Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

| Citation       | Emdin, C. A., A. V. Khera, M. Chaffin, D. Klarin, P. Natarajan, K. Aragam, M. Haas, et al. 2018. “Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease.” Nature Communications 9 (1): 1613. doi:10.1038/s41467-018-03911-8. http://dx.doi.org/10.1038/s41467-018-03911-8. |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Published Version | doi:10.1038/s41467-018-03911-8                                                                                                                                                                                                                                                                             |
| Citable link    | http://nrs.harvard.edu/urn-3:HUL.InstRepos:37068149                                                                                                                                                                                                                                                             |
| Terms of Use    | This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA |
ARTICLE

DOI: 10.1038/s41467-018-03911-8

Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease

Connor A. Emdin, Amit V. Khera et al.#

Less than 3% of protein-coding genetic variants are predicted to result in loss of protein function through the introduction of a stop codon, frameshift, or the disruption of an essential splice site; however, such predicted loss-of-function (pLOF) variants provide insight into effector transcript and direction of biological effect. In >400,000 UK Biobank participants, we conduct association analyses of 3759 pLOF variants with six metabolic traits, six cardiometabolic diseases, and twelve additional diseases. We identified 18 new low-frequency or rare (allele frequency < 5%) pLOF variant-phenotype associations. pLOF variants in the gene GPR151 protect against obesity and type 2 diabetes, in the gene IL33 against asthma and allergic disease, and in the gene IFIH1 against hypothyroidism. In the gene PDE3B, pLOF variants associate with elevated height, improved body fat distribution and protection from coronary artery disease. Our findings prioritize genes for which pharmacologic mimics of pLOF variants may lower risk for disease.

Correspondence and requests for materials should be addressed to S.K. (email: skathiresan1@mgh.harvard.edu).
#A full list of authors and their affiliations appears at the end of the paper.
focused investigation of predicted loss-of-function (pLOF) variants provides several advantages when compared with analysis of other types of variants. First, analysis of pLOF variants may allow for the direct identification of a gene rather than a locus containing many candidate genes. Second, pLOF variants provide directionality of effect, unlike non-coding regulatory variants which may increase or decrease expression of a given gene. Third, identification of pLOF variants which protect against disease may aid with prioritization of therapeutic target genes (e.g., the recent development of inhibitors of PCSK9 or ANGPTL3 which mimic human pLOF mutations protective against cardiovascular disease).

Here, we analyse pLOF variants in UK Biobank and other datasets to identify genes for which pharmacologic inhibition may protect against disease. We identify associations of pLOF variants with cardiometabolic and immune disease, prioritizing the therapeutic targets.

**Results**

**Analysis of loss-of-function variants.** In 405,569 individuals in UK Biobank (335,660 individuals of European ancestry and 69,909 individuals of Non-European ancestry, Supplementary Table 1), we analyzed the association of 3759 pLOF variants with six metabolic traits [body mass index (BMI), waist-to-hip ratio adjusted for body mass index (WHRadjBMI), height, systolic blood pressure (SBP), diastolic blood pressure (DBP), forced expiratory volume to forced vital capacity ratio (FEV1/FVC)], six cardiometabolic diseases (coronary artery disease, type 2 diabetes, atrial fibrillation, stroke, heart failure, venous thromboembolism) and twelve diseases with more than 5000 cases (allergic rhinitis, asthma, anxiety, breast cancer, cataract, cholelithiasis, depression, hypothyroidism, gastric reflux, osteoporosis, osteoarthritis, and psoriasis). The Variant Effect Predictor and associated LOFTEE plugin algorithms were used to annotate variants which were pLOF (1) nonsense mutations that resulted in early termination of a protein; (2) frameshift mutations due to insertions or deletions of DNA; or (3) splice-site mutations which result in an incorrectly spliced protein. For coronary artery disease, we pooled datasets to identify genes for which pharmacologic inhibition may protect against disease. We identify associations of pLOF variants with cardiometabolic and immune disease, prioritizing the therapeutic targets.

**Table 1 Predicted loss-of-function variants with minor allele frequency <5% which are significantly associated with traits or diseases in UK Biobank**

