**Genome-Wide Association Study of Circulating Interleukin 6 Levels Identifies Novel Loci**

Tarunveer S. Ahluwalia1#, Mohhamedreza Abbdalahi2, Nicola J. Armstrong3, Stella Aslibekyan4, Lisa Bain5, Barbara Jefferis6, Jens Baumert7, Marian Beekman8, Yoav Ben-Shlomo9, Joshua C. Bis10, Yu-Ching Cheng11, Eco deGeus12,13, Graciela E. Delgado14, Marek Diana15, Joel Eriksson16, Eero Kajantie17,18, Stavroula Kanoni19, John P. Kemp20,21, Chen Lu22, Riccardo E. Marioni23,24, Stela McLachlan25, Yuri Milaneschi26, Ilja M. Nolte2, Alexandros M. Petrelis27, Eleonora Porcu28, Maria Sabater-Lleal29,30, Ilkka Seppälä31, Tina Shah32, Gaurav Singhal33, Marie Standl34, Alexander Teumer35, Anbupalam Thalamuthu36,37, Elisabeth Thiering34,38, Stella Trompet39,40, Christie M. Ballantyne41, Emelia J. Benjamin42,43, Juan P Casas44, Catherine Toben33, George Dedoussis45, Joris Deelen8,46, Peter Durda47, Jorgen Engmann32, Mary F. Feitosa48, Harald Grallert34,49, Ann Hammarstedt50, Sarah E. Harris24,51, Georg Homuth52, Jouke-Jan Hottenga12,13, Sirpa Jalkanen53,54, Yalda Jamshidi55, Magdalene C. Jawahar33, Tine Jess56, Mika Kivimaki57, Marcus E. Kleber14, Jari Lahti58, Yongmei Liu59, Pedro Marques-Vidal60, Dan Mellström16, Simon P. Mooijaart40, Martina Müller-Nurasyid61,62, Brenda Penninx26, Katri Räikkönen58, Naveed Sattar63, Hubert Scharnagl64, Bengt Sennblad29,65, Angela Silveira29, Beate St Pourcain21,66,67, Nicholas J. Timpson21, Julian Trollor68,69, Jenny van Dongen12,13, Diana Van Heemst70, Sophie Visvikis-Siest27, Peter Vollenweider60, Uwe Völker52, Melanie Waldenberger61, Gonneke Willemsen12,13, Delilah Zabaneh71, Richard W. Morris72, Donna K. Arnett4, Bernhard T. Baune73,74,75, Dorret I. Boomsma12,13, Yen-Pei C. Chang11, Ian J. Deary24,51, Panos Deloukas19,76, Johan G. Eriksson77,78, David M. Evans20,21, Manuel A. Ferreira5, Tom Gaunt79,80, Vilmundur Gudnason81,82, Anders Hamsten29, Joachim Heinrich34,83,84, Aroon Hingorani32, Steve E. Humphries85, J.W. Jukema39,86, Wolfgang Koeing7, Meena Kumari57,87, Zoltan Kutalik15,88, Deborah A. Lawler79,80, Terho Lehtimäki31, Winfried März14,64,89, Karen Mather68,90, Silvia Naitza28, Mathias Nauck91, Claes Ohlsson16, Jacqueline F. Price25, Olli Raitakari92,93, Kenneth Rice94, Perminder S. Sachdev37, Eline Slagboom8,46, Thorkild I.A. Sørensen95,96, Tim Spector97, David Stacey98, Maria Stathopoulou27, Toshiko Tanaka99, S. Goya Wannamethee6, Peter Whincup100, Jerome I. Rotter101,102, Abbas Dehghan103, Eric Boerwinkle104,105, Harold Sneider2, Bruce M. Psaty10,106,107, Bram P. Prins2\*, Behrooz A. Alizadeh2\*#

**\*Equal contributions**

**#Corresponding authors**

1. Tarunveer S. Ahluwalia, PhD, Senior Researcher, Steno Diabetes Center Copenhagen, Gentofte, Denmark

Email: tarun.veer.singh.ahluwalia@regionh.dk

1. Behrooz A. Alizadeh, PhD, Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Email: b.z.alizadeh@umcg.nl

**Affiliations**

1Steno Diabetes Center Copenhagen, Gentofte, Denmark

2Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

3Mathematics and Statistics, Murdoch University, Perth, Australia

4Department of Epidemiology, University of Alabama at Birmingham School of Public Health, Birmingham, Alabama, USA

5QIMR Berghofer Medical Research Institute, Brisbane, 4006, Australia

6Department of Primary Care & Population Health, UCL Institute of Epidemiology & Health Care, University College London, London, United Kingdom

7Deutsches Herzzentrum München, Technische Universität München, Munich, Germany, DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany and Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm, Germany.

8Section of Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

9Population Health Sciences, University of Bristol, Bristol, United Kingdom

10Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA

11Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

12Department of Biological Psychology, Behaviour and Movement Sciences, Vrije Universiteit Amsterdam

13Amsterdam Public Health research institute, Amsterdam University Medical Center, The Netherlands

14Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany

15SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

16Centre for Bone and Arthritis Research (CBAR), Department of Internal Medicine and Clinical Nutrition, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

17Chronic Disease Prevention Unit, National Institute for Health and Welfare, PO Box 30, 00271 Helsinki, Finland

18Hospital for Children and Adolescents, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

19William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London, London EC1M 6BQ, UK

20University of Queensland Diamantina Institute, University of Queensland, Brisbane, Queensland, Australia

21MRC Integrative Epidemiology Unit, Population Health Science, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Beistol BS8 2BN, United Kingdom

22Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States.

23Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK, EH4 2XU

24Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK, EH8 9JZ

25Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom

26Department of Psychiatry, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands

27Université de Lorraine, Inserm, IGE-PCV, 54000, Nancy, France

28Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Cagliari, Italy

29Cardiovascular Medicine, Department of Medicine Solna, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden

30Unit of Genomics of Complex Diseases, Institut d'Investigació Biomèdica Sant Pau (IIB-Sant Pau), Barcelona, Spain

31Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland

32Institute of Cardiovascular Science, University College London, London, United Kingdom

33Discipline of Psychiatry, Adelaide Medical Schhool, University of Adelaide, Adelaide Australia

34Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

35Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

36Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, Australia

37Neuropsychiatric Institute, Prince of Wales Hospital, Sydney, Australia

38Ludwig-Maximilians-University of Munich, Dr. von Hauner Children's Hospital, Division of Metabolic Diseases and Nutritional Medicine, Munich, Germany

39Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands.

40Section of Gerontology and Geriatrics, department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands

41Baylor College of Medicine, 6565 Fannin Street, MS A-601, Houston, TX 77030

42National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, United States

43 Section of Cardiovascular Medicine and Preventive Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, United States

44Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC),VA Boston Healthcare System, Boston, Massachusetts, USA

45Department of Nutrition-Dietetics/Harokopio University, Athens, Greece

46Max Planck Institute for Biology of Ageing, Cologne, Germany

47Department of Pathology, University of Vermont College of Medicine, Burlington, VT, USA

48Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

49German Center for Diabetes Research (DZD), Neuherberg, Germany

50The Lundberg Laboratory for Diabetes Research, Department of Molecular and Clinical Medicine, Sahlgrenska Academy at the University of Gothenburg, SE-41345 Gothenburg, Sweden

51Department of Psychology, University of Edinburgh, Edinburgh, UK, EH8 9JZ

52Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany

53MediCity Research Laboratory, University of Turku, Turku, Finland.

54Department of Medical Microbiology and Immunology, University of Turku, Turku, Finland.

