

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

van Schouwenburg, PA;Davenport, EE;Kienzler, A-K;Marwah, I;Wright, B;Lucas, M;Malinauskas, T;Martin, HC;Lockstone, HE;Cazier, J-B;Chapel, HM;Knight, JC;Patel, SY

Title:

Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders

Date:

2015-10

Citation:

van Schouwenburg, P. A., Davenport, E. E., Kienzler, A. -K., Marwah, I., Wright, B., Lucas, M., Malinauskas, T., Martin, H. C., Lockstone, H. E., Cazier, J. -B., Chapel, H. M., Knight, J. C. & Patel, S. Y. (2015). Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders. *CLINICAL IMMUNOLOGY*, 160 (2), pp.301-314. <https://doi.org/10.1016/j.clim.2015.05.020>.

Persistent Link:

<https://hdl.handle.net/11343/250122>

License:

[CC BY](#)

Published in final edited form as:

Clin Immunol. 2015 October ; 160(2): 301–314. doi:10.1016/j.clim.2015.05.020.

Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders

Pauline A. van Schouwenburg^{a,b,1}, Emma E. Davenport^{a,c,1}, Anne-Kathrin Kienzler^{a,b}, Ishita Marwah^{a,b}, Benjamin Wright^{a,c}, Mary Lucas^a, Tomas Malinauskas^d, Hilary C. Martin^{a,c}, WGS500 Consortium^{a,c,2}, Helen E. Lockstone^c, Jean-Baptiste Cazier^{c,e}, Helen M. Chapel^{a,b}, Julian C. Knight^{a,c,*3}, and Smita Y. Patel^{a,b,**3}

^aNuffield Department of Medicine, University of Oxford, Oxford, UK

^bOxford NIHR Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

^cWellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

^dCold Spring Harbor Laboratory, W. M. Keck Structural Biology Laboratory, Cold Spring Harbor, NY 11724, USA

^eCentre for Computational Biology, University of Birmingham, Haworth Building, B15 2TT Edgbaston, UK

Abstract

Common Variable Immunodeficiency Disorders (CVIDs) are the most prevalent cause of primary antibody failure. CVIDs are highly variable and a genetic causes have been identified in <5% of patients. Here, we performed whole genome sequencing (WGS) of 34 CVID patients (94% sporadic) and combined them with transcriptomic profiling (RNA-sequencing of B cells) from three patients and three healthy controls. We identified variants in CVID disease genes *TNFRSF13B*, *TNFRSF13C*, *LRBA* and *NLRP12* and enrichment of variants in known and novel disease pathways. The pathways identified include B-cell receptor signalling, non-homologous end-joining, regulation of apoptosis, T cell regulation and ICOS signalling. Our data confirm the polygenic nature of CVID and suggest individual-specific aetiologies in many cases. Together our data show that WGS in combination with RNA-sequencing allows for a better understanding of CVIDs and the identification of novel disease associated pathways.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*Correspondence to: J. C. Knight, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. julian@well.ox.ac.uk (J.C. Knight). **Correspondence to: S. Y. Patel, Oxford NIHR Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK. smita.patel@ndm.ox.ac.uk (S.Y. Patel) .

¹The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors.

²A full list of members is provided in the Supplementary Material.

³The authors wish it to be known that, in their opinion, the last 2 authors should be regarded as joint Senior Authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.clim.2015.05.020>.

Keywords

Common variable immunodeficiency; Whole genome sequencing; Transcriptome; Polygenic; B-cell

1. Introduction

Common Variable Immunodeficiency Disorders (CVIDs) are the most clinically prevalent primary antibody deficiencies, present in about 1 in 25,000 people [1]. CVID is characterised by recurrent infections, low serum levels of IgG, low IgA and/or IgM and poor specific antibody responses [2]. CVID is currently a diagnosis of exclusion leaving a highly heterogeneous patient group in terms of clinical features and complications. Some patients have a relatively mild phenotype, comprising of recurrent bacterial infections and infection-related complications while others suffer from disease related complications such as autoimmune cytopenias, polyclonal lymphoproliferation, enteropathy and lymphoid malignancy, indicating underlying immune dysregulation [3].

The heterogeneous disease phenotype of CVID has made resolving genetic aetiologies challenging. Causative variants of CVID-like primary antibody failure have been described in *CD19* [4–6], *CD21* [7], *CD81* [8], *CD20* [9], *ICOS* [10,11] and *TNFRSF13C* [12], conditions now classified as specific deficiencies in these genes (Table S1). Mutations in *CD27* [13, 14], *PLCG2* [15,16], *LRBA* [17,18], *NFKB2* [19,20], *PIK3CD* [21,22] and *NLRP12* [23] cause CVID-like symptoms often combined with a more extensive clinical phenotype (Table S1). Variants in *TNFRSF13B* [24–26], *TNFRSF13C* [27], *FCGR2A* [28] and HLA [29] have been described to predispose to CVID (Table S1). Together these variants only explain the genetic cause of CVID-like diseases in very few patients and all genes were identified in familial cases of CVID, while the vast majority of CVID patients are sporadic. The wide variety in genes implicated in CVID further underlines the heterogenic nature of the disease. Further unravelling of the underlying genetic causes of sporadic CVID would give additional insight into the disease, opportunities for better patient stratification and novel insights into treatment opportunities.

In 2011 Orange et al. published the first genome-wide association study (GWAS) of CVID to identify genomic regions associated with CVID development [30]. Analysis of 363 patients and 3031 controls led to the conclusion that CVID is likely to be a polygenic disease with multiple novel susceptibility loci implicated. However, as of yet this has not resulted in further identification and elucidation of genes or variants that cause or predispose for sporadic CVID emphasizing the difficulties in studying this highly variable disease.

The development of next generation sequencing techniques has transformed the identification of the genetic basis of Mendelian diseases. In contrast, identification of the genetic basis remains challenging in polygenic conditions. Here, we present the first whole genome sequencing (WGS) data for a cohort of CVID patients to investigate novel underlying aetiologies. We further leveraged the potential of WGS by combining the results with global transcriptomic profiling through RNA-sequencing (RNA-seq). Because of the complex and probable polygenic nature of CVID, we combine the identification of genes of

interest with pathway-based analysis and focus on combining these results to identify pathways dysregulated in CVID.

2. Material and methods

2.1. Samples

Patients were recruited into the study through the Clinical Immunology Department at the Oxford University Hospital, Oxford. All patients gave informed written consent and the studies were performed according to the Declaration of Helsinki. All 34 patients were of Caucasian origin and met the ESID diagnostic criteria at the time of enrollment [2]. The majority of patients were regularly followed in the Clinical Immunology clinic at 6 monthly intervals over a period of up to 30 years with detailed clinical information entered into the local database that enabled accurate clinