVW}$ = 0.88, 0.85-0.91 per doubling prevalence of HCM] and suggestive evidence that higher left ventricular ejection fraction is a risk factor for ni-HFpEF was observed [$OR_{MR\text{-}IVW}$ = 1.28, 0.99-1.67 per SD LVEF]. These findings are consistent with divergent risk effects of subnormal and supranormal contractility DCM / HFrEF and HCM / HFpEF[44,51]. Despite the widely held hypothesis that diastolic dysfunction is the key driver for HFpEF, we found inconsistent evidence to support a causal association [$OR_{MR\text{-}IVW}$ for ni-HFpEF = 1.16, 0.91-1.48 per SD PDSR-r].

Discussion

The processes of evolution to animal life on land led to exquisite systems of body water regulation but an inherent susceptibility to systemic and pulmonary oedema[52]. Syndromes of fluid overload are a common clinical problem described since antiquity and are associated with elevated morbidity and mortality[53]. In the 20th century, cardiac dysfunction was identified as an important cause of fluid overload and formed the basis for the definition of heart failure. Increasingly, however, the importance of non-cardiac tissue in HF aetiology is emerging. By leveraging the de-confounding properties of germline genetic variation, we systematically appraise the aetiology of HF, highlighting

key phenotypic traits, tissue, pathways and genes. Our findings generate new insights into the drivers of risk in particular HF subtypes, including for patients with preserved ejection fraction where therapeutic options are limited.

The insights from our study have several important translational implications. First, our appraisal of cardiovascular risk factor effects on HF subtypes yields several notable findings. Whilst clear effects of epicardial CAD and T2D on HF were observed, no effect on HF patients who had not undergone revascularisation or suffered from myocardial infarction was identified. Similarly, while T2D and HF share common genetic and environmental risk factors, we did not find strong evidence supporting a causal association between T2D and HF. These findings highlight the importance of managing upstream risk factors for diabetes and CAD, for the prevention of HF. SGLT2 inhibitors may exert their beneficial effects on HF risk and progression through modification of upstream physiology for both T2D and HF [54], such as higher body mass index; the effects of which are consistent across subtypes and may be independent of established risk factors[55].

Second, our findings highlight the importance of cardiac contractility traits in HF. Non-ischaemic HFrEF was the most heritable subtype (h^2_{SNP} = 11.8% ± 2.6%) and was highly correlated with DCM (r_g = 0.57 ± 0.12), as expected. We demonstrate opposing associations between contractility-related traits and ni-HFrEF and ni-HFpEF risk: higher baseline contractility decreased the risk of HFrEF, but likely increased the risk of HFpEF. Our findings do not provide positive evidence to support the widely held hypothesis that diastolic function is the major cause of HFpEF.

579 Third, our heritability enrichment analysis suggests that kidney, vascular, and
580 metabolic tissues play a central role in the aetiology of HFpEF. The kidney is the
581 primary organ for managing body fluid, and renal impairment is common in heart
582 failure patients across the phenotypic spectrum. CKD, however, co-occurs more
583 frequently with HFpEF than HF with mid-range or reduced ejection fraction[56]. It is
584 possible that suggestive evidence of a paradoxical protective effect of CKD on ni-
585 HFpEF reflects these diagnoses being alternative diagnoses for overlapping or
586 indistinguishable clinical presentations. These findings motivate future studies to
587 further investigate extra-cardiac drivers of fluid congestion in this patient population.

588 Fourth, our results reveal novel molecular mechanisms, both cardiac and extracardiac,
589 across the heart failure spectrum. Tissue homeostasis emerges as a major
590 mechanism underlying risk across the phenotypic spectrum of heart failure. Cell
591 intrinsic mechanisms, including through modulation of *BACH1-YAP, KLF12-CDKN1A*
592 activity, may influence myocardial tissue homeostasis through effects on cardiac
593 growth. We found enrichment of the senescence-associated secretory phenotype
594 components in pathway analysis and identify *IGFBP7* as a novel prioritised HFpEF
595 gene, reflecting cell-autonomous effects of senescent cells on organismal aging[57].
596 *IGFBP7* is synthesised by the vasculature in response to tissue damage and may be
597 an upstream driver of cell cycle arrest, fibrosis, and vascular rarefaction that are
598 observed in HFpEF. *IGFBP7* is an established clinical biomarker of acute renal failure
599 and may represent a target for novel therapeutics[37,58].

600 In summary, by harnessing the de-confounding qualities of human germline variation,
601 we map the aetiology of the heart failure spectrum, from organs to molecules. Our
602 findings provide genetic evidence to bolster the concept of HFpEF as a multi-system

603 disorder and yield insights into the biological mechanisms underlying subtypes of HF
604 that may inform the development of more effective therapeutics to improve outcomes
605 in those affected.

606

607

Online Methods

Study design

The present meta-analysis included 1,946,349 individuals from 42 studies participating in the HERMES Consortium (40 studies) or publicly released GWAS data of heart failure (Biobank Japan and FINNGEN r3). Sample genotyping was performed locally per study per phenotype per ancestry (one dataset) using high-density genotyping arrays and imputed against Haplotype Reference Consortium[59] (75 datasets), 1000 Genomes Project[60] (6 datasets), TOPMED[61] (15 datasets), or population-specific reference panels (10 datasets).

Study-level GWAS was performed locally per dataset using logistic regression (for prevalent cases) or cox proportional hazard (for incident cases) assuming additive genetic effect with adjustment for sex, age at DNA draw, genetic principal components, and study-specific covariates. All participating studies were ethically approved by local institutional review boards and all study participants provided written informed consent. The meta-analysis was performed centrally at the coordinating centre in accordance with guidelines for study procedures provided by the University College London Research Ethics Committee. Details of study-level participant characteristics, genetic analysis, and cohort description are provided in **Supplementary Tables 16-18**.

Phenotype definition

The present analysis investigated 4 phenotypes: 1) Heart failure (HF), 2) Non-ischaemic HF (ni-HF), 3) Non-ischaemic HFrEF (ni-HFrEF), and 4) Non-ischaemic HFpEF (ni-HFpEF). The HF phenotype includes any diagnosis of HF based on physician's adjudication, hospital record review, or diagnosis codes. Non-ischaemic HF was defined by excluding antecedent ischaemic, valvular, and congenital heart diseases. Non-ischaemic HFrEF was defined as non-ischaemic HF with LVEF <50% based on cardiac imaging or diagnosis of left ventricular systolic dysfunction (LVSD) at any point. Non-ischaemic HFpEF was defined as non-ischaemic HF with LVEF ≥50% based on cardiac imaging without record of LVSD at any point. Phenotyping was performed separately in each participating study, using a harmonised multi-modal phenotyping algorithm (**Supplementary Note**) as a guide. Phenotype definition for FINGENN (release 3)[62] and Biobank Japan[63] follows the original study case definition. Details of study-level phenotype definition are provided in **Supplementary Table 19**.

Study-level summary statistics quality control

GWAS summary statistics from each participating study were processed centrally with a quality control (QC) workflow implemented in *Snakemake*[64] following the procedure described by Winkler T, *et al*[65]. Variants with more than two alleles, regression coefficients (log odds ratio or log hazard ratios) >10, standard error of the coefficients >10, minor allele frequency (MAF) <1%, imputation (INFO) score <0.6, or effective allele count (, calculated as $2 \times MAF \times (1 - MAF) \times N_{sample} \times INFO$, of <50 were excluded from the analysis. Remaining variants with allele frequency difference >0.2

as compared against ancestry-specific reference panels were further excluded. The reference panels were derived from whole-genome sequence data from HRC and 1000 genomes project phase 3 version 5 (1000Gp3v5) for European ancestry, or 1000Gp3v5 for East Asian, South Asian, African, and Admixed American (Hispanic) ancestries. Genomic inflation adjustment was applied for summary statistics with genomic control coefficient (λ_{GC}) > 1.1. In addition, the reported *P* values and quantile-quantile (QQ) plots from all studies were inspected to check for consistency and to identify spurious associations. Full details of the quality control procedure and results are provided in **Supplementary Materials.**

Genome-wide association meta-analysis

Study-level GWAS results were meta-analysed using a fixed-effect inverse variance weighted model implemented in METAL[66]. For each phenotype, we calculated variant-level and summary-level effective sample size as $4 / (1/N_{case} + 1/N_{control})$ where N_{case} and $N_{control}$ represent the number of cases and controls included in the meta-analysis. Variants with an estimated overall MAF <0.01, variant-level effective sample size < 10% of the meta-analysis effective sample size, or reported in only one study were further removed. The final results consist of genetic association estimates of 10,199,961 common genetic variants (MAF >1%) with HF; 9,414,975 variants with non-ischaemic HF; 9,198,919 variants with non-ischaemic HFrEF; and 8,277,415 variants with non-ischaemic HFpEF.