55Genetics Research Centre, Molecular and Clinical Sciences Institute, St George's University of London, London, United Kingdom

56Department of Epidemiology Research, Statens Serum Institut, Denmark

57Department of Epidemiology & Public Health, UCL Institute of Epidemiology & Health Care, University College London, London, United Kingdom

58Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

59Department of Epidemiology and Prevention, Wake Forest School of Medicine, Winston-Salem, NC, 27157, USA.

60Department of Internal Medicine, Lausanne University Hospital (CHUV) and University of Lausanne, Rue du Bugnon 46, Lausanne, 1011, Switzerland.

61Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany62Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany

63BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow, United Kingdom

64Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria

65Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

66Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

67Donders Institute for Brain, Cognition and Behaviour, Radboud University, The Netherlands

68Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, Australia

69Department of Developmental Disability Neuropsychiatry, School of Psychiatry, University of New South Wales, Sydney, Australia

70Section of Geriatrics and Gerontology, dept of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands

71University College London Genetics Institute, Department of Genetics, Environment and Evolution, London, UK

72Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

73Department of Psychiatry, Melbourne Medical School, The University of Melbourne, Parkville, Australia

74Department of Psychiatry and Psychotherapy, University of Muenster, Albert Schweitzer-Campus 1, A9, 48149 Muenster, Germany

75The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, Australia

76Centre for Genomic Health, Queen Mary University of London, London EC1M 6BQ, UK

77National Institute for Health and Welfare, University of Helsinki, Helsinki, Finland

78Department of General Practice and Primary health Care, Tukholmankatu 8 B , University of Helsinki, Finland

79MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, BS6 2BN

80Population Health Science, Bristol Medical School, University of Bristol

81Icelandic Heart Association, Kópavogur, Iceland

82Faculty of Medicine, University of Iceland, Reykjavik, Iceland

83Institute and Clinic for Occupational, Social and Environmental Medicine, University Hospital, LMU Munich, Munich, Germany

84Allergy and Lung Health Unit, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia

85Institute of Cardiovascular Science, Rayne Building, University College London, London, UK

86Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands

87Institute for Social and Economic Research, University of Essex, Colchester, UK

88University Center for Primary Care and Public Health, University of Lausanne, Route de la Corniche 10, Lausanne 1010, Switzerland

89SYNLAB Academy, SYNALB Holding Deutschland GmbH, Mannheim, Germany

90Neuroscience Research Australia, Sydney Australia

91Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany92Centre for Population Health Research, University of Turku and Turku University Hospital; Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku

93Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

94Department of Biostatistics, University of Washington, Seattle, WA, USA

95Novo Nordisk Foundation Center For Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

96Department of Public Health, Section of Epidemiology, University of Copenhagen, Copenhagen, Denmark

97Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom

98MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK

99Translational Gerontology Branch, National Institute on Aging, Baltimore, MD, USA.

100Population Health Research Institute, St George’s, University of London, London, United Kingdom

101Departments of Pediatrics and Medicine, Harbor-UCLA Medical Center, Los Angeles, CA, USA

102Institute for Translational Genomics and Population Sciences, Lundquist Institute (formerly LABiomed) at Harbor-UCLA Medical Center, Los Angeles, CA, USA

103Department of Epidemiology, Erasmus MC, Rotterdam, TheNetherlands

104Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX 77030, USA

105Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

106Departments of Epidemiology and Health Services, University of Washington, Seattle, WA, USA

107Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA

# ABSTRACT

Interleukin-6 (IL-6) is a multifunctional cytokine with both pro- and anti-inflammatory properties, synthesized by a wide range of tissues and cell types. Increased levels of circulating IL-6 in blood is associated with the pathophysiology of complex disorders like type 2 diabetes, cardiovascular and autoimmune diseases. Albeit, IL-6 levels are heritable with estimates up to 61%, only a few common genetic loci associated with circulating IL-6 levels have been identified. We therefore conducted a 2 stage (discovery & replication) meta genome-wide association study (GWAS) of circulating serum IL-6 concentrations comprising up to 67,428 individuals of European ancestry. We identified three IL-6 associated, independent signals on chromosomes (Chr) 2p14, and 6p21 and 1q21, reaching genome-wide significance in the combined meta-analyses (GWS; *p* < 5.0 × 10-8). Among the identified single nucleotide polymorphisms (SNPs), *IL1F10/IL1RN* rs6734238 (Chr 2q14, *p* = 1.8 × 10-11), and *HLA-DRB1/DRB5* rs660895 (Chr 6p21, *p* =1.5 × 10-10) were novel while the *IL6R* rs4537545 (Chr 1q21, *p*= 1.2 × 10-122) is a known locus.

Our study identifies 2 novel loci for circulating IL-6 levels uncovering new immunological and inflammatory pathways that may influence IL-6 pathobiology.

**INTRODUCTION**

Interleukin-6 (IL-6) is a multifunctional cytokine, which is involved in a wide range of immunomodulatory processes, from cellular migration and adhesion to proliferation and maturation ([1](#_ENREF_1), [2](#_ENREF_2)). Specifically in regard to the immune system, interleukins are involved in cell differentiation and activation([3](#_ENREF_3)). IL-6 is synthesized by a variety of different cell types, both immune system related such as monocytes([4](#_ENREF_4)), B-cells([5](#_ENREF_5)) and T-cells([6](#_ENREF_6)) and non-immune cells such as epithelial and smooth muscle cells([7](#_ENREF_7)), adipocytes([8](#_ENREF_8)), endothelial cells([9](#_ENREF_9)), and osteoblasts([10](#_ENREF_10)).

We have previously demonstrated that IL-6 levels decrease with age in children and increase with age in adults ([11](#_ENREF_11)). Increased levels of IL-6 have been observed in various disease classes, not surprisingly in autoimmune diseases such as rheumatoid arthritis ([12](#_ENREF_12)) and systemic juvenile idiopathic arthritis ([13](#_ENREF_13)), but also in cardio-metabolic outcomes like heart failure and coronary heart disease ([14](#_ENREF_14)), atherosclerosis ([15](#_ENREF_15)), and also in cancers ([16](#_ENREF_16)) and psychological disorders (like depression) ([17](#_ENREF_17)). Given its possibly critical role in the pathogenesis of different disorders, it forms an appropriate choice for drug targeting ([18](#_ENREF_18)). The most illustrious IL-6 inhibitor is tocilizumab ([19](#_ENREF_19)), a monoclonal antibody binding the IL-6 receptor proved to improve patients with rheumatoid arthritis ([20](#_ENREF_20)) and systemic juvenile idiopathic arthritis ([21](#_ENREF_21)) with high efficacy.

IL-6 baseline levels are heritable with estimates from twin studies ranging between 15-61% ([22-25](#_ENREF_22)). To date few, relatively small-scale GWAS ([26-29](#_ENREF_26)) or sequencing based association studies ([30](#_ENREF_30)) have been performed for IL-6, identifying the IL-6 receptor gene (*IL-6R*) and the gene encoding histo-blood group ABO system transferase (*ABO*) variants reaching GWS. A genetic risk score constructed of variants identified in the study by Shah and colleagues explained up to 2% of variation in IL-6 levels ([29](#_ENREF_29)), leaving a substantial part of the heritability unexplained. These seemingly sparse results could be due to power limitation issues. Substantially increasing sample sizes of GWASs would very likely lead to the identification of additional variants explaining IL-6 levels ([31-33](#_ENREF_31)).

In this study, we greatly increase the sample size, compared to the previous studies, through inclusion of more cohorts and identify additional genetic variants explaining IL-6 levels. Thus, the current study is a meta-analysis of GWASs of 66,341 individuals of European ancestry.

# METHODS

## Discovery Stage

### Study populations The discovery stage included 52,654 individuals from 26 cohorts of European ancestry described in **Supp. Text S1** with summary characteristics in **Supp.** **Table S1**. Only population-based samples or healthy controls from case-control studies were included in the analyses.

