The impact of rare and common genetic variation in the Interleukin-1 pathway for human cytokine responses

Rosanne C. van Deuren[1,2,3], Peer Arts[2,4], Giulio Cavalli[1,5,6], Martin Jaeger[1,3], Marloes Steehouwer[2], Maartje van de Vorst[2], Christian Gilissen[2,3], Leo A.B. Joosten[1,3,7], Charles A. Dinarello[1,6], Musa M. Mhlanga[8,9], Vinod Kumar[1,10], Mihai G. Netea[1,3,11], Frank L. van de Veerdonk[1,3], Alexander Hoischen[1,2,3]

1. Department of Internal Medicine, Radboud University Medical Center, Nijmegen, the Netherlands
2. Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands
3. Radboud Institute of Molecular Life Sciences (RIMLS), Radboud University Medical Center, Nijmegen, the Netherlands
4. Department of Genetics and Molecular Pathology, Centre for Cancer Biology, SA Pathology and the University of South Australia, Adelaide, South Australia, Australia
5. Unit of Immunology, Rheumatology, Allergy and Rare Diseases, IRCCS San Raffaele Hospital and Vita-Salute San Raffaele University, Milan, Italy
6. Department of Medicine, University of Colorado, Aurora, Colorado, USA
7. Department of Medical Genetics, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania
8. Division of Chemical Systems & Synthetic Biology, Institute for Infectious Disease & Molecular Medicine (IDM), Department of Integrative Biological & Medical Sciences, University of Cape Town, Cape Town, South Africa
9. Faculty of Health Sciences, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
10. University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands
11. Department for Genomics & Immunoregulation, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany

ABSTRACT

Interleukin(IL)-1 signaling is of major importance in human innate cytokine responses. Common variants in underlying genes have been linked to various inflammation-mediated diseases and stimulation induced cytokine responses, but the role of rare variants remains to be elucidated.

In this study, we characterize the role of rare and common genetic variation, as identified by molecular inversion probe-based sequencing, in 48 genes related to the IL-1 pathway. We examined the inter-individual variability in *in vitro* stimulation-specific human cytokine responses from 463 healthy individuals of the Human Functional Genomics Project and assessed the role of rare and common genetic variants, separately and combined, by means of the Sequence Kernel Association Test with various variant grouping strategies.

We identified strong *NCF4* and *CASP1* rare genetic variant associations with IL-6 cytokine production in response to PHA ($P=7.2E^{-05}$) and LPS ($P=3.0E^{-05}$) stimulation. In addition, common variants in *IL36A* and *IL38* were associated to both *C. albicans* induced IL-1β and IL-6 (*IL36A* P=0.0442 and 0.0037; *IL38* P=0.0092 and 0.0082), an effect that was magnified on the respective IL-30 subpathway level (IL-1β P=0.0017; IL-6 $P=1.8E^{-04}$). The inflammatory-phenotype level analysis confirmed the common variant signature for the immunological response to *C. albicans* by an association with the anti-inflammatory group (IL-1β $P=1.87E^{-03}$; IL-6 $P=5.75E^{-04}$), and we validated this finding for non-coding common variants. Lastly, we identified a rare variant signature for LPS-induced IL-6 production with the pro-inflammatory group ($P=2.42E^{-04}$), and we detected a new role for anti-inflammatory rare variants on *S. aureus* stimulated IL-6 cytokine ($P=6.71E^{-03}$).

In conclusion, we show that both common and rare genetic variation in genes of the IL-1 pathway, separately and combined, differentially influence *in vitro* cytokine responses to various stimuli in healthy individuals. This study therefore provides insight into potential mechanisms that are translatable into new hypothesis-driven characterization of common and rare genetic variant involvement in a wide variety of inflammatory and immunological mechanisms and diseases.

INTRODUCTION

The innate immune system is our first line of defense against invading pathogens such as viruses, bacteria, fungi and parasites. It consists of a variety of cell populations (*e.g.* myeloid cells, natural killer (NK) cells, innate lymphocytes) and humoral factors (*e.g.* complement), each fulfilling a targeted function. Collectively, these components shape the innate immune response. The induction of inflammation in response to *e.g.* infection is therefore a naturally occurring and necessary process. Notwithstanding, both stimulatory and inhibitory mechanisms are required to induce effective elimination of the pathogens on the one hand, and to limit any collateral damage to the tissues on the other, *i.e.* prevent auto-inflammation.

The Interleukin-1 (IL-1) family of cytokines and receptors play a major role in the induction and regulation of host defense and inflammation[**1**]. The IL-1 family comprises pro-inflammatory cytokines (*e.g.* IL-1α/β, IL-36α/β/γ), anti-inflammatory cytokines (*e.g.* IL-37, IL-38), activating receptors (*e.g.* IL1-R1, IL-36R), decoy receptors (*e.g.* IL-1R2, IL-18BP), and additional regulators, kinases and phosphatases that altogether are responsible for the IL-1 mediated response[**2**]. Next to core IL-1 family effectors, members of the inflammasome and autophagy are important contributors to regulating the IL-1-induced inflammation. For instance, activation of the inflammasome allows for cleavage and activation of caspase-1, with subsequent release of the pro-inflammatory cytokines IL-1β and IL-18 as a consequence, and autophagy is an important process involved in cell homeostasis, but is also able to directly inhibit the inflammatory response by removing inflammasome components and damaged mitochondria[**3**].

Defects in IL-1 pathway signaling and its specific members have been linked to various inflammation-mediated diseases[**1, 4**]. Pro-inflammatory members of the IL-1 family, like IL-1β and IL-18, play an important role in a variety of (auto) inflammatory/immune diseases. For instance, a loss of balance in processing and secretion of IL-1β is an important characteristic of chronic inflammatory conditions like gout, systemic-onset juvenile idiopathic arthritis (sJIA), adult-onset Still's disease (AoSD) and osteoarthritis[**4**]. Moreover, IL-18 plasma concentrations are moderately elevated in systemic lupus erythematodes (SLE) and in rheumatoid arthritis (RA), while being severely increased in patients with macrophage activation syndrome (MAS)[**4**]. Protein levels of anti-inflammatory members of the IL-1 family are characteristically increased in numerous inflammatory conditions. Specifically, plasma concentrations of IL-1Ra, the anti-inflammatory competitor of IL-1α and IL-1β encoded by the *IL1RN* gene, are systemically increased in various pathologies such as sepsis, Crohn's disease and ulcerative colitis, and locally increased in the joints from patients with RA and osteoarthritis[**5**]. The IL-1 type 2 receptor, which is a decoy receptor that binds IL-1 but does not induce an intracellular pro-inflammatory signal, has also been found increased in sepsis and even suggested as a candidate biomarker in patients with acute respiratory distress syndrome[**5**].

For a number of (auto-)inflammatory diseases, the clinical presentation clearly points towards deregulated activity of the IL-1 pathway. A targeted search, based on clinical presentation, for genetic defects in genes related to the IL-1 pathway, has therefore been proven most useful in for instance CAPS (cryopyrin associated periodic syndromes) and DIRA (deficiency of IL-1 receptor antagonist) with the identification of mutations in *NLRP3* and *IL1RN* respectively[**4, 6**]. On the other hand, even though AoSD, Behcet's and Schnitzler disease share clinical similarities with CAPS, no single causal genetic defect has been identified to date despite the fact that subsets of patients have presented with mutations in related genes indicating that the likely genetic basis is close[**6, 7**]. In addition to rare

variants, a number of Genome Wide Association Studies (GWAS) have identified common variants in the IL-1 pathway to be associated with inflammatory diseases and stimulation induced cytokine responses[8-11]. The remaining unexplained genetic basis could partially be explained by the complex interplay between members of the IL-1 pathway that orchestrate a specific immune response, combined with the fact that this is highly context specific as components can have pro- or anti-inflammatory characteristics depending on the cascade that is triggered. To illustrate a hypothetical example; the location of a genetic variant in a pro-inflammatory ligand is indispensable considering the different target sites for binding either it's activating or decoy receptor, and as such the clinical presentation will be obviously different, a phenomenon also termed allelic series. Moreover, a second mutation in either one of the receptors could aggravate or negate the first mutation, likewise resulting in a completely distinct phenotype, which can be viewed as genetic modifiers. Understanding the mechanisms behind specific clinical symptoms can aid in a targeted search for genetic defects of 'unsolved' inflammatory conditions.

Characteristic of an inflammatory disease is abnormal inflammation in terms of an aberrant immune response, and understanding how the immune response becomes apparent in general is therefore key in understanding the basis of an inflammatory disease. For this reason, in the past few decades various studies have focused on determining the genetic variation that contributes to the inter-individual variability in immune responses[12-15]. In summary, these studies show the separate and shared contribution of host and environmental factors to an immunological response after a specific stimulus. Despite these efforts, a considerable percentage of immune response variation between individuals remains unexplained. One important shortcoming is that most studies to date have focused on common genetic variants, while the impact of rare or private variants remains poorly or not at all understood. With recent advancements in sequencing technologies, the ability to study the role of rare and low-frequency variants has remarkably improved, and its value has been proven in several studies. For instance, increasing evidence is showing that the combined effect of common- and rare variants could partially explain the missing heritability problem in complex diseases[16-18]. The relatively small-to-moderate effects of common variants can be significantly modified by the presence or absence of (multiple) rare variants[19], and thus could partially explain the 'missing heritability' in the inter-individual variability in immune responses. Molecular Inversion Probe (MIP) based re-sequencing is a targeted sequencing technology that can identify common, low-frequency, rare and even private variants within a region of interest, and as such allows to cost-effectively study intermediate sized gene panels (*e.g.* 50 genes) in relatively large cohorts[20-23].