673 **Cross-ancestry allelic effect heterogeneity assessment**

674 To account for a possible bias due to heterogeneity of allelic effect across ancestries,
675 we performed a sensitivity analysis using a meta-regression technique to model
676 ancestry-specific allelic effect by incorporating axes of genetic variation derived from
677 genome-wide metrics of diversity between populations, as implemented in MR-
678 MEGA[67]. Using this technique, we also estimated overall *P* value for associations
679 accounting for allelic effect heterogeneity across ancestries, *P* value for heterogeneity
680 correlated with ancestry, and *P* value for residual heterogeneity. Concordance
681 between *P* values from the fixed-effect meta-analysis with METAL and those from MR-
682 MEGA were checked by calculating Pearson's correlation coefficient and plotting the
683 two vectors on a two-dimensional plane across conditionally independent variants.
684 Given the limited non-European ancestry sample for non-ischaemic HF subtypes, this
685 comparison was only performed using the overall HF phenotype.

686

687 **Identification of genomic susceptibility loci across HF subtypes**

688 Genomic susceptibility loci for HF were identified using sets of conditionally
689 independent variants identified through a chromosome-wide stepwise conditional-joint
690 analysis using Genome-wide Complex Trait Analysis (GCTA) software[68]. Conditionally
691 independent variants across 4 HF phenotypes with joint *P* value $< 5 \times 10^{-8}$ that are
692 physically located within 500 kilobases of each other were aggregated into one set. A
693 genomic locus was then defined as the genomic region within 500 kilobases upstream
694 and downstream of the farthest variants of each aggregated set of conditionally
695 independent variants.

The identified genomic loci were labelled with an incremental one-based integer sequence based on phenotype order (HF, ni-HF, ni-HFrEF, and ni-HFpEF), chromosome, and base pair positions. A genomic locus was declared as novel if all conditionally independent variants within the locus and any of the sentinel variants reported at $P < 5 \times 10^{8}$ in previous GWAS of HF[1–5] were physically located more than 250 kb away and not in LD ($R^2 < 0.2$). To model LD between genetic variants, we used a reference panel derived from individual-level imputed genotype data of a randomly sampled 10,000 UK Biobank participants with an ancestry composition proportionally matched to the present meta-analysis (Hispanic ancestry was not included due to unavailability in UK Biobank and small proportion in the overall sample).

Further, association between a locus and a HF subtype was categorised based on *P* values at cut-off values of $P < 5 \times 10^{-8}$ (genome-wide significant), $P < 0.05$ / number of identified loci (replicated at Bonferroni-adjusted threshold), $P < 0.05$ (nominally significant), and $P \geq 0.05$ (no evidence of association). For presentation, loci associated with any non-ischaemic HF subtype at Bonferroni-adjusted or genome-wide significance thresholds were labelled as non-ischaemic HF loci, and the remainings were labelled as secondary HF loci. Non-ischaemic HF loci that are associated with ni-HFrEF at Bonferroni-adjusted or genome-wide significance thresholds were labelled as ni-HFrEF loci, and the remainings were labelled as ni-HFpEF loci.

Genetic architecture assessment

To assess genetic architecture and polygenicity across HF phenotypes, we compared quantiles of the expected and observed genome-wide genetic association *P*-values using quantile-quantile (QQ plot) and calculated genomic control coefficient (λ_{GC}). An early inflation on QQ plot with λ_{GC} > 1.1 suggests genome-wide genomic inflation, which may be due to polygenic genetic architecture or confounding by population stratification. To distinguish between the two, we calculated λ_{GC} assuming 1000 participants ($\lambda_{GC-1000}$), and estimated LD score regression slope using LDSC software[69], whereby values close to 1 indicate that any observed genomic inflation is likely due to polygenic genetic architecture. To minimise confounding by population structure, LDSC regression was performed using European ancestry meta-analysis subset and LD reference panel. We investigated the distribution of allelic effect on HF across the spectrum of allele frequency by plotting the risk ratio per additional minor allele as a function of minor allele frequency of conditionally independent variants. To assess the shape of the relationship, we fitted two separate local polynomial regression with locally estimated scatterplot smoothing (LOESS) for groups of variants associated with increased and decreased risk of HF. To increase precision of the fitted regression, a sub-genome-wide threshold association at a false discovery rate (FDR) <1%, estimated using the *qvalue* package in R[70], was used to identify conditionally independent variants.

738

739 **SNP-based heritability estimation**

740 The proportion of variance in HF risk explained by common SNPs, i.e. SNP-based
741 heritability (h^2_{SNP}), was estimated from GWAS meta-analysis summary statistics using
742 Linkage-Disequilibrium Adjusted Kinships (LDAK) SumHer software[71] with LDAK-Thin
743 and BLD-LDAK heritability models[72]. The h^2_{SNP} estimates were calculated on a liability
744 scale, which assumes that a binary phenotype has an underlying continuous liability,
745 and above a certain liability threshold an individual becomes affected[73]. To obtain a
746 more accurate estimate, we used the European ancestry meta-analysis summary
747 statistics with pre-computed tagging files derived from 2,000 individuals White British
748 individuals, and used a correction for sample prevalence by calculating the effective
749 sample size assuming an equal number of cases and controls[74]. The conversion to
750 liability scale was calculated using population prevalence derived from the total
751 number of cases divided by the total number of cases and controls included in the
752 meta-analysis.

753

754 **Polygenic risk score analysis**

755 To explore the potential utility of SNP-based heritability for prediction, we explored the
756 association between polygenic score of HF (PGS_{HF}) and HF risk in 346,667 UKB
757 European participants, including 13,824 prevalent and incident HF cases. The PGS_{HF}
758 was constructed as a weighted sum of allelic count of 1,012,059 genetic variants
759 selected using the LDpred2-auto model[75], with weights derived from the present

GWAS meta-analysis of HF excluding UK Biobank. Odds ratio of HF per standard deviation of PGS_{HF} was estimated using logistic regression with binary HF status as response variable, standardised PGS_{HF} as predictor, and sex, age and first 10 ten genetic principal components (PC)s as covariates. To estimate risks of HF in individuals with high PGS_{HF}, we grouped participants into deciles of PGS_{HF}, and calculated odds ratios of the top decile as compared against the fifth and the first (bottom) decile using logistic regression.

Fine mapping of causal variants and functional assessment

We performed functionally-informed fine-mapping using PolyFun[7] and SuSiE[76] to prioritise likely causal variants in each identified genomic locus for HF. We used precomputed functionally-informed prior causal probabilities of 19 million imputed UK Biobank SNPs with MAF>0.1%, based on a meta-analysis of 15 UK Biobank traits[7] and genome-wide association estimates from the current analysis to calculate prior causal-probability proportional to per-SNP heritability of phenotype under analysis. The resulting estimates were passed on to the sum-of-single-effects fine-mapping model implemented in SuSiE, assuming at most 5 causal variants per locus, to calculate per-SNP posterior inclusion probability (PIP) and to construct 95% credible sets of likely causal variants. To minimise bias due to differential ancestry composition, we used effect estimates from meta-analysis of European ancestry GWAS and LD reference panel from 10,000 randomly selected UK Biobank participants of European ancestry. To assess functions and consequences of fine-mapped variants, we extracted variant-level information on nearest gene(s), genic functions, and Combined

783 Annotation-Dependent Depletion (CADD)[8] Phred score from ANNOVAR[77] and
784 OpenTargets Genetics[12].
785
786 **Prioritisation of effector genes for HF**
787 To identify effector genes for HF, we implemented a two-step in silico gene
788 prioritisation approach. In step 1, we identified *candidate* gene set using a combination
789 of three *predictors*:
790 1. Polygenic priority score (PoPS)
791 Polygenic enrichment of gene features derived from cell-type specific gene
792 expression, biological pathways, and protein-protein interactions[13]
793 2. Variant to gene score (V2G)
794 Highest V2G score across variants in the 95% credible set and variants in high
795 LD (r^2 >0.8) with the conditionally independent variants within the locus. The
796 V2G score was extracted from the OpenTargets Genetics[12,78], and was derived
797 from in silico functional prediction, expression and protein quantitative trait loci,
798 chromatin interactions, and distance to gene's canonical transcription start site
799 portal
800 3. Transcriptome-wide association study (TWAS)
801 *P* value for association of overall predicted gene expression across tissues with
802 HF calculated using S-MulTiXcan[14] with GTEx[79] v8 MASHR models.
803 Genes with highest PoPS, highest V2G score, or lowest TWAS *P* value within a locus
804 were considered as *candidate* effector genes for HF. In step 2, we prioritised these
805 genes using 3 boolean (True / False) *classifiers*:
806 1. Top ≥2

Whether a gene ranked top by 2 of the 3 *predictors* (PoPS, V2G, TWAS)

2. Mendel

   Whether a gene is associated with at least one Mendelian disease term that is enriched at FDR <1% as estimated from Mendelian gene set enrichment analysis using MendelVar[16].

3. Coloc

   Whether gene expression level in tissue with lowest *P*-value from the multi-tissue TWAS analysis colocalised with at least 1 HF phenotype under study (posterior probability of shared causal variants >0.8). The colocalization analysis was performed with the R *coloc* package allowing for multiple causal variants[15] using gene expression data from GTEx v8[79].

In addition, we derived an *Overall* predictor score based on weighted average of PoPS, highest V2G, and -$\log_{10}$ *P* MulTiXcan values with 2:2:1 weight ratio, scaled to 0-100 value using a quantile transformation with uniform output distribution as implemented in python *scikit-learn*[80] library. Finally, for each locus, we ranked genes based on total classifier score (sum of *True* values) and the overall predictor score. Genes that are top-ranked or have a total classifier score ≥2 were *prioritised* as most likely effector genes for HF.