### Serum IL-6 measurements Each study typically collected venous blood samples stored below -80°C until the time of measurement using various types of immunoassays and expressed as pg/ml (**Supp.** **Table S1).** IL-6 measurement methods are available in **Supp. Table S2;** the trait transformation and phenotype data quality control (QC) have been stated under **Supp. Text S3.** (**Supp. Text** **S3.1** and **S3.2**).

### Genotyping and imputation Each participating cohort performed genome-wide genotyping and subsequent quality control (QC), and imputation using a variety of genotyping platforms, and using haplotypes from the Hapmap Phase II reference panel (NCBI Build 36), using IMPUTE ([34](#_ENREF_34)), MACH ([35](#_ENREF_35)), Minimac ([36](#_ENREF_36)), or BIMBAM ([37](#_ENREF_37)), to infer unobserved genotypes, resulting in a per-study set of ~2.5 million SNPs **(Supp. Table S3)**.

### **Statistical methods**

### Genetic association testing GWAS result files from each study cohort underwent QC with the QCGWAS package in R ([38](#_ENREF_38)) **(Supp. Text S3.2)**, prior to meta-analysis. Each cohort tested the association between SNPs and natural log transformed values of serum IL-6 levels (**Supp. Text S3.3**) using linear regression model, or linear mixed effect models to account for familial correlation when warranted, with additive genetic effects accounting for imputation uncertainty while adjusting for age, sex, population sub-structure through study-specific principal components and/or study specific site, when necessary. An additional QC measure in the genetic association testing comprised, filtering a known false positive GWAS signal (ABO gene) ([24](#_ENREF_24), [26](#_ENREF_26)) while using a specific cytokine measurement kit. After identifying the 4 cohorts in the current study using this IL-6 measurement kit additional QC (**Supp. Text S3.2**) measures were performed before proceeding to the discovery meta-analyses.

### GWAS meta analyses

### Individual GWAS results from 26 European studies were meta-analysed using the inverse variance weighted, fixed-effects method applying the double genomic control (GC) correction for population stratification as implemented in GWAMA([39](#_ENREF_39)). We performed an approximate joint conditional analysis to identify distinct signals in a specific chromosomal region as implemented in GCTA([40](#_ENREF_40)) using a high quality imputed genetic dataset (NEtherlands Study of Depression and Anxiety (NESDA) study) to estimate linkage disequilibrium (LD) ([41](#_ENREF_41)). Furthermore, regional association plots for the discovered loci were generated through the LocusZoom ([38](#_ENREF_38)) tool. We used SNAP tool([42](#_ENREF_42)) to perform the pairwise LD checks (HapMap release 22 data) and to verify low LD with secondary signals. SNPs selected for the replication stage had to fulfil the following criteria: (i) having an association pdiscovery ≤ 1×10-5, (ii) statistically distinct based on LD (r2>0.25), and (iii) available in at least 50% of study cohorts.

## Replication Stage

### Study population A total of 12 distinct SNPs were taken forward for replication. Replication analyses were performed using a combination of in silico and de novo genotyping in 14,774 individuals from European ancestry from 12 cohorts described in **Supp. Text S2**, and following the same QC and statistical procedures as for the discovery phase **(Supp. Text S3 and Supp. Table S3)**. Venous blood samples (serum or plasma) were collected and stored at -80°C until various immunoassay measurements were performed. IL-6 was measured as pg/ml. Each cohort tested the selected SNPs using the same statistical model as for the discovery association analyses. Effect size estimates of all replication SNPs from each replication study were compared with the effect size estimates from the discovery meta-analyses.

### To account for the inter-study IL6 assay sensitivity differences in the replication association analyses, we combined results across the replication studies using a fixed-effects, sample-size weighted Z-score meta-analysis as implemented in the METAL package([43](#_ENREF_43)). The summary GWAS meta-analyses results from the discovery and replication stages were then used to perform a combined (discovery + replication) GWAS meta-analysis using a sample-size weighted Z-score method. Variants that were significant in the replication meta-analysis (p<0.05) with an overall pcombined < 5×10-8 in the combined meta-analysis were considered genome-wide significant.

### Heritability estimates We approximated the variance explained by all distinct lead SNPs from the meta-analysis using the following formula:

where EAF is the effect allele frequency, and the effect size of the individual variants and n is the total number of lead variants. The current formula may overestimate the variance to a small extent as some level of SNP correlation was existent (LD r2<0.25). The variance of the residuals of ln(IL-6) was calculated using data from the NESDA cohort (N=2,517). The total common SNP heritability of serum IL-6 levels explained by all GWAS variants was estimated using the observed Z-statistics from the discovery analyses for a subset of pruned SNPs. Following the original method (SumVg) ([44](#_ENREF_44)), we pruned the imputed (based on the 1000G *Phase1 Integrated Release, Version 3, 2012.04.30* reference panel) genotypes of the NESDA cohort using PLINK ([45](#_ENREF_45)), by removing correlated SNPs at r2>0.25 within a 100-SNP sliding window, and with a step size of 25 SNPs per forward slide. This resulted in a pruned set of 1, 63, 459 SNPs.

### Fine mapping and putative causal risk variants and candidate genes We searched for variants in high LD (r2>0.8) within a 1 Mb region on either side of the lead SNPs using 1000 Genomes sequence data (Phase1 Integrated Release, Version 3, 2012.04.30), by means of tools available in Liftover([46](#_ENREF_46)), VCFtools([47](#_ENREF_47)), and PLINK([45](#_ENREF_45)). We subsequently annotated these variants using ANNOVAR([48](#_ENREF_48)) with the RefSeq([49](#_ENREF_49)) database for variant function and genic residence or distance.

*Mining known SNP-trait associations* We used the PhenoScanner([50](#_ENREF_50)) data mining tool to identify existing SNP-trait associations for GWAS (*p*<5×10-8), eQTLs (*p*< 1 × 10-4), and metabolites (*p*<=0.01)for the GWAS significant findings in the current study.

# RESULTS

A total of 52,654 individuals of European descent from 26 cohorts were included in the discovery GWAS meta-analysis with up to 2,835,074 autosomal SNPs passing quality control. The final genomic control inflation factor (λGC after correction) at the discovery stage meta-analysis was 1.0. Two common genetic loci on chromosomes 1q21 (*IL6R*), and 6p21 (*HLA-DRB1*/*HLA-DRB5*), associated with IL-6 levels at genome wide significance (*p* < 5×10−8) while a third locus 2q14 (*IL1F10*/*IL1RN*) was suggestively close (*p* < 1.0×10−5 **Figure 1**). The minor alleles of *IL6R* rs4537545 (T allele, β= 0.091; *p*=8.39×10-85), *IL1F10*/*IL1RN* rs6734238 (G allele, β=0.025; *p*=1.45×10-7) and rs660895 *HLA-DRB1*/*5* rs660895 (G allele, β=0.036; *p*=1.80×10-9) associated with increased circulating IL-6 levels, respectively (**Table 1**). LD based conditional analysis identified two additional genome-wide significant distinct SNPs rs11265618 and rs10796927, (r2<0.25 with the lead variant, rs4537545) on the chromosome 1q21 locus.

A total of 12 SNPs, with  p < 1 × 10−5 in the discovery analyses were selected for further replication (**Supp. Table S4**). Additionally 3 SNPs as negative controls and 3 SNPs in LD (r2>0.25) with the chromosome 1 index SNP (to control for possible variability of index SNP across replication cohorts) were also added to the replication list.