Considering the importance of the IL-1 pathway for inflammation and innate immune responses, we aimed to identify and characterize common and rare genetic variants in 48 genes of the IL-1 pathway, and their impact on the inter-individual variability in cytokine responses between healthy individuals. A complete overview of the study workflow can be found in **Figure 1**.

RESULTS

Study Cohort

Here, we studied healthy individuals from the Human Functional Genomics Project (FG500-cohort, http://www.humanfunctionalgenomics.org/site/), by making use of the publicly available demographic data and stimuli-specific *in vitro* cytokine measurements (https://hfgp.bbmri.nl/). The gender distribution over 463 included individuals for analysis shows a mild overrepresentation of females as compared to males (Male n=201, Female n=262), whereas the mean- and median age distribution for these groups separately is comparable (**Supplementary Figure S1A**).

Raw IL-1β and IL-6 cytokine production by Peripheral Blood Mononuclear Cells (PBMCs) in response to stimulation with either 100ng/mL Lipopolysaccharide (LPS), 10µg/mL Phytohemagglutinin (PHA), heat-killed *Candida albicans* 10^6CFU/mL (*C. albicans*) and $1x10^6$/mL *Staphylococcus aureus* (*S. aureus*), were likewise evenly distributed between females and males (**Supplementary Figure S1B**), and were log-transformed prior to analysis. Based on the above-mentioned distributions in combination with the fact that previous research has shown that both age and gender can influence cytokine responses[**12-15**], both variables were included as covariates in our analyses.

Sequencing

Molecular Inversion Probe (MIP)-based re-sequencing of all coding exons of the 48 genes in our IL-1 pathway MIP-panel generated sequencing data from 520 healthy individuals (for all MIPs see **Supplementary Table S1**). Overlapping the sequencing data with the available immunophenotyping data we managed to obtain complete datasets from 463 individuals for analysis. The average coverage depth for these 463 individuals over all MIPs was 830x (**Supplementary Figure S2**). Five genes in our panel (*SIGIRR*, *PYCARD*, *CYBA*, *RAC2* and *MAP1LC3A*) were unfavorably covered (<100x average coverage of the entire coding part of the gene) for more than half of the samples, and one gene (*NCF1*) lost all coverage in our extensive quality filtering due to homology regions (**Supplementary Figure S2**). We identified 201 non-synonymous variants in the coding exon regions, out of which 35 were common, 166 were rare (based on cohort allele frequencies (AFs) using a threshold of ≥ 5% for common variants). Furthermore, of the 166 rare variants, we identified 18 variants to be novel (never observed before in public databases). For a complete variant list see **Supplementary Table S2**.

Variant Analysis

The role of common and rare variants, separately and combined on stimuli-specific cytokine responses was assessed by performing a Rare Variant Burden Analysis (RVBA) using the Sequence Kernel Association Test (SKAT)[**18, 24, 25**]. The main issue with rare variant association testing lies in the low-frequency of rare variants, and thus appropriate grouping is required. We performed the SKAT using three different grouping strategies (**Figure 1D.c** and **Supplementary Table S3**): I) **gene-level**, where all variants within the coding region of each gene were considered; II) **subpathway-level**, where all coding variants within genes that belong to the corresponding subpathway were considered; and III) **inflammatory-phenotype level**, where genes were classified as either pro-inflammatory or anti-inflammatory based on encoded protein function, and all variants from genes in either group were considered. Each level was assessed for the role of genetic variants on stimuli-specific IL-1β and IL-6 cytokine production, through the full spectrum of allele frequencies (from rare- to common variants separately and combined) by performing four different tests (**Figure 1E**): i.SKAToC; testing only common

variants, iii.SKAToR; testing only rare variants, ii.SKATjoint; testing common and rare variants combined, where common-rare cut-off was based on cohort AF 5%. To verify potentially missed strong unidirectional signals, the rare variants were subjected to an additional test, the SKAT-O (iv.SKATO). Output from all SKATs performed in this study can be found in **Supplementary Table S4**.

Association landscapes show similarities and differences in IL-1ß and IL-6 response

We created holistic heatmap overviews termed 'association landscapes', for the purpose of summarizing rare- and common variant associations, separately and combined, both on the gene and subpathway level, in an organized fashion. **Figure 2** shows these landscapes of gene- and subpathway based associations for IL-1ß (**2A**) and IL-6 (**2B**) production by whole blood. **Figure 3** shows the inflammatory-phenotype based associations for IL-1ß (**3A**) and IL-6 (**3B**) production by whole blood in classic rectangular heatmaps.

The **gene-level** analysis revealed a significant association between rare genetic variants in *NCF4* and both IL-1ß and IL-6 cytokine production in response to *in vitro* PHA stimulation (SKAToR $_{adj}$P-value=0.0126 and 7.2E^{-05} respectively). In addition, we detected a strong association between rare *CASP1* variants and IL-6 cytokine production in response to LPS (SKAToR $_{adj}$P-value=3.0E^{-05}). When common variants in *NCF4* were considered as covariates in the rare variant analysis, the signal sustained for both cytokines with only a minor decrease in significance (SKATjoint $_{adj}$P-value=0.0185 and 1.3E^{-04} for IL-1ß and IL-6 respectively). Common variants in *IL36A* and *IL38* were significantly associated with both IL-1ß and IL-6 cytokine production after *C. albicans* stimulation (*IL36A* SKAToC $_{adj}$P-value=0.0442 and 0.0037; *IL38* SKAToC $_{adj}$P-value=0.0092 and 0.0082). In the combined analysis of common and rare variants these associations substantially weakened, with the exception of the *IL36A* association with IL-6 that remained comparable to the rare variant result (SKATjoint $_{adj}$P-value=0.0070).

The **subpathway-level** analysis identified a significant association between rare genetic variants in IL-1 subpathway genes combined and IL-1ß cytokine production in response to LPS stimulation (SKAToR $_{adj}$P-value=0.0030), that maintained in the combined analysis, where common variants were considered covariates to the rare variant effect, with only a minor decrease in significance (SKATjoint $_{adj}$P-value=0.0056). Specific to IL-6, we identified associations between rare variants in ROS-production genes after PHA stimulation (SKAToR $_{adj}$P-value=0.0295), and rare variants in Inflammasome genes combined after LPS stimulation (SKAToR $_{adj}$P-value=3.7E^{-04}), while only the latter remained in the joint analysis (SKATjoint $_{adj}$P-value=4.7E^{-04}). Finally, common variants in IL-30 subpathway genes were significantly associated with both IL-1ß and IL-6 cytokine production in response to *C. albicans* stimulation (SKAToC $_{adj}$P-value=0.0017 and 1.8E^{-04} respectively), both slightly weaker in the joint analysis.

The **inflammatory-phenotype level** analysis (**Figure 3**), revealed a strong association between rare variants in genes with pro-inflammatory effects and IL-6 cytokine production in response to LPS stimulation (SKAToR $_{adj}$P-value=2.42E^{-04} and SKATO $_{adj}$P-value=1.99E^{-03}), that was recurrent when common variants were considered as covariates (SKATjoint $_{adj}$P-value=3.60E^{-04}). Rare variants in anti-inflammatory genes on the other hand, were found to be significantly associated with IL-6 cytokine production in response to *S. aureus* stimulation (SKATO $_{adj}$P-value=6.71E^{-03}). We detected additional associations between common variants in anti-inflammatory genes and both IL-1ß and IL-6 cytokine production in response to *C. albicans* stimulation (SKAToC $_{adj}$P-value=1.87E^{-03} and 5.75E-^{04} respectively). For IL-1ß this signal persisted when common variants were considered as covariates over the rare variant effect (SKATjoint $_{adj}$P-value=7.11E^{-03}), whereas in the case of IL-6 most of the signal was lost

(SKATjoint $_{adj}$P-value=0.0240). Finally, unique to *C. albicans* stimulated IL-6 cytokine production, we detected an association with common variants in pro-inflammatory genes (SKAToC $_{adj}$P-value=0.0153).

Gene-level rare variant associations of cytokine-production outliers imply cytokine/stimulus-specific features

The associations discussed thus far were produced by using stimulus-specific cytokine profiles of all included individuals as continuous measurements. In order to uncover rare variant gene associations that may reflect strong stimulus-specific mechanisms, we supplemented these continuous trait associations (CT), with extreme binary trait associations (BT). The BT associations were conducted by categorizing individuals into two groups, based on 1% highest and lowest cytokine production respectively, in response to a certain *in vitro* stimulus, from here on referred to as top-producers ($_{TOP}$BT) and low-producers ($_{LOW}$BT). Identifying recurrent rare variant associations (that is, nominally significant in CT and either $_{TOP}$BT or $_{LOW}$BT), allowed us to give meaning in terms of direction of the effect of our continuous association results. **Figure 4** shows recurrent gene-level rare variant association (SKAToR) P-values, separately for IL-1ß (**A**) and IL-6 (**B**).

The analysis of extreme cytokine producers confirmed the continuous association between rare variants in *NCF4* and IL-1ß and IL-6 cytokine production after PHA stimulation, by the identification of a recurrent signal in the low-producers (IL-1ß $_{LOW}$BT P-value=2.18E^{-04}; IL-6 $_{LOW}$BT P-value=5.94E^{-05}), along with the association between *CASP1* and IL-6 low-producers after LPS stimulation ($_{LOW}$BT P-value=0.0299). In addition to that, we identified three recurrent associations in IL-1ß top-producers; rare variants in *IL1R2* after LPS- (P-value=0.0429) and PHA stimulation (P-value=0.0493), and *IL18BP* rare variants after *S. aureus* stimulation (P-value=0.0038), and four recurrent association in LPS-stimulated IL-1ß low-producers; rare variants in *NCF4* (P-value=8.27E^{-03}), *IL1R1* (P-value=0.0194), *IL1RL2* (P-value=8.54E^{-04}) and *IL36G* (P-value=3.03E^{-03}) (**Figure 4A**). In the case of IL-6 (**Figure 4B**) we detected one additional recurrent signal, that is *IL33* rare variants and top-producers after *C. albicans* stimulation (P-value=0.0143).

Immunological response to C. albicans reflects a common variant signatures

We detected strong common coding variant set signals over various levels of magnitude both on IL-1ß and IL-6 cytokine production after *C. albicans* stimulation, reflecting a common variant signature in this immunological response. For the separate common coding variants in *IL36A* (rs895497) and *IL38* (rs6761276 and rs6743376), we observed that the alternative allele for each variant separately presented with 1) a higher frequency as compared to the ancestral allele, and 2) a higher cytokine production (residual, after correcting for co-variates age and gender). **Figure 5A** shows the IL-1ß and IL-6 *in vitro* cytokine production in response to *C. albicans* stimulation, for each variant separately over the different genotype categories, with the homozygous reference (minor allele) categories on the right. Complementary Wilcoxon-rank-sum tests confirmed that for all three variants homozygous reference (dark-blue and dark-red for IL-1ß and IL-6 respectively) versus homozygous alternative (light-blue and light-red for IL-1ß and IL-6 respectively) and heterozygous (mid-blue and mid-red for IL-1ß and IL-6 respectively) versus homozygous alternative, was significantly associated with higher IL-1ß and IL-6 cytokine production, and homozygous reference versus heterozygous only for *IL38* rs6743376 (**Figure 5A**).

For the purpose of replicating these common variant signals, we repeated our significantly associated common variant sets in this study with previously published genotyping data from the same (FG500) cohort containing common genome-wide genetic variation (https://hfgp.bbmri.nl/).

Considering accumulating evidence for a role of non-coding common genetic variation in human health in general[**26, 27**], as well as in inflammatory responses in specific[**28, 29**]. All common variants (cohort AF>0.05) within a set were pruned for Linkage Disequilibrium (LD), using within cohort R^2 metrics. **Figure 5B** shows that by subjecting these expanded sets to the same SKAT, we identified associations between non-coding common variants and *C. albicans*-stimulated IL-6 production in *IL36A* (P-value=0.049), *IL38* (P-value=0.007) and pro-inflammatory phenotype (P-value=0.019), and *C. albicans* stimulated IL-1ß production in *IL38* (P-value=0.046). To visualize these associations, we calculated the allelic score for all significant sets in **5B**, by incorporating single SNP directions and MAF-based variant weights. In **Figure 5C** the weighted directional allelic score is displayed in correlation with cytokine production, demonstrating that for each set there is an increasing cytokine production with weighted directional allelic score, with R^2 (0.025) for *IL38*.

DISCUSSION

In this study we identify and characterize genetic variation in genes of the IL-1 pathway that can explain the inter-individual variability in cytokine responses in healthy individuals. In addition, we determine the relative contribution of rare variants as compared to common variants, as well as the combined effect of rare and common variants on *in vitro* cytokine responses measured in whole blood in response to various stimuli.

Over the past decades, various studies have identified a role for common genetic variation on cytokine level and response, however a significant proportion of inter-individual variability remains to be determined[12-15]. Considering the increasing evidence that specific combinations of variants with variable frequencies can influence phenotypic variability[16-18], in particular for a combination of phenotypic characteristics that do not fit one specific clinical diagnosis[30], we hypothesized that the inter-individual variability in cytokine responses might be subjectable to same concept. For this purpose, we sequenced the coding regions of 48 genes related to the IL-1 pathway in almost 500 healthy individuals, and assessed IL-1ß and IL-6 levels in whole blood in response to *in vitro* LPS, PHA, *C. albicans* and *S. aureus.* By means of the Sequence Kernel Association Test (SKAT), we tested for association between cytokine production and only common- (SKAToC), only rare- (SKAToR), and common and rare variants combined (SKATjoint), over various levels of grouping strategies; gene-, subpathway, and inflammatory-phenotype groups. A complete overview of the study workflow is shown in **Figure 1**.

The strongest rare variant association detected in this study was between rare genetic variants in *CASP1* and LPS induced IL-6 cytokine production. *CASP1* encodes the similarly named CASP-1 protein, which is responsible for cleavage of the inactive mediators IL-1ß, IL-18 and IL-33 into their active form. Inhibition of CASP-1 has been shown to block active IL-1ß levels, thereby making it a potent therapeutic target. The association in this study was based on five rare variants; two private variants - one of which is new to public databases, two variants identified in two individuals in heterozygous state, and one variant that was identified homozygous in five individuals (see **Supplementary Table S2**). The two private variants are located in a shorter transcript form of *CASP1* that is apoptotically inactive, thereby potentially affecting the inflammatory balance. Previously, mutations in *CASP1* have been linked mostly to IL-1ß-mediated disorders like CAPS[5]. The fact that we observed a lower burden of rare variants not with IL-1ß but with IL-6 production, may suggest an unknown effect of CASP-1 on IL-6 production and/or release, although this is a speculation and remains to be demonstrated. Alternatively, since we used 24h cytokine measurements, it could also reflect an unnoticed effect of these variants on preceding IL-1ß production, that subsequently influence the induction of IL-6.

We detected an almost equally strong decreased rare variant burden between *NCF4* and both IL-1ß and IL-6 production after PHA stimulation. This association was based on three private rare variants, one of which is new to public databases. Two variants are located in the canonical splice-acceptor sites, thereby possibly affecting the splicing of this gene, which most likely leads to loss-of-function variants with reduced RNA and possibly protein levels as a consequence. The *NCF4* gene encodes the NCF4 protein which is part of the cytoplasmic unit of the NADPH (nicotinamide adenine dinucleotide phosphate) oxidase enzyme system that plays an important role in phagocytosis[31]. Mutations in NADPH related genes have been linked to chronic granulomatous disease, a primary immunodeficiency, and common variants have been associated with immune related diseases like Crohn's disease and systemic lupus erythematodes (SLE)[32, 33]. Knock-out of *Ncf4* in mice causes a

defective NADPH-complex, with reduced ROS-production and subsequent increased cytokine production as a consequence[**34, 35**]. These findings are inconsistent with the association detected in this study between NCF4 rare variants and decreased cytokine production. However, considering the splice-affecting characteristics of two of these rare variants, the resulting effect on it is debatable what the result for the NADPH oxidase system remains inconclusive and requires further investigation.

Finally, the association between IL-6 production after *S. aureus* stimulation and anti-inflammatory phenotype rare variants detected by the SKATO, had a unidirectional origin (as presented by rho-value of 1). Upon more in-depth inspection, we observed that individuals with rare variants presented with a higher IL-6 cytokine, as confirmed by Wilcoxon rank-sum P-value=0.0054 (**Figure 3C**). The fact that more than half of the genes in the anti-inflammatory group are autophagy genes, is in line with the finding that defective autophagy has been associated with increased cytokine production in general, and IL-1ß in particular, as well as a more severe colitis and CGD[**34**].

An intrinsic issue with rare variants is their low-frequency and thus the main issue in association testing is limited power[**36**]. Combining rare variants within regions and subsequently using an appropriate test for association can improve power, yet requires prior knowledge on the biological effects of the genes studied for proper grouping. Our IL-1 subpathway rare variant association with LPS-stimulated IL-1ß production confirms that properly grouping rare variants and testing them combined produces a significant association (SKAToR $_{adj}$P-value=0.003), even though the individual underlying genes did not (**Figure 2A**). The most likely reason for this is insufficient power, and indeed a thorough inspection of the raw data showed us that the individual underlying gene-level association P-values were significant, but did not survive the Bonferroni-adjustment. This substantiates a role for rare variants in cytokine responses, and the necessity for a well-defined hypothesis-driven target region for combining them when testing for association with a certain phenotype. Moreover, the comparable IL-1 subpathway SKATjoint P-value indicates that common variants are not influencing the rare variant effects in LPS stimulated IL-1ß production.

The subpathway level association between common variants in genes of the IL-30 subpathway and IL-6 production in response to *C. albicans* illustrates that even for common variants appropriate grouping can increase power. For instance, the IL-30 subpathway association with *C. albicans* stimulated IL-6 cytokine we detected, was more significant than could be explained by the gene-level associations with *IL36A* and *IL38*. As other members of the respective subpathway like *SIGIRR* and *IL37* also contain common variants, the most likely explanation for this is likewise insufficient power in *SIGIRR* and *IL37* separately to detect an association, yet their common variants do contribute to the subpathway signal.

