Biobank-based papers included in

"The Emerging Landscape of Health Research Based on Biobanks Linked to Electronic Health Records: Existing Resources, Analytic Challenges and Potential Opportunities"

Authors: Lauren J Beesley, Maxwell Salvatore, Lars G. Fritsche, Anita Pandit, Arvind Rao, Chad Brummett, Cristen J. Willer, Lynda D. Lisabeth, Bhramar Mukherjee

Here, we provide a list of papers published using biobank data included in this review. These papers provide the basis for **"Recent Major Biobank-Based Literature"** (subsection of **Section 2**). The paper identification algorithm is described in **Supplementary Section S1**. The papers are summarized by biobank and by year of publication in **Table S1**. The common outcomes in biobank-based research were identified using the research topics from these papers and are described in **Supplementary Section S2**.

Abana, C. O., Bingham, B. S., Cho, J. H., Graves, A. J., Koyama, T., Pilarski, R. T., … Xia, F. (2017). IL-6 variant is associated with metastasis in breast cancer patients. *PLOS ONE*, *12*(7), e0181725. https://doi.org/10.1371/journal.pone.0181725

Ahram, M., Othman, A., Shahrouri, M., & Mustafa, E. (2014). Factors influencing public participation in biobanking. *European Journal of Human Genetics*, *22*(4), 445–451. https://doi.org/10.1038/ejhg.2013.174

Al Kuwari, H., Al Thani, A., Al Marri, A., Al Kaabi, A., Abderrahim, H., Afifi, N., … Elliott, P. (2015). The Qatar Biobank: Background and methods Chronic Disease epidemiology. *BMC Public Health*, *15*(1), 1–9. https://doi.org/10.1186/s12889-015-2522-7

Allen, R. J., Porte, J., Braybrooke, R., Flores, C., Fingerlin, T. E., Oldham, J. M., … Jenkins, R. G. (2017). Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *The Lancet. Respiratory Medicine*, *5*(11), 869–880. https://doi.org/10.1016/S2213-2600(17)30387-9

Amaral, A. F. S., Strachan, D. P., Burney, P. G. J., & Jarvis, D. L. (2017). Female Smokers Are at Greater Risk of Airflow Obstruction Than Male Smokers. UK Biobank. *American Journal of Respiratory and Critical Care Medicine*, *195*(9), 1226–1235. https://doi.org/10.1164/rccm.201608-1545OC

Anderson, J. J., Darwis, N. D. M., Mackay, D. F., Celis-Morales, C. A., Lyall, D. M., Sattar, N., … Pell, J. P. (2018). Red and processed meat consumption and breast cancer: UK Biobank cohort study and meta-analysis. *European Journal of Cancer*, *90*, 73–82. https://doi.org/10.1016/j.ejca.2017.11.022

Arora, T., Jiang, C. Q., Thomas, G. N., Lam, K. -b. H., Zhang, W. Sen, Cheng, K. K., … Taheri, S. (2011). Self-Reported Long Total Sleep Duration Is Associated With Metabolic Syndrome: The Guangzhou Biobank Cohort Study. *Diabetes Care*, *34*(10), 2317–2319. https://doi.org/10.2337/dc11-0647

Astley, C. M., Todd, J. N., Salem, R. M., Vedantam, S., Ebbeling, C. B., Huang, P. L., … Florez, J. C. (2018). Genetic Evidence That Carbohydrate-Stimulated Insulin Secretion Leads to Obesity. *Clinical Chemistry*, *64*(1), 192–200. https://doi.org/10.1373/clinchem.2017.280727

Au Yeung, S. L., Jiang, C., Cheng, K. K., Adab, P., Lam, K. B., Liu, B., … Schooling, C. M. (2014). Aldehyde dehydrogenase 2 - a potential genetic risk factor for lung function among southern Chinese: evidence from the Guangzhou Biobank Cohort Study. *Annals of Epidemiology*, *24*(8), 606–611. https://doi.org/10.1016/j.annepidem.2014.05.015

Awadalla, P., Boileau, C., Payette, Y., Idaghdour, Y., Goulet, J. P., Knoppers, B., … Laberge, C. (2013). Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *International Journal of Epidemiology*, *42*(5), 1285–1299. https://doi.org/10.1093/ije/dys160

Bastarache, L., Hughey, J. J., Hebbring, S., Marlo, J., Zhao, W., Ho, W. T., … Denny, J. C. (2018). Phenotype risk scores identify pations with unrecognized Mendelian disease patterns. *Science*, *359*, 1233–1239. https://doi.org/10.1126/science.aal4043

Beaumont, R. N., Warrington, N. M., Cavadino, A., Tyrrell, J., Nodzenski, M., Horikoshi, M., … Freathy, R. M. (2018). Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Human Molecular Genetics*, *27*(4), 742–756. https://doi.org/10.1093/hmg/ddx429

Birdwell, K. A., Grady, B., Choi, L., Xu, H., Bian, A., Denny, J. C., … Haas, D. W. (2012). The use of a DNA biobank linked to electronic medical records to characterize pharmacogenomic predictors of tacrolimus dose requirement in kidney transplant recipients. *Pharmacogenetics and Genomics*, *22*(1), 32–42. https://doi.org/10.1097/FPC.0b013e32834e1641

Bjørngaard, J. H., Nordestgaard, A. T., Taylor, A. E., Treur, J. L., Gabrielsen, M. E., Munafò, M. R., … Davey Smith, G. (2017). Heavier smoking increases coffee consumption: findings from a Mendelian randomization analysis. *International Journal of Epidemiology*, *46*(6), 1958–1967. https://doi.org/10.1093/ije/dyx147

Cadzow, M., Merriman, T. R., & Dalbeth, N. (2017). Performance of gout definitions for genetic epidemiological studies: analysis of UK Biobank. *Arthritis Research & Therapy*, *19*(1), 181. https://doi.org/10.1186/s13075-017-1390-1

Cai, Y., Hodgson, S., Blangiardo, M., Gulliver, J., Morley, D., Fecht, D., … Hansell, A. L. (2018). Road traffic noise, air pollution and incident cardiovascular disease: A joint analysis of the HUNT, EPIC-Oxford and UK Biobank cohorts. *Environment International*, *114*(November 2017), 191–201. https://doi.org/10.1016/j.envint.2018.02.048

Carey, D. J., Fetterolf, S. N., Davis, F. D., Faucett, W. A., Kirchner, H. L., Mirshahi, U., … Ledbetter, D. H. (2016). The Geisinger MyCode community health initiative: an electronic health record–linked biobank for precision medicine research. *Genetics in Medicine*, *18*(9), 906–913. https://doi.org/10.1038/gim.2015.187

Casalone, E., Tachmazidou, I., Zengini, E., Hatzikotoulas, K., Hackinger, S., Suveges, D., … Zeggini, E. (2018). A novel variant in GLIS3 is associated with osteoarthritis. *Annals of the Rheumatic Diseases*, *77*(4), 620–623. https://doi.org/10.1136/annrheumdis-2017-211848

Casey, J. A., Schwartz, B. S., Stewart, W. F., & Adler, N. E. (2016). Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. *Annual Review of Public Health*, *37*(1), 61–81. https://doi.org/10.1146/annurev-publhealth-032315-021353

Celis-Morales, C. A., Welsh, P., Lyall, D. M., Steell, L., Petermann, F., Anderson, J., … Gray, S. R. (2018). Associations of grip strength with cardiovascular, respiratory, and cancer outcomes and all cause mortality: prospective cohort study of half a million UK Biobank participants. *BMJ (Clinical Research Ed.)*, *361*, k1651. https://doi.org/10.1136/BMJ.K1651

Chan, K., Pu, X., Sandesara, P., Poston, R. N., Simpson, I. A., Quyyumi, A. A., … Patel, R. S. (2017). Genetic Variation at the ADAMTS7 Locus is Associated With Reduced Severity of Coronary Artery Disease. *Journal of the American Heart Association*, *6*(11), e006928. https://doi.org/10.1161/JAHA.117.006928

Chen, H., Gottweis, H., & Starkbaum, J. (2013). Public Perceptions of Biobanks in China: A Focus Group Study. *Biopreservation and Biobanking*, *11*(5), 267–271. https://doi.org/10.1089/bio.2013.0016

Chen, W.-C., Wang, W.-C., Lu, H.-F., Okada, Y., Chang, W.-P., Chou, Y.-H., … Chang, W.-C. (2017). rs2841277 (*PLD4*) is associated with susceptibility and rs4672495 is associated with disease activity in rheumatoid arthritis. *Oncotarget*, *8*(38), 64180–64190. https://doi.org/10.18632/oncotarget.19419

Chen, Z., Chen, J., Collins, R., Guo, Y., Peto, R., Wu, F., & Li, L. (2011). China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *International Journal of Epidemiology*, *40*(6), 1652–1666. https://doi.org/10.1093/ije/dyr120

Chisholm, R. L., Denny, J., Fridsma, D., Kheterpal, S., Masys, D., & Ohno-Machado, L. (2015). *Opportunities and Challenges related to the use of Electronic Health Records data for research*.

Clarke, T.-K., Adams, M. J., Davies, G., Howard, D. M., Hall, L. S., Padmanabhan, S., … McIntosh, A. M. (2017). Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Molecular Psychiatry*, *22*(10), 1376–1384. https://doi.org/10.1038/mp.2017.153

Cohn, E. G., Hamilton, N., Larson, E. L., & Williams, J. K. (2017). Self-reported race and ethnicity of US biobank participants compared to the US Census. *Journal of Community Genetics*, *8*(3), 229–238. https://doi.org/10.1007/s12687-017-0308-6

Cortes, A., Dendrou, C. A., Motyer, A., Jostins, L., Vukcevic, D., Dilthey, A., … McVean, G. (2017). Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. *Nature Genetics*, *49*(9), 1311–1318. https://doi.org/10.1038/ng.3926

Cowie, M. R., Blomster, J. I., Curtis, L. H., Duclaux, S., Ford, I., Fritz, F., … Zalewski, A. (2017). Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*, *106*(1), 1–9. https://doi.org/10.1007/s00392-016-1025-6

Croarkin, P. E., Luby, J. L., Cercy, K., Geske, J. R., Veldic, M., Simonson, M., … Biernacka, J. M. (2017). Genetic Risk Score Analysis in Early-Onset Bipolar Disorder. *The Journal of Clinical Psychiatry*, *78*(9), 1337–1343. https://doi.org/10.4088/JCP.15m10314

Cronin, R. M., Field, J. R., Bradford, Y., Shaffer, C. M., Carroll, R. J., Mosley, J. D., … Denny, J. C. (2014). Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. *Frontiers in Genetics*, *5*(AUG), 1–14. https://doi.org/10.3389/fgene.2014.00250

Dand, N., Mucha, S., Tsoi, L. C., Mahil, S. K., Stuart, P. E., Arnold, A., … Barker, J. N. (2017). Exome-wide association study reveals novel psoriasis susceptibility locus at TNFSF15 and rare protective alleles in genes contributing to type I IFN signalling. *Human Molecular Genetics*, *26*(21), 4301–4313. https://doi.org/10.1093/hmg/ddx328

De Souza, Y. G., & Greenspan, J. S. (2013). Biobanking past, present and future. *AIDS*, *27*(3), 303–312. https://doi.org/10.1097/QAD.0b013e32835c1244

Deary, V., Hagenaars, S. P., Harris, S. E., Hill, W. D., Davies, G., Liewald, D. C. M., … Deary, I. J. (2018). Genetic contributions to self-reported tiredness. *Molecular Psychiatry*, *23*(3), 609–620. https://doi.org/10.1038/mp.2017.5

Delaney, J. T., Ramirez, A. H., Bowton, E., Pulley, J. M., Basford, M. A., Schildcrout, J. S., … Denny, J. C. (2012). Predicting Clopidogrel Response Using DNA Samples Linked to an Electronic Health Record. *Clinical Pharmacology & Therapeutics*, *91*(2), 257–263.

https://doi.org/10.1038/clpt.2011.221

Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., … Crawford, D. C. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*, *26*(9), 1205–1210. https://doi.org/10.1093/bioinformatics/btq126

Du Rietz, E., Coleman, J., Glanville, K., Choi, S. W., O'Reilly, P. F., & Kuntsi, J. (2017). Association of Polygenic Risk for Attention-Deficit/Hyperactivity Disorder With Co-occurring Traits and Disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1–9. https://doi.org/10.1016/j.bpsc.2017.11.013

Dudding, T., Johansson, M., Thomas, S. J., Brennan, P., Martin, R. M., & Timpson, N. J. (2018). Assessing the causal association between 25-hydroxyvitamin D and the risk of oral and oropharyngeal cancer using Mendelian randomization. *International Journal of Cancer*. https://doi.org/10.1002/ijc.31377

Emdin, C. A., Khera, A. V., Natarajan, P., Klarin, D., Zekavat, S. M., Hsiao, A. J., & Kathiresan, S. (2017). Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA*, *317*(6), 626. https://doi.org/10.1001/jama.2016.21042

Feder, S. L. (2018). Data Quality in Electronic Health Records Research: Quality Domains and Assessment Methods. *Western Journal of Nursing Research*, *40*(5), 753–766. https://doi.org/10.1177/0193945916689084

Ferreira, M. A., Vonk, J. M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J. D., … Paternoster, L. (2018). Eleven loci with new reproducible genetic associations with allergic disease risk. *Journal of Allergy and Clinical Immunology*, 1–9. https://doi.org/10.1016/j.jaci.2018.03.012

Fritsche, L. G., Gruber, S. B., Wu, Z., Schmidt, E. M., Zawistowski, M., Moser, S. E., … Mukherjee, B. (2018). Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative. *The American Journal of Human Genetics*, 205021. https://doi.org/10.1016/j.ajhg.2018.04.001

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., … Allen, N. E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *American Journal of Epidemiology*, *186*(9), 1026–1034. https://doi.org/10.1093/aje/kwx246

Ganna, A., & Ingelsson, E. (2015). 5 year mortality predictors in 498 103 UK Biobank participants: a prospective population-based study. *The Lancet*, *386*(9993), 533–540. https://doi.org/10.1016/S0140-6736(15)60175-1

Ge, T., Chen, C.-Y., Neale, B. M., Sabuncu, M. R., & Smoller, J. W. (2017). Phenome-wide heritability analysis of the UK Biobank. *PLOS Genetics*, *13*(4), e1006711. https://doi.org/10.1371/journal.pgen.1006711

Gibson, J., Russ, T. C., Adams, M. J., Clarke, T.-K., Howard, D. M., Hall, L. S., … McIntosh, A. M. (2017). Assessing the presence of shared genetic architecture between Alzheimer's disease and major depressive disorder using genome-wide association data. *Translational Psychiatry*, *7*(4), e1094. https://doi.org/10.1038/tp.2017.49

Gill, D., Brewer, C. F., Del Greco M, F., Sivakumaran, P., Bowden, J., Sheehan, N. A., & Minelli, C. (2018). Age at menarche and adult body mass index: a Mendelian randomization study. *International Journal of Obesity*. https://doi.org/10.1038/s41366-018-0048-7

Glicksberg, B. S., Johnson, K. W., & Dudley, J. T. (2018). The next generation of precision

medicine: observational studies, electronic health records, biobanks and continuous monitoring. *Human Molecular Genetics*, *27*(R1), R56–R62. https://doi.org/10.1093/hmg/ddy114

Greely, H. T. (2007). The Uneasy Ethical and Legal Underpinnings of Large-Scale Genomic Biobanks. *Annual Review of Genomics and Human Genetics*, *8*(1), 343–364. https://doi.org/10.1146/annurev.genom.7.080505.115721

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C. M., Ritchie, S. J., … Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Molecular Psychiatry*, *21*(11), 1624–1632. https://doi.org/10.1038/mp.2015.225

Häggström, C., Van Hemelrijck, M., Garmo, H., Robinson, D., Stattin, P., Rowley, M., … Holmberg, L. (2018). Heterogeneity in risk of prostate cancer: A Swedish population-based cohort study of competing risks and type 2 diabetes mellitus. *International Journal of Cancer*. https://doi.org/10.1002/ijc.31587

Hall, L. S., Adams, M. J., Arnau-Soler, A., Clarke, T.-K., Howard, D. M., Zeng, Y., … McIntosh, A. M. (2018). Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Translational Psychiatry*, *8*(1), 9. https://doi.org/10.1038/s41398-017-0034-1

Hatlen, P., Grønberg, B. H., Langhammer, A., Carlsen, S. M., & Amundsen, T. (2011). Prolonged Survival in Patients with Lung Cancer with Diabetes Mellitus. *Journal of Thoracic Oncology*, *6*(11), 1810–1817. https://doi.org/10.1097/JTO.0b013e31822a75be

