BENJAMIN F. VOIGHT
CURRICULUM VITAE

Assistant Professor
Department of Systems Pharmacology and Translational Therapeutics
and Department of Genetics
University of Pennsylvania – Perelman School of Medicine
10-126 Smilow Center for Translational Research
3400 Civic Center Boulevard
Philadelphia, PA, 19104

web: http://coruscant.itmat.upenn.edu
email: bvoight@upenn.edu
twitter: @bvoight28
CV last updated: March, 2017

EMPLOYMENT HISTORY

2011–Present	Assistant Professor, Department of Systems Pharmacology and Translational Therapeutics
2012–Present	Assistant Professor, Department of Genetics University of Philadelphia - Perelman School of Medicine Philadelphia, PA
2006–2011	Postdoctoral Research Fellow and Research Scientist Advised by Drs. Mark Daly and David Altshuler Massachusetts General Hospital and The Broad Institute of Harvard and MIT Cambridge, MA

EDUCATION

2006	Ph.D., Advised by Drs. Jonathan Pritchard and Nancy Cox Human Genetics, University of Chicago
2001	B.S. in Biology and B.A. Mathematics University of Washington, Seattle [Transferred from Gonzaga University, Spokane, WA, in 1999]

RESEARCH SUPPORT[†]

2014–2018	NIH R01DK101478: *"Algorithms to identify non-coding mutational burden and disease-relevant pathways"*
2016–2018	ITMAT Maturational Human Biology Pilot Grant: *"Characterizing the genetic determinants of pubertal timing and body weight regulation"*

[†] Listing grant support where BFV is the PI

COMPLETED SUPPORT (PREVIOUS 3 YEARS)

As PI:

2012–2016 AHA 13SDG14330006: "*Human genetics of high-density lipoprotein to elucidate the etiology of heart disease*"

2013–2014 W.W. Smith Charitable Trust H1201: "*Identifying the etiological basis for heightened risk of cardiovascular disease in the context of glycemic disorder*"

2012–2014 Alfred P. Sloan Foundation Fellowship

Additional:

2015–2016 March of Dimes Preterm Birth Research Center Grant. (Role: Co-Investigator).

PENDING GRANTS

2017–2022 R01 DK113153, "*Deciphering T2D and CAD associated loci via computational and functional validation,*" Role: MPI. (PIs: Voight, Musunuru) [A0: 18th percentile]

2017–2022 R01 GM124209, "*Human genomic mutation rate: models, methods, and applications to disease,"* Role: PI. [A0: 24th percentile]

2017–2022 R01 GM119076, "*The role of essential gene dosage in human disease,"* Role: Co-I (PIs: Bucan, Brown).

2017–2021 X01, responding to TOPMED resequencing PAR-16-021, *"Leveraging a highly consanguineous cohort to discover risk factors for MI,"* Role: Co-I (PIs: Rader, Saleheen).

AWARDS AND HONORS

2017 Recipient of the 2014 Presidential Early Career Award for Scientist and Engineers (PECASE), Department of Health and Human Services

2012 Selected Alfred P. Sloan Research Fellow (see also Grants)

2009 Semi-finalist, Trainee Research Award, 59th Meeting of the American Society of Human Genetics

2007 Team Award for Outstanding Research, Clinical Research Day, Massachusetts General Hospital

2006 My Ph.D. Dissertation was awarded Best in the Biological Sciences Division, University of Chicago

2006 New York Times Front Page News article, "Still evolving, human genes tell new story" (Voight et al, see Ref #3)

2000-2006 Numerous travel awards and scholarships (MSRI/PMMB, Keystone Symposium, NHGRI, etc.)