| Outcome   | Gene          | pLOF variant | Location | EA       | RA     | AA change          | Freq (%) | Beta   | SE     | P-value | Novel? | MHC locus? |
|-----------|---------------|--------------|----------|----------|--------|--------------------|----------|--------|--------|---------|--------|------------|
| Asthma    | FLG           | rs61817661   | 1:152285861 | A        | G      | p.Arg50Ter         | 1.51     | 0.21   | 0.03   | 1.51 × 10⁻¹⁵ | Yes   | No         |
| Asthma    | HLA-DQBI1     | rs28668207   | 6:32628660 | C        | T      | Splice Acceptor c.773–1A>G | 3.14     | -0.17  | 0.02   | 3.11 × 10⁻⁸   | Yes   | No         |
| Asthma    | IL33          | rs146590786  | 9:6255967  | C        | G      | Splice Acceptor c.487–1G>C | 0.44     | -0.54  | 0.06   | 7.79 × 10⁻¹⁷ | Yes   | Yes        |
| BMI       | GPR151        | rs144285050  | 5:14559394 | A        | G      | p.Arg50Ter         | 0.78     | -0.07  | 0.01   | 4.89 × 10⁻⁸   | Yes   | No         |
| BMI       | PKHD1L1       | rs533623778  | 8:10152313 | T        | C      | p.Arg769Ter        | 1.0 × 10⁻⁴ | 5.30   | 0.99   | 9.45 × 10⁻⁸   | Yes   | No         |
| DBP       | ENPEP         | rs33966350   | 4:11413444 | A        | G      | p.Trp743Ter        | 1.19     | 0.06   | 0.01   | 8.12 × 10⁻¹⁰ | No    | No         |
| DBP       | BTDNA2       | rs83675938   | 6:26370833 | G        | T      | Splice Donor c.715+2T>G | 3.75     | 0.03   | 0.01   | 2.03 × 10⁻⁸   | Yes   | Yes        |
| DBP       | TMCS3         | rs150843763  | 15:18612492 | T        | G      | p.Arg1043Ter       | 2.14     | 0.05   | 0.01   | 8.16 × 10⁻⁹   | No    | No         |
| Height    | PDE11A        | rs76380111   | 17:287918 | A        | G      | p.Arg577Ter        | 0.52     | 0.07   | 0.01   | 6.20 × 10⁻¹¹ | Yes   | No         |
| Height    | CLH1C         | rs114931154  | 5:254407644 | T        | A      | Splice Donor c.1384+2T>A | 1.26     | -0.05  | 0.01   | 1.54 × 10⁻⁷   | Yes   | No         |
| Height    | CCD3C6        | rs150364083  | 11:14865399 | C        | G      | p.Arg427Ter        | 0.58     | 0.05   | 0.01   | 2.09 × 10⁻⁷   | Yes   | No         |
| Height    | DAP           | rs201354802  | 5:10761531 | C        | A      | p.Cys101Arg        | 0.24     | 0.18   | 0.02   | 1.68 × 10⁻⁸   | No    | No         |
| Height    | TRIM40        | rs15651142   | 6:30153120 | T        | G      | Splice Donor c.602+1G>T | 0.63     | -0.08  | 0.01   | 1.16 × 10⁻⁹   | Yes   | No         |
| Height    | MICA          | rs184130930  | 6:31738755 | A        | G      | Splice Donor c.286+1G>A | 0.26     | -0.12  | 0.02   | 7.87 × 10⁻⁸   | Yes   | No         |
| Height    | PDE3B         | rs15009666   | 11:14865399 | A        | C      | p.Asp783Ter        | 0.06     | 0.24   | 0.04   | 9.32 × 10⁻⁸   | Yes   | No         |
| Height    | APOLD1        | rs202116412  | 12:18790931 | T        | G      | Splice Donor c.96+1G>A | 0.03     | 0.12   | 0.02   | 3.06 × 10⁻⁸   | Yes   | No         |
| Hypothyroidism | IFIH1     | rs33357453  | 2:16313650 | G        | C      | Splice Donor c.1641+1G>C | 1.45     | -0.27  | 0.04   | 2.95 × 10⁻⁹   | Yes   | No         |
| Psoriasis | ZKSCAN3       | rs73387810   | 6:28318166 | A        | G      | Splice Donor c.–63+1G>A | 0.86     | 0.55   | 0.08   | 4.18 × 10⁻¹⁰ | Yes   | No         |
| Psoriasis | EGF1B         | rs140826798  | 6:32134395 | G        | A      | p.Arg74Ter         | 0.53     | 0.90   | 0.08   | 2.19 × 10⁻⁶   | Yes   | No         |
| SBP       | ENPEP         | rs33966350   | 4:11413444 | T        | C      | p.Trp743Ter        | 1.19     | 0.06   | 0.01   | 3.46 × 10⁻⁹   | No    | No         |
| SBP       | GEM           | rs33852164   | 8:95624625 | A        | G      | p.Arg199Ter        | 0.04     | 0.30   | 0.06   | 1.93 × 10⁻⁷   | No    | No         |
| WHRadjBMI | PYGM          | rs16987552   | 11:64527223 | A        | G      | p.Arg50Ter        | 0.39     | 0.09   | 0.02   | 1.32 × 10⁻⁷   | Yes   | No         |

**Table 1 Predicted loss-of-function variants with minor allele frequency <5% which are significantly associated with traits or diseases in UK Biobank**

Beta in terms of standard deviations and reported for the effect allele

pLOF predicted loss-of-function, EA effect allele, RA reference allele, AA Change amino acid change, Freq(%) Frequency(%), BMI body mass index, DBP diastolic blood pressure, SBP systolic blood pressure, WHRadjBMI waist-to-hip ratio adjusted for body mass index.
BMI (beta −0.14, P = 0.04; pooled beta −0.07, P = 9.8 × 10⁻⁹; Fig. 1). UK Biobank participants who carry one copy of p.Arg95Ter were at 12% lower odds of clinical obesity (BMI ≥30 kg/m²). As obesity is a causal risk factor for type 2 diabetes and coronary artery disease, we examined whether p.Arg95Ter may provide protection against both diseases. p.Arg95Ter was associated with 14% lower odds of type 2 diabetes (OR 0.86; P = 0.006) and 9% lower odds of coronary artery disease (OR 0.91; P = 0.01; Fig. 1). Although GPR151 encodes a G-protein coupled receptor of unknown function whose expression is limited to the central nervous system, recent studies tracing the lineage of neurons expressing GPR151 have localized connections to hypothalamic neurons, a region of the brain important in the control of appetite.

**IL33, GSDMB, and asthma.** We identified several pLOF variants that associated with lower risk of asthma (Table 1). At GSDMB encoding gasdermin B, splice acceptor variant c.662-2A > G (rs11078928, allele frequency 46% in European ancestry) protected against asthma (OR 0.87 CI 0.84, 0.91, P = 6.7 × 10⁻⁵⁰; Supplementary Table 3). This variant is in tight linkage disequilibrium (r² = 0.99) with a previously identified non-coding variant in the GSDMB locus (rs2305480) associated with lower risk of asthma (P = 9.6 × 10⁻⁸)¹². GSDMB c.662-2A > G is associated with lower expression of GSDMB transcripts.¹³ Furthermore, overexpression of GSDMB causes airway remodeling and asthma symptoms in a mouse model,¹⁴ suggesting that loss of GSDMB function may protect against asthma.

At the IL33 gene, a splice acceptor site variant c.487-1G > C (rs146597587, allele frequency 0.004 in European ancestry) was observed to protect against asthma (OR 0.58 CI 0.51, 0.66, P = 7.8 × 10⁻¹⁷). This variant was recently identified as associated with lower blood eosinophil concentration at genome-wide significance and with lower risk of asthma at more modest levels of significance (P = 1.8 × 10⁻⁴)¹⁵. To further replicate the association of IL33 c.487-1G > C with asthma, we examined the association of IL33 c.487-1G > C with asthma in individuals from three additional studies (Partners Biobank, the Vanderbilt eMERGE network, and the Women’s Genome Health Study). IL33 c.487-1G > C was associated with a protective effect of asthma in each data set. Overall, IL33 c.487-1G > C was associated with 43% lower odds of asthma (OR 0.57 CI 0.51, 0.66, P = 9.6 × 10⁻¹⁹; Fig. 2), suggesting that IL33 inhibition may be a useful approach for treatment of asthma. Of note, an inhibitor of IL33 is currently under development for treatment of asthma.¹⁶

**IFIH1 and autoimmune disorders.** A splice donor variant in IFIH1 (interferon induced with helicase C domain 1), c.1641 + 1G > C (rs35337543, allele frequency 1.5% in European ancestry), is associated with a reduced risk of hypothyroidism in UK
Biobank participants (OR 0.77 CI 0.70, 0.85; \( P = 5 \times 10^{-9} \); Table 1). A gene-based test combining four additional pLOF variants in IFIH1 (rs35732034, rs201026962, rs35744605, rs148590996) similarly demonstrated protection against hypothyroidism in UK Biobank (OR 0.79 CI 0.72, 0.86; \( P = 4.4 \times 10^{-8} \)). Carriers of pLOF variants in IFIH1 were also protected against hyperthyroidism (OR 0.84 CI 0.73, 0.96; \( P = 0.01 \); Fig. 4).

Common variants in the IFIH1 locus have previously been associated with psoriasis\(^{18}\) and vitiligo\(^{19}\), while rare pLOF variants in IFIH1 have been associated with a reduced risk of type 1 diabetes\(^{20}\). We therefore examined whether carriers of pLOF variants in IFIH1 in UK Biobank were protected against these diseases. Carriers of pLOF variants in IFIH1 were protected against type 1 diabetes, psoriasis, and vitiligo in UK Biobank (Fig. 4). These results suggest that IFIH1 pLOF variants alter risk for a range of autoimmune diseases.

In addition, an exploratory analysis demonstrated a nominally lower risk of coronary artery disease among IFIH1 pLOF carriers. Pooling UK Biobank and MiGen, IFIH1 pLOF carriers were protected against coronary artery disease (OR 0.92 CI 0.87, 0.98; \( P = 0.009 \)). To examine whether this may be a chance finding, we examined whether the common IFIH1 missense variant rs1990760, previously identified as associated with autoimmune disorders, also associated with risk of coronary artery disease. The T allele of rs1990760 (frequency 41%) associated with a reduced risk of hypothyroidism (OR 0.92 CI 0.90, 0.94; \( P = 9.3 \times 10^{-15} \)) in UK Biobank. Pooling UK Biobank and CARDIOGRAM Exome, the T allele of rs1990760 also associated with a lower risk of coronary artery disease (OR 0.97 CI 0.96, 0.99; \( P = 2.5 \times 10^{-5} \)), providing complementary evidence that IFIH1 may influence coronary artery disease risk.