Another method to increase power in rare variant association studies, is to use extreme phenotypes, for the purpose of enriching rare variants with strong(er) effects in these outlier individuals[**37, 38**]. In this study, we used 1% extreme cytokine producers in a binary association and overlapped the results with our continuous tests, for the purpose of 1) characterization of stimuli-specific mechanisms, and 2) providing an indication of direction. We therefore were able to give direction to the identified continuous rare variant associations, as we identified recurrence in low-responders for our strongest rare variant associations (*CASP1* with LPS induced IL-6 cytokine and *NCF4* with PHA induced IL-1ß and IL-6 cytokine). In addition, we identified recurrent associations between LPS stimulated IL-1ß cytokine production and five genes of the IL-1 subpathway substantiating the subpathway level association and previously discussed lack of power to identify individual gene associations. Interestingly, we identified an association between *S. aureus* stimulated IL-1ß cytokine

production and rare variants in *IL18BP*, recurrent in the top-responders. The encoded protein (IL-18BP) inhibits IL-18 signaling[39], and previous studies have shown that IL-18BP levels are negatively correlated with cytokine production by lymphocytes[12]. The identified burden in top-responders here, suggests that these rare variants in *IL18BP* negatively affect IL-18BP levels, allowing for high *S. aureus* induced cytokine production. The role of IL-18/IL-18BP pathway in cytokine response to staphylococci needs to be further investigated, as it has the potential to provide important new insights in severe diseases caused by this pathogenic microorganism (*e.g.* sepsis).

Notwithstanding our attempts to compensate for known issues in rare variant association testing, this study has several important limitations. First of all, the relatively small sample size and restricted cohort characteristics in terms of age distribution and residency limited our power for association testing. However, we would like to emphasize the labor-intensity of the extensive immunophenotyping experiments, and that this is one of the largest cohorts with this magnitude of dynamic cytokine responses to date. Secondly, MIP -sequencing is a cost-efficient strategy to identify genetic variation from common to private in a region of interest (*e.g.* ± €25,- per sample for the IL-1 panel), and has been proven highly valuable in various diagnostic settings[21, 22]. However, they do not yet allow to include the even larger intronic or non-coding part of genes. Future investigations using whole genome sequencing data to investigate a role for non-coding genetic variation in addition to coding, is therefore vital. Thirdly, the SKAT is powerful, but computes only set-wise association P-values and does not provide direction in terms of positive/negative effect size or increased/decreased risk. In addition, when testing multiple variants combined in a set, the single variant effects (individually or possibly in interaction with each other) are unknown. Consequently, questions on the effect of a single variant are valid from the clinical perspective, but remain difficult to estimate. We have attempted to account for these limitations by performing additional extreme phenotype binary associations and by using the direction extracted from single-variant tests in the allelic score for the combined test, but we cannot exclude that interaction effects between the variants in the entire set also may influence the phenotype. Finally, it was a well-considered decision to first test for association in a subset of available immunophenotyping data that are obviously distinctive in terms of clinical effects (LPS, PHA, *C. albicans* and *S. aureus*), and cytokine production by whole blood as it is most comparable to a real-life situation. However, for future studies with larger sample sizes, it would be interesting to also investigate whether more similar stimuli present with specific characteristics, and whether specific cell-subtypes also show differential associations, as these data are likewise publicly available for this cohort (https://hfgp.bbmri.nl/).

In conclusion, we show that common and rare genetic variation in genes of the IL-1 pathway, separately and combined, differentially influence the IL-1ß and IL-6 cytokine response after various stimuli *in vitro*. Not only do we add to the knowledge on the role of common variants in the IL-1 pathway, we additionally highlight the important role of rare variants (alone or in combination) on immune variability. By identifying specific genetic variants responsible for the orchestration of a particular inflammatory response, we contribute to the targeted development of appropriate functional hypotheses for a wide variety of inflammatory and immunological mechanisms and diseases. On a broader perspective, this study provides insight into potential mechanisms that are translatable to new hypothesis-driven identification of rare variant involvement in inflammatory and immunological phenotypes. Considering the fact that the framework used in this study is expandable to a wide variety of (non-immune) complex phenotypes, it can lead the way to new insights and theories for any

phenotype of interest, and therefore has the potential to contribute to better understanding of unresolved, complex diseases.

MATERIAL AND METHODS

Study Cohort

Cohort characteristics

The study was conducted using healthy individuals from the Human Functional Genomics Project (500FG-cohort, http://www.humanfunctionalgenomics.org/site/). The entire 500FG-cohort consists of 534 healthy individuals from the Netherlands (296 females and 237 males) with an age range 18-75, from which we were able to obtain DNA from 520 individuals for sequencing.

Immunophenotyping

In this study we made use of the publicly available extensive immunophenotyping data that was generated as part of the Human Functional Genomics Project (https://hfgp.bbmri.nl/). Specifically, Interleukin-1ß (IL-1ß) and Interleukin-6 (IL-6) production by whole blood stimulated with either lipopolysaccharide (LPS, 100ng/mL), phytohaemagglutinin (PHA, 10µg/mL), heat-killed *Candida albicans* (*C. albicans* 10^6 CFU/mL) or *Staphylococcus aureus* (*S. aureus* 1 x 10^6/mL). A detailed description of these experiments can be found elsewhere[**13**]. In short, blood was drawn from participants and 100 µL of heparin blood was stimulated with 400 µL of stimulus and subsequently incubated for 48hours at 37°C and 5% CO_2. Supernatants were collected and stored in -20°C until used for ELISA.

Sequencing

MIP-panel design

We sequenced all coding exons of 48 genes of the IL-1 pathway in 520 healthy individuals by Molecular Inversion Probe (MIP)-based re-sequencing. MIPs are a targeted re-sequencing technology, that allows for the identification of both common- and rare genetic variation in regions of interest. A detailed description of MIP-probe design and sequencing methods can be found elsewhere[**20, 23, 40**]. In short, 1285 MIP-probes were designed to cover all coding exons of 48 IL-1 pathway related genes and sequencing was performed using the Illumina NextSeq500 system. These 48 IL-1 pathway related genes can be further functionally subclassified into 6 subpathways that represent sub-mechanisms of the IL-1-mediated inflammatory response; IL-1 subpathway, IL-18 subpathway, IL-30 subpathway, Inflammasome, (reactive oxygen species) ROS-production, and Autophagy. An additional functional subclassification into two inflammation phenotype groups, was based on the gene-encoded protein function and its pro- or anti-inflammatory effect. A full explanation on the subclassifications can be found in **Supplementary Table S3**.

Data processing

A carefully developed filtering pipeline, based on Sanger Sequencing validations, was applied to ensure high sensitivity and specificity in our final variant set. First, the reads were aligned using BWA-MEM[**41**] and subsequently filtered on Mapping Quality ≥ 60, no soft-clipping, properly paired and not more than five variations from the reference per read, with the exception of multi-basepair insertions and deletions. Variants were then called using the GATK unified genotyper[**42**], which uses a Bayesian genotype likelihood model to estimate the most likely genotypes. Rare variants (here defined as absent in dbSnp build 150 common[**43**], or defines as rare by our custom annotator as explained below), were

further filtered on the QUAL parameter ≥ 1000 in the vcf. Additionally, the percentage of alternative alleles for each variant position using samtools mpileup[**44**], with maximum read depth 10000, no BAQ and a minimal base quality of 30. Homozygous rare variants required an alternative allele percentage of ≥ 90%, heterozygous rare variants required an alternative allele percentage of 25% and < 90%, and rare variants with an alternative allele percentage of < 25% were considered false positive. Samples with an average coverage depth of all MIPs ≥ 100x were included for analysis. The final variant set was annotated using our custom annotator, which makes use of several annotation sources, among others the Variant Effect Predictor (VEP) from Ensembl[**45**], Combined Annotation Dependent Depletion (CADD) score[**46**], and several population based variant databases (*e.g.* dbSnp, ExAc and gnomAD[**47**]) and an "inHouse" database consisting of > 25,000 clinical exomes run at the diagnostic division of the Department of Human Genetics of the Radboudumc. Based on this annotation information, variants were classified as common if either dbSNP-, ExAc-, gnomAD or inHouse allele frequency (AF) was ≥ 1%.

Variant Analysis

Continuous Trait

A Rare Variant Burden Analysis (RVBA) was performed on the log-transformed cytokine levels by using the Sequence Kernel Association Test (SKAT)[**18, 24**] in R version 3.5.2. The SKAT is a kernel-based test method, that aggregates weighted individual variant-score test statistics whilst allowing variant-variant interactions and is extremely powerful when a genetic region has both protective and deleterious variants or many non-causal variants[**18, 24, 25**]. The SKAT was performed using three different grouping strategies: I) gene region; where all variants within the gene region were considered, II) subpathway region; where all variants within genes that belong to the corresponding subpathway were considered, and III) inflammatory-phenotype region; where based on encoded protein function genes were classified to have either a pro-inflammatory or anti-inflammatory effect and all variants from genes in either groups were considered (**Figure 1D.c**). Furthermore, for each region the 'normal' SKAT was run to determine the effect of only common (i.SKAToC), only rare (iii.SKAToR) and common and rare variants combined (ii.SKATjoint), based on a cohort MAF of 0.05 as cut-off, using the SKAT_CommonRare function with default weights. However, since the SKAT can be less powerful than burden tests when rare variants in set are truly causal or influencing the phenotype in the same direction[**25**], we applied the SKAT-O function as an additional test on the rare variants (SKATO), appropriately weighing the variants using a logistic function based on the cohort AF, and extracted accompanying rho-values do asses the contribution of SKAT versus Burden Test for significant sets. In the case of rare- and joint tests only output based on > 1 variant was considered, and in the case of joint tests the presence of both rare and common variants in the set was an additional requirement. P-values were Bonferroni-adjusted based on the number of regions tested within each grouping separately.

Binary Trait

In addition to the associations of continuous cytokine levels as described above, we performed a binary association analysis on outlier individuals, here defined as extreme cytokine producers. As research has shown that individuals with outlier expression patterns are likely to be enriched in nearby rare variants[**48, 49**], we hypothesized that outlier individuals in terms of cytokine production could similarly be enriched in rare variants in specific genes, thereby favoring the identification of stimulus-specific mechanisms. For this purpose, we defined for each cytokine-stimulus the 1% extreme cytokine

producers (rounded up, so generally ±5 individuals), which categorized the individuals in two groups that were subjected to binary trait association. Specifically, 1% highest cytokine producers versus all other individuals ($_{TOP}$BT) and 1% lowest cytokine producers versus all other individuals ($_{LOW}$BT). In two cases, *C. albicans* stimulated IL-1ß production lowest cytokine producers and LPS stimulated IL-6 production highest cytokine producers, no distinctive categories could be created due to equal cytokine measurements at the 1% cut-off, and as such the groups were extended to 7 and 9 respectively. We followed up on associations based on >1 variant, that were nominally significant (unadjusted P-values) in the continuous associations and recurrent in either $_{TOP}$BT or $_{LOW}$BT.

Follow Up

In order to give meaning to our detected associations, we extracted the residual (corrected for covariates age and gender) cytokine production from the null-model and correlated those to the genotype categories, where applicable. For all plots, correlations and cytokine levels mentioned from hereon, the cytokine production therefore has been corrected for age and gender. In the case of set-based common variant associations we correlated cytokine production to the three separate genotype categories; homozygous reference, heterozygous, homozygous alternative, whereas in the case of set-based rare variant associations we correlated cytokine production to only two categories; absence or presence of one of the rare variants in the set.

Additionally, considering accumulating evidence for a role of non-coding genetic variation in health and disease[**26, 27**], we followed-up on common variant associations by using the publicly available genotype data from the FG500-cohort, generated with the commercially available SNP chip Illumina HumanOmniExpressExome-8v1.0 (for further details we refer to previously published work[**13, 50**]). We extracted all common variants (based on cohort AF > 0.05) within NCBI RefSeq 'Whole Gene' gene regions and extended the start position by 50kB[**51**] for the following sets: *IL36A*, *IL38*, IL-30 subpathway, pro-inflammatory phenotype and anti-inflammatory phenotype. Variant sets were pruned for Linkage Disequilibrium (LD) based within cohort metrics and the commonly used R^2 cut-off of >0.8, by means of the snpStats package in R. The final pruned set of variants, termed tagSnps, were subjected to the same SKAT with default weights to test for association with continuous IL-1ß (N=428) and IL-6 (N=425) cytokine production. Finally, for the purpose of correlating significant non-coding common variant sets to cytokine levels, we calculated an allelic score for all variants in the set. An allelic score is a way to collapse multidimensional genetic data associated with a risk factor into a single variable[**52**]. We slightly adapted the allelic score calculation to our SKAT-based test results, into a weighted (based on AF-based Beta.Weights function from SKAT package), directional (based on increasing or decreasing cytokine production over the genotype categories) allelic score. Specifically, we inferred the direction of each variant in a set, and combined this with the manually computed variant weight, by negating the weight only for variants with decreasing cytokine production over the genotype categories. Common variant genotypes were converted to dosages and multiplied by their directional weight, which was summed up to an allelic score per set of variants. The final weighted, directional allelic score was plotted versus the cytokine production, and a linear model equation and R^2 metrics were extracted.

ACKNOWLEDGEMENTS

The authors thank the Radboud Genomics Technology Center (Radboud University Medical Center Nijmegen) for support in sequencing. CAD was supported by NIH AI-15614. MMM was supported by the South African Medical Research Council, Department of Science & Technology of South Africa, Chan Zuckerberg Initiative and Bill & Melinda Gates Foundation. VK was supported by Hypatia tenure track fellowship. MGN was supported by the Nederlandse Organisatie voor Wetenschappelijk onderzoek (Spinoza grant), and the European Research Council (grant agreement No 310372). FLV and AH were supported by the Interleukin Foundation. AH was supported by the SOLVE-RD project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 779257.

DATA AVAILABILITY

All demographic-, immunophenotyping- and genotyping- data from the FG500-cohort used in this study is publicly available on the BBMRI-NL archive (https://hfgp.bbmri.nl/). All variants called in MIP-sequencing data based on the IL-1 panel from the same cohort are published here, as well as all association results.

CODE AVAILABILITY

Code for processing and filtering MIP-based sequencing data are extensively explained in the methods section of this manuscript and will be made available upon reasonable request. The source code from the R packages used in this study are freely available online. Code for processing the output and generating the figures will likewise be made available upon reasonable request.

URLS

Human Functional Genomics Project home site: http://www.humanfunctionalgenomics.org/site/
Human Functional Genomics Project publicly available data: https://hfgp.bbmri.nl/