Hirata, M., Nagai, A., Kamatani, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., … Yoshimori, K. (2017). Overview of BioBank Japan follow-up data in 32 diseases. *Journal of Epidemiology*, *27*, S22–S28. https://doi.org/10.1016/j.je.2016.12.006

Hoffman, J. D., Graff, R. E., Emami, N. C., Tai, C. G., Passarelli, M. N., Hu, D., … Witte, J. S. (2017). Cis-eQTL-based trans-ethnic meta-analysis reveals novel genes associated with breast cancer risk. *PLOS Genetics*, *13*(3), e1006690. https://doi.org/10.1371/journal.pgen.1006690

Howard, D. M., Adams, M. J., Clarke, T.-K., Wigmore, E. M., Zeng, Y., Hagenaars, S. P., … McIntosh, A. M. (2017). Haplotype-based association analysis of general cognitive ability in Generation Scotland, the English Longitudinal Study of Ageing, and UK Biobank. *Wellcome Open Research*, *2*(0), 61. https://doi.org/10.12688/wellcomeopenres.12171.1

Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., Davies, G., … McIntosh, A. M. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nature Communications*, *9*(1), 1470. https://doi.org/10.1038/s41467-018-03819-3

Howard, D. M., Hall, L. S., Hafferty, J. D., Zeng, Y., Adams, M. J., Clarke, T.-K., … McIntosh, A. M. (2017). Genome-wide haplotype-based association analysis of major depressive disorder in Generation Scotland and UK Biobank. *Translational Psychiatry*, *7*(11), 1263. https://doi.org/10.1038/s41398-017-0010-9

Jiang, C. Q., Xu, L., Lam, T. H., Lin, J. M., Cheng, K. K., & Thomas, G. N. (2010). Smoking cessation and carotid atherosclerosis: the Guangzhou Biobank Cohort Study--CVD. *Journal of Epidemiology & Community Health*, *64*(11), 1004–1009. https://doi.org/10.1136/jech.2009.092718

Jiang, C., Thomas, G. N., Lam, T. H., Schooling, C. M., Zhang, W., Lao, X., … Cheng, K. K. (2006). Cohort Profile: The Guangzhou Biobank Cohort Study, a Guangzhou–Hong Kong–

Birmingham collaboration. *International Journal of Epidemiology*, *35*(4), 844–852. https://doi.org/10.1093/ije/dyl131

Kemp, J. P., Morris, J. A., Medina-Gomez, C., Forgetta, V., Warrington, N. M., Youlten, S. E., … Evans, D. M. (2017). Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nature Genetics*, *49*(10), 1468–1475. https://doi.org/10.1038/ng.3949

Klarin, D., Zhu, Q. M., Emdin, C. A., Chaffin, M., Horner, S., McMillan, B. J., … Kathiresan, S. (2017). Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nature Genetics*, *49*(9), 1392–1397. https://doi.org/10.1038/ng.3914

Krokstad, S., Langhammer, A., Hveem, K., Holmen, T., Midthjell, K., Stene, T., … Holmen, J. (2013). Cohort Profile: The HUNT Study, Norway. *International Journal of Epidemiology*, *42*(4), 968–977. https://doi.org/10.1093/ije/dys095

Kunzmann, A. T., Thrift, A. P., Cardwell, C. R., Lagergren, J., Xie, S., Johnston, B. T., … Coleman, H. G. (2018). Model for Identifying Individuals at Risk for Esophageal Adenocarcinoma. *Clinical Gastroenterology and Hepatology*, (April). https://doi.org/10.1016/j.cgh.2018.03.014

Lam, K. H., Jiang, C. Q., Jordan, R. E., Miller, M. R., Sen Zhang, W., Cheng, K. K., … Adab, P. (2010). Prior TB, Smoking, and Airflow Obstruction. *Chest*, *137*(3), 593–600. https://doi.org/10.1378/chest.09-1435

Lee, C. I., Bassett, L. W., Leng, M., Maliski, S. L., Pezeshki, B. B., Wells, C. J., … Naeim, A. (2012). Patients' willingness to participate in a breast cancer biobank at screening mammogram. *Breast Cancer Research and Treatment*, *136*(3), 899–906. https://doi.org/10.1007/s10549-012-2324-x

Lee, J. J., & Chow, C. C. (2017). LD score regression as an estimator of confounding and genetic correlations in genome-wide association studies. *BioRXiv*, 1–24.

Leitsalu, L., Haller, T., Esko, T., Tammesoo, M. L., Alavere, H., Snieder, H., … Metspalu, A. (2015). Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *International Journal of Epidemiology*, *44*(4), 1137–1147. https://doi.org/10.1093/ije/dyt268

Li, X., Meng, X., Spiliopoulou, A., Timofeeva, M., Wei, W.-Q., Gifford, A., … Theodoratou, E. (2018). MR-PheWAS: exploring the causal effect of SUA level on multiple disease outcomes by using genetic instruments in UK Biobank. *Annals of the Rheumatic Diseases*, annrheumdis-2017-212534. https://doi.org/10.1136/annrheumdis-2017-212534

Lin, E., Tsai, S.-J., Kuo, P.-H., Liu, Y.-L., Yang, A. C., & Kao, C.-F. (2017). Association and interaction effects of Alzheimer's disease-associated genes and lifestyle on cognitive aging in older adults in a Taiwanese population. *Oncotarget*, *8*(15), 24077–24087. https://doi.org/10.18632/oncotarget.15269