**PDE3B and body fat distribution.** At PDE3B encoding the gene phosphodiesterase 3B, p.Arg783Ter (rs150090666, allele frequency 0.0006 in European ancestry) associated with elevated height (beta 0.24, \( P = 9.3 \times 10^{-4} \)). Targeted deletion of Pde3b in mice leads to white adipose tissue gaining characteristics of lymphocytes\(^{21}\), a reduction in adipocyte size\(^{22}\), smaller fat deposits\(^{23}\) and reduced atherosclerosis\(^{24}\). We therefore studied the association of PDE3B p.Arg783Ter with metabolic phenotypes in UK Biobank and/or MiGen, where 36,581 individuals have been sequenced for the 16 exons of the PDE3B gene. In UK Biobank, which lacks direct measurements of blood lipids, carriers of p.Arg783Ter carriers were at reduced risk of physician-diagnosis of hypercholesterolemia (OR 0.52, \( P = 0.0002 \). Pooling UK Biobank and MiGen, pLOF carriers in PDE3B had reduced WHRadjBMI (beta –0.15, \( P = 0.0005 \). As genetic predisposition to improved body fat distribution has been associated with a lower risk of coronary artery disease\(^{25}\), we examined whether loss of PDE3B function protects against coronary artery disease. We aggregated rare PDE3B pLOFs in cases and compared this count with that controls. Across 14,805 cases in UK Biobank, the carrier frequency of pLOF in cases was 0.1% and in controls 0.2%. Across 20,186 cases in MiGen, the carrier frequency of pLOF was 0.05% and 0.1% in controls. Collectively, carrier status for PDE3B

---

**Fig. 2** Association of IL33 c.487-1G > C with asthma in UK Biobank, Partners Biobank, Vanderbilt eMERGE network and Women’s Genome Health Study. UK Biobank estimates were derived using logistic regression, adjusted for age, sex, ten principal components of ancestry, and array type. Partners Biobank and Vanderbilt estimates were derived using logistic regression, adjusted for age, sex, and principal components of ancestry. Women’s Genome Health Study estimates were derived using logistic regression, adjusted for age and principal components of ancestry

| Data source                  | Cases | Controls | OR (95% CI)   | \( p \)-value |
|------------------------------|-------|----------|---------------|--------------|
| UK Biobank                   | 47179 | 358390   | 0.58 [0.51; 0.66] | 7.8 \times 10^{-17} |
| Fixed effect model           |       |          | 0.58 [0.51; 0.66] |              |

**Replication**

| Partners Biobank             | 224   | 2318     | 0.67 [0.11; 3.98] | 0.66       |
| Vanderbilt                   | 1886  | 23477    | 0.33 [0.15; 0.72] | 0.0057     |
| Women’s Genome Health Study | 1503  | 21115    | 0.49 [0.22; 1.10] | 0.084      |
| Fixed effect model           |       |          | 0.42 [0.24; 0.72] | 0.0015     |

| Fixed effect model           |       |          | 0.57 [0.51; 0.65] | 9.6 \times 10^{-19} |

---

**Fig. 3** Association of predicted loss-of-function variant in GSDMB with allergic disease in UK Biobank. Estimates were derived in UK Biobank using logistic regression, adjusted for age, sex, ten principal components of ancestry, and array type

| Variant          | Cases | Controls | OR (95% CI)   | \( p \)-value |
|------------------|-------|----------|---------------|--------------|
| IL33, rs146597587|       |          |               |              |
| Asthma           | 47179 | 358390   | 0.58 [0.51; 0.66] | 7.8 \times 10^{-17} |
| Allergic rhinitis| 25489 | 380080   | 0.68 [0.57; 0.80] | 5.7 \times 10^{-6} |
| Atopic dermatitis| 10274 | 395295   | 0.95 [0.76; 1.19] | 0.66       |
| Food allergy     | 1841  | 403728   | 0.62 [0.32; 1.20] | 0.16       |

**GSDMB: rs11078928**

| Asthma           | 47179 | 358390   | 0.90 [0.89; 0.91] | 6.7 \times 10^{-50} |
| Allergic rhinitis| 25489 | 380080   | 0.96 [0.94; 0.98] | 1.7 \times 10^{-5}  |
| Atopic dermatitis| 10274 | 395295   | 1.00 [0.97; 1.03] | 0.8        |
| Food allergy     | 1841  | 403728   | 0.98 [0.91; 1.06] | 0.63       |
pLOFs was associated with reduced risk for coronary artery disease (OR 0.65 CI 0.43, 0.97; P = 0.03; Fig. 5).

Presence of homozygote individuals for pLOF variants in target genes may provide an in vivo demonstration of safety of pharmacologic inhibition of target genes. We therefore examined whether homozygotes for these pLOF variants were present in UK Biobank and in the gnomAD database. At least one individual homozygous for a pLOF variant was identified in UK Biobank or in the gnomAD database for the genes GPR151, GSDMB, IL33 and IFIH1, and PDE3B (Supplementary Table 6).

**Discussion**

In this study, we identified pLOF variants that protect against obesity (GPR151), asthma (GSDMB, IL33), autoimmune disorders (IFIH1), and coronary artery disease (PDE3B), prioritizing genes and pathways for which pharmacologic attempts to mimic these protective mutations might ameliorate disease.

Identification of protective loss-of-function variants has led to the development of therapeutics. Most notably, the discovery of missense and loss-of-function variants in PCSK9 that lower LDL cholesterol and protect against coronary artery disease suggested that inhibition of PCSK9 may be a useful therapeutic strategy for prevention and treatment of cardiovascular disease. These genetic studies were validated by a large-scale randomized trial demonstrating that a monoclonal antibody directed against PCSK9 reduced the risk of recurrent cardiovascular events. More recently, the discovery of loss-of-function variants in ANGPTL3 that lower blood triglyceride levels and protect against coronary artery disease has suggested that ANGPTL3 inhibition may reduce blood triglyceride levels and risk of coronary artery disease. ANGPTL3 inhibitors are in clinical development and have now been demonstrated to reduce blood triglyceride levels.