Three loci (chromosome (Chr.) 1q21 *IL6R*, Chr 6p21 *HLA-DRB1*/*5* and Chr 2q14 *ILF10/IL1RN*) replicated (*p*<0.05) reaching GWS (1q21 rs4537545, *p*combined= 1.20×10−122; 6p21 rs660895, *p*combined= 1.55×10−10; 2q14 rs6734238, *p*combined= 1.84×10−11) in the combined meta-analyses analysis (discovery and replication stage, **Table 1**). The Chr. 2 *ILF10/IL1RN* and 6 *HLA-DRB1*/*HLA-DRB5* loci are novel whereas the Chr. 1 *IL6R* is a known locus for IL6 levels. In both, discovery and replication association analyses the effect sizes were generally consistent across individual studies for genome-wide significant variants, and we did not observe evidence of heterogeneity (I2 p value > 0.05). The imputation quality scores (r2) for the GWAS significant (Index) SNPs for each cohort (Discovery and Replication) are available in **Supp. Table S5.** The 9 SNPs that showed suggestive association in the discovery stage did not reach genome-wide significance in the combined meta-analyses (**Supp. Table S4**).

The three GWAS index SNPs when combined, explained approximately 1.06% of the variance in circulating levels of IL-6 (using data from the NESDA cohort). The phenotypic variance explained by all the common variants was estimated to be 4.45% using the SumVg method([44](#_ENREF_44)).

***Replication of other known/suggestive loci for IL-6*** *IL6R* was the only IL6 known locus that we replicate at genome wide significance. *IL1RN* and *HLA-DRB1*, our primary findings have been reported as suggestive loci (*p*>1×10-6 and *p* <1×10-4) by Shah *et al.* while some known/suggestive IL6 loci (*ABO*, *BUD13*, *TRIB3*, and *SEZ6L*) did not replicate (*p*>0.05) in the current study.

# *SNP- other trait associations*

# Significant SNP-trait associations at GWAS, eQTL and metabolite levels were mined for the identified index SNPs using the Pheno Scanner database (accessed, July 2018)

*GWAS based IL1F10*/*IL1RN* rs6734238 G allele has been associated with increased levels of serum C reactive protein (CRP) and decreased fibrinogen levels in recent GWAS reports ([51](#_ENREF_51), [52](#_ENREF_52)). *HLA-DRB1/DRB5*rs660895 G allele is associated with increased risk of Rheumatoid arthritis (RA) in Europeans and Asians([53](#_ENREF_53)), IgA nephropathy in Asians([54](#_ENREF_54)), while decreased risk of Ulcerative colitis and Inflammatory Bowel Disease (IBD)([55](#_ENREF_55)). *IL-6R* rs4537545 T allele is known to associate with increased circulating CRP levels ([56](#_ENREF_56)) and a decreased risk of RA([53](#_ENREF_53)) in mixed ancestries (Supp. Table S6.1).

*eQTL based* rs6734238 is associated with expression levels in peripheral blood, skin, brain cerebellar hemisphere, adipose subcutaneous tissue and thyroid tissue (p<1.0×10-4; Supp. Table S6.2). rs660895 has been associated with expression levels in multiple tissues including lung, peripheral blood, whole blood, thyroid, tibial artery, skeletal muscle, tibial nerve, adipose tissue, stomach, aorta, heart left ventricle, skin sun exposed lower leg, testis and small intestine, lymphoblastoid cell lines (p<1.0 × 10-7; Supp. Table S6.2). *IL6R* SNP is also associated with IL6R expression levels in peripheral blood tissue, tibial artery, and adrenal gland (p<1.0 × 10-4; Supp. Table S6.2). Some metabolites associated with these SNPs (p< 0.01) are also listed under table S6.3.

# DISCUSSION

We performed a large GWAS meta-analysis for circulating IL-6 levels, which includes 66,341 individuals of European ancestry. We identified three loci associated with circulating concentrations of IL-6 in the general population amongst which two are novel (Chr 6p21and Chr 2q14), located in/nearby genes (*HLA-DRB1* and *IL1RN/IL-38*) with inflammatory roles explaining up to 1.06 % variance.

The strongest SNP Interleukin-6 Receptor (*IL-6R)* rs4537545 at the 1q21 locus is in high LD (r2=0.95) with a missense *IL-6R* SNP rs2228145 (D358A). The missense SNP is known to impair the responsiveness of cells targeted by IL-6 ([57](#_ENREF_57)) by reducing IL-6R expression on cell surfaces ([58](#_ENREF_58)), and increasing levels of soluble IL-6R in individuals homozygous for this mutation ([59](#_ENREF_59), [60](#_ENREF_60)). Recently it has been demonstrated that increased levels of sIL-6R induced by this variant, can be explained by ectodomain shedding off IL-6R, a mechanism in which membrane-associated proteins are rapidly converted into soluble effectors whereby simultaneously cell surface expression of the same protein is reduced ([61](#_ENREF_61)). Increased levels of sIL-6R may act as a counter-balance to limit exaggerated IL-6 signalling, and may explain the protective effect of the 358A allele for various cardiovascular diseases including coronary artery disease (CAD)([62-64](#_ENREF_62)), atrial fibrillation([65](#_ENREF_65)), and abdominal aortic aneurysm([66](#_ENREF_66)) as well as RA ([67](#_ENREF_67)). However in contrast with this finding, the IL-6-sIL-6R complex itself is capable of transducing IL-6 signalling to non-IL-6R expressing cells, known as trans-signalling([68](#_ENREF_68)), and it is this mechanism, as opposed to classic signalling, that is linked to chronic inflammatory disorders including IBD and RA([69](#_ENREF_69)). Blocking IL-6 signalling cascades can be achieved by using an IL-6R specific inhibitor in the form of a monoclonal antibody, tocilizumab, which is a widely used therapy in the treatment of RA. Several variants in IL-6R, including rs2228145, may assist in the prediction of patient response to tocilizumab in rheumatoid arthritis([70](#_ENREF_70)). rs4537545 T allele which associates with IL6 levels is known to associate with increased circulating CRP levels([56](#_ENREF_56)) and a decreased risk of RA([53](#_ENREF_53)) in studies comprising mixed ancestries. Moreover this SNP has been associated with IL6R expression in peripheral blood tissue, tibial artery, and adrenal gland (Supp Table 6.2).

At the identified chromosome 2 locus the lead SNP, rs6734238, is intergenic and has also been reported to associate with circulating CRP and fibrinogen([51](#_ENREF_51), [52](#_ENREF_52), [71](#_ENREF_71))levels. The nearest genes to this locus are the Interleukin 1 Family Member 10 (*IL1F10*, distance=7.6 kB, currently known as *IL-38*) and Interleukin 1 Receptor Antagonist (*IL1RN*, distance=34.4 kB). *IL1F10/IL-38* and *IL1RN* variants (rs6759676 and rs4251961) in partial LD with the lead SNP (r2LD:0.10 and 0.61) have been recently reported to be protective against development of insulin resistance (PMID: 24969107). This further supports the molecular mechanisms behind IL-6 mediated insulin secretion via glucagon-like peptide 1 (GLP-1) (PMID: 22037645) contributing to type 2 diabetes (T2D) pathophysiology. For IL-6 specifically it has been found that synthesis increases when dendritic cells are stimulated by bacterial lipopolysaccharides (LPS) in the presence of *IL1F10* ([72](#_ENREF_72)). *IL-1RN* is another member of the interleukin 1 cytokine family, with suggestive evidence for involvement in determining IL-6 levels in blood. One study found significant associations of one variant residing in this gene, rs4251961, with plasma CRP and IL-6 levels, albeit not independently replicated and not genome-wide significant (*p*=1×10-4 and *p*=0.004) ([73](#_ENREF_73)). Our lead SNP was not in high LD (r2 < 0.8) with variants in either neighbouring genes, and therefore in conjunction with its intergenic position, identifying a causal variant in this locus remains non-trivial.

The 6p21 locus rs660895 that was identified resides within the HLA region, which forms one of the most complex genomic regions to study due to its large LD blocks and sequence diversity. This region has some population substructure in Europeans which may have influenced the results however 1) each cohort population substructure adjustment was applied, followed by genomic correction for overall discovery stage meta analyses. Thus we reduced the chances that the population substructure may have had on this locus.

The nearest genes to the index SNP, *HLA-DRB1* (distance=19.8 kB) and *HLA-DQA1* (distance=27.8 kB) are both histocompatibility complex genes encoding proteins that form cell surface complexes for certain immune system cells helping in antigen presentation to trigger an immune response. It is noteworthy that variations at this locus code for antigen presenting complexes (APCs) which has been previously associated with diseases having dysfunctional immune system; while we report for the first time that there exist also a strong association of this locus with circulating cytokine levels. Therefore the association of this locus with the disease may corroborate through its effect through IL6 levels. One high-LD SNP (rs9272422, r2=0.82 with our index SNP, rs660895) residing in the promoter region of *HLA-DQA1*, support this hypothesis and has been identified previously for Systemic Lupus Erythematosus ([74](#_ENREF_74)) and Ulcerative Colitis ([75](#_ENREF_75)). rs660895 G allele is associated with increased risk of Rheumatoid arthritis (RA) in Europeans and Asians ([53](#_ENREF_53)), IgA nephropathy in Asians([54](#_ENREF_54)), while decreased risk of Ulcerative colitis and Inflammatory Bowel Disease (IBD)([55](#_ENREF_55))

Various studies aimed to identify genetic variation underlying levels of IL-6 (22-26 have found genome-wide significant associations in the *IL-6R* and *ABO* genes. The study performed by Shah and colleagues25 found suggestive evidence (not genome-wide significant) for additional loci, including *ABO*, *BUD13*, *TRIB3* and *SEZ6L*, none of which replicate in the current study (*p*>0.05) indicating that these might be false positive findings due to reduced sample size or loci with sex specific effects (associations based on all women population) or due to technical shortcomings (*ABO* locus).

Several other genes encompass the 1q21 locus, including Src Homology 2 Domain Containing E (*SHE*) and Tudor Domain containing 10 (*TDRD10*), but after careful analysis there does not seem to be other signals within this locus25.

It is surprising that even with increased statistical power (*n*discovery:52,654; *n*replicaiton: 14,774) in the current study (compared to the previous IL6 GWAS) ([29](#_ENREF_29)), we could identify three genetic loci (1q21, 2q14, and 6p21) accounting ~ 1% of the genetic variance for circulating IL-6 levels. According to the current estimates, the heritability levels for IL6 levels range between 15 to 61%, suggesting that an enormous increases in sample sizes would be required to identify additional variants explaining this remaining heritability. Multiple explanations for this so-called missing heritability phenomenon have been proposed in the past68, which can be sought in different classes of genetic variation such as rare variants69, or can be explained by non-additive effects which may cause inflated estimates of heritability70. Plausible evidence for other sources of unexplained heritability that have been found are epigenetic changes71, and haplotypes of common SNPs72. Collectively, our results provided additional insights into the biology of IL-6 synthesis. Albeit we identify new loci, our main limitation is that we investigate only common variants. Future studies are recommended to aim for identification of additional common but also rare variants with increasingly smaller effects, by firstly using deeper imputation panels, (UK10K project73 or the Haplotype Reference Consortium), a strategy that holds great promise69, and secondly by making use of genetically isolated populations74. Thirdly, we would like to stress the importance of phenotype harmonisation. As we identified genome-wide variants in the *ABO* locus, in four studies participating in the discovery, but not in the remaining 22 cohorts, there is a strong indication that this locus is assay-specific. Also, conventionally increasing sample sizes without correction for population substructures may raise heterogeneity within populations([76](#_ENREF_76)), likely concealing the SNPs that have an effect in particular subgroups. Future specific studies should counter the widely held assumption of unconditional risk alleles of complex traits and focus on the importance of studying more homogenous subgroups in order to, for example, investigate the age-dependant effect of genetic variants([77](#_ENREF_77), [78](#_ENREF_78)). Here while further exploring the pleiotropic effect of IL-6-related variants, we identified phenotypes differentially regulated by diverse variants in the 1q21 locus. Biologic systems are dynamic complex networks and are evolving through lifespan and investigating the interrelationships existing between phenotypes as well as between genetic variations and phenotypic variations has the potential for uncovering the complex mechanisms. This is the case here for IL-6 and tailored methodologies should be devoted to the study of such traits, hopefully resulting in clinically significant breakthroughs. Future collaborative efforts therefore should strive to use well-calibrated assays, standardised protocols for sample handling and processing75, though this will be difficult to achieve in practice.

In conclusion we identify two novel common genetic variants associated with circulating IL-6 levels that may influence pathophysiology of complex cardio-metabolic, psychiatric and immunological traits, among individuals of European ancestry. This may further unravel additional pathways and therapeutic targets that can be developed for the aforementioned disorders while suggesting for looking deeper into the genome for variants (rare and common) with greater individual effects.

**Acknowledgments**

We acknowledge the participants of this study and the funding resources that have contributed in achieving this effort. Detailed acknowledgements can be found in the **Supplementary table S7.**

**Funding resources**

Detailed Funding statements can be found in the **Supplementary Table S7.**

**Conflict of Interest**

There is no conflict of interest from the co-authors.

# WEB RESOURCES

QCGWAS, https://cran.r-project.org/web/packages/QCGWAS/index.html

GWAMA, http://www.well.ox.ac.uk/gwama/

METAL, http://csg.sph.umich.edu//abecasis/metal/

GCTA, http://www.complextraitgenomics.com/software/gcta/

LocusZoom, http://csg.sph.umich.edu/locuszoom/

1000 Genomes, ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/

PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/

VCFtools, http://vcftools.sourceforge.net/

ANNOVAR, http://www.openbioinformatics.org/annovar/

PhenoScanner, http://www.phenoscanner.medschl.cam.ac.uk/phenoscanner

**REFERENCES**

1 Kishimoto, T. (2010) IL-6: from its discovery to clinical applications. *International immunology*, **22**, 347-352.

2 Nishimoto, N. and Kishimoto, T. (2006) Interleukin 6: from bench to bedside. *Nature clinical practice. Rheumatology*, **2**, 619-626.

3 Brocker, C., Thompson, D., Matsumoto, A., Nebert, D.W. and Vasiliou, V. (2010) Evolutionary divergence and functions of the human interleukin (IL) gene family. *Human genomics*, **5**, 30-55.

4 Gelinas, L., Falkenham, A., Oxner, A., Sopel, M. and Legare, J.F. (2011) Highly purified human peripheral blood monocytes produce IL-6 but not TNFalpha in response to angiotensin II. *Journal of the renin-angiotensin-aldosterone system : JRAAS*, **12**, 295-303.

5 Kitani, A., Hara, M., Hirose, T., Harigai, M., Suzuki, K., Kawakami, M., Kawaguchi, Y., Hidaka, T., Kawagoe, M. and Nakamura, H. (1992) Autostimulatory effects of IL-6 on excessive B cell differentiation in patients with systemic lupus erythematosus: analysis of IL-6 production and IL-6R expression. *Clinical and experimental immunology*, **88**, 75-83.

6 Li, T. and He, S. (2006) Induction of IL-6 release from human T cells by PAR-1 and PAR-2 agonists. *Immunology and cell biology*, **84**, 461-466.

7 Ng, E.K., Panesar, N., Longo, W.E., Shapiro, M.J., Kaminski, D.L., Tolman, K.C. and Mazuski, J.E. (2003) Human intestinal epithelial and smooth muscle cells are potent producers of IL-6. *Mediators of inflammation*, **12**, 3-8.