1998-1999 McDonald's Fellowship for Biological Studies, Gonzaga University

1997-1999 Academic Merit Scholarship and Academic Debate Scholarship, Gonzaga University

PROFESSIONAL ACTIVITIES

2016–Present Member, American Heart Association's Institute for Precision Cardiovascular Medicine Data Science and Technology Committee

2015–Present	Editorial Board Member, *Circulation: Cardiovascular Genetics*
2015–Present	Vice Chairman, Genomics and Computational Biology Graduate Group, University of Pennsylvania
2012–Present	Editorial Review Board, *Frontiers* (Pop Gen, Stat Gen)
2012–Present	Member: The American Diabetes Association Member: The American Heart Association
2011–Present	Co-organizer of the Penn Bioinformatics Forum (with Yoseph Barash)
2011–Present	Member of numerous Penn internal committees (e.g. Curriculum Committee for GCB and GGR, task force for graduate biostatistics, CTSA KL2/ITMAT fellowship reviewer, IBI Faculty Search, Genetics Faculty Search, PennOmics Governance, etc.).
2002–Present	External referee for numerous journals (*Nature*, *Nature Genetics, Lancet, JAMA, PLoS Medicine, Science Translational Medicine*, *Bioinformatics, ATVB, JACC, AJHG,* several others.)
2001–Present	Member: American Society of Human Genetics
2016	Reviewer, NIH Special Emphasis Panel for RFA-DK-15-025, NIH/NIDDK
2016	Program Committee, RECOMB 2016 Satellite Meeting on Computational Methods in Genetics
2012–2015	Reviewer for American Heart Association GTOE Study Section
2013–2014	Associate Scientific Advisor, *Science Translational Medicine*
2013	Ad hoc reviewer, NIH Special Emphasis Panel for K23/K99 Career Awards, NIH/NIEHS

INVITED LECTURES (OUTSIDE PHILADELPHIA, SINCE 2011)

2017	Five Points Lecture Series, New York Genome Center, New York, NY
2017	American Diabetes Association's 77th Annual Meeting, San Diego, CA
2017	Genome Sciences Seminar Series, Center for Public Health Genomics, University of Virginia, Charlottesville, VA
2016	Program in Quantitative Genomics Seminar Series, Harvard School of Public Health, Boston, MA
2016	Frontiers in Bioinformatics and Systems Biology Seminar Series, University of California, San Diego, CA
2016	Genetics Institute Seminar Series, Vanderbilt University, TN
2016	Institute for Personalized Medicine seminar Series, Icahn School of Medicine, Mount Sinai, NY
2016	Seminar Series, University of California at Los Angeles, CA
2016	CIHR – Strategic Training for Advanced Genetic Epidemiology International Speaker Seminar Series, University of Toronto, Canada
2016	New York Area Population Genomics Workshop, Princeton University
2015	Department of Genetics Seminar Series, Yale University, NH
2013	Bioscience Conference on Genomics in Medicine, Copenhagen, Denmark
2013	Keynote Lecture, American Heart Association Epi|NPAM Council's Spring Conference, New Orleans, LA
2013	73rd Meeting of the American Diabetes Association, Chicago, IL
2012	62nd Annual Meeting of the American Society of Human Genetics, San Francisco, CA

2012	Medical Population Genetics Seminar Series, The Broad Institute of Harvard and MIT, Cambridge, MA
2012	Department of Biology Seminar Series, University of Vermont, Burlington, VT
2012	Session Co-Chair, 25th Annual Cold Spring Harbor Meeting, Biology of Genomes, Cold Spring Harbor, NY
2011	1st Annual Illumina America's Scientific Summit, Clearwater Beach, FL
2011	Botnia 20th Anniversary Symposium, Lund University, Vaasa, Finland
2011	The 2011 European Human Genetics Conference, Amsterdam RAI, The Netherlands
2011	National Institute of Genomic Medicine in Mexico, Ciudad de México, Mexico