Analyses which require GWAS summary statistics (PoPS, TWAS, and Coloc) were performed separately for each of the 4 HF phenotypes using meta-analysis results from the European subset. Locus-specific results for gene prioritisation were extracted from the overall HF phenotype for loci with $P_{\text{HF-GWAS}} < 5 \times 10^{-8}$ (locus 1-56), or from HF phenotype with the lowest $P_{GWAS} < 5 \times 10^{-8}$ for subtype-specific loci (ni-HF for locus 57-61, ni-HFrEF for locus 62-64, ni-HFpEF lor locus 65-66).

831

832 **Heritability enrichment analysis**

833 To identify relevant tissues and cell types involved in HF pathology, we estimated
834 heritability enrichment of specifically expressed genes using LDSC-SEG[81]. We used
835 specifically expressed gene sets for 206 tissues and cell types from GTEx and Franke
836 Lab[82,83], classified into 12 organ / system categories (**Supplementary Table 20**). For
837 each set, LDSC-SEG tested the enrichment for per-SNP heritability attributed to the
838 given set, conditional on the set that includes all genes and a baseline model
839 consisting of 52 genomic annotations (including genic regions, enhancer regions and
840 conserved regions) using stratified LDSC regression framework[84]. A positive
841 regression coefficient, statistically tested using a one-sided *P*-value test, represents a
842 positive contribution of a given tissue / cell type to trait heritability. For this analysis,
843 we used the European ancestry meta-analysis subset with pre-computed LD score
844 weights derived from the 1000 Genomes European reference panel[84].

845 Further, based on the observation that heart tissues are the major contributors to SNP
846 heritability in almost all HF phenotypes, we extended the heritability enrichment
847 analysis to 15 cell types derived from single-nucleus RNA sequencing (snRNA-seq)
848 experiment on 185,185 cell nuclei from 16 non-failing human heart donors[39]
849 (**Supplementary Table 21**). To identify specifically expressed genes for each cell
850 type, we followed the approach described in Finucane et al. (2018)[81] by taking top
851 10% genes with the highest *t*-statistic for specific expression in a given cell type
852 (compared to other cell types). The sets of specifically expressed genes for each
853 cardiac cell type were used to perform the heritability enrichment analysis with LDSC-
854 SEG as described above.

855

856 **Single-nucleus differential gene expression in failing vs. non-failing heart**

857 To assess the extent to which transcriptional pattern of risk genes for HF changes in
858 failing heart, we performed single-nucleus differential gene expression analysis of
859 candidate effector genes within HF loci by comparing transcription level in cell nuclei
860 from non-failing heart as described above with cell nuclei from 28 failing heart donors
861 diagnosed with end-stage dilated cardiomyopathy (DCM) or hypertrophic
862 cardiomyopathy (HCM)[39].

863 Samples were collected from Myocardial Applied Genetics Network (MAGNet;
864 www.med.upenn.edu/magnet), with snRNA-seq processed using CellBender[85] and
865 Cell Ranger[86] as previously described[39,87,88]. Where available, for each gene in a given
866 cell type, we computed statistics for differential gene expression between failing and
867 non-failing samples using limma–voom[89,90] model adjusting for age and sex. To
868 account for the correlation in expression among nuclei from a given individual, we
869 summed counts for genes across nuclei for each patient within each cell type, requiring
870 a minimum of 25 nuclei. We excluded mitochondrial genes and ribosomal genes,
871 removed genes in less than 1% of nuclei in the given cell type of both groups being
872 compared, and applied an additional filter for lowly expressed genes using the
873 filterByExpr(group=group) function. Two-sided P-values were calculated and A
874 Benjamini–Hochberg correction was applied for multiple testing correction.

875 A gene was declared to be differentially expressed if it survived multiple-testing
876 correction at FDR-adjusted P-value <0.01 and showed a concordant differential
877 expression sign in both CellBender and Cell Ranger quantifications and had no

878 background contamination as estimated in CellBender. To test whether differentially
879 expressed genes in a given cell type are overrepresented amongst prioritised HF
880 genes, we performed a Fisher exact test implemented in *R*, using all genes tested in
881 the given cell type as the background set.

882

883 **Pathway enrichment analysis**

884 To identify biological pathways and processes that are relevant to HF pathology, we
885 tested for overrepresentation of biological terms in prioritised genes for HF using
886 g:Profiler[91]. To account for aetiological differences underlying HF subtypes and
887 uncertainty in gene prioritisation, we tested 6 gene sets constructed from combinations
888 of 3 phenotypic classifications of loci (all loci, secondary HF loci, non-ischaemic HF
889 loci); and 2 gene prioritisation categories (*Candidate* gene set: genes that are ranked
890 top by PoPS, TWAS, or V2G predictor and *Prioritised* gene set: candidate genes with
891 classifier score ≥2 or highest overall predictor score). For each gene set, we performed
892 an unordered enrichment analysis of biological terms in Kyoto Encyclopedia of Genes
893 and Genomes (KEGG)[92], Reactome (REAC)[93]; Wiki Pathways (WP)[94]; and Gene
894 Ontology (GO)[95,96]. For presentation (**Figure 3a**), we excluded terms with more than
895 2,000 genes (representing ~10% of protein coding genes in the human genome) and
896 included terms with adjusted $P < 0.05$ following g:Profiler correction for multiple testing.
897 GO terms were summarised by highlighting only driver terms as identified by the
898 g:Profiler clustering algorithm. Terms that are enriched in both *candidate* and
899 *prioritised* gene sets were collapsed by presenting the median enrichment *P* value on
900 $-\log_{10}$ scale.

Phenome-wide association studies in UK Biobank

To characterise pleiotropic effects of HF genetic loci across human diseases, we performed a phenome-wide association study (PheWAS) by testing the associations of the lead variants at 66 HF loci with 294 disease phenotypes in 408,480 UK Biobank participants of European ancestries. Phenotypes were derived from clinical events recorded in data from linked hospital admission, death certificate, primary care visit, and self-reported cancer diagnosis, non-cancer diagnosis, and procedure history; using expert-curated phenotype definition and category described elsewhere[43]. For a given phenotype, participants were classified as cases if they ever had a record of at least one matching code in any data source; otherwise as controls. For each phenotype - lead variant pair, we ran a case-control genetic association analysis implemented in PLINK2, requiring at least 100 cases and 100 controls per phenotype. The association analysis was performed using a logistic regression model assuming an additive allelic effect and adjusted for sex, genotyping array, and first ten genetic principal components. For presentation, allelic effect estimates were aligned to represent the effect of one additional HF risk-increasing allele.

Pleiotropy network analysis

Using results from the identified genotype-phenotype associations from the locus PheWAS, we performed a network analysis to evaluate the connections between HF genetic loci based on their pleiotropic effects across human diseases. First, we

924 constructed a network dataset with nodes (vertices) representing loci or phenotypes;
925 and edges representing unidirectional associations of the locus lead variant with a
926 given phenotype. The resulting network was then visualised using the Davidson-Harel
927 layout algorithm[97]. To help with interpretation, edges on the graph were coloured by
928 the effect direction of HF risk-increasing allele, with thickness of the line representing
929 absolute Z statistics for the association. Phenotype nodes were labelled by disease
930 name and coloured by disease category; whereas locus nodes were labelled by most
931 likely effector genes. Eigenvector centrality measure was calculated to estimate node
932 influence within the network (**Supplementary Figure 20**, **Supplementary Table 22**),
933 and then mapped to the node size so that a more influential node appears larger.

934 Further, we performed a data-driven community detection analysis on the graph using
935 a walktrap algorithm[98] weighted by absolute Z score of the edges to identify clusters
936 of densely connected nodes. Overall modularity[99] of the resulting clusters was
937 calculated to quantify the strength of the division, where a higher value reflects a
938 connection between nodes that is denser for nodes within the same cluster but sparser
939 for nodes belonging to different clusters. The network analysis and visualisation was
940 performed using the *igraph*, *tidygraph* and *ggraph* package in R.

941

942 **Locus-specific pleiotropy assessment**

943 We investigated the local pleiotropic effects of sentinel variants in 66 genetic loci on
944 24 HF-relevant GWAS traits (**Supplementary Table 23**). Sentinel variants were
945 defined as conditionally independent variants showing lowest *P*-value for association
946 with HF phenotypes within a locus. For loci associated with both all-comer HF and at

least one non-ischaemic HF phenotype, sentinel variants were taken from all-comer HF. Estimates for sentinel variant that were not reported in a target GWAS were substituted with those of a proxy variant with highest LD r^2 (requiring at least LD r^2 >0.8), identified using *plink*[100] with LD reference panel derived from 10,000 random samples of UK Biobank European participants. For presentation (**Supplementary Figure 15**), effect estimates were converted to Z-scores derived from the regression coefficients (betas) divided by their standard errors, with direction of effect aligned to reflect the additive effect of HF-risk-increasing allele. Loci and GWAS traits were ordered using hierarchical agglomerative clustering with average linkage method, implemented with the *hclust* function in R 4.2.0. Distances between loci / traits for clustering were calculated with the Euclidean method using vectors of absolute Z-scores (reflecting allele-agnostic effects across tested loci / traits).

Cross-trait colocalisation

To test whether locus-specific genetic associations with HF and a given trait derived from common causal variants, we performed pairwise cross-trait colocalisation analysis implemented in R *coloc* package[15], allowing multiple causal variants per locus by integrating Sum of Single Effects (SuSiE) regression framework[76]. This approach first runs SuSiE to detect credible set(s) L_1 and L_2 corresponding to trait 1 and trait 2 respectively; followed by colocalisation to estimate posterior probabilities of common causal variants (hypothesis 4, H_4) for each pairwise combination of elements of L_1 x L_2[15]. To run *coloc*, we used the default marginal prior probabilities for association with trait 1 (p_1) and trait 2 (p_2) of 10^{-4}, and a prior probability of joint causal association (p_{12})

of 10^{-5}. A posterior probability of H_4 ($PP_{coloc\ H4}$) >0.8 in at least one pair of credible sets was considered as evidence of shared causal variants. The colocalisation analysis was performed for each identified HF locus using an overlapping set of variants available in both GWAS of HF and trait under analysis situated within the locus. Genetic association estimates for HF were extracted from GWAS of the overall HF phenotype for loci with $P_{HF} < 5 \times 10^{-8}$ (locus 1-56); or otherwise from GWAS of non-ischaemic HF phenotypes for subtype-specific loci: ni-HF for locus 57-61, ni-HFrEF for locus 62-64, and ni-HFpEF for locus 65-66.

Genetic correlation

To assess the extent to which genetic associations are shared amongst HF subtypes, we performed genetic correlation analysis across pairs of 4 HF subtypes under analysis, and genetic correlation between pairwise combinations of each HF subtype and 24 HF-related GWAS traits. The genetic correlation analysis was performed using bivariate linkage disequilibrium score (LDSC) regression[101] with input GWAS summary statistics taken from the present study (for HF phenotypes) and publicly available data (for other traits). Genetic correlation of HF and other binary traits was estimated on a liability scale, with sample prevalence estimated as the number of cases divided by total number of samples in the GWAS and population prevalence taken from literatures if available, or assumed to equate the sample prevalence otherwise (**Supplementary Table 23**).

Mendelian randomisation

We estimated the causal effects of each of the 24 HF-related traits (**Supplementary Table 23**) on each of the 4 HF subtypes analysed in the present study using two-sample Mendelian randomization (MR) as implemented in the *MendelianRandomization* package[102]. MR instruments for the exposure traits (phenotypes other than HF) were selected from genetic variants available in both GWAS of exposure and outcome (HF phenotypes) traits using LD-based clumping algorithm with *P* value threshold of 5×10^{-8} and LD r^2 threshold of 0.05 implemented in PLINK[100]. For each exposure trait, we estimated the causal association with each HF phenotype using inverse-variance weighted (IVW) MR estimator; and performed sensitivity analyses with MR-Egger and weighted median estimators (WME)[103,104]. Traits that survived the multiple testing adjustment at FDR <1% in IVW analysis and showed consistent direction of effect in sensitivity analyses were considered causal.

Multiple testing adjustment

Where mentioned in the text, we performed multiple testing adjustments by controlling false discovery rate (FDR) using the Benjamini-Hochberg procedure[105] as implemented in the *p.adjust(method = "BH")* function in *R*. For each analysis, adjusted *P*-values less than a tolerable type-I error rate (alpha) of 0.01, corresponding to false discovery rate of 1%, were considered to survive the multiple testing correction.

Figures

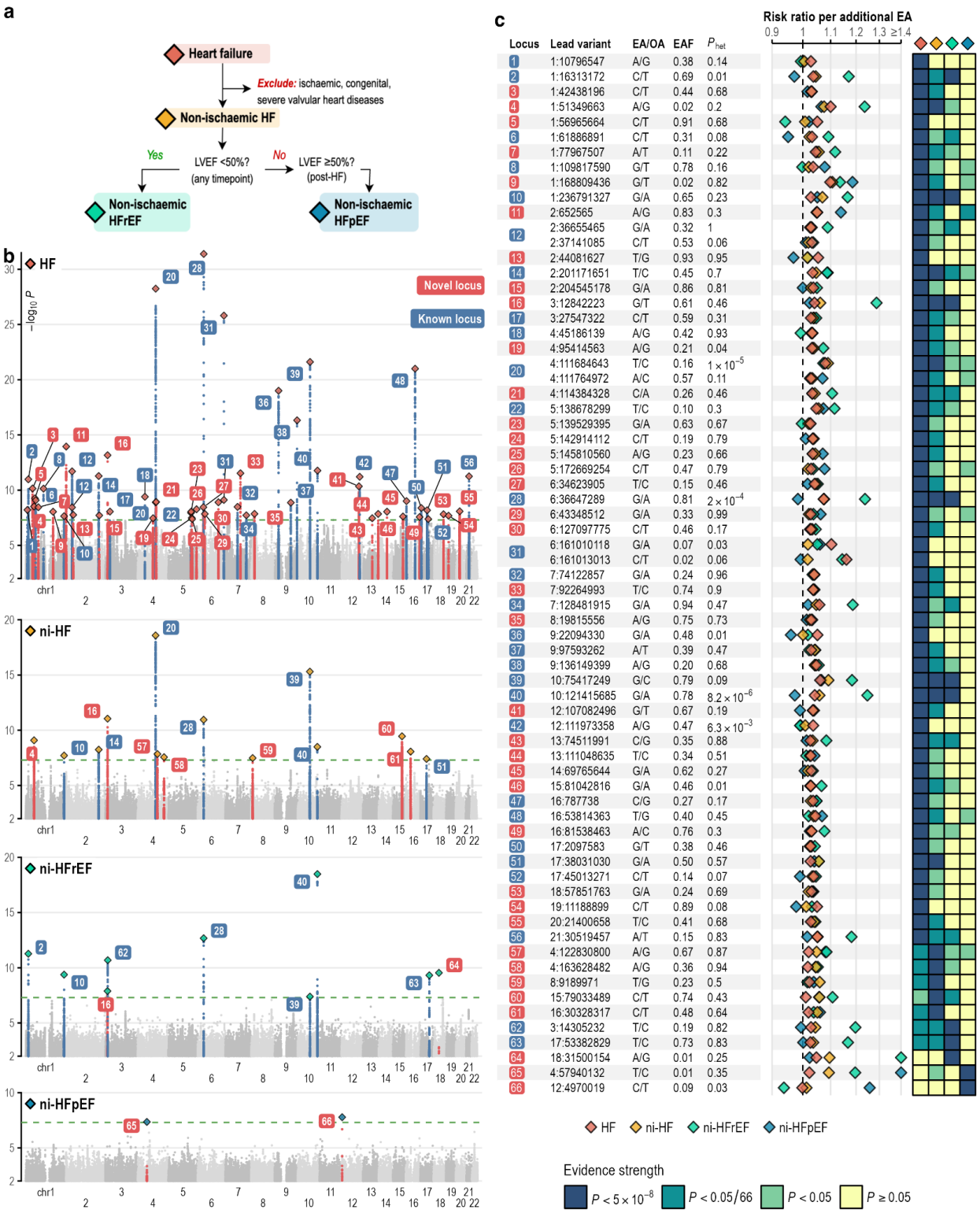

Figure 1.a. Phenotyping schema **b.** Manhattan plots of 4 HF subtypes. **c.** Summary of conditionally independent lead variants across HF phenotypes. Lead variants are denoted using chromosome and base pair position according to GRCh37 assembly. RR = risk ratio, EA = effect allele, aligned to risk-increasing allele in HF phenotype with lowest *P* value for association; OA = other (non-effect) allele; EAF = effect allele frequency, P_{het} = *P* value for effect estimate heterogeneity across studies. EAF and P_{het} estimates were taken from the first HF phenotype with genome-wide significant lead variant.