Our findings identify putative therapeutic targets that, similar to PCSK9 and ANGPTL3, may be useful for prevention of disease. For obesity and type 2 diabetes, these results highlight GPR151 as a potential therapeutic target. GPR151 encodes a largely uncharacterized G protein-coupled receptor. The mechanism by which it influences risk of obesity and type 2 diabetes is unclear. However, it is highly expressed in the hypothalamus, a region of the brain known to be involved in appetite regulation. Indeed, genetic variation in MC4R, which encodes the melanocortin 4 receptor, strongly influences obesity risk at a population level. Similar to GPR151, MC4R is highly expressed in the hypothalamus and is involved in appetite regulation.

PDE3B inhibition may be a useful therapeutic strategy to improve body fat distribution and reduce risk of coronary artery disease. Unlike GPR151, for which no pharmacologic inhibitor is currently in clinical use, an inhibitor of PDE3B is available. Cilostazol is a non-selective pharmacologic inhibitor of both phosphodiesterase 3B and the related isoform phosphodiesterase 3A. In a small randomized trial including 211 participants, cilostazol significantly reduced restenosis after percutaneous coronary balloon angioplasty. The association of PDE3B pLOFs with improved body fat distribution, reduced risk of hypercholesterolemia and reduced risk of coronary artery disease suggests that selective inhibition of PDE3B may be useful for multiple features of metabolic syndrome.

We identified pLOF variants in IL33 encoding interleukin 33 and GSDMB encoding gasdermin B as associated with a lower risk of asthma. The IL33 variant rs146597587 was recently found to be associated with lower blood eosinophil concentration at a genome-wide significance level with lower risk of asthma at a modest level of significance (P = 1.8 × 10^{-4}). Consistent with these findings, induction of antibodies against IL33 by vaccination induces protection against airway inflammation in a mouse model of asthma. An inhibitor of IL33 is currently under development for treatment of asthma. In contrast to IL33, no inhibitor for GSDMB is in clinical development. These findings suggest that IL33 and GSDMB inhibition may both be useful therapeutic strategies for treatment of asthma and allergic disease.

Although our restriction of the present analysis to pLOF variants increases the likelihood of identifying causal variants substantially, it remains possible that a highly correlated nearby
variant could be driving the association in some cases. Future functional studies may permit additional validation of causal variants.

In summary, we associated pLOF variants with a range of biomarker and disease phenotypes in a large, national biobank and identified several new genes in which pLOF variants protect against disease, prioritizing these genes for therapeutic targeting. More generally, large-scale analysis of pLOF variants is emerging as a useful tool for therapeutic target identification and validation.

Methods
Gene and variant annotation. Variants in hg19 coordinates were annotated with information from Ensembl release 82 using Variant Effect Predictor (VEP)7. Only pLOFs were collapsed as premature stop (nonsense), canonical splice-sites (splice-donor or splice-acceptor) or insertion/deletion variants that shifted frame (frameshift) were annotated as predicted loss-of-function (pLOF), using the “–pick-allele” annotation. pLOFs as defined by VEP were then merged with publicly available data from the Exome Aggregation Consortium (ExAC), Version 0.3.1, to confirm consistency in variant annotation25.

We identified 3759 pLOF variants in UK Biobank with an info score greater than 0.3 (Supplementary Table 1). We used a Bonferroni corrected P value of 5.5 × 10−5 to denote significance [0.05/(3759 variants × 24 outcomes) = 5.5 × 10−7] in our primary pLOF analysis.

Study design. We analyzed the association of pLOF variants with 24 phenotypes: cardiovascular, metabolic and pulmonary phenotypes: six metabolic traits (body mass index, waist-to-hip ratio, height, systolic blood pressure, diastolic blood pressure, and forced expiratory volume to forced vital capacity ratio), six cardio-metabolic diseases (coronary artery disease, type 2 diabetes, atrial fibrillation, stroke, heart failure, and venous thromboembolism) and 12 diseases with more than 5000 cases (allergic rhinitis, asthma, anxiety, breast cancer, cataract, cholerais, depression, hypothyroidism, gastric reflux, osteoporosis, osteoarthritis, and psoriasis; Supplementary Table 4). All six metabolic traits were inverse normalized prior to analysis, with adjustment for age and sex. Forced expiratory volume to forced vital capacity ratio was additionally adjusted for height. To adjust for the presence of antihypertensive medication, we added 15 mm Hg to systolic blood pressure and 10 mm Hg to diastolic blood pressure of individuals on anti-hypertensive medication at baseline, as in the International Consortium for Blood Pressure genome-wide association study25. Definitions for disease outcomes in UK Biobank are provided (Supplementary Table 7).

In UK Biobank, analysis was performed separately in unrelated individuals of European and Non-European ancestry. Estimates for variants were then pooled using inverse-variance weighted-fixed effects meta-analysis. For height, estimates of variants in UK Biobank were pooled with the GIANT Height Exome consortium using inverse variance weighted-fixed effects meta-analysis10. For height, estimates of variants in UK Biobank were pooled with the GIANT Height Exome consortium using inverse variance weighted-fixed effects meta-analysis10.