Lin, Y.-Y., Yu, M.-W., Lin, S.-M., Lee, S.-D., Chen, C.-L., Chen, D.-S., & Chen, P.-J. (2017). Genome-wide association analysis identifies a GLUL haplotype for familial hepatitis B virus-related hepatocellular carcinoma. *Cancer*, *123*(20), 3966–3976. https://doi.org/10.1002/cncr.30851

Liu, D. J., Peloso, G. M., Yu, H., Butterworth, A. S., Wang, X., Mahajan, A., … Kathiresan, S. (2017). Exome-wide association study of plasma lipids in >300,000 individuals. *Nature Genetics*, *49*(12), 1758–1766. https://doi.org/10.1038/ng.3977

Liu, J. Z., Erlich, Y., & Pickrell, J. K. (2017). Case–control association mapping by proxy using family history of disease. *Nature Genetics*, *49*(3), 325–331. https://doi.org/10.1038/ng.3766

Luciano, M., Hagenaars, S. P., Cox, S. R., Hill, W. D., Davies, G., Harris, S. E., … Bates, T. C. (2017). Single Nucleotide Polymorphisms Associated with Reading Ability Show Connection to Socio-Economic Outcomes. *Behavior Genetics*, *47*(5), 469–479. https://doi.org/10.1007/s10519-017-9859-x

Lyall, D. M., Celis-Morales, C., Ward, J., Iliodromiti, S., Anderson, J. J., Gill, J. M. R., … Pell, J. P. (2017). Association of Body Mass Index With Cardiometabolic Disease in the UK Biobank. *JAMA Cardiology*, *2*(8), 882. https://doi.org/10.1001/jamacardio.2016.5804

Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., … Dichgans, M. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature Genetics*, *50*(4), 524–537. https://doi.org/10.1038/s41588-018-0058-3

Mancini, J., Pellegrini, I., Viret, F., Vey, N., Daufresne, L.-M., Chabannon, C., & Julian-Reynier, C. (2011). Consent for Biobanking: Assessing the Understanding and Views of Cancer Patients. *JNCI: Journal of the National Cancer Institute*, *103*(2), 154–157. https://doi.org/10.1093/jnci/djq498

Manousaki, D., Paternoster, L., Standl, M., Moffatt, M. F., Farrall, M., Bouzigon, E., … Richards, J. B. (2017). Vitamin D levels and susceptibility to asthma, elevated immunoglobulin E levels, and atopic dermatitis: A Mendelian randomization study. *PLOS Medicine*, *14*(5), e1002294. https://doi.org/10.1371/journal.pmed.1002294

Marioni, R. E., Harris, S. E., Zhang, Q., McRae, A. F., Hagenaars, S. P., Hill, W. D., … Visscher, P. M. (2018). GWAS on family history of Alzheimer's disease. *Translational Psychiatry*, *8*(1), 99. https://doi.org/10.1038/s41398-018-0150-6

Marouli, E., Graff, M., Medina-Gomez, C., Lo, K. S., Wood, A. R., Kjaer, T. R., … Lettre, G. (2017). Rare and low-frequency coding variants alter human adult height. *Nature*, *542*(7640), 186–190. https://doi.org/10.1038/nature21039

Márquez-Luna, C., Loh, P.-R., & Price, A. L. (2017). Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genetic Epidemiology*, *41*(8), 811–823. https://doi.org/10.1002/gepi.22083

Mc Menamin, Ú. C., Kunzmann, A. T., Cook, M. B., Johnston, B. T., Murray, L. J., Spence, A. D., … Cardwell, C. R. (2018). Hormonal and reproductive factors and risk of upper gastrointestinal cancers in men: A prospective cohort study within the UK Biobank. *International Journal of Cancer*, *143*(4), 831–841. https://doi.org/10.1002/ijc.31375

McCoy, T. H., Castro, V. M., Hart, K. L., Pellegrini, A. M., Yu, S., Cai, T., & Perlis, R. H. (2018). Genome-wide Association Study of Dimensional Psychopathology Using Electronic Health Records. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2017.12.004

McElroy, S. L., Winham, S. J., Cuellar-Barboza, A. B., Colby, C. L., Ho, A. M.-C., Sicotte, H., … Biernacka, J. M. (2018). Bipolar disorder with binge eating behavior: a genome-wide association study implicates PRR5-ARHGAP8. *Translational Psychiatry*, *8*(1), 40. https://doi.org/10.1038/s41398-017-0085-3

Meng, W., Adams, M. J., Hebert, H. L., Deary, I. J., McIntosh, A. M., & Smith, B. H. (2018). A Genome-Wide Association Study Finds Genetic Associations with Broadly-Defined Headache in UK Biobank (N = 223,773). *EBioMedicine*, *28*, 180–186. https://doi.org/10.1016/j.ebiom.2018.01.023

Morris, J. S., Bradbury, K. E., Cross, A. J., Gunter, M. J., & Murphy, N. (2018). Physical activity, sedentary behaviour and colorectal cancer risk in the UK Biobank. *British Journal of Cancer*, *118*(6), 920–929. https://doi.org/10.1038/bjc.2017.496

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., … Yanai, H. (2017). Overview of the BioBank Japan Project: Study design and profile. *Journal of Epidemiology*, *27*(3), S2–S8. https://doi.org/10.1016/j.je.2016.12.005

Nelson, C. P., Goel, A., Butterworth, A. S., Kanoni, S., Webb, T. R., Marouli, E., … Deloukas, P. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics*, *49*(9), 1385–1391. https://doi.org/10.1038/ng.3913

Nielsen, J. B., Fritsche, L. G., Zhou, W., Teslovich, T. M., Holmen, O. L., Gustafsson, S., … Willer, C. J. (2018). Genome-wide Study of Atrial Fibrillation Identifies Seven Risk Loci and Highlights Biological Pathways and Regulatory Elements Involved in Cardiac Development. *The American Journal of Human Genetics*, *102*(1), 103–115. https://doi.org/10.1016/j.ajhg.2017.12.003

Ohno-Machado, L., Kim, J., Gabriel, R. A., Kuo, G. M., & Hogarth, M. A. (2018). Genomics and electronic health record systems. *Human Molecular Genetics*, *27*(R1), R48–R55. https://doi.org/10.1093/hmg/ddy104

Okada, E., Ukawa, S., Nakamura, K., Hirata, M., Nagai, A., Matsuda, K., … Ogata, H. (2017). Demographic and lifestyle factors and survival among patients with esophageal and gastric cancer: The Biobank Japan Project. *Journal of Epidemiology*, *27*(3), S29–S35. https://doi.org/10.1016/j.je.2016.12.002

Ong, J.-S., An, J., Law, M. H., Whiteman, D. C., Neale, R. E., Gharahkhani, P., & MacGregor, S. (2018). Height and overall cancer risk and mortality: evidence from a Mendelian randomisation study on 310,000 UK Biobank participants. *British Journal of Cancer*, *118*(9), 1262–1267. https://doi.org/10.1038/s41416-018-0063-4

Pang, Y., Kartsonaki, C., Turnbull, I., Guo, Y., Clarke, R., Chen, Y., … Chen, Z. (2018). Diabetes, plasma glucose and incidence of fatty liver, cirrhosis and liver cancer: A prospective study of 0.5 million people. *Hepatology*, *2*(April), 1–28. https://doi.org/10.1002/hep.30083

Paré, G., Mao, S., & Deng, W. Q. (2017). A machine-learning heuristic to improve gene score prediction of polygenic traits. *Scientific Reports*, *7*(1), 12665. https://doi.org/10.1038/s41598-017-13056-1

Peters, S. A. E., Bots, S. H., & Woodward, M. (2018). Sex Differences in the Association Between Measures of General and Central Adiposity and the Risk of Myocardial Infarction: Results From the UK Biobank. *Journal of the American Heart Association*, *7*(5), e008507. https://doi.org/10.1161/JAHA.117.008507

Pietraszkiewicz, A., van Asten, F., Kwong, A., Ratnapriya, R., Abecasis, G., Swaroop, A., & Chew, E. Y. (2018). Association of Rare Predicted Loss-of-Function Variants in Cellular Pathways with Sub-Phenotypes in Age-Related Macular Degeneration. *Ophthalmology*, *125*(3), 398–406. https://doi.org/10.1016/j.ophtha.2017.10.027

Pillai, U., Phillips, K., Wilkins, G., Baxter, R. C., Benn, D. E., Parker, N. R., … Marsh, D. J. (2014). Factors that May Influence the Willingness of Cancer Patients to Consent for Biobanking. *Biopreservation and Biobanking*, *12*(6), 409–414. https://doi.org/10.1089/bio.2014.0039

Pilling, L. C., Kuo, C.-L., Sicinski, K., Tamosauskaite, J., Kuchel, G. A., Harries, L. W., … Melzer, D. (2017). Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging*, *9*(12), 2504–2520. https://doi.org/10.18632/aging.101334

Ramirez, A. H., Shi, Y., Schildcrout, J. S., Delaney, J. T., Xu, H., Oetjens, M. T., … Denny, J. C. (2012). Predicting warfarin dosage in European–Americans and African–Americans using

DNA samples linked to an electronic health record. *Pharmacogenomics*, *13*(4), 407–418. https://doi.org/10.2217/pgs.11.164

Rask-Andersen, M., Karlsson, T., Ek, W. E., & Johansson, Å. (2017). Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLOS Genetics*, *13*(9), e1006977. https://doi.org/10.1371/journal.pgen.1006977

Ren, M., Ng, F. L., Warren, H. R., Witkowska, K., Baron, M., Jia, Z., … Caulfield, M. J. (2018). The biological impact of blood pressure-associated genetic variants in the natriuretic peptide receptor C gene on human vascular smooth muscle. *Human Molecular Genetics*, *27*(1), 199–210. https://doi.org/10.1093/hmg/ddx375

Reus, L. M., Shen, X., Gibson, J., Wigmore, E., Ligthart, L., Adams, M. J., … McIntosh, A. M. (2017). Association of polygenic risk for major psychiatric illness with subcortical volumes and white matter integrity in UK Biobank. *Scientific Reports*, *7*(January), 42140. https://doi.org/10.1038/srep42140

Ritchie, M. D., Denny, J. C., Crawford, D. C., Ramirez, A. H., Weiner, J. B., Pulley, J. M., … Roden, D. M. (2010). Robust Replication of Genotype-Phenotype Associations across Multiple Diseases in an Electronic Medical Record. *The American Journal of Human Genetics*, *86*(4), 560–572. https://doi.org/10.1016/j.ajhg.2010.03.003

Rutten-Jacobs, L. C. A., Tozer, D. J., Duering, M., Malik, R., Dichgans, M., Markus, H. S., & Traylor, M. (2018). Genetic Study of White Matter Integrity in UK Biobank (N=8448) and the Overlap With Stroke, Depression, and Dementia. *Stroke*, STROKEAHA.118.020811. https://doi.org/10.1161/STROKEAHA.118.020811

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J. L., Dotinga, A., Vonk, J. M., … Stolk, R. P. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *44*(4), 1172–1180. https://doi.org/10.1093/ije/dyu229

Schooling, C. M., Jiang, C. Q., Lam, T. H., Zhang, W. S., Cheng, K. K., & Leung, G. M. (2009). Alcohol use and fasting glucose in a developing southern Chinese population: The Guangzhou Biobank Cohort Study. *Journal of Epidemiology and Community Health*, *63*(2), 121–127. https://doi.org/10.1136/jech.2008.077602

Shadrina, A., Tsepilov, Y., Smetanina, M., Voronina, E., Seliverstov, E., Ilyukhin, E., … Filipenko, M. (2018). Polymorphisms of genes involved in inflammation and blood vessel development influence the risk of varicose veins. *Clinical Genetics*, (April), 1–23. https://doi.org/10.1111/cge.13362

Shweikh, Y., Ko, F., Chan, M. P. Y., Patel, P. J., Muthy, Z., Khaw, P. T., … Foster, P. J. (2015). Measures of socioeconomic status and self-reported glaucoma in the UK Biobank cohort. *Eye*, *29*(10), 1360–1367. https://doi.org/10.1038/eye.2015.157

Smeland, O. B., Frei, O., Kauppi, K., Hill, W. D., Li, W., Wang, Y., … Andreassen, O. A. (2017). Identification of Genetic Loci Jointly Influencing Schizophrenia Risk and the Cognitive Traits of Verbal-Numerical Reasoning, Reaction Time, and General Cognitive Function. *JAMA Psychiatry*, *74*(10), 1065. https://doi.org/10.1001/jamapsychiatry.2017.1986

Sodini, S. M., Kemper, K. E., Wray, N. R., & Trzaskowski, M. (2018). Comparison of Genotypic and Phenotypic Correlations: Cheverud's Conjecture in Humans. *Genetics*, genetics.300630.2017. https://doi.org/10.1534/genetics.117.300630

Song, R. J., Nguyen, X.-M. T., Quaden, R., Ho, Y.-L., Justice, A. C., Gagnon, D. R., … Heinz. (2018). Alcohol Consumption and Risk of Coronary Artery Disease (from the Million Veteran Program). *The American Journal of Cardiology*.

https://doi.org/10.1016/j.amjcard.2018.01.042

Starkbaum, J., Gottweis, H., Gottweis, U., Kleiser, C., Linseisen, J., Meisinger, C., … Wichmann, H.-E. (2014). Public Perceptions of Cohort Studies and Biobanks in Germany. *Biopreservation and Biobanking*, *12*(2), 121–130. https://doi.org/10.1089/bio.2013.0071

Strawbridge, R. J., Ward, J., Cullen, B., Tunbridge, E. M., Hartz, S., Bierut, L., … Smith, D. J. (2018). Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. *Translational Psychiatry*, *8*(1), 39. https://doi.org/10.1038/s41398-017-0079-1

Tamakoshi, A., Nakamura, K., Ukawa, S., Okada, E., Hirata, M., Nagai, A., … Kokuto, H. (2017). Characteristics and prognosis of Japanese colorectal cancer patients: The BioBank Japan Project. *Journal of Epidemiology*, *27*(3), S36–S42. https://doi.org/10.1016/j.je.2016.12.004

Taylor, A. E., Davey Smith, G., & Munafò, M. R. (2018). Associations of coffee genetic risk scores with consumption of coffee, tea and other beverages in the UK Biobank. *Addiction*, *113*(1), 148–157. https://doi.org/10.1111/add.13975

Taylor, M., Rode, L., Bjørngaard, J., Taylor, A. E., Bojesen, S. E., Åsvold, B. O., … Munafò, M. R. (2018). Is smoking heaviness causally associated with alcohol use? A Mendelian randomization study in four European cohorts. *International Journal of Epidemiology*, (May), 1–8. https://doi.org/10.1093/ije/dyy027

Teleka, S., Haggstrom, C., Nagel, G., Bjorge, T., Manjer, J., Ulmer, H., … Stocks, T. (2018). Risk of bladder cancer by disease severity in relation to metabolic factors and smoking: a prospective pooled cohort study of 800,000 men and women. *International Journal of Cancer*. https://doi.org/10.1111/joms.

Thériault, S., Gaudreault, N., Lamontagne, M., Rosa, M., Boulanger, M.-C., Messika-Zeitoun, D., … Bossé, Y. (2018). A transcriptome-wide association study identifies PALMD as a susceptibility gene for calcific aortic valve stenosis. *Nature Communications*, *9*(1), 988. https://doi.org/10.1038/s41467-018-03260-6

Tikkanen, E., Gustafsson, S., Amar, D., Shcherbina, A., Waggott, D., Ashley, E. A., & Ingelsson, E. (2018). Biological Insights Into Muscular Strength: Genetic Findings in the UK Biobank. *Scientific Reports*, *8*(1), 6451. https://doi.org/10.1038/s41598-018-24735-y

Tran, K. T., McMenamin, Ú. C., Hicks, B., Murchie, P., Thrift, A. P., Coleman, H. G., … Cardwell, C. R. (2018). Proton pump inhibitor and histamine-2 receptor antagonist use and risk of liver cancer in two population-based studies. *Alimentary Pharmacology & Therapeutics*, *48*(1), 55–64. https://doi.org/10.1111/apt.14796

Turcot, V., Lu, Y., Highland, H. M., Schurmann, C., Justice, A. E., Fine, R. S., … Loos, R. J. F. (2018). Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nature Genetics*, *50*(1), 26–41. https://doi.org/10.1038/s41588-017-0011-x

Tyrrell, J., Wood, A. R., Ames, R. M., Yaghootkar, H., Beaumont, R. N., Jones, S. E., … Frayling, T. M. (2017). Gene–obesogenic environment interactions in the UK Biobank study. *International Journal of Epidemiology*, 559–575. https://doi.org/10.1093/ije/dyw337