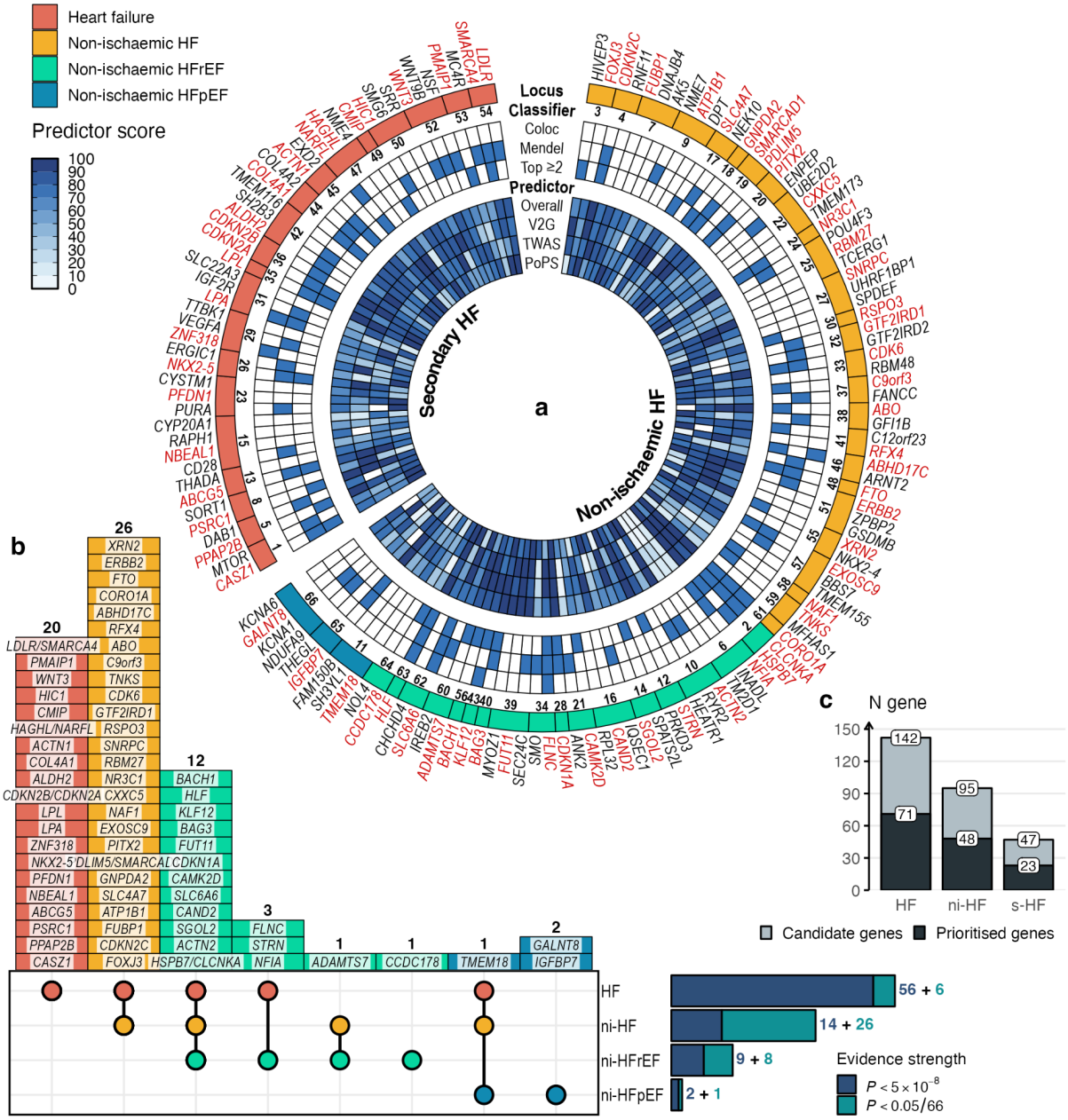

Figure 2. a. Sunburst chart of Candidate genes (black label) and Prioritised genes (red label) for HF identified through gene prioritisation strategy across 66 loci showing heatmap of predictor score and boolean classifier score (blue tile indicates a 'True' value). **b.** Upset plot showing number of loci categorised by phenotypic associations across HF subtypes, colour coded by the most specific association at $P < 0.05 / 66$ and labelled with Prioritised genes. **c.** Number of Candidate and Prioritised genes across HF phenotype gene set.

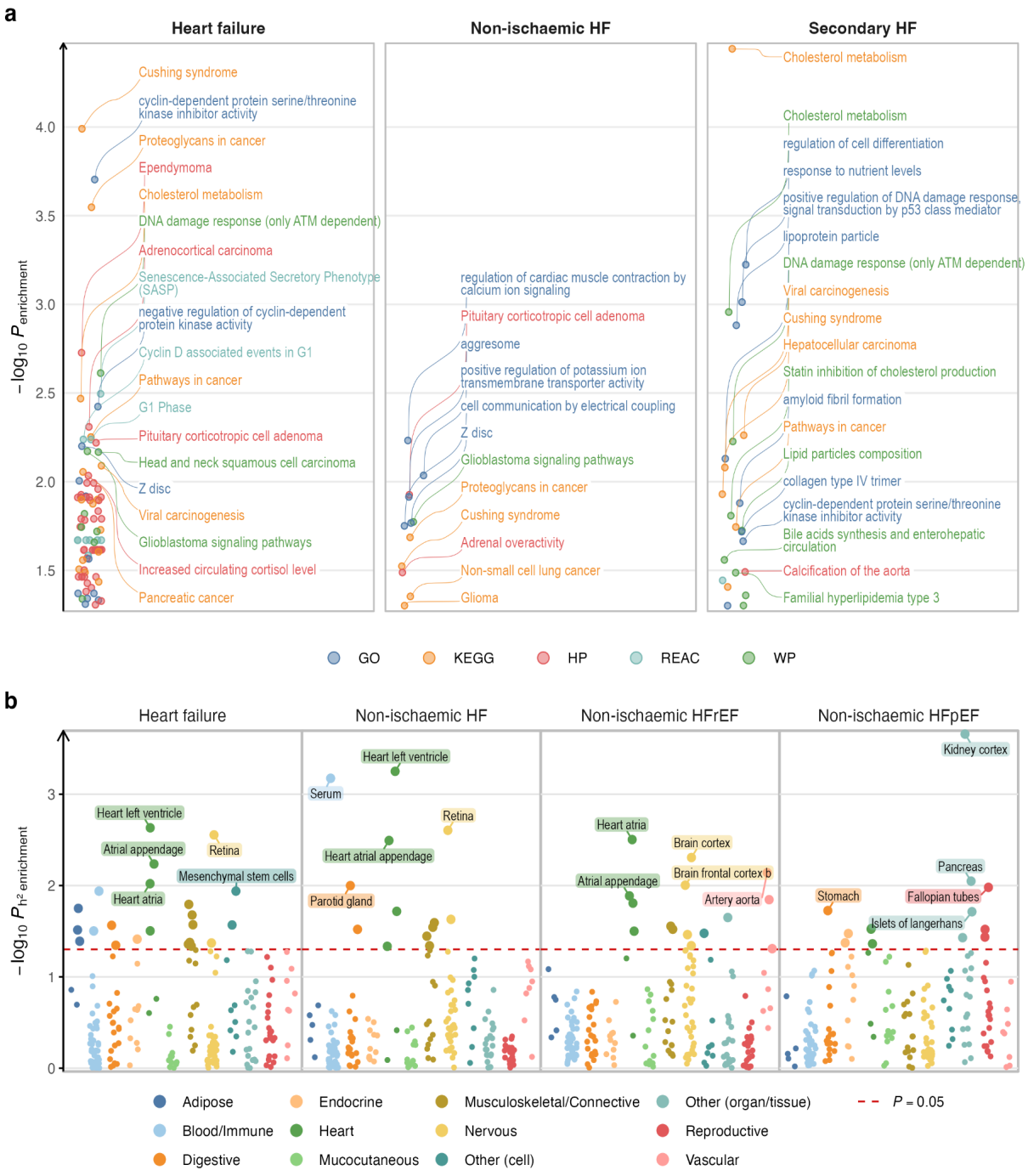

Figure 3. a. Enriched terms (Penrichment < 0.05) from Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Human Phenotype Ontology (HP), Reactome (REAC), and Wiki Pathways (WP). Up to top 20 enriched terms in each HF phenotype gene set are labelled. **b.** Heritability enrichment of 206 tissues and cell types, classified into 12 system / organ categories, across 4 HF phenotypes. Top 5 tissues / cell types per phenotype are labelled.