Genotyping and quality control. Phasing and imputation were performed centrally, by UK Biobank, using a reference panel combining UK10k and 1000 Genomes samples. 39,235,157 variants included in the Haplotype Reference Consortium were imputed. As recommended by UK Biobank, we excluded any samples with an imputation quality <0.3 as well as pLOF variants which were not included in the Haplotype Reference Consortium. Mitochondrial genetic data and sex chromosomes were excluded from this analysis. Individual level genetic data was available from individuals in UK Biobank, after excluding one related individual of each related pair of individuals, individuals whose genetic sex did not match self-reported sex and individuals with an excess of missing genotype calls or more heterozygosity than expected.

We analysed 3759 variants identified as pLOF variants in UK Biobank. PLINK 2 software was used to examine the association of these variants with traits and disease in UK Biobank under the assumption of additive effects (https://www.cog-genomics.org/plink/2.0/). Adjustment was performed for age, sex, ten principal components of ancestry, and array type.

Conditional analysis. A locus-wide conditional analysis (±500 kb of the pLOF variant) was performed to determine the extent to which the identified pLOF variant signal was independent of other genetic variation at the locus. We iteratively performed association analyses conditional on the top variants at each locus, until no variants were below the Bonferroni-adjusted threshold for significance (P < 5.5 × 10−7). A statistically significant and independent signal for the pLOF variant provides increased confidence for a causal association.

Analysis of PDE3B association with coronary artery disease. We aimed to analyse the association of pLOF variants in PDE3B with coronary artery disease in a combined analysis of UK Biobank and the Myocardial Infarction Genetics Consortium (MiGen). Replication was performed in MiGen rather than the CARDIOGRAM Exome consortium as rs15009666 was not included in the exome chip analysis of the CARDIOGRAM Exome consortium2. Estimates of the association of rs15009666 with coronary artery disease in UK Biobank were derived as described above, using logistic regression with adjustment for age, sex, ten principal components of ancestry, and a dummy variable for array type. An additional pLOF variant, rs535108921, present in UK Biobank, was also analysed for association with coronary artery disease, as above.

pLOFs in variants in PDE3B were identified in the MiGen Consortium using exome sequencing or whole genome sequencing, as previously described26–30. Studies included in the MiGen consortium were: (1) the Italian Atherosclerosis Thrombosis and Vascular Biology (ATVB) study (dbGaP Study Accession phs000814.v1.p1); (2) the Exome Sequencing Project Early-Onset Myocardial Infarction (ESP-OMI) study9; (3) a nested case-control cohort from the Jackson Heart Study (JHS); (4) the South German Myocardial Infarction study (dbGaP Study Accession phs000916.v1.p1); (5) the Ottawa Heart Study (OHS) (dbGaP Study Accession phs000806.v1.p1); (6) the Precocious Coronary Artery Disease (PROCARDIS) study (dbGaP Study Accession phs000883.v1.p1); (7) the Pakistan Risk of Myocardial Infarction Study (PROMIS) (dbGaP Study Accession phs000917.v1.p1); (8) the Registre Gironi del COR (Gerona Heart Registry or REGICOR) study (dbGaP Study Accession phs000992.v1.p1); (9) the Leicester Myocardial Infarction study (dbGaP Study Accession phs001000.v1.p1); (10) the Bioimage study (dbGaP Study Accession phs001058.v1.p1); (11) the North German Myocardial Infarction study (dbGaP Study Accession phs000990.v1.p1); (12) Multi-Ethnic Study of Atherosclerosis (dbGaP Study Accession: phs000292.v2.p1); (13) Variation In Recovery: Role of Gender on Outcomes of Young AMI cohort and (14) Taiwan Metabolism Cohort.

The Burrows–Wheeler Aligner algorithm was used to align reads from participants to the reference genome (hg19). The GATK HaplotypeCaller was used to jointly call variants. Metrics including Variant Quality Score recalibration (VQSR), quality over depth, and strand bias were then used to filter variants. We excluded samples which were related to other samples, which had high ratios of heterozygous to non-reference homozygous genotypes, which had high missing genotypes, which had a discordant genetic gender relative to reports gender, and samples which were discordant relative to genotype data.

After variant calling and quality control, the Variant Effect Predictor,2,3 was used to annotate variants which were pLOF: (1) nonsense mutations that resulted in early termination of PDE3B (2) frameshift mutations due to insertions or deletions of DNA; or (3) splice-site mutations which result in an incorrectly spliced protein (Supplementary Table 8). For analysis of rare pLOF variants, we pooled rare pLOF variants in MiGen, testing for the association of a pLOF with coronary artery disease using logistic regression, after adjustment for age, sex, cohort, and five principal components of ancestry. We meta-analysed the association of pLOFs with coronary artery disease in MiGen combined with UK Biobank.

Replication of IL33 finding. To replicate the association of rs146597587, a splice site variant in IL33, with asthma, we pooled estimated values of the association of rs146597587 with asthma from Partners Biobank, from the Vanderbilt eMERGE network and from the Women’s Genome Health Study. In Partners Biobank, rs146597587 was imputed (info score of 0.77) in 25,363 individuals using the Illumina Exome BeadChip. The association of rs146597587 with asthma (hospitalization for ICD9 code 493) was estimated using logistic regression, adjusted for age, sex, and five principal components of ancestry. In the Vanderbilt eMERGE network, rs146597587 was genotyped in 25,363 individuals using the Illumina Exome BeadChip. The association of rs146597587 with asthma (hospitalization for ICD9 code 493) was estimated using logistic regression, adjusted for age, sex, and principal components of ancestry. In Women’s Genome Health Study, rs146597587 was genotyped in 22,618 individuals using the Illumina Exome. The association of rs146597587 with asthma (hospitalization for ICD9 code 493 or ICD10 code J45) was estimated using logistic regression, adjusted for age and principal components of ancestry.

Data availability. All individual-level data from UK Biobank can be accessed by applying to the UK Biobank Central Access Committee (http://www.ukbiobank.ac.uk/register-apply/).

Received: 3 October 2017 Accepted: 21 March 2018 Published online: 24 April 2018

References
1. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. 93, 779–797 (2013).

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-03911-8 | www.nature.com/naturecommunications
2. Abifadel, M. et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat. Genet. 34, 154–156 (2003).

3. Cohen, J. C., Boeckwinkel, E., Mosley, T. H. & Hobbs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N. Engl. J. Med. 354, 1264–1272 (2006).

4. Sabatine, M. S. et al. Evolocumab and clinical outcomes in patients with hypercholesterolemia. Nat. Genet. 48, 1134–1144 (2016).

5. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. Nature 542, 186–190 (2017).

6. Graham, M. J. et al. Cardiovascular and metabolic effects of ANGPTL3 antisense oligonucleotides. Nat. Med. 17, 1713–1722 (2011).

7. Gangwar, R. S., Landolina, N., Arpinati, L. & Levi-Schaffer, F. Mast cell and eosinophil counts and protects from asthma. Am. J. Respir. Cell Mol. Biol. 51, 634–639 (2014).

8. Ahmad, F. et al. Phosphodiesterase 3B (PDE3B) regulates NLRP3 inflammasome in adipose tissue. Proc. Natl Acad. Sci. USA 111, 31393–31398 (2014).

9. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia. Coding variation in ANGPTL3, LPL, and SVEP1 and the risk of coronary disease. N. Engl. J. Med. 374, 1134–1144 (2016).

10. Emdin, C. et al. Genetic association of waist-to-hip ratio with cardiometabolic characteristics. Nat. Genet. 48, 1561–1567 (2016).

11. Turcot, V. et al. Protein-altering variants associated with body mass index and in the parabrachial nucleus. Nat. Genet. 50, 26–41 (2018).

12. Shah, B. P. et al. MC4R-expressing glutamatergic neurons in the paraventricular hypothalamus regulate feeding and are synaptically connected to the parabrachial nucleus. Proc. Natl Acad. Sci. USA 111, 13193–13198 (2014).

13. Chung, Y. W. et al. Targeted disruption of PDE3B, but not PDE3A, protects murine heart from ischemia/reperfusion injury. Proc. Natl Acad. Sci. USA 112, 22553–22562 (2015).

14. Musunuru, K. et al. Exome sequencing, ANGPTL3 mutations, and familial combined hyperlipidemia. N. Engl. J. Med. 366, 2202–2227 (2010).

15. Sztuzkier, N. O. et al. ANGPTL3 deficiency and protection against coronary artery disease. J. Am. Coll. Cardiol. 69, 2054–2063 (2017).

16. Do, R. et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature 518, 102–106 (2015).

17. Hara, A. V. et al. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. J. Am. Coll. Cardiol. 67, 2578–2589 (2016).

18. Khera, A. V. et al. Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease. Preprint at bioRxiv https://doi.org/10.1101/218388 (2017).

19. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. Nat. Genet. 49, 403–415 (2017).

Acknowledgements
This research has been conducted using the UK Biobank resource, application 7089. The Vega data were supported by the National Heart, Lung, and Blood Institute (HL04351 and HL080467) and the National Cancer Institute (CA047988 and U01CA128931) with funding for genotyping provided by Amgen. The VYRGO study was supported by grant R01 HL08153-01A1K from the National Heart, Lung, and Blood Institute. The TAICHI study was supported by the National Health Research Institutes, Taiwan (PH-099-PP-03, PH-108-PP-03, PH-101-PP-03), the National Science Council, Taiwan (Grant Nos NSC 109-2314-B-075A-006-MY3, MOST 104-2314-B-075A-007, MOST 105-2314-B-075A-003), the Taichung Veterans General Hospital, Taiwan (TCVGH-1021010C, TCVGH-1021020D, TCVGH-1031020B, TCVGH-1031020D, TCVGH-1031020C, TCVGH-1033010D, TCVGH-1033010D, TCVGH-10401010C, TCVGH-1041020D, TCVGH-1043010C, TCVGH-1043104B), and the National Center for Advancing Translational Sciences, CTSA grant UL1TR001881. The MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, N01-HC-95170, TR-001079, UL1-TR-001079, UL1-TR-001079, UL1-TR-001079, UL1-TR-001079, DK063491. Whole genome sequencing of the VYRGO and TAICHI cohorts was funded by grant 5UM1HG008895-02 from the National Human Genome Research Institute’s Center for Common Disease Genomics. Whole genome sequencing of the MESA cohort was funded through the Trans-Omics for Precision Medicine (TOPMed) Program of the National Heart, Lung, and Blood Institute. The TOPMed Data Coordinating Center (3R01HL12393-02S1). The contributions of the investigators of the NHLBI TOPMed Consortium (https://www.nhlbiworks.org/topmed-banned-authorship) are gratefully acknowledged. The Atherosclerosis Risk in Communities study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201500005, HHSN268201500008, HHSN268201500008, HHSN268201500009, HHSN268201500010, HHSN268201500010, HHSN268201500011, HHSN268201500012). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for “Building on GWAS for NHLBI diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5R01HL104219).