Ukawa, S., Nakamura, K., Okada, E., Hirata, M., Nagai, A., Yamagata, Z., … Yoshida, T. (2017). Clinical and histopathological characteristics of patients with prostate cancer in the BioBank Japan project. *Journal of Epidemiology*, *27*(3), S65–S70. https://doi.org/10.1016/j.je.2016.12.008

Usher-Smith, J. A., Harshfield, A., Saunders, C. L., Sharp, S. J., Emery, J., Walter, F. M., …

Griffin, S. J. (2018). External validation of risk prediction models for incident colorectal cancer using UK Biobank. *British Journal of Cancer*, *118*(5), 750–759. https://doi.org/10.1038/bjc.2017.463

van der Harst, P., & Verweij, N. (2018). Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery DiseaseNovelty and Significance. *Circulation Research*, *122*(3), 433–443. https://doi.org/10.1161/CIRCRESAHA.117.312086

van Duijnhoven, F. J. B., Jenab, M., Hveem, K., Siersema, P. D., Fedirko, V., Duell, E. J., … Bueno-de-Mesquita, H. B. (2018). Circulating concentrations of vitamin D in relation to pancreatic cancer risk in European populations. *International Journal of Cancer*, *142*(6), 1189–1201. https://doi.org/10.1002/ijc.31146

van Zon, S. K. R., Reijneveld, S. A., van der Most, P. J., Swertz, M. A., Bültmann, U., & Snieder, H. (2018). The interaction of genetic predisposition and socioeconomic position with type 2 diabetes mellitus. *Psychosomatic Medicine*, (April), 1. https://doi.org/10.1097/PSY.0000000000000562

Venkateswaran, S., Prince, J., Cutler, D. J., Marigorta, U. M., Okou, D. T., Prahalad, S., … Kugathasan, S. (2018). Enhanced Contribution of HLA in Pediatric Onset Ulcerative Colitis. *Inflammatory Bowel Diseases*, *24*(4), 829–838. https://doi.org/10.1093/ibd/izx084

Verweij, N., van de Vegte, Y. J., & van der Harst, P. (2018). Genetic study links components of the autonomous nervous system to heart-rate profile during exercise. *Nature Communications*, *9*(1), 898. https://doi.org/10.1038/s41467-018-03395-6

Wain, L. V, Shrine, N., Artigas, M. S., Erzurumluoglu, A. M., Noyvert, B., Bossini-Castillo, L., … Tobin, M. D. (2017). Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nature Genetics*, *49*(3), 416–425. https://doi.org/10.1038/ng.3787

Warren, H. R., Evangelou, E., Cabrera, C. P., Gao, H., Ren, M., Mifsud, B., … Elliott, P. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nature Genetics*, *49*(3), 403–415. https://doi.org/10.1038/ng.3768

Wigmore, E. M., Clarke, T.-K., Howard, D. M., Adams, M. J., Hall, L. S., Zeng, Y., … McIntosh, A. M. (2017). Do regional brain volumes and major depressive disorder share genetic architecture? A study of Generation Scotland (n=19 762), UK Biobank (n=24 048) and the English Longitudinal Study of Ageing (n=5766). *Translational Psychiatry*, *7*(8), e1205. https://doi.org/10.1038/tp.2017.148

Wolford, B. N., Willer, C. J., & Surakka, I. (2018). Electronic health records: The next wave of complex disease genetics. *Human Molecular Genetics*, *27*(R1), R14–R21. https://doi.org/10.1093/hmg/ddy081

Wood, A. M., Kaptoge, S., Butterworth, A. S., Willeit, P., Warnakula, S., Bolton, T., … Danesh, J. (2018). Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *The Lancet*, *391*(10129), 1513–1523. https://doi.org/10.1016/S0140-6736(18)30134-X

Xu, H., Jiang, M., Oetjens, M., Bowton, E. A., Ramirez, A. H., Jeff, J. M., … Denny, J. C. (2011). Facilitating pharmacogenetic studies using electronic health records and natural-language processing: a case study of warfarin. *Journal of the American Medical Informatics Association*, *18*(4), 387–391. https://doi.org/10.1136/amiajnl-2011-000208

Yaghjyan, L., Rich, S., Mao, L., Mai, V., & Egan, K. M. (2018). Interactions of coffee

consumption and postmenopausal hormone use in relation to breast cancer risk in UK Biobank. *Cancer Causes & Control*, *29*(6), 519–525. https://doi.org/10.1007/s10552-018-1028-x

Yokomichi, H., Nagai, A., Hirata, M., Tamakoshi, A., Kiyohara, Y., Kamatani, Y., … Ueyama, M. (2017). Statin use and all-cause and cancer mortality: BioBank Japan cohort. *Journal of Epidemiology*, *27*(3), S84–S91. https://doi.org/10.1016/j.je.2016.12.011

Yu, C., Guo, Y., Bian, Z., Yang, L., Millwood, I. Y., Walters, R. G., … Li, L. (2018). Association of low-activity ALDH2 and alcohol consumption with risk of esophageal cancer in Chinese adults: A population-based cohort study. *International Journal of Cancer*. https://doi.org/10.1002/ijc.31566

Zengini, E., Hatzikotoulas, K., Tachmazidou, I., Steinberg, J., Hartwig, F. P., Southam, L., … Zeggini, E. (2018). Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nature Genetics*, *50*(4), 549–558. https://doi.org/10.1038/s41588-018-0079-y

Zhao, W., Rasheed, A., Tikkanen, E., Lee, J.-J., Butterworth, A. S., Howson, J. M. M., … Saleheen, D. (2017). Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nature Genetics*, *49*(10), 1450–1457. https://doi.org/10.1038/ng.3943