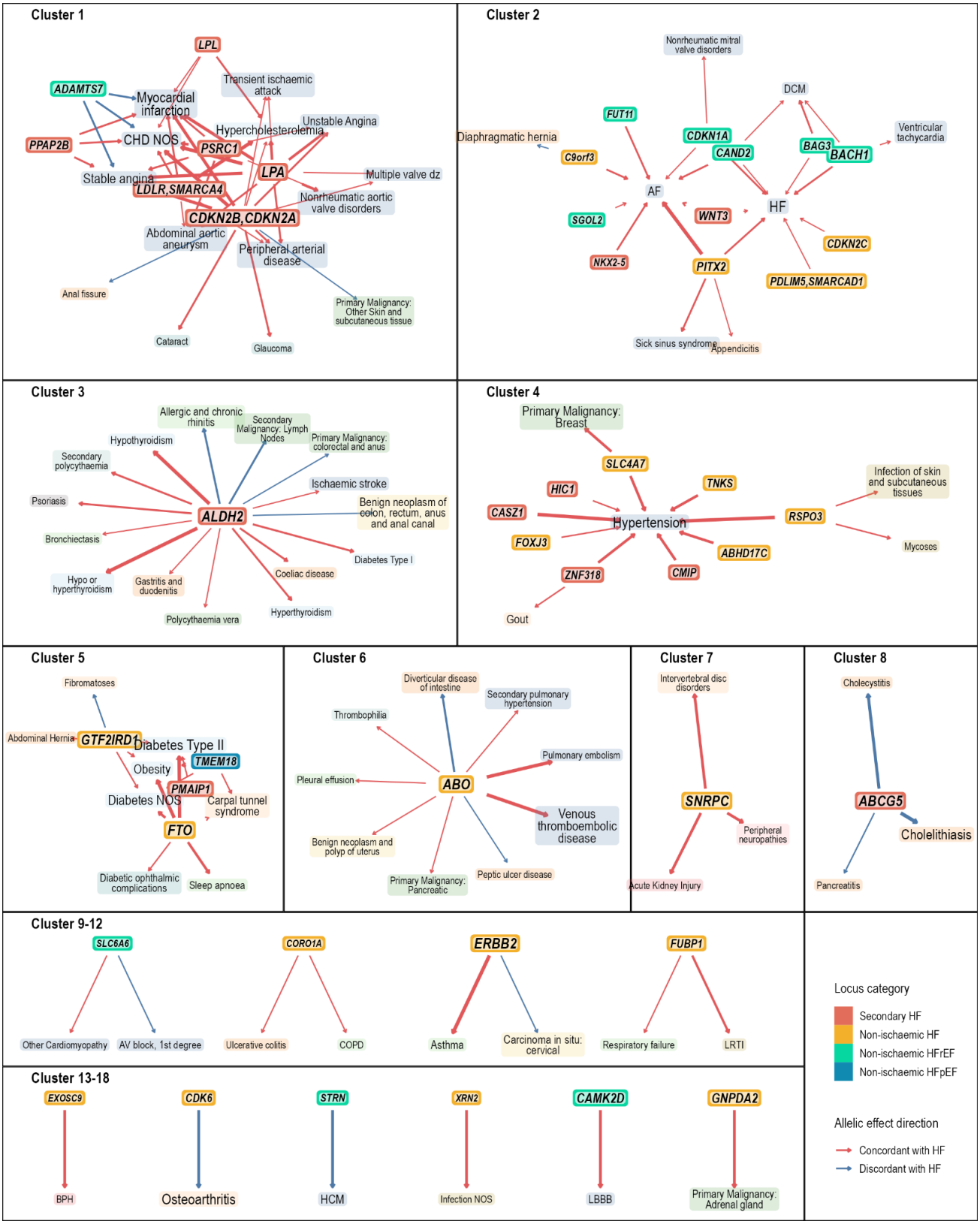

Figure 4. Aetiologic clusters identified through locus-phenotype pleiotropy network analysis of the phenome-wide association estimates. Nodes represent loci (solid background) and phenotype (translucent background), with size representing centrality measure. Edges represent locus-phenotype association, with thickness representing strength of association measured by absolute Z score.

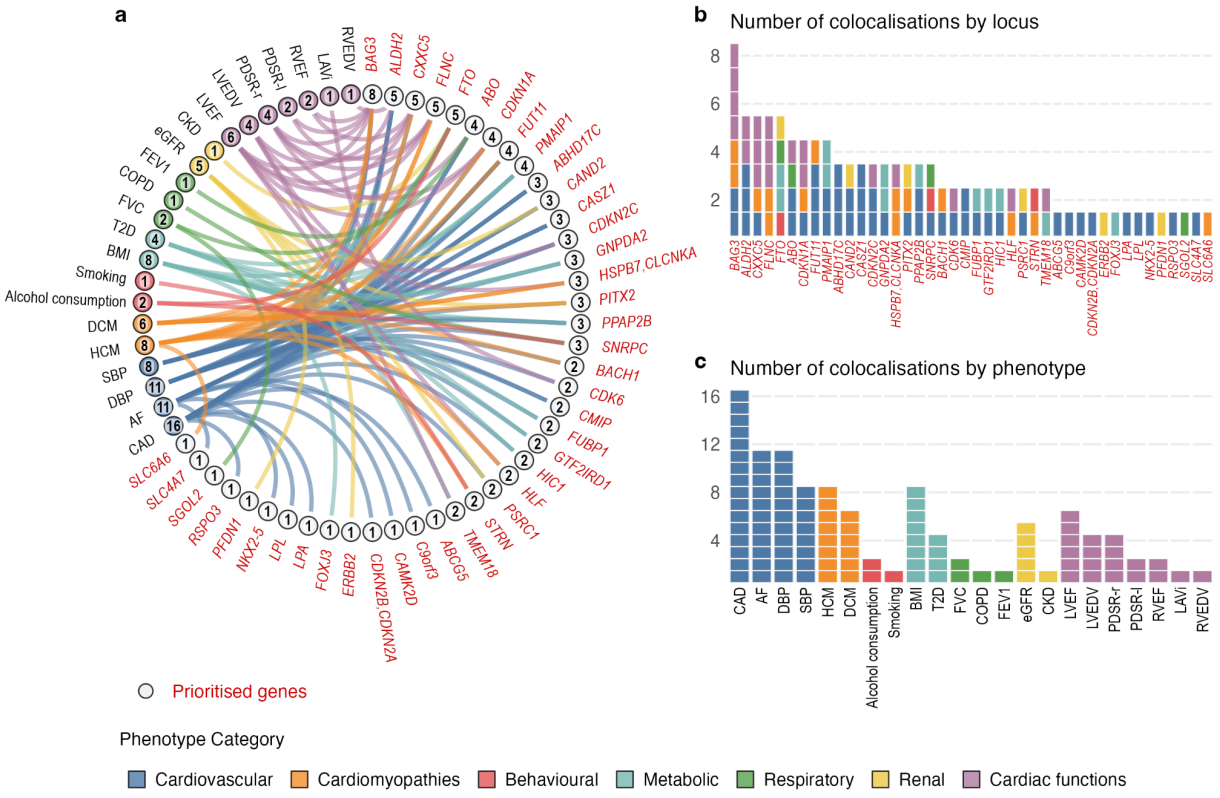

Figure 5. a. Colocalisations between HF and 22 GWAS phenotypes (out of 24 tested) across 42 (out of 66) loci with posterior probability of shared causal variants ($PP_{coloc\ H4}$) > 0.8. Each band connects a pair of locus - phenotype nodes, and represents sharing of causal variants between the phenotype and HF at the locus. Annotated number represents total number of colocalisations by locus and phenotype, which is displayed in panel **b. & c.**

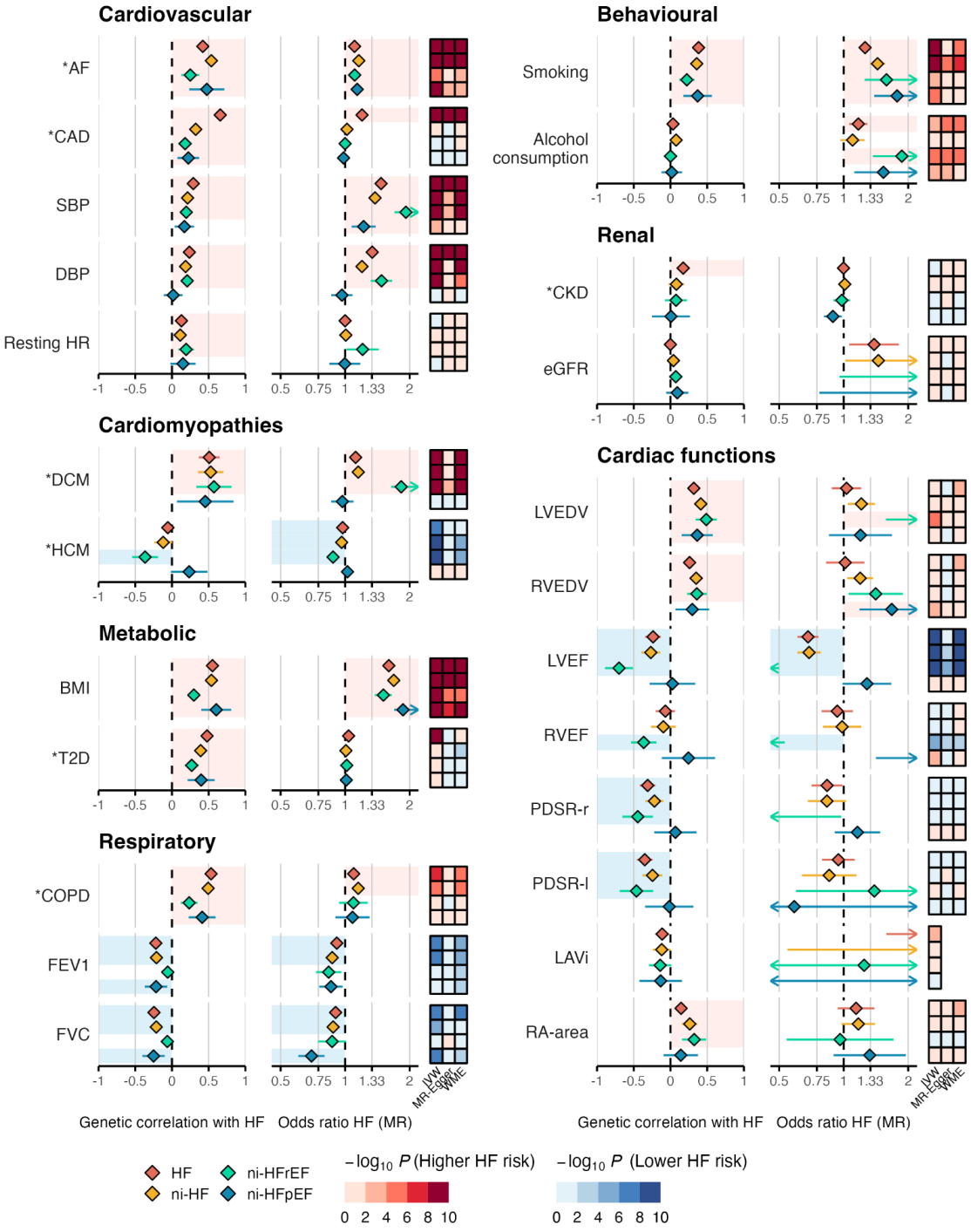

Figure 6. Bivariate genetic correlation (r_g) and Mendelian randomisation (MR) estimates across 24 traits and 4 HF phenotypes. Asterisks (*) indicate binary traits. MR effects estimates are reported as odds ratio (OR_{MR}) per doubling prevalence for binary traits; or per standard deviation increase for quantitative traits. Estimates which were robust to multiple testing adjustment and sensitivity analyses were indicated by light blue shade (for $r_g < 0$ and $OR_{MR} < 1$) or light red shade (for $r_g > 0$ and $OR_{MR} > 1$). The heat maps represent *P-values* for different MR models, colour-coded with direction of MR estimates and strength of associations.