8 Fain, J.N. (2006) Release of interleukins and other inflammatory cytokines by human adipose tissue is enhanced in obesity and primarily due to the nonfat cells. *Vitamins and hormones*, **74**, 443-477.

9 Podor, T.J., Jirik, F.R., Loskutoff, D.J., Carson, D.A. and Lotz, M. (1989) Human endothelial cells produce IL-6. Lack of responses to exogenous IL-6. *Annals of the New York Academy of Sciences*, **557**, 374-385; discussion 386-377.

10 Sanchez, C., Gabay, O., Salvat, C., Henrotin, Y.E. and Berenbaum, F. (2009) Mechanical loading highly increases IL-6 production and decreases OPG expression by osteoblasts. *Osteoarthritis and cartilage*, **17**, 473-481.

11 Haddy, N., Sass, C., Maumus, S., Marie, B., Droesch, S., Siest, G., Lambert, D. and Visvikis, S. (2005) Biological variations, genetic polymorphisms and familial resemblance of TNF-alpha and IL-6 concentrations: STANISLAS cohort. *European journal of human genetics : EJHG*, **13**, 109-117.

12 Madhok, R., Crilly, A., Watson, J. and Capell, H.A. (1993) Serum interleukin 6 levels in rheumatoid arthritis: correlations with clinical and laboratory indices of disease activity. *Annals of the rheumatic diseases*, **52**, 232-234.

13 de Benedetti, F., Massa, M., Robbioni, P., Ravelli, A., Burgio, G.R. and Martini, A. (1991) Correlation of serum interleukin-6 levels with joint involvement and thrombocytosis in systemic juvenile rheumatoid arthritis. *Arthritis and rheumatism*, **34**, 1158-1163.

14 Cesari, M., Penninx, B.W., Newman, A.B., Kritchevsky, S.B., Nicklas, B.J., Sutton-Tyrrell, K., Rubin, S.M., Ding, J., Simonsick, E.M., Harris, T.B. *et al.* (2003) Inflammatory markers and onset of cardiovascular events: results from the Health ABC study. *Circulation*, **108**, 2317-2322.

15 Qu, D., Liu, J., Lau, C.W. and Huang, Y. (2014) IL-6 in diabetes and cardiovascular complications. *British journal of pharmacology*, **171**, 3595-3603.

16 Mauer, J., Denson, J.L. and Bruning, J.C. (2015) Versatile functions for IL-6 in metabolism and cancer. *Trends in immunology*, **36**, 92-101.

17 Zhang, C., Wu, Z., Zhao, G., Wang, F. and Fang, Y. (2016) Identification of IL6 as a susceptibility gene for major depressive disorder. *Scientific reports*, **6**, 31264.

18 Calabrese, L.H. and Rose-John, S. (2014) IL-6 biology: implications for clinical targeting in rheumatic disease. *Nature reviews. Rheumatology*, **10**, 720-727.

19 Scheinecker, C., Smolen, J., Yasothan, U., Stoll, J. and Kirkpatrick, P. (2009) Tocilizumab. *Nature reviews. Drug discovery*, **8**, 273-274.

20 Yazici, Y., Curtis, J.R., Ince, A., Baraf, H., Malamet, R.L., Teng, L.L. and Kavanaugh, A. (2012) Efficacy of tocilizumab in patients with moderate to severe active rheumatoid arthritis and a previous inadequate response to disease-modifying antirheumatic drugs: the ROSE study. *Annals of the rheumatic diseases*, **71**, 198-205.

21 Yokota, S., Imagawa, T., Mori, M., Miyamae, T., Aihara, Y., Takei, S., Iwata, N., Umebayashi, H., Murata, T., Miyoshi, M. *et al.* (2008) Efficacy and safety of tocilizumab in patients with systemic-onset juvenile idiopathic arthritis: a randomised, double-blind, placebo-controlled, withdrawal phase III trial. *Lancet*, **371**, 998-1006.

22 Worns, M.A., Victor, A., Galle, P.R. and Hohler, T. (2006) Genetic and environmental contributions to plasma C-reactive protein and interleukin-6 levels--a study in twins. *Genes and immunity*, **7**, 600-605.

23 Sas, A.A., Jamshidi, Y., Zheng, D., Wu, T., Korf, J., Alizadeh, B.Z., Spector, T.D. and Snieder, H. (2012) The age-dependency of genetic and environmental influences on serum cytokine levels: a twin study. *Cytokine*, **60**, 108-113.

24 Neijts, M., van Dongen, J., Kluft, C., Boomsma, D.I., Willemsen, G. and de Geus, E.J. (2013) Genetic architecture of the pro-inflammatory state in an extended twin-family design. *Twin research and human genetics : the official journal of the International Society for Twin Studies*, **16**, 931-940.

25 Amaral, W.Z., Krueger, R.F., Ryff, C.D. and Coe, C.L. (2015) Genetic and environmental determinants of population variation in interleukin-6, its soluble receptor and C-reactive protein: insights from identical and fraternal twins. *Brain, behavior, and immunity*, **49**, 171-181.

26 Melzer, D., Perry, J.R., Hernandez, D., Corsi, A.M., Stevens, K., Rafferty, I., Lauretani, F., Murray, A., Gibbs, J.R., Paolisso, G. *et al.* (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS genetics*, **4**, e1000072.

27 Naitza, S., Porcu, E., Steri, M., Taub, D.D., Mulas, A., Xiao, X., Strait, J., Dei, M., Lai, S., Busonero, F. *et al.* (2012) A genome-wide association scan on the levels of markers of inflammation in Sardinians reveals associations that underpin its complex regulation. *PLoS genetics*, **8**, e1002480.

28 Comuzzie, A.G., Cole, S.A., Laston, S.L., Voruganti, V.S., Haack, K., Gibbs, R.A. and Butte, N.F. (2012) Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PloS one*, **7**, e51954.

29 Shah, T., Zabaneh, D., Gaunt, T., Swerdlow, D.I., Shah, S., Talmud, P.J., Day, I.N., Whittaker, J., Holmes, M.V., Sofat, R. *et al.* (2013) Gene-centric analysis identifies variants associated with interleukin-6 levels and shared pathways with other inflammation markers. *Circulation. Cardiovascular genetics*, **6**, 163-170.

30 Sidore, C., Busonero, F., Maschio, A., Porcu, E., Naitza, S., Zoledziewska, M., Mulas, A., Pistis, G., Steri, M., Danjou, F. *et al.* (2015) Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nature genetics*, **47**, 1272-1281.

31 Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M. and Crawford, D.C. (2010) PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*, **26**, 1205-1210.

32 Pendergrass, S.A., Brown-Gentry, K., Dudek, S.M., Torstenson, E.S., Ambite, J.L., Avery, C.L., Buyske, S., Cai, C., Fesinmeyer, M.D., Haiman, C. *et al.* (2011) The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genetic epidemiology*, **35**, 410-422.

33 Pendergrass, S.A., Brown-Gentry, K., Dudek, S., Frase, A., Torstenson, E.S., Goodloe, R., Ambite, J.L., Avery, C.L., Buyske, S., Buzkova, P. *et al.* (2013) Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. *PLoS genetics*, **9**, e1003087.

34 Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics*, **39**, 906-913.

35 Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology*, **34**, 816-834.

36 Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. and Abecasis, G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature genetics*, **44**, 955-959.

37 Servin, B. and Stephens, M. (2007) Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics*, **3**, e114.

38 van der Most, P.J., Vaez, A., Prins, B.P., Munoz, M.L., Snieder, H., Alizadeh, B.Z. and Nolte, I.M. (2014) QCGWAS: A flexible R package for automated quality control of genome-wide association results. *Bioinformatics*, **30**, 1185-1186.