Author contributions
C.A.E., A.V.K., and S.K. designed the study. C.A.E., A.V.K., and S.K. acquired, analysed, and interpreted data. C.A.E., A.V.K., and S.K. drafted the manuscript. All authors revised the manuscript for important intellectual content. S.K. supervised the study.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-03911-8.

Competing interests: A.V.K. is supported by a John S. LaDue Memorial Fellowship at Harvard Medical School, and a KL2/Catalyst Medical Research Investigator Training award from Harvard Catalyst funded by the National Institutes of Health (NIH).
Connor A. Emdin1,2,3, Amit V. Khera1,2,3, Mark Chaffin4,5, Derek Klarin1,3,4, Pradeep Natarajan1,2,3, Krishna Aragam1,2,3, Mary Haas1,2,3, Alexander Bick1,3, Seyedeh M. Zekavat1,3,5, Akihiro Nomura1,2,3, Diego Ardissino6,7, James G. Wilson8, Heribert Schunkert9, Ruth McPherson10, Hugh Watkins11,12, Roberto Elosua13,14,15, Matthew J. Bown16, Nilesh J. Samani16, Usman Baber17, Jeanette Erdmann18, Namrata Gupta3, John Danesh19,20,21, Paul Ridker22, Joshua Denny23, Lisa Bastarache23, Judith H. Lichtman24, Gail D’Onofrio25, Jennifer Mattera26, John A. Spertus27, Wayne H.-H. Sheu28, Kent D. Taylor29, Bruce M. Psaty30,31, Stephen S. Rich32, Wendy Post33, Jerome I. Rotter29, Yii-Der Ida Chen29, Harlan Krumholz26, Danish Saleheen34,35, Stacey Gabriel3 & Sekar Kathiresan1,2,3

1Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. 2Department of Medicine, Massachusetts General Hospital, Cardiology Division, Harvard Medical School, Boston, MA 02114, USA. 3Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA. 4Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. 5Department of Computational Biology & Bioinformatics, Yale Medical School, Yale University, New Haven, CT 06510, USA. 6Division of Cardiology, Azienda Ospedaliero-Universitaria di Parma, Parma 43121, Italy. 7Associazione per lo Studio Della Trombosi in Cardiologia, Pavia 27100, Italy. 8Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA. 9Deutsches Herzzentrum München, Technische Universität München, Deutsches Zentrum für Herz-Kreislauß-Forschung, München 80333, Germany. 10University of Ottawa Heart Institute, Ottawa, ON K1Y4W7, Canada. 11Radcliffe Department of Medicine, Division of Cardiovascular Medicine, University of Oxford, Oxford OX1 2JD, UK. 12Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX1 2JD, UK. 13Cardiovascular Epidemiology and Genetics, Hospital del Mar Research Institute, Barcelona 08003, Spain. 14CIBER Enfermedades Cardiovasculares (CIBERCV), Barcelona 28029, Spain. 15Facultad de Medicina, Universitat de Vic-Central de Catalunya, Barcelona, VIC 08500, Spain. 16Department of Cardiovascular Sciences, University of Leicester, and NIHR Leicester Biomedical Research Centre, Leicester LE1 7RJ, UK. 17The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York 10029 NY, USA. 18Institute for Integrative and Experimental Genomics, University of Lübeck, Lübeck 23562, Germany. 19Department of Public Health and Primary Care, Cardiovascular Epidemiology Unit, University of Cambridge, Cambridge CB2 0SR, UK. 20Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. 21National Institute of Health Research Blood and Transplant; Research Unit in Donor Health and Genetics, University of Cambridge, Cambridge CB2 1TN, UK. 22Center for Cardiovascular Disease Prevention, Brigham and Women’s Hospital, Boston 02115, USA. 23Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37235, USA. 24Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT 06510, USA. 25Department of Emergency Medicine, Yale University, New Haven, CT 06520, USA. 26Center for Outcomes Research and Evaluation, Yale–New Haven Hospital, New Haven, CT 06510, USA. 27Department of Biomedical & Saint Luke’s Mid America Heart Institute and the Health Informatics, Division of Endocrinology and Metabolism, University of Missouri-Kansas City, Kansas City, MO 64110, USA. 28Department of Internal Medicine, Taichung Veterans General Hospital, Taichung 40705, Taiwan. 29The Institute for Translational Genomics and Population Sciences LABioMed and Department of Pediatrics at Harbor-UCLA Medical Center, Torrance, CA 90095, USA. 30Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle 98195 WA, USA. 31Cardiovascular Health Research Unit, Kaiser Permanente Washington Health Research Institute, 98101 Seattle, WA, USA. 32Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA. 33Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. 34Center for Non-Communicable Diseases, Karachi 74800, Pakistan. 35Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. These authors contributed equally: Connor A. Emdin, Amit V. Khera.