FIGURES

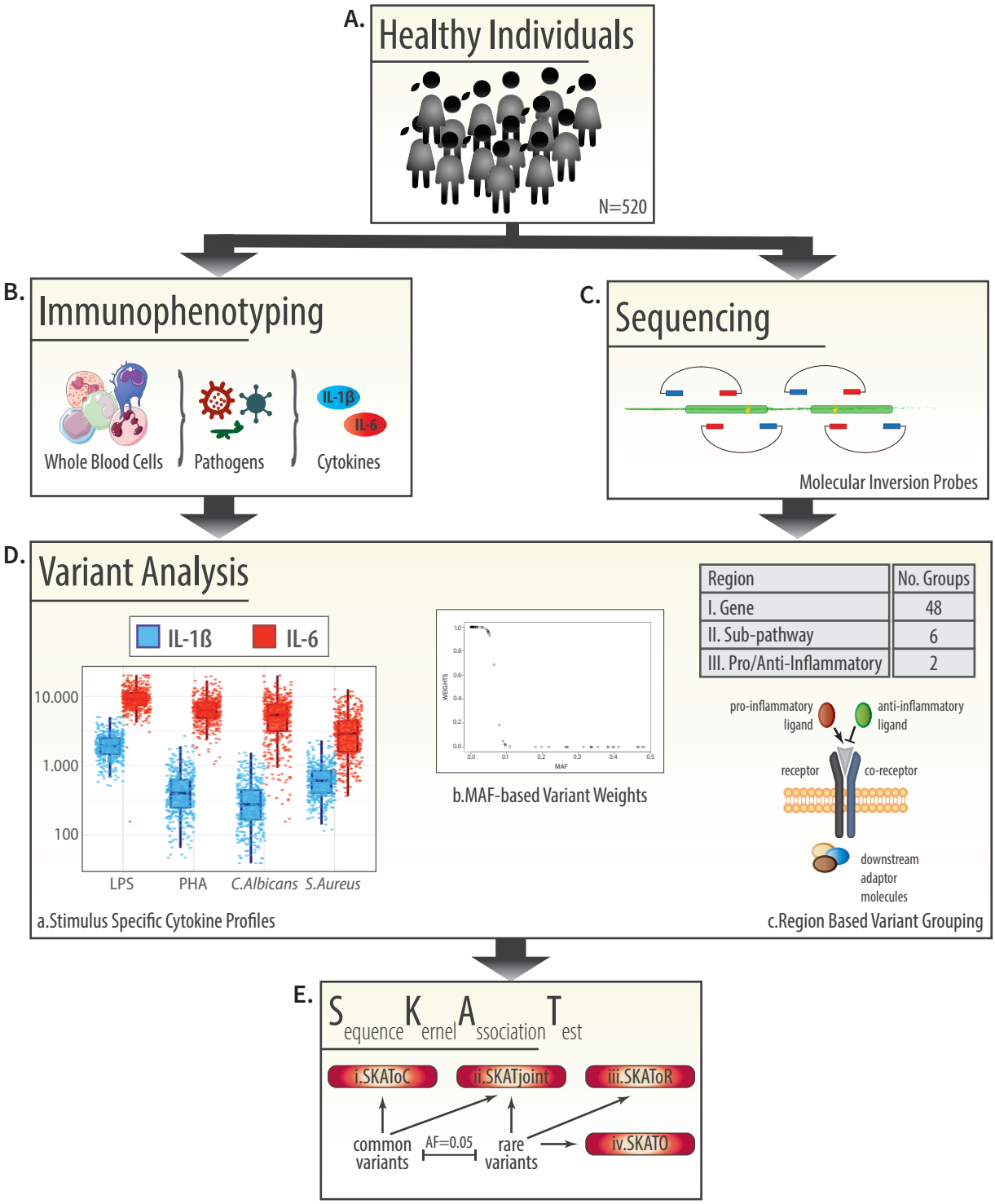


Figure 1. Flowchart of the study workflow method. Figure orientation from top to bottom. Blood was extracted from 520 healthy individuals (A) on which extensive immunophenotyping was performed (B) and simultaneously Moleclar Inversion Probe sequencing data was produced from coding regions of 48 Interleukin-1 pathway genes (C). The resulting cytokine production after stimulation was measured and used for analysis (D-a.Stimulus Specific Cytokine Profiles). Similarly, the identified variants were classified based on their Minor Allele Frequency (MAF), which determined their weighing factor (D-b.MAF-based variant weights). Variants were grouped over three different regions: I.Gene-level; where all variants within the gene-region were considered, II.Subpathway-level; where all variants within the genes that belong to the corresponding subpathway were considered, and III.Pro/Anti-inflammatory level; where based on known gene function genes were classified as either pro- or anti-inflammatory and all variants within pro-inflammatory group and anti-inflammatory group were considered separately (D-c.Region Based Variant Grouping). Finally, each of these sets of variants were subjected to four different versions of the Sequence Kernel Association Test (SKAT), based on cohort allele frequency classification: SKAToC; SKAT with only common variants (E-i.SKAToC), SKATjoint; SKAT with common and rare variants (E-ii.SKATjoint), SKAToR; SKAT with only rare variants (E-iii.SKAToR), and SKATO; best combination of the SKAT and Burden Test with only rare variants (E-iv.SKATO).

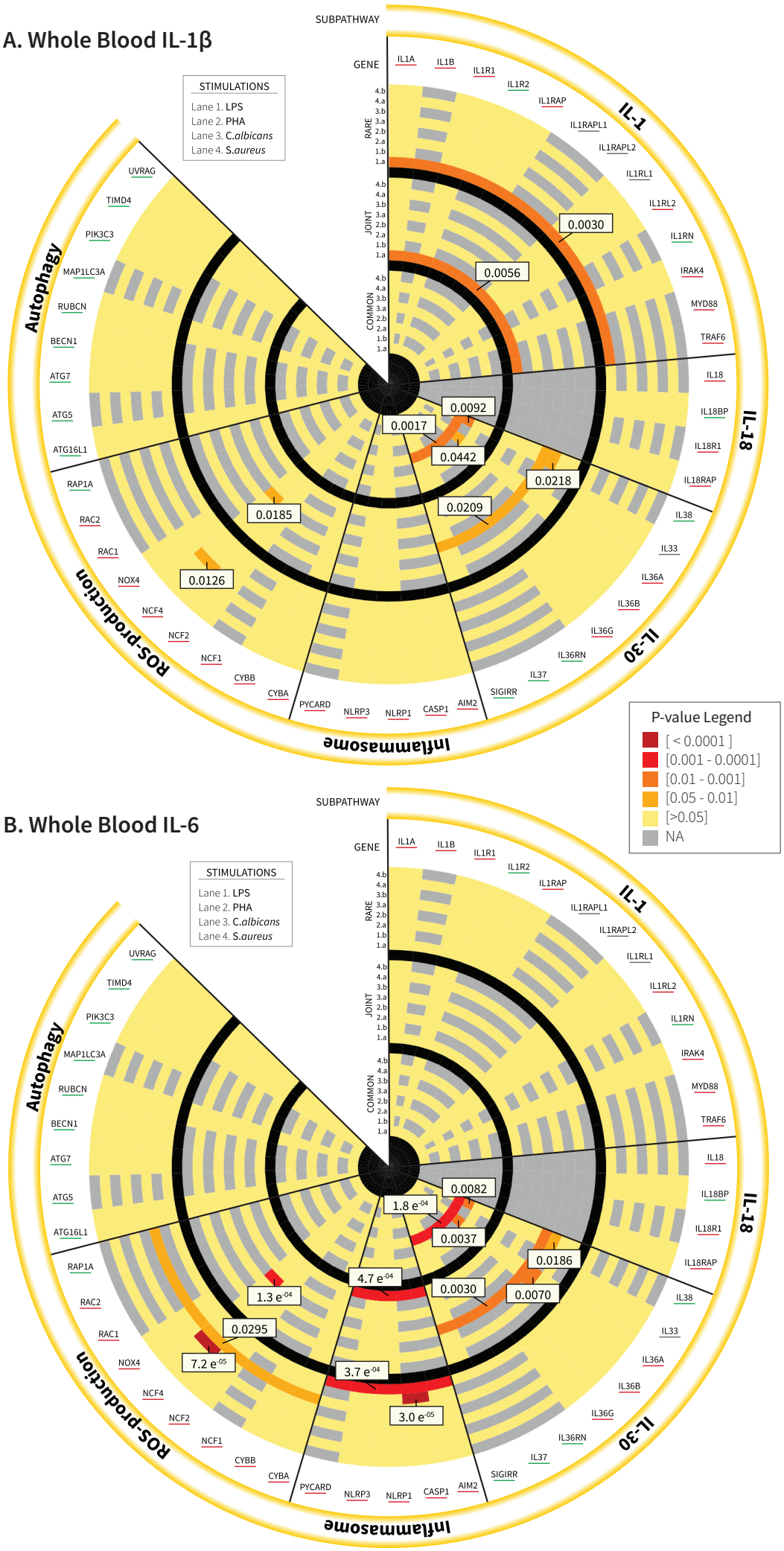


Figure 2. Circular heatmaps of SKAT association $_{adj}$P-values. The circular heatmaps consist of three righs separated by black lanes, representing from inner to outer the SKAToC (SKAT with only common variants), SKATjoint (SKAT with common and rare variants), and SKAToR (SKAT with only rare variants) association $_{adj}$P-values with log-transformed IL-1β and IL-6 cytokine production respectively. Each ring consists of 8 lanes, that represent different stimuli; 1) LPS 100ng/mL, 2) PHA 10 µg/mL, 3) *C.albicans* 10^6CFU/mL, 4) *S.aureus* $1x10^6$/mL, with sublane a) showing the subpathway result and sublane b) showing the gene level result. The genes annotated at the surface of the heatmap are underlined with a red, green or grey colour, representing it's classification in the inflammatory-phenotype groups (**Figure 3**).