TEACHING ACTIVITIES

2012–Present	Director, *Introduction to Bioinformatics* (GCB 535) Undergrad/grad/post-doc/MD/MD+PhD, 50+ students Includes lectures and administrative responsibilities. 100+ hours total, 50+ direct contact
2012-2013	Lecturer, Medical School Module One, *Genetic Foundations of Disease* First year MD trainees, ~100 students 2 hours of lecture (2012, 2013), along with 16 additional hours in small group discussion (2013)
2015	Guest Lecture, CTSA Summer Internship Seminar
2012	Guest Lecture, *Introduction to Genome Sciences* (GCB534)
2013	Guest Lecture, *Advanced Computational Biology* (GCB537)
2014	Guest Lecture, *Biology of Human Disease* (BIOL015)
2006–2011	Organizer of several workshops in statistical genetic analysis

STUDENT AND POSTDOCTORAL MENTORSHIP

CURRENT STUDENTS

2016–Present	Katerina Gawronski (PhD Student, GGR) - *Selected for Genetics T32 (2016)* - *American Polish Cultural Society Scholarship (2016)*
2016–Present	Onur Yörük (PhD Student, GCB)
2015–Present	Diana Cousminer (Post-doc, joint with Struan Grant) - *Young Investigator Travel award, ASBMR Bone-omics Symp. (2016)* - *ADA Postdoc Fellowship Award Recipient (2016)*
2015–Present	Rachael ("Rocky") Aikens (Swarthmore Undergrad) - *Penn Summer Undergrad Internship Program (2016)* - *Penn CTSA Summer Internship (2015)*
2015–Present	Kim Lorenz (Post-doc) - *Selected for a Diabetes/Endocrine Post-doc T32 (2015)*
2015–Present	Chris Thom (MD/PhD Visiting Research Scholar)
2014–Present	Kelsey Johnson (PhD Student, GGR)

- *Selected for Genetics T32 (2014)*

2014–Present Katie Siewert (PhD Student, GCB)
- *Selected for Genomics and Computational Biology T32 (2015)*

2012–Present Paul Babb (Post-doc)
- *Post-doc Symposium Poster Award (2014)*

PREVIOUS STUDENTS

2012–2016 Varun Aggarwala (PhD Student, GCB)
- *Semi-finalist for the ASHG Charles J. Epstein Trainee Award (2015)*
- *Penn Genetics Retreat Poster Award (2015)*
- *Now: Post-doc, Bushman Lab, Univ. of Penn*

2015–2016 David Nicholson (Post-Bac)
- *Selected for the PennPrep Program (2015)*
- *Selected for the Penn Summer Undergrad Intership Program (2014)*
- *Now: PhD Program – University of Pennsylvania*

2013–2015 Peter Yin (Undergrad, now Technical Staff at Transcriptic, Menlo Park)
- *Undergraduate Research and Fellowship Recipient (2015)*
- *Now: Technical Staff at Transcriptic, Menlo Park*

RESEARCH INTERESTS

To fully take advantage of discoveries of susceptibility loci for human complex disease, translating insights into actionable intelligence for clinical and therapeutic applications, several key questions need to be addressed. How many genes and alleles underlie complex disease, and what are their effects and frequencies? For a given susceptibility locus, what are the causal genes, functional variants, and alleles which modify risk to disease? What are the contributing genetic mechanisms? What pathways or gene networks do these genes modulate? What direction – overexpression or knockdown – do causal genes or networks need to be perturbed in order to obtain therapeutic benefits? How can current genetics data for causal and predictive biomarkers further improve clinical risk prediction and patient stratification?

As a statistical and population geneticist my lab focuses on development of statistical methods grounded in principles of population biology, applying them to genetic data collected in large, human genomics data sets. Primarily, I am motivated to understand the genetic, biological, and evolutionary basis of metabolic and cardiovascular phenotypes in human populations. Past and current research efforts in the lab focuses include (are not limited to) the following areas:

- ***Disease mapping and analysis of complex disease.*** My work at Harvard and the Broad Institute, from 2006 to 2011, has included developing novel methods and best practices for mapping studies and other genetic analyses for common human disease. My first accomplishment there was in leading the construction of the first robust statistical pipelines for conducting genome-wide association studies for common SNP variation. Subsequently, these genetic assays were extended to facilitate the characterization of structural polymorphisms, and because a previous pipeline to analyze such polymorphisms had not been described, I developed the first robust analytical pipeline for these data. My efforts to date have resulted in >100 loci associated with type 2 diabetes, cardiovascular disease, cholesterol levels, and

numerous anthropometric traits, as well as pinpointed individual genetic variants contributing susceptibility to multiple auto-immune diseases. One theme that underlies this effort is the development and application of new technologies to dissect complex traits, e.g. custom array genotyping (the *Metabochip*), as well as planning, implementation, and analysis of data generated by targeted or whole-genome resequencing studies. Work in the lab continues advancing locus discovery through multi-ethnic analysis, characterization of the genetic architecture of disease, as well as performing genetic studies to dissect the genetic basis of comorbidity between cardiovascular and glycemic diseases and traits.

- ***Causal inference studies.*** Causal inference is one of the most challenging problems in biology and medicine, requiring evidence from experimental, observational, and interventional studies. Under specific assumptions, statistical methodology (called *Mendelian Randomization*) can take advantage of human genetics data to perform this inference. This efficient approach can provide one piece of evidence in the causal inference puzzle with potentially big impact in guiding which expensive clinical trial or experimental follow-up studies to perform next. My leading efforts here have demonstrated that genetic perturbations that naturally raise serum HDL levels are not protective against heart disease, in contrast to LDL cholesterol, where causality is support by genetic, epidemiologic, and pharmacologic studies. Since 2011, I continue to pioneer applications to new traits, where we confirm causal evidence for elevated levels of serum urate levels in gout, but not in T2D, heart disease, ischemic stroke, or heart failure, the largest studies for these traits to date. This is a critical finding, as clinical trials have not yet been initiated for urate-lowering therapies in these endpoints. I have also developed a Monte Carlo approach to model arbitrarily complex causal graphs. Realistic scenarios in biology violate the assumptions underlying simple causal-inference tests. Consequently, this tool will be essential to evaluate the robustness of existing inferences and essential to create new inference procedures that handle increasing complexity, a focus of my future efforts. Work in the lab continues to apply this approach to additional clinical endpoints (*e.g.*, migraine headache and serum calcium levels; type 2 diabetes and systolic blood pressure), by taking advantage of the wealth of genetic association data in the public domain.

- ***Population genetic inference toward disease applications.*** My lab has adopted a population-based paradigm in developing statistical tools to guide the search for disease-causing variants in the human genome. Consider genes that cannot be dispensed with in human populations (*i.e.*, are likely *essential*). To identify a set of candidate essential genes, we identified a set of ~2,500 human orthologs of genes that are embryonic or post-natal lethal in mouse, and demonstrated that these genes are subject to strong, selective constraint in human populations with fewer loss-of-function mutations than the genome-wide average consistent with their essential status. We then demonstrated that *de novo* mutations found in patients with Autism are more likely to occur in this collection of genes. The implication is that non-lethal but deleterious mutations in essential genes influence risk to complex disease.

  We have recently focused the problem of quantifying the rate of mutation genome-wide. The rate of mutation varies substantially by position in human and mammalian genomes and fundamentally influences evolution and incidence of genetic disease.

Using a novel statistical framework we developed and applied to large-scale human population genomics data, we have recently showed that the three nucleotides of sequence context that flank a polymorphic site – a seven nucleotide window in total – explains >81% of variability in substitution probabilities and highlights new mutation-promoting motifs. Using our model, we show that candidate *de novo* mutations found in patients with Autism could be further prioritized for follow-up studies. Our model provides a much-need hypothesis-testing framework to this important problem. A genome-wide map of mutation rate at base-pair resolution can be envisioned, a focus of future work for my lab. Current work focuses on models for other types of mutations (*e.g.*, in/dels) as well as other sources of variation (*e.g.*, replication timing or population-specific mutation rates). We are also developing new tools for the analysis of *de novo* mutations, based on our model.