39 Magi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC bioinformatics*, **11**, 288.

40 Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*, **42**, 565-569.

41 Penninx, B.W., Beekman, A.T., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W., Assendelft, W.J. *et al.* (2008) The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *International journal of methods in psychiatric research*, **17**, 121-140.

42 Pei, Y.F., Zhang, L., Li, J. and Deng, H.W. (2010) Analyses and comparison of imputation-based association methods. *PloS one*, **5**, e10827.

43 Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190-2191.

44 So, H.C., Li, M. and Sham, P.C. (2011) Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genetic epidemiology*, **35**, 447-456.

45 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, **81**, 559-575.

46 Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F. *et al.* (2006) The UCSC Genome Browser Database: update 2006. *Nucleic acids research*, **34**, D590-598.

47 Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

48 Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, **38**, e164.

49 Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, **33**, D501-504.

50 Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J. *et al.* (2016) PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics*, **32**, 3207-3209.

51 Sabater-Lleal, M., Huang, J., Chasman, D., Naitza, S., Dehghan, A., Johnson, A.D., Teumer, A., Reiner, A.P., Folkersen, L., Basu, S. *et al.* (2013) Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation*, **128**, 1310-1324.

52 Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P. *et al.* (2011) Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*, **123**, 731-738.

53 Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. *et al.* (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376-381.

54 Yu, X.Q., Li, M., Zhang, H., Low, H.Q., Wei, X., Wang, J.Q., Sun, L.D., Sim, K.S., Li, Y., Foo, J.N. *et al.* (2011) A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nature genetics*, **44**, 178-182.

55 Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T. *et al.* (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*, **47**, 979-986.

56 Elliott, P., Chambers, J.C., Zhang, W., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund, P., Engert, J.C., Bennett, D. *et al.* (2009) Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA*, **302**, 37-48.

57 Ferreira, R.C., Freitag, D.F., Cutler, A.J., Howson, J.M., Rainbow, D.B., Smyth, D.J., Kaptoge, S., Clarke, P., Boreham, C., Coulson, R.M. *et al.* (2013) Functional IL6R 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS genetics*, **9**, e1003444.

58 Stone, K., Woods, E., Szmania, S.M., Stephens, O.W., Garg, T.K., Barlogie, B., Shaughnessy, J.D., Jr., Hall, B., Reddy, M., Hoering, A. *et al.* (2013) Interleukin-6 receptor polymorphism is prevalent in HIV-negative Castleman Disease and is associated with increased soluble interleukin-6 receptor levels. *PloS one*, **8**, e54610.

59 Rafiq, S., Frayling, T.M., Murray, A., Hurst, A., Stevens, K., Weedon, M.N., Henley, W., Ferrucci, L., Bandinelli, S., Corsi, A.M. *et al.* (2007) A common variant of the interleukin 6 receptor (IL-6r) gene increases IL-6r and IL-6 levels, without other inflammatory effects. *Genes and immunity*, **8**, 552-559.

60 Galicia, J.C., Tai, H., Komatsu, Y., Shimada, Y., Akazawa, K. and Yoshie, H. (2004) Polymorphisms in the IL-6 receptor (IL-6R) gene: strong evidence that serum levels of soluble IL-6R are genetically influenced. *Genes and immunity*, **5**, 513-516.

61 Hayashida, K., Bartlett, A.H., Chen, Y. and Park, P.W. (2010) Molecular and cellular mechanisms of ectodomain shedding. *Anatomical record*, **293**, 925-937.

62 Collaboration, I.R.G.C.E.R.F., Sarwar, N., Butterworth, A.S., Freitag, D.F., Gregson, J., Willeit, P., Gorman, D.N., Gao, P., Saleheen, D., Rendon, A. *et al.* (2012) Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. *Lancet*, **379**, 1205-1213.

63 Interleukin-6 Receptor Mendelian Randomisation Analysis, C., Swerdlow, D.I., Holmes, M.V., Kuchenbaecker, K.B., Engmann, J.E., Shah, T., Sofat, R., Guo, Y., Chung, C., Peasey, A. *et al.* (2012) The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. *Lancet*, **379**, 1214-1224.

64 Consortium, C.A.D., Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T.L., Thompson, J.R., Ingelsson, E., Saleheen, D., Erdmann, J. *et al.* (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics*, **45**, 25-33.

65 Schnabel, R.B., Kerr, K.F., Lubitz, S.A., Alkylbekova, E.L., Marcus, G.M., Sinner, M.F., Magnani, J.W., Wolf, P.A., Deo, R., Lloyd-Jones, D.M. *et al.* (2011) Large-scale candidate gene analysis in whites and African Americans identifies IL6R polymorphism in relation to atrial fibrillation: the National Heart, Lung, and Blood Institute's Candidate Gene Association Resource (CARe) project. *Circulation. Cardiovascular genetics*, **4**, 557-564.

66 Harrison, S.C., Smith, A.J., Jones, G.T., Swerdlow, D.I., Rampuri, R., Bown, M.J., Aneurysm, C., Folkersen, L., Baas, A.F., de Borst, G.J. *et al.* (2013) Interleukin-6 receptor pathways in abdominal aortic aneurysm. *European heart journal*, **34**, 3707-3716.

67 Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K. *et al.* (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature genetics*, **44**, 1336-1340.

68 Scheller, J., Ohnesorge, N. and Rose-John, S. (2006) Interleukin-6 trans-signalling in chronic inflammation and cancer. *Scandinavian journal of immunology*, **63**, 321-329.

69 Chalaris, A., Schmidt-Arras, D., Yamamoto, K. and Rose-John, S. (2012) Interleukin-6 trans-signaling and colonic cancer associated with inflammatory bowel disease. *Digestive diseases*, **30**, 492-499.

70 Enevold, C., Baslund, B., Linde, L., Josephsen, N.L., Tarp, U., Lindegaard, H., Jacobsen, S. and Nielsen, C.H. (2014) Interleukin-6-receptor polymorphisms rs12083537, rs2228145, and rs4329505 as predictors of response to tocilizumab in rheumatoid arthritis. *Pharmacogenetics and genomics*, **24**, 401-405.

71 Reiner, A.P., Beleza, S., Franceschini, N., Auer, P.L., Robinson, J.G., Kooperberg, C., Peters, U. and Tang, H. (2012) Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. *American journal of human genetics*, **91**, 502-512.

72 van de Veerdonk, F.L., Stoeckman, A.K., Wu, G., Boeckermann, A.N., Azam, T., Netea, M.G., Joosten, L.A., van der Meer, J.W., Hao, R., Kalabokis, V. *et al.* (2012) IL-38 binds to the IL-36 receptor and has biological effects on immune cells similar to IL-36 receptor antagonist. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 3001-3005.

73 Reiner, A.P., Wurfel, M.M., Lange, L.A., Carlson, C.S., Nord, A.S., Carty, C.L., Rieder, M.J., Desmarais, C., Jenny, N.S., Iribarren, C. *et al.* (2008) Polymorphisms of the IL1-receptor antagonist gene (IL1RN) are associated with multiple markers of systemic inflammation. *Arteriosclerosis, thrombosis, and vascular biology*, **28**, 1407-1412.

74 Hom, G., Graham, R.R., Modrek, B., Taylor, K.E., Ortmann, W., Garnier, S., Lee, A.T., Chung, S.A., Ferreira, R.C., Pant, P.V. *et al.* (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *The New England journal of medicine*, **358**, 900-909.

75 Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A. *et al.* (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, **491**, 119-124.

76 MacRae, C.A. and Vasan, R.S. (2011) Next-generation genome-wide association studies: time to focus on phenotype? *Circulation. Cardiovascular genetics*, **4**, 334-336.