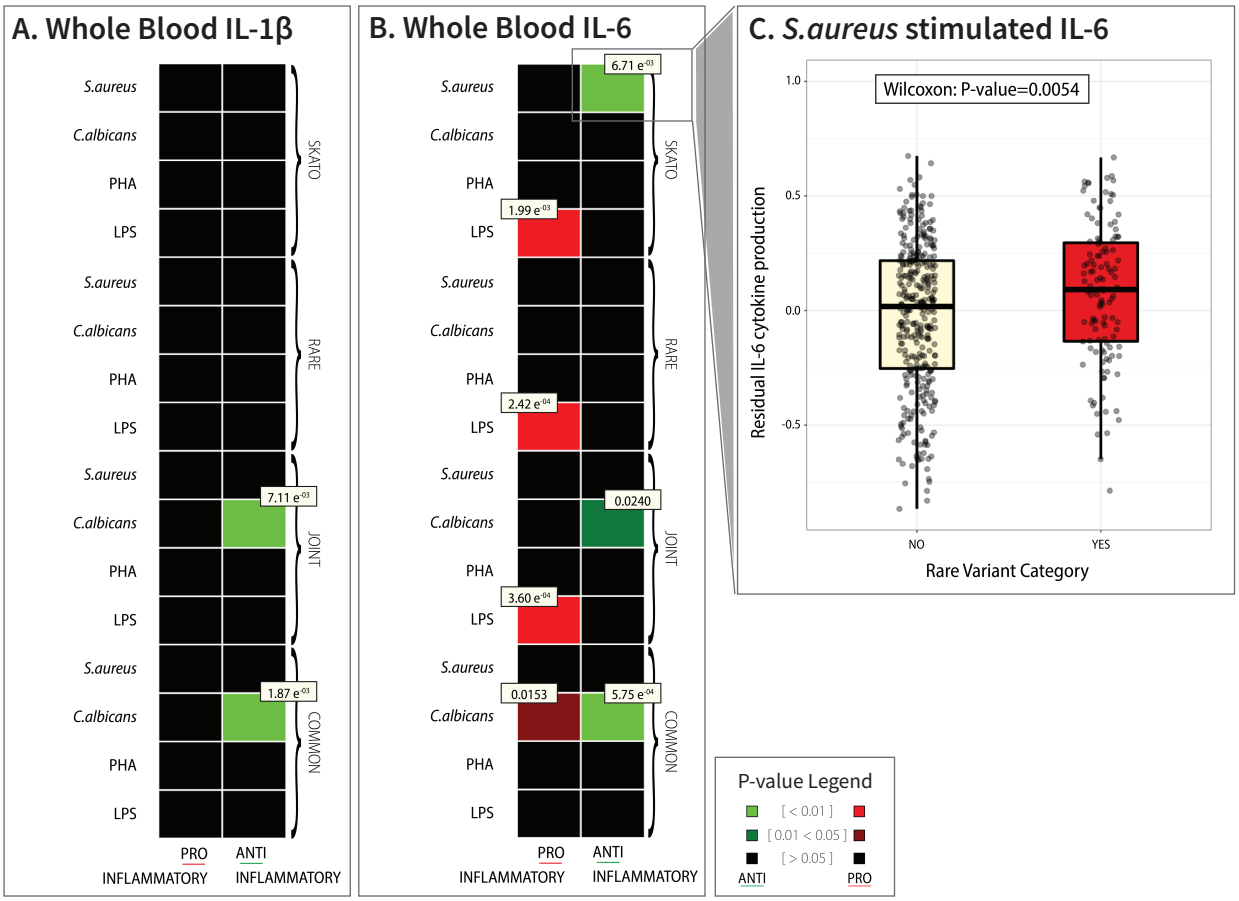


Figure 3. Inflammatory-phenotype level SKAT association adjP-values. A heatmap representation of association between inflammatory-phenotype groups and IL-1β (A) and IL-6 (B) cytokine production in response to four different stimuli; LPS 100ng/mL, PHA 10µg/mL, heat-killed *C. albicans* 10^6 CFU/mL, and *S. aureus* $1x10^6$/mL. Variants in genes categorized as pro- or anti-inflammatory were subjected to four different association tests; SKAToC (common variants only), SKATjoint (common and rare variants), SKAToR (rare variants only), and SKATO (combination of SKAT and Burden Test with rare variants only). (C) shows a boxplot of *S. aureus* stimulated residual IL-6 cytokine production (corrected for age and gender) over rare variant categories (NO=individuals without rare variant; YES=individuals carrying a rare variant) within the anti-inflammatory phenotype group. Wilcoxon rank-sum test P-value reveals a significant difference between the two categories.

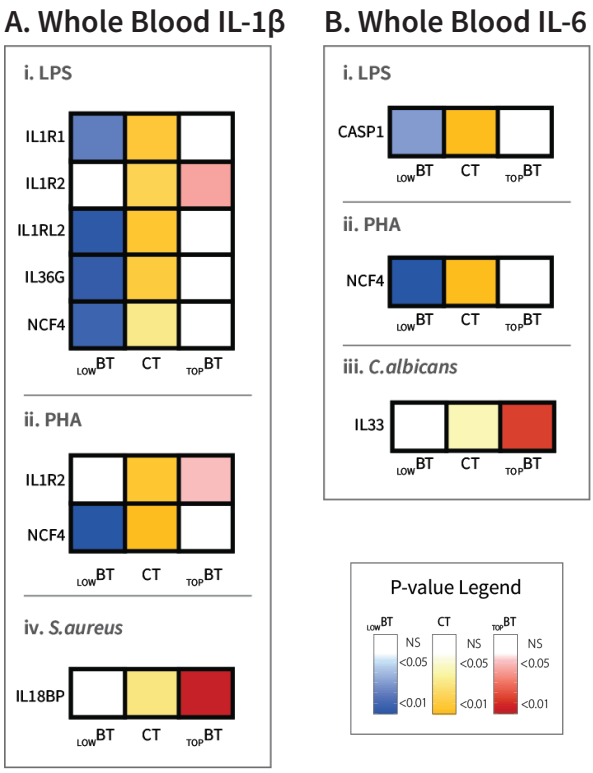
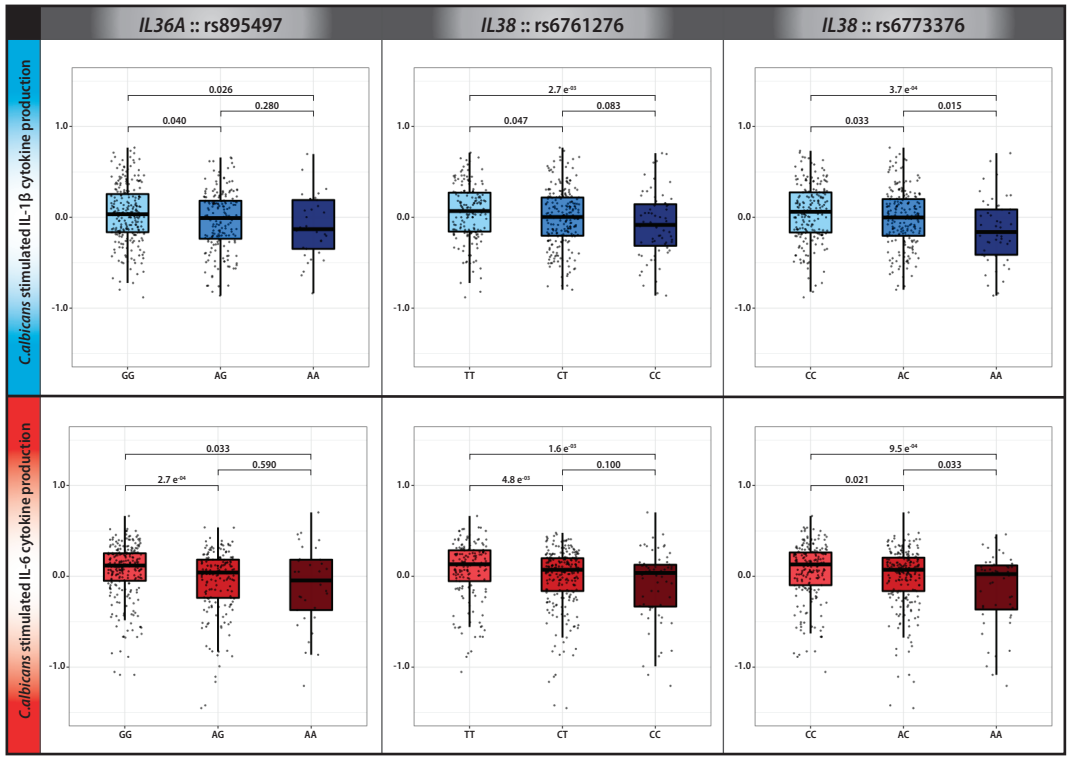


Figure 4. Recurrent gene-level rare variant SKAT P-values in extreme cytokine producers reflect direction of continuous associations and identify distinct stimulus/cytokine-specific signatures. A heatmap representation of all recurrent gene-level rare variant SKAT P-values with IL-1β (A) and IL-6 (B) cytokine production in response to LPS 100ng/mL (i.), PHA 10µg/mL (ii.), *C.albicans* 10^6CFU/mL, and *S.aureus* $1x10^6$/mL stimulation (iv.).

Annotation: $_{LOW}$BT = SKATBinary with lowest 1% producers; CT = SKAT with log-transformed continuous cytokine producers; $_{TOP}$BT = SKATBinary with highest 1% cytokine producers.

A. Coding *IL36A* and *IL38* common variant genotype categories with IL-1β and IL-6



B. Non-coding and coding common variant set SKAT P-values

Set	all SNPs (n)	tag SNPs (n)	IL-1β SKAT P-value	IL-6 SKAT P-value
IL36A	103	17	0.248	0.049*
IL38	207	41	0.046*	0.007**
IL-30 subpathway	886	283	0.337	0.427
Pro-inflammatory phenotype	4541	1436	*NA*	0.019*
Anti-inflammatory phenotype	2996	994	0.222	0.280

C. Allelic score of significant non-coding and coding common variant sets

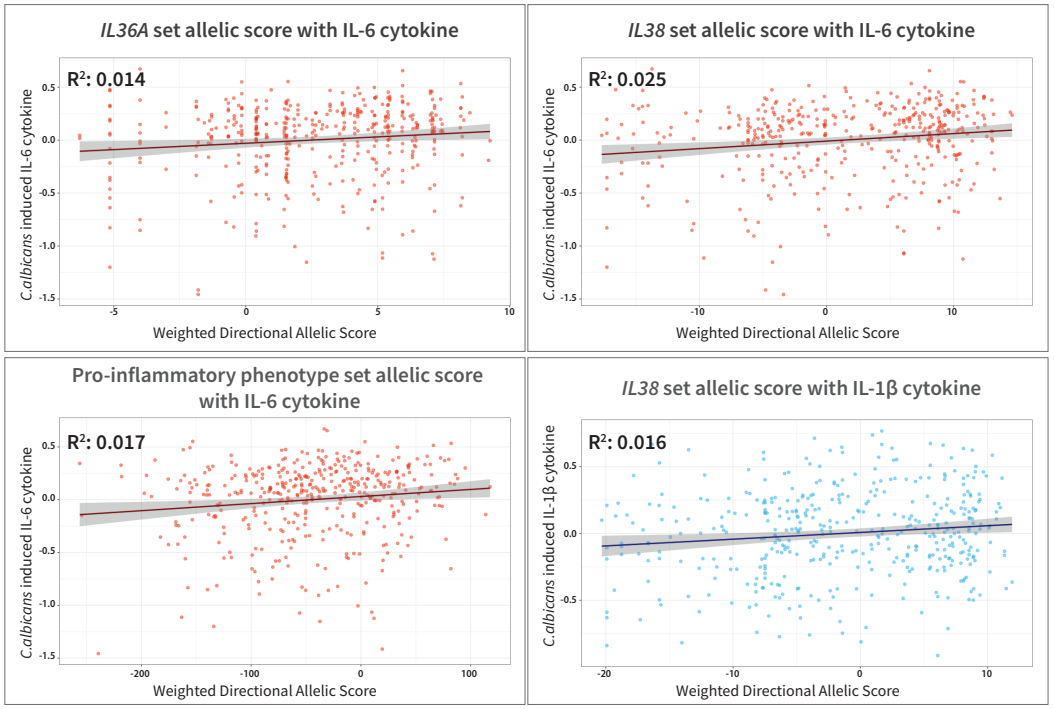


Figure 5. Coding and non-coding common variant sets are associated with *C. albicans* induced cytokine production. (A) shows that the residual IL-1β (top-panel in blue) and IL-6 (lower-panel in red) cytokine production for coding SNPs in *IL36A* and *IL38* decreases over the genotype categories. For all plots the ancestral allele is the minor allele and thus cytokine production is lowest in the homozygous ancestral category. (B) confirms that next to coding common variants, also non-coding common variants in *IL36A*, *IL38* and Pro-inflammatory phenotype sets are associated with *C.albicans* induced residual IL-6 cytokine production, and IL38 set with *C.albicans* induced residual IL-1β cytokine production. (C) shows the Beta-weighted, directional, allelic score in correlation with *C.albicans* induced residual IL-1β or IL-6 cytokine production. The straight line represents the linear model equation using method 'lm' with standard error of 0.95, and the R2 is displayed in the plot.