- ***Discovery and characterization of selective sweeps in Humans*.** I have a long-standing interest in identifying and characterizing genomic locations whose patterns of genetic variation are consistent the action of recent positive selection in humans. Previous work addressed the fundamental question, are humans still evolving? The preponderance of evidence suggests this is true, consequently spawning additional questions about the nature of these sweeps: When did these events occur, are they convergent across populations, what pathways do they implicate, what are the genetic mechanisms that underlie selective sweeps, are the distribution of selective events and gene targets focused on specific regions of the genome, and how do these signatures relate to phenotypic association data? My efforts along this research vein aim to design and apply novel statistical and computational methodology to address these (and other) questions, using large publicly available human population genetic data. Current work is focused on developing new statistics to detect balancing selection, and to investigate signatures of overlapping positive selective sweeps in human populations.

- ***Genomics in non-human species (Order Araneae)*.** Despite their historical interest, unusual sex-dimorphic trait distributions, and production of unusual macromolecules (venom and silk), little genetically is known about species of spider at a high-resolution molecular level. To overcome this gap in knowledge, and in collaboration with Drs. Linden Higgins. Ingi Agnarsson (Univ. of Vermont), and Cheryl Hiyashi (Univ of California – Riverside), we are generating *de novo* assembly at high coverage (100x) using diverse fragment libraries, in two species of the Order Araneae: *Nephila clavipes* (a species of Golden orb-web spider) and *Caerostris darwini* (Darwin's bark spider). This assembly will be further augmented and annotated with transcriptomic profiles obtained by RNA-seq in whole-body as well as specific tissues of interest. One overall deliverable from this effort will be the construction of a robust and systematic pipeline that leads to the construction of *extremely* high-quality draft genomes, in contrast to 'first version' assemblies, generally of usable, but less "finished", quality.

PUBLICATIONS (FROM >90 PAPERS, H-INDEX=61, CITATIONS=36,245)

LINK TO FULL PUBLICATION LIST:

http://www.ncbi.nlm.nih.gov/myncbi/browse/collection/50165572/?sort=date&direction=descending

PREPRINTS

[1] Siewert KM and **Voight BF**. Detecting Long-term Balancing Selection Using Allele Frequency Correlation. BioRxiv, March 1st, 2017. doi: https://doi.org/10.1101/112870.

[2] Johnson KE and **Voight BF**. Patterns of shared signatures of recent positive selection across human populations. BioRxiv, February 17th, 2017. doi: https://doi.org/10.1101/109371.

SELECTED PUBLICATIONS

[1] **Voight BF**, Pritchard JK. (2005). Confounding from cryptic relatedness in case-control association studies. *PLoS Genet.* 1(3): e32.

[2] **Voight BF***, Adams AA*, Frisse L, Quan Y, Hudson RR, Di Rienzo A. (2005). Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci USA* 102(51):18508-18513.

[3] **Voight BF***, Kudaravalli S*, Wen X, Pritchard JK. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4(3): e72. PMCID: PMC1382018

[4] Tishkoff SA, Reed FA, Ranciaro A, **Voight BF**, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39(1): 31-40.

[5] Saxena R, **Voight BF**, Lyssenko V, Burtt NP, …, Ricke D, Purcell S. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316(5829): 1331-1336.

[6] de Bakker PIW, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, **Voight BF**. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*. 17(R2): R122-R128.

[7] Zeggini E*, Scott LJ*, Saxena R*, **Voight BF*** on behalf of the DIAGRAM Consortium. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat. Genet. 40(5): 638-645.

[8] Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, **Voight BF**, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008 Feb;40(2):189-97.

[9] Kathiresan S, **Voight BF**, Purcell S, Musunuru K, …, Salomaa V, Schwartz SM. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 41(3): 334-341.

[10] **Voight BF***, Scott LJ*, Steinthorsdottir V*, Morris AP*, Dina C* on behalf of the DIAbetes Genome-wide Replication and Meta-Analysis (DIAGRAM) Consortium. (2010).

Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*. 42(7): 579-589.