Data availability

GWAS summary statistics from the meta-analysis will be made available on the Cardiovascular Disease Knowledge Portal (https://cvd.hugeamp.org/) and GWAS Catalog (https://www.ebi.ac.uk/gwas/summary-statistics) upon publication.

Code availability

A sample code to define heart failure phenotypes in UK Biobank is available on: https://github.com/ihi-comp-med/ukb-hf-phenotyping. Other codes to perform key analyses presented in this work will be made available on https://github.com/ihi-comp-med/hermes2-gwas

Acknowledgements

A.Henry was supported by the BHF Cardiovascular Biomedicine PhD studentship (FS/18/65/34186). R.T.L. and A.Henry are partly supported by a Pfizer Innovative Targets Exploration Network Grant. The project was additionally supported by BigData@Heart Consortium funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074, the UCL British Heart Foundation Accelerator (AA/18/6/34223), National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR203328), and Health Data Research UK (MR/S003754/1). Analyses using the UK Biobank resource presented in this work were conducted under Application Numbers 9922, 15422, 12113, and 47602. The authors thank all research participants included in the presented work. The views expressed in this work are those of the authors and not necessarily those of the funders. Additional study-level acknowledgements are provided in **Supplementary Table 18**.

Author contributions

Conceptualisation: R.S.V., J.G.S., H.Holm, Sonia Shah, P.T.E., A.D.H., Q.W., R.T.L.

Methodology: A.Henry, M.D.C., D.S., Sonia Shah

Project administration: A.Henry, H.I., D.M., R.T.L.

Formal analysis: A.Henry, X.M.

Software: A.Henry, C.F., D.S., S.C., A.F.S.

Data Curation: A.Henry, H.I.

Writing - Original Draft: A.Henry, R.T.L.

Visualisation: A.Henry

Resources: C.F., M.D.C., S.D., J.Gratton, S.C., J.G.S., H.Holm, P.T.E., A.D.H.

Investigation: M.D.C., I.B., C.R., D.M.

Supervision: F.W.A., T.P.C., M.P.D., M.E.D., C.C.L., N.J.S., R.S.V., J.G.S., H.Holm, Sonia Shah, P.T.E., A.D.H., Q.W., R.T.L.

Funding acquisition: R.T.L.

F.W.A., T.P.C., M.P.D., M.E.D., C.C.L., N.J.S., Svati Shah, R.S.V., J.G.S., H.Holm, P.T.E., A.D.H., Q.W., R.T.L. are members of the HERMES Executive Committee who provided additional supervision of the work.

Other co-authors not listed above contributed to data generation, funding acquisition, formal analysis, and supervision at individual study level.

Lists of contributors from Genes & Health Research Team, DBDS Genomic Consortium, and HERMES Consortium are provided in **Supplementary Note**.

All authors have reviewed and approved the final version of the manuscript.

Competing interests

A.Henry and R.T.L. received funding from Pfizer. J.S.W. have acted as a consultant for MyoKardia, Pfizer, Foresite Labs, and Health Lumen, and received institutional support from Bristol-Myers Squibb and Pfizer. S.d.D. was supported through grants from AstraZeneca, Roche Molecular Science/DalCor. J.R.K. declares stock ownership in AbbVie, Abbott, Bristol Myers Squibb, Johnson & Johnson, Medtronic, Merck, Pfizer. N.A.M. received speaking honoraria from Amgen and is involved in clinical trials with Ionis, Amgen, Pfizer, and Novartis. B.M.P. serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. C.T.R. received honoraria for scientific advisory boards and consulting from Anthos, Bayer, Bristol Myers Squibb, Daiichi Sankyo, Janssen, Pfizer, and received institutional research grants from Anthos, AstraZeneca, Daiichi Sankyo, Janssen and Novartis. M.S.S. received significant research grant support from Abbott Laboratories, Amgen, AstraZeneca, Bayer, Critical Diagnostics, Daiichi-Sankyo, Eisai, Genzyme, Gilead, GlaxoSmithKline, Intarcia, Janssen Research and Development, The Medicines Company, MedImmune, Merck, Novartis, Poxel, Pfizer, Quark Pharmaceuticals, Roche Diagnostics, and Takeda; and has received consulting fees from Alnylam, AstraZeneca, Bristol-Myers Squibb, CVS, Amgen. A.A.V. received consultancy fees and/or research support from AnaCardia, AstraZeneca, Bayer, BMS, Boehringer Ingelheim, Corteria, Cytokinetics, EliLilly, Moderna, Novartis, NovoNordisk, Roche Diagnostics. M-P.D. declares holding equity in Dalcor Pharmaceuticals, unrelated to this work.

Members of the TIMI Study Group (ENGAGE, FOURIER, PEGASUS, SAVOR, SOLID) have received institutional research grant support through Brigham and Women's Hospital from: Abbott, Amgen, Anthos Therapeutics, ARCA Biopharma, Inc., AstraZeneca, Bayer HealthCare Pharmaceuticals, Inc., Daiichi-Sankyo, Eisai, Intarcia, Ionis Pharmaceuticals, Inc., Janssen Research and Development, LLC, MedImmune, Merck, Novartis, Pfizer, Quark Pharmaceuticals, Regeneron Pharmaceuticals, Inc., Roche, Siemens Healthcare Diagnostics, Inc., Softcell Medical Limited, The Medicines Company, Zora Biosciences, Caremark, Dyrnamix, Esperon, IFM Pharmaceuticals, MyoKardia.

The authors who are affiliated with deCODE genetics/Amgen Inc. and the authors affiliated with Pfizer Inc. declare competing financial interests as employees.

The remaining authors declare no competing interests.

References

1. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
2. Levin, M. G. *et al.* Genome-wide association and multi-trait analyses characterize the common genetic architecture of heart failure. *Nat. Commun.* **13**, 6914 (2022).
3. Joseph, J. *et al.* Genetic architecture of heart failure with preserved versus reduced ejection fraction. *Nat. Commun.* **13**, 1–14 (2022).
4. Arvanitis, M. *et al.* Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. *Nat. Commun.* **11**, 1122 (2020).
5. Zhou, W. *et al.* Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genomics* **2**, (2022).
6. Garnier, S. *et al.* Genome-wide association analysis in dilated cardiomyopathy reveals two new players in systolic heart failure on chromosomes 3p25.1 and 22q11.23. *Eur. Heart J.* **42**, 2000–2011 (2021).
7. Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).
8. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
9. Ortiz-Genga, M. F. *et al.* Truncating FLNC Mutations Are Associated With High-Risk Dilated and Arrhythmogenic Cardiomyopathies. *J. Am. Coll. Cardiol.* **68**, 2440–2451 (2016).
10. Domínguez, F. *et al.* Dilated Cardiomyopathy Due to BLC2-Associated Athanogene 3 (BAG3) Mutations. *J. Am. Coll. Cardiol.* **72**, 2471–2481 (2018).
11. Wu, T. *et al.* HSPB7 is indispensable for heart development by modulating actin filament assembly. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11956–11961 (2017).
12. Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
13. Weeks, E. M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *bioRxiv* (2020) doi:10.1101/2020.09.08.20190561.
14. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.* **15**, e1007889 (2019).
15. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440 (2021).
16. Sobczyk, M. K., Gaunt, T. R. & Paternoster, L. MendelVar: gene prioritization at GWAS loci using phenotypic enrichment of Mendelian disease genes. *Bioinformatics* **37**, 1–8 (2021).
17. Sanan, D. A. *et al.* Low density lipoprotein receptor-negative mice expressing human apolipoprotein B-100 develop complex atherosclerotic lesions on a chow diet: no accentuation by apolipoprotein(a). *Proc. Natl. Acad. Sci. U. S. A.* **95**, 4544–4549 (1998).
18. Steinberg, D., Parthasarathy, S., Carew, T. E., Khoo, J. C. & Witztum, J. L. Beyond cholesterol. Modifications of low-density lipoprotein that increase its atherogenicity. *N. Engl. J. Med.* **320**, 915–924 (1989).
19. Berge, K. E. *et al.* Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science* **290**, 1771–1775 (2000).
20. Lee, M. H. *et al.* Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. *Nat. Genet.* **27**, 79–83 (2001).
21. Kronenberg, F. *et al.* Lipoprotein(a) in atherosclerotic cardiovascular disease and aortic stenosis: a European Atherosclerosis Society consensus statement. *Eur. Heart J.* **43**, 3925–3946 (2022).
22. Congrains, A. *et al.* Genetic variants at the 9p21 locus contribute to atherosclerosis through

modulation of ANRIL and CDKN2A/B. *Atherosclerosis* **220**, 449–455 (2012).