Annotation: * = P-value < 0.05; ** = P-value < 0.01.

REFERENCES

1. Dinarello, C.A., *Overview of the IL-1 family in innate inflammation and acquired immunity.* Immunol Rev, 2018. **281**(1): p. 8-27.
2. Garlanda, C., C.A. Dinarello, and A. Mantovani, *The interleukin-1 family: back to the future.* Immunity, 2013. **39**(6): p. 1003-18.
3. Sun, Q., et al., *Inflammasome and autophagy regulation - a two-way street.* Mol Med, 2017. **23**: p. 188-195.
4. Dinarello, C.A., *The IL-1 family of cytokines and receptors in rheumatic diseases.* Nat Rev Rheumatol, 2019. **15**(10): p. 612-632.
5. Palomo, J., et al., *The interleukin (IL)-1 cytokine family--Balance between agonists and antagonists in inflammatory diseases.* Cytokine, 2015. **76**(1): p. 25-37.
6. Jesus, A.A. and R. Goldbach-Mansky, *IL-1 blockade in autoinflammatory syndromes.* Annu Rev Med, 2014. **65**: p. 223-44.
7. van Deuren, R.C., *in preparation*.
8. Wjst, M., M. Sargurupremraj, and M. Arnold, *Genome-wide association studies in asthma: what they really told us about pathogenesis.* Curr Opin Allergy Clin Immunol, 2013. **13**(1): p. 112-8.
9. Korver, S.K., et al., *Toll-like receptor/interleukin-1 domain innate immune signalling pathway genetic variants are candidate predictors for severe gastrointestinal toxicity risk following 5-fluorouracil-based chemotherapy.* Cancer Chemother Pharmacol, 2019. **83**(2): p. 217-236.
10. Tekola Ayele, F., et al., *Genome-wide associated loci influencing interleukin (IL)-10, IL-1Ra, and IL-6 levels in African Americans.* Immunogenetics, 2012. **64**(5): p. 351-9.
11. He, M., et al., *Genome-wide association study identifies variants at the IL18-BCO2 locus associated with interleukin-18 levels.* Arterioscler Thromb Vasc Biol, 2010. **30**(4): p. 885-90.
12. Bakker, O.B., et al., *Integration of multi-omics data and deep phenotyping enables prediction of cytokine responses.* Nat Immunol, 2018. **19**(7): p. 776-786.
13. Li, Y., et al., *A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans.* Cell, 2016. **167**(4): p. 1099-1110 e14.
14. Piasecka, B., et al., *Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges.* Proc Natl Acad Sci U S A, 2018. **115**(3): p. E488-E497.
15. Ter Horst, R., et al., *Host and Environmental Factors Influencing Individual Human Cytokine Responses.* Cell, 2016. **167**(4): p. 1111-1124 e13.
16. Chung, R.H. and C.Y. Kang, *A Powerful Gene-Based Test Accommodating Common and Low-Frequency Variants to Detect Both Main Effects and Gene-Gene Interaction Effects in Case-Control Studies.* Front Genet, 2017. **8**: p. 228.
17. Hsieh, A.R., et al., *A non-threshold region-specific method for detecting rare variants in complex diseases.* PLoS One, 2017. **12**(11): p. e0188566.
18. Ionita-Laza, I., et al., *Sequence kernel association tests for the combined effect of rare and common variants.* Am J Hum Genet, 2013. **92**(6): p. 841-53.
19. Bodmer, W. and C. Bonilla, *Common and rare variants in multifactorial susceptibility to common diseases.* Nat Genet, 2008. **40**(6): p. 695-701.
20. Hiatt, J.B., et al., *Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation.* Genome Res, 2013. **23**(5): p. 843-54.
21. Jansen, S., et al., *A genotype-first approach identifies an intellectual disability-overweight syndrome caused by PHIP haploinsufficiency.* Eur J Hum Genet, 2018. **26**(1): p. 54-63.
22. Neveling, K., et al., *BRCA Testing by Single-Molecule Molecular Inversion Probes.* Clin Chem, 2017. **63**(2): p. 503-512.
23. O'Roak, B.J., et al., *Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders.* Science, 2012. **338**(6114): p. 1619-22.
24. Wu, M.C., et al., *Rare-variant association testing for sequencing data with the sequence kernel association test.* Am J Hum Genet, 2011. **89**(1): p. 82-93.
25. Lee, S., et al., *Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies.* Am J Hum Genet, 2012. **91**(2): p. 224-37.
26. Gorkin, D.U., et al., *Common DNA sequence variation influences 3-dimensional conformation of the human genome.* Genome Biol, 2019. **20**(1): p. 255.
27. Quinn, J.P., A.L. Savage, and V.J. Bubb, *Non-coding genetic variation shaping mental health.* Curr Opin Psychol, 2019. **27**: p. 18-24.
28. Meddens, C.A., et al., *Non-coding DNA in IBD: from sequence variation in DNA regulatory elements to novel therapeutic potential.* Gut, 2019. **68**(5): p. 928-941.
29. Ramsuran, V., et al., *Variation in the Untranslated Genome and Susceptibility to Infections.* Front Immunol, 2018. **9**: p. 2046.
30. Tcheandjieu, C., et al., *A phenome-wide association study of four syndromic genes reveals pleiotropic effects of common and rare variants in the general population.* medRxiv, 2019: p. 19010736.
31. Tarazona-Santos, E., et al., *Evolutionary dynamics of the human NADPH oxidase genes CYBB, CYBA, NCF2, and NCF4: functional implications.* Mol Biol Evol, 2013. **30**(9): p. 2157-67.

32. Jacob, C.O., et al., *Lupus-associated causal mutation in neutrophil cytosolic factor 2 (NCF2) brings unique insights to the structure and function of NADPH oxidase.* Proc Natl Acad Sci U S A, 2012. **109**(2): p. E59-67.
33. Rioux, J.D., et al., *Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis.* Nat Genet, 2007. **39**(5): p. 596-604.
34. de Luca, A., et al., *IL-1 receptor blockade restores autophagy and reduces inflammation in chronic granulomatous disease in mice and in humans.* Proc Natl Acad Sci U S A, 2014. **111**(9): p. 3526-31.
35. Winter, S., et al., *A Reduction in Intracellular Reactive Oxygen Species Due to a Mutation in NCF4 Promotes Autoimmune Arthritis in Mice.* Antioxid Redox Signal, 2016. **25**(18): p. 983-996.
36. Sun, J., Y. Zheng, and L. Hsu, *A unified mixed-effects model for rare-variant association in sequencing studies.* Genet Epidemiol, 2013. **37**(4): p. 334-44.
37. Barnett, I.J., S. Lee, and X. Lin, *Detecting rare variant effects using extreme phenotype sampling in sequencing association studies.* Genet Epidemiol, 2013. **37**(2): p. 142-51.
38. Bjornland, T., et al., *Powerful extreme phenotype sampling designs and score tests for genetic association studies.* Stat Med, 2018. **37**(28): p. 4234-4251.
39. Novick, D., et al., *Interleukin-18 binding protein: a novel modulator of the Th1 cytokine response.* Immunity, 1999. **10**(1): p. 127-36.
40. Boyle, E.A., et al., *MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing.* Bioinformatics, 2014. **30**(18): p. 2670-2.
41. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
42. Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Curr Protoc Bioinformatics, 2013. **43**: p. 11 10 1-11 10 33.
43. Smigielski, E.M., et al., *dbSNP: a database of single nucleotide polymorphisms.* Nucleic Acids Res, 2000. **28**(1): p. 352-5.
44. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
45. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.
46. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet, 2014. **46**(3): p. 310-5.
47. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
48. Hernandez, R.D., et al., *Ultrarare variants drive substantial cis heritability of human gene expression.* Nat Genet, 2019. **51**(9): p. 1349-1355.
49. Li, X., et al., *The impact of rare variation on gene expression across tissues.* Nature, 2017. **550**(7675): p. 239-243.
50. Li, Y., et al., *Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi.* Nat Med, 2016. **22**(8): p. 952-60.
51. Võsa, U., et al., *Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis.* bioRxiv, 2018: p. 447367.
52. Burgess, S. and S.G. Thompson, *Use of allele scores as instrumental variables for Mendelian randomization.* Int J Epidemiol, 2013. **42**(4): p. 1134-44.