[11] Pulit SL, **Voight BF**, de Bakker PI. (2010). Multiethnic genetic association studies improve power for locus discovery. *PLoS ONE* 5(9): e12600.

[12] Guey LT, Kravic J, Melander O, Burtt NP, Laramie JM, Lyssenko V, Jonsson A, Lindholm E, Tuomi T, Isomaa B, Nilsson P, Almgren P, Kathiresan S, Groop L, Seymour AB, Altshuler D, **Voight BF**. (2011). Power in the phenotypic extremes: A simulation study of power in discovery and replication of rare variants. *Gen Epidemiol*. 35(4): 236-246.

[13] Neale BM, Rivas MA, **Voight BF**, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. (2011). Testing for an unusual distribution of rare variants. *PLoS Genet*. 7(3): e1001322.

[14] Cotsapas C*, **Voight BF***, Rossin E, Lage K, Neale BM, Wallace C, Abecasis GR, Barrett JC, Behrens T, Cho J, De Jager PL, Elder JT, Graham RR, Gregersen P, Klareskog L, Siminovitch KA, van Heel DA, Wijmenga C, Worthington J, Todd JA, Hafler DA, Rich SS, Daly MJ; on behalf of the FOCiS Network of Consortia. (2011). Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet*. 7(8): e1002254.

[15] Bumgarner SL, Neuert G, **Voight BF**, Symbor-Nagrabska A, Grisafi P, van Oudenaarden A, Fink GR. (2012) Single-Cell Analysis Reveals that Noncoding RNAs Contribute to Clonal Heterogeneity by Modulating Transcription Factor Recruitment. *Mol. Cell* Feb 24; 45(4):470-82.

[16] Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, **Voight BF**, Kraft P, Chen R, Kallberg HJ, Kurreeman FA; Diabetes Genetics Replication and Meta-analysis Consortium; Myocardial Infarction Genetics Consortium, Kathiresan S, Wijmenga C, Gregersen PK, Alfredsson L, Siminovitch KA, Worthington J, de Bakker PI, Raychaudhuri S, Plenge RM. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet*. Mar 25;44(5):483-9.

[17] Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfelser R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, Lewis L, Han Y, **Voight BF**, …, Gibbs RA, Roeder K, Schellenberg GD, Sutcliffe JS, Daly MJ. (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* Apr 4;485(7397):242-5.

[18] **Voight BF***, Peloso GM*, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, …, O'Donnell CJ, Salomaa V, Rader DJ, Peltonen L, Schwartz SM, Altshuler D, Kathiresan S. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* Aug 11;380(9841):572-802.

[19] **Voight BF***, Kang HM*, Ding J, Palmer CD, Sidore C, Chines PS, Burtt NP, Fuchsberger C, Li Y, Erdmann J, et al. (2012). The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet*. Aug;8(8):e1002793.

[20] Perry JR, **Voight BF**, Yengo L, Amin N, Dupuis J, …, Frayling TM, Cauchi S. (2012). Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet*. May;8(5):e1002741.

[21] Morris AP*, **Voight BF***, Teslovich TM*, Ferreira T*, Segrè AV*, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet*. Aug 12;44(9):981-990.

[22] **Voight BF**, Cotsapas C (2012). Human genetics offers an emerging picture of common pathways and mechanisms in autoimmunity. *Curr Opin Immunol*. Oct;24(5):552-7.

[23] Georgi B, **Voight BF**, Bućan M (2013). From mouse to human: evolutionary genomics analysis of human orthologs of essential genes. *PLoS Genet*. May;9(5):e1003484.

[24] Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burtt NP, Mahajan A, Fuchsberger C, Atzmon G, Benediktsson R, …, **Voight BF**, Wilson JG, Boehnke M, McCarthy MI, Njølstad PR, Pedersen O, Groop L, Cox DR, Stefansson K, Altshuler D. (2014). Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nature Genet.* Apr 4;46(4):357-363.