23. Chiu, C. *et al.* Mutations in alpha-actinin-2 cause hypertrophic cardiomyopathy: a genome-wide analysis. *J. Am. Coll. Cardiol.* **55**, 1127–1135 (2010).
24. Ling, H. *et al.* Requirement for Ca2+/calmodulin-dependent kinase II in the transition from pressure overload-induced cardiac hypertrophy to heart failure in mice. *J. Clin. Invest.* **119**, 1230–1240 (2009).
25. Meurs, K. M. *et al.* Association of dilated cardiomyopathy with the striatin mutation genotype in boxer dogs. *J. Vet. Intern. Med.* **27**, 1437–1440 (2013).
26. Wang, J. *et al. Pitx2* prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9753–9758 (2010).
27. Sotoodehnia, N. *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat. Genet.* **42**, 1068–1076 (2010).
28. Pfeufer, A. *et al.* Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat. Genet.* **41**, 407–414 (2009).
29. Górska, A. A. *et al.* Muscle-specific Cand2 is translationally upregulated by mTORC1 and promotes adverse cardiac remodeling. *EMBO Rep.* **22**, e52170 (2021).
30. Benson, D. W. *et al.* Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J. Clin. Invest.* **104**, 1567–1573 (1999).
31. Pashmforoush, M. *et al.* Nkx2-5 pathways and congenital heart disease; loss of ventricular myocyte lineage specification leads to progressive cardiomyopathy and complete heart block. *Cell* **117**, 373–386 (2004).
32. Sveinbjornsson, G. *et al.* Variants in NKX2-5 and FLNC Cause Dilated Cardiomyopathy and Sudden Cardiac Death. *Circ Genom Precis Med* **11**, e002151 (2018).
33. Copeland-Halperin, R. S., Liu, J. E. & Yu, A. F. Cardiotoxicity of HER2-targeted therapies. *Curr. Opin. Cardiol.* **34**, 451–458 (2019).
34. Christensen, G., Herum, K. M. & Lunde, I. G. Sweet, yet underappreciated: Proteoglycans and extracellular matrix remodeling in heart disease. *Matrix Biol.* **75-76**, 286–299 (2019).
35. Johnston, J. A., Ward, C. L. & Kopito, R. R. Aggresomes: a cellular response to misfolded proteins. *J. Cell Biol.* **143**, 1883–1898 (1998).
36. Martin, T. G. *et al.* Cardiomyocyte contractile impairment in heart failure results from reduced BAG3-mediated sarcomeric protein turnover. *Nat. Commun.* **12**, 2942 (2021).
37. Zhang, L. *et al.* Insulin-like growth factor-binding protein-7 (IGFBP7) links senescence to heart failure. *Nature Cardiovascular Research* **1**, 1195–1214 (2022).
38. Oakley, R. H. & Cidlowski, J. A. Glucocorticoid signaling in the heart: A cardiomyocyte perspective. *J. Steroid Biochem. Mol. Biol.* **153**, 27–34 (2015).
39. Chaffin, M. *et al.* Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature* (2022) doi:10.1038/s41586-022-04817-8.
40. Oakley, R. H. *et al.* Cardiomyocyte glucocorticoid and mineralocorticoid receptors directly and antagonistically regulate heart disease in mice. *Sci. Signal.* **12**, (2019).
41. Rossignol, P. *et al.* Eplerenone survival benefits in heart failure patients post-myocardial infarction are independent from its diuretic and potassium-sparing effects. Insights from an EPHESUS (Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study) substudy. *J. Am. Coll. Cardiol.* **58**, 1958–1966 (2011).
42. Tan, W. L. W. *et al.* Epigenomes of Human Hearts Reveal New Genetic Variants Relevant for Cardiac Disease and Phenotype. *Circ. Res.* **127**, 761–777 (2020).
43. Kuan, V. *et al.* A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit Health* **1**, e63–e77 (2019).
44. Tadros, R. *et al.* Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect. *Nat. Genet.* **53**, 128–134 (2021).
45. Jia, M. *et al.* Deletion of BACH1 Attenuates Atherosclerosis by Reducing Endothelial

Inflammation. *Circ. Res.* **130**, 1038–1055 (2022).

46. Xin, M. *et al.* Hippo pathway effector Yap promotes cardiac regeneration. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13839–13844 (2013).
47. Blake, J. A. *et al.* Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. *Nucleic Acids Res.* **49**, D981–D987 (2021).
48. Ochoa, D. *et al.* The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2023).
49. Ware, J. S. *et al.* Genetic Etiology for Alcohol-Induced Cardiac Toxicity. *J. Am. Coll. Cardiol.* **71**, 2293–2302 (2018).
50. Banerjee, A. *et al.* A population-based study of 92 clinically recognized risk factors for heart failure: co-occurrence, prognosis and preventive potential. *Eur. J. Heart Fail.* **24**, 466–480 (2022).
51. Wehner, G. J. *et al.* Routinely reported ejection fraction and mortality in clinical practice: where does the nadir of risk lie? *Eur. Heart J.* **41**, 1249–1257 (2020).
52. Harris, P. Evolution and the cardiac patient. *Cardiovasc. Res.* **17**, 437–445 (1983).
53. Bianucci, R. *et al.* Forensic Analysis Reveals Acute Decompensation of Chronic Heart Failure in a 3500-Year-Old Egyptian Dignitary. *J. Forensic Sci.* **61**, 1378–1381 (2016).
54. Vaduganathan, M. *et al.* SGLT2 inhibitors in patients with heart failure: a comprehensive meta-analysis of five randomised controlled trials. *Lancet* **400**, 757–767 (2022).
55. Lumbers, R. T. *et al.* Body mass index and heart failure risk: a cohort study in 1.5 million individuals and Mendelian randomisation analysis. *medRxiv* 2020.09.23.20200360 (2020).
56. Löfman, I., Szummer, K., Dahlström, U., Jernberg, T. & Lund, L. H. Associations with and prognostic impact of chronic kidney disease in heart failure with preserved, mid-range, and reduced ejection fraction. *Eur. J. Heart Fail.* **19**, 1606–1614 (2017).
57. Di Micco, R., Krizhanovsky, V., Baker, D. & d'Adda di Fagagna, F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat. Rev. Mol. Cell Biol.* **22**, 75–95 (2021).
58. Bracun, V. *et al.* Insulin-like growth factor binding protein 7 (IGFBP7), a link between heart failure and senescence. *ESC Heart Fail* (2022) doi:10.1002/ehf2.14120.
59. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
60. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
61. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
62. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
63. Ishigaki, K. *et al.* Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat. Genet.* **52**, 669–679 (2020).
64. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
65. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
66. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
67. Mägi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).
68. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–75, S1–3 (2012).
69. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

70. Dabney, A., Storey, J. D. & Warnes, G. R. qvalue: Q-value estimation for false discovery rate control. *R package version* **1**, (2010).
71. Speed, D. & Balding, D. J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* **51**, 277–284 (2019).
72. Speed, D., Holmes, J. & Balding, D. J. Evaluating and improving heritability models using summary statistics. *Nat. Genet.* **52**, 458–462 (2020).
73. Ojavee, S. E., Kutalik, Z. & Robinson, M. R. Liability-scale heritability estimation for biobank studies of low-prevalence disease. *Am. J. Hum. Genet.* **109**, 2009–2017 (2022).
74. Grotzinger, A. D., Fuente, J. de la, Privé, F., Nivard, M. G. & Tucker-Drob, E. M. Pervasive Downward Bias in Estimates of Liability-Scale Heritability in Genome-wide Association Study Meta-analysis: A Simple Solution. *Biol. Psychiatry* (2022) doi:10.1016/j.biopsych.2022.05.029.
75. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2020).
76. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 1273–1300 (2020).
77. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
78. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
79. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
80. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
81. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
82. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
83. Fehrmann, R. S. N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
84. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
85. Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *bioRxiv* 791699 (2022) doi:10.1101/791699.
86. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
87. Tucker, N. R. *et al.* Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**, 466–482 (2020).
88. Ma, Y. *et al.* Cardiomyocyte d-dopachrome tautomerase protects against heart failure. *JCI Insight* **4**, (2019).
89. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
90. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
91. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
92. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–62 (2016).
93. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).

94. Slenter, D. N. *et al.* WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–D667 (2018).
95. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
96. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
97. Davidson, R. & Harel, D. Drawing graphs nicely using simulated annealing. *ACM Trans. Graph.* **15**, 301–331 (1996).
98. Pons, P. & Latapy, M. Computing Communities in Large Networks Using Random Walks. in *Computer and Information Sciences - ISCIS 2005* 284–293 (Springer Berlin Heidelberg, 2005).
99. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8577–8582 (2006).
100. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
101. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
102. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
103. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
104. Burgess, S., Dudbridge, F. & Thompson, S. G. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat. Med.* **35**, 1880–1906 (2016).
105. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).