77 Dumitrescu, L., Carty, C.L., Franceschini, N., Hindorff, L.A., Cole, S.A., Buzkova, P., Schumacher, F.R., Eaton, C.B., Goodloe, R.J., Duggan, D.J. *et al.* (2013) Post-genome-wide association study challenges for lipid traits: describing age as a modifier of gene-lipid associations in the Population Architecture using Genomics and Epidemiology (PAGE) study. *Annals of human genetics*, **77**, 416-425.

78 Kulminski, A.M., Culminskaya, I., Arbeev, K.G., Arbeeva, L., Ukraintseva, S.V., Stallard, E., Wu, D. and Yashin, A.I. (2015) Birth Cohort, Age, and Sex Strongly Modulate Effects of Lipid Risk Alleles Identified in Genome-Wide Association Studies. *PloS one*, **10**, e0136319.

**Table 1: Novel and replicated loci associated with circulating IL-6 levels at *p* < 5.0 × 10-8 in the combined GWAS meta analyses**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Chr | Lead SNP  (rsID) | BP  (Hg19) | Effect/Other allele | EAF | Beta(SE) | *P*discovery | *P*replication | *P*combined | Annotation | Nearest Genes |
| *Novel Loci* | | | | | | | | | | |
| 2q14 | rs6734238 | 113841030 | G/A | 0.42 | 0.025(0.005) | 1.45×10-7 | 3.24×10-5 | 1.84×10-11 | intergenic | *IL1F10/IL1RN* |
| 6p21 | rs660895 | 32577380 | G/A | 0.19 | 0.036(0.006) | 1.80×10-9 | 3.38×10-2 | 1.55×10-10 | intergenic | *HLA-DRB5/DRB1* |
| *Replicated Known Locus* | | | | | | | | | | |
| 1q21 | rs4537545 | 154418879 | T/C | 0.39 | 0.091(0.005) | 8.39×10-85 | 7.88×10-37 | 1.20×10-122 | intronic | *IL6R* |

Index SNPs that reached *p* < 5×10−8 in the combined analysis and are independent signals are reported here. Sample sizes: discovery cohorts, n=52,654; replication cohorts, n=14,774; combined, n=67,428. The effect sizes (β) in the discovery phase, given for the effect allele. EAF: Effect Allele Frequency; Effect sizes and standard error (SE) values are based on natural log transformed IL6 (pg /ml) levels.

**Figure Legends:**

**Figure 1.**

**Manhattan, QQ and LocusZoom plots of the discovery GWAS meta-analyses.**

a) Manhattan plot showing the association of SNPs with IL-6. Loci coloured in red or blue, three in total, represent those for which the lead SNPs reached genome-wide significance (P=5×10−8). Horizontal axis: relative genomic position of variants on the genome, vertical axis : -log10 p-value of each SNP; b) Quantile-quantile plot for p-values obtained from the meta-analysis. The horizontal and vertical axes represents the expected distribution of -log10 (P-values) under the null hypothesis of no association, whereas the vertical axis shows the observed -log10 (P-values). The blue dashed line represents the null, and λgc value represents the genomic inflation factor lambda. Each data point represents the observed versus the expected p-value of a variant included in the association analyses; c-e) Regional association plots for each of the three genome-wide significant loci, 1q21, 2q14, and 6p21, respectively. Pairwise LD (r2) with the lead SNP is indicated following a color-coded scale. Horizontal axis: relative genomic position of variants within the locus, vertical axis: -log10 p-value of each SNP.

**Figure 1:**

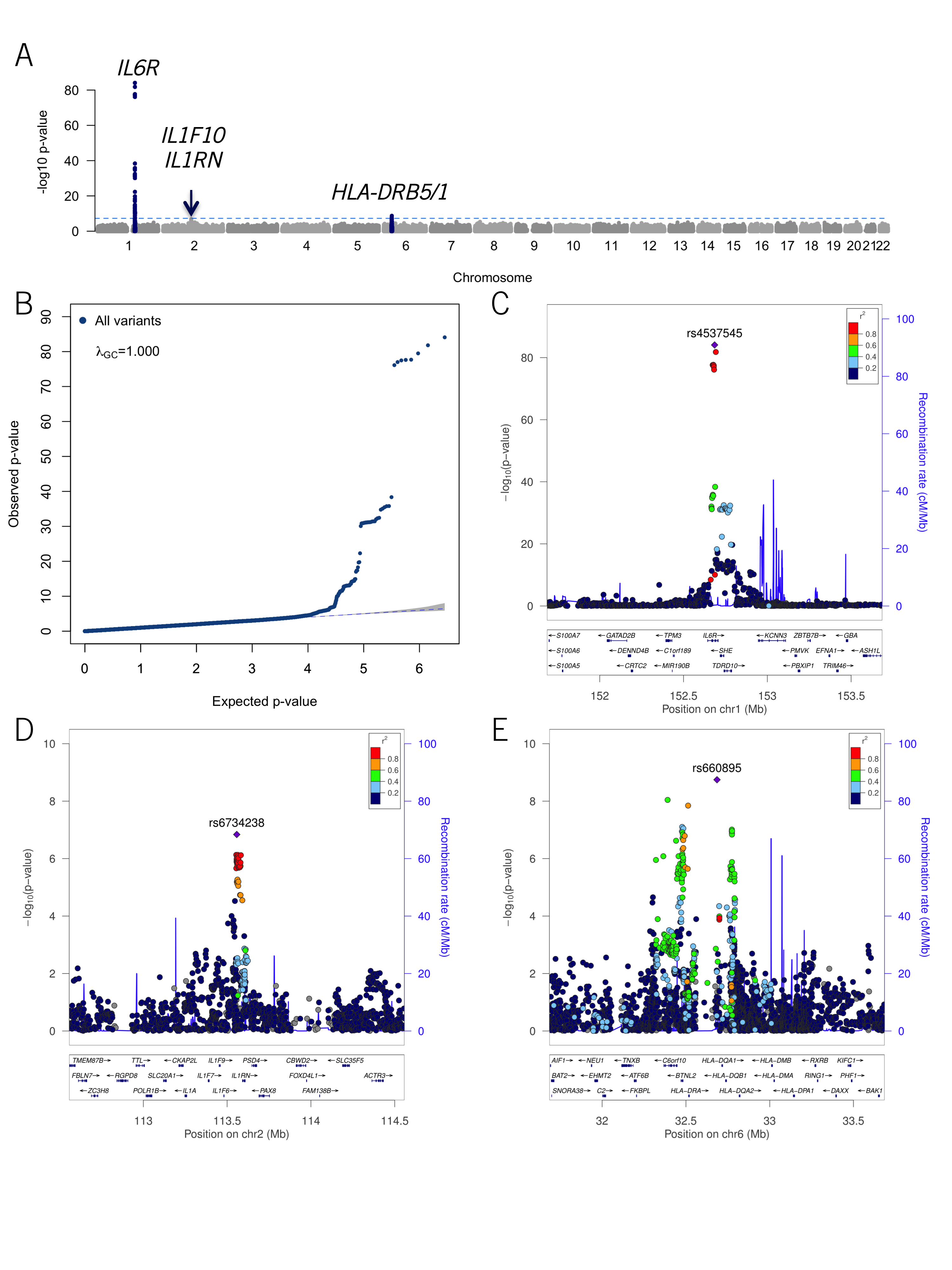


Figure 2. Combined discovery and replication forrest plots for the GWAS Index SNPs

Forrest plots for *IL6R* rs4537545 (chr. 1q21), b) IL1RN rs6734238 (chr. 2q14) , c) *HLA-DRB5* rs660895 (chr. 6p21) with discovery, replication and combined effect estimates, 95%CI and weights based on the fixed effects inverse variance meta-analyses