[25] Prokopenko I, Poon W, Mägi R, Prasad B R, Salehi SA, Almgren P, Osmark P, Bouatia-Naji N, Wierup N, Fall T, …, **Voight BF**, et al. (2014). A central role for GRB10 in regulation of islet function in man. *PLoS Genet* Apr 3;10(4):e1004235.

[26] MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, **Voight BF**, Winckler W, Gunter C (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature* Apr 24;508:469-476.

[27] **Voight, BF** (2014). MR_predictor: a simulation engine for Mendelian Randomization studies. *Bioinformatics.* Dec 1; 30(23):3432-4.

[28] Yin, P and **Voight, BF** (2015). MeRP: a high-throughput pipeline for Mendelian Randomization Analysis. *Bioinformatics.* Mar 15;31(6):957-9.

[29] Jansen H, Loley C, Lieb W, Pencina MJ, Nelson CP, Kathiresan S, Peloso GM, **Voight BF**, Reilly MP, Assimes TL, Boerwinkle E, Hengstenberg C, Laaksonen R, McPherson R, Roberts R, Thorsteinsdottir U, Peters A, Gieger C, Rawal R, Thompson JR, König IR; CARDIoGRAM consortium, Vasan RS, Erdmann J, Samani NJ, Schunkert H. (2015) Genetic variants primarily associated with type 2 diabetes are related to coronary artery disease risk. *Atherosclerosis* Jun 3;241(2):419-426.

[30] Soccio RE, Chen ER, Rajapurkar SR, Safabakhsh P, Marinis JM, Dispirito JR, Emmett MJ, Briggs ER, Fang B, Everett LJ, Lim HW, Won KJ, Steger DJ, Wu Y, Civelek M, **Voight BF**, Lazar MA. (2015) Genetic Variation Determines PPARγ Function and Anti-diabetic Drug Response in Vivo. *Cell* Jul 2;162(1):33-44.

[31] Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Mägi R, …, **Voight BF**, et al. (2015) Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet.* 2015 Dec;47(12):1415-25.

[32] Keenan T, Zhao W, Rasheed A, Ho WK, Malik R, …, Rader DJ, **Voight BF***, Saleheen D*. Causal Assessment of Serum Urate Levels in Cardiometabolic Diseases Through a Mendelian Randomization Study. J Am Coll Cardiol. 2016 Feb 2;67(4):407-16.

[33] Aggarwala V, and **Voight BF**. An expanded sequence context model broadly explains variability in polymorphism levels across the human genome. *Nat Genet*. 2016 Apr;48(4):349-55.

[34] Cousminer DL, Arkader A, **Voight BF**, Pacifici M, Grant SF. Assessing the general population frequency of rare coding variants in the EXT1 and EXT2 genes previously implicated in hereditary multiple exostoses. Bone. 2016 Sep 9;92:196-200.

[35] Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, …, **Voight BF** et al. The genetic architecture of type 2 diabetes. *Nature*. 2016 Aug 4;536(7614):41-7.

[36] Cousminer DL, Arkader A, **Voight BF**, Pacifici M, Grant SF. Assessing the general population frequency of rare coding variants in the EXT1 and EXT2 genes previously implicated in hereditary multiple exostoses. Bone. 2016 Nov;92:196-200.

[37] Aikens RC, Zhao W, Saleheen D, Reilly MP, Epstein SE, Tikkanen E, Salomaa V, **Voight BF**. Systolic Blood Pressure and Risk of Type 2 Diabetes: a Mendelian Randomization Study. Diabetes. 2017 Feb;66(2):543-550.

[38] Yin P, Anttila V, Siewert KM, Palotie A, Smith GD, **Voight BF**. Serum calcium and risk of migraine: a Mendelian randomization study. Hum Mol Genet. 2016 [Epub ahead of print]

[39] Aggarawala V, Ganguly A, **Voight BF**. *De novo* mutational profile in RB1 clarified using a mutation rate modeling algorithm. *BMC Genomics.* 2017 Feb 14;18(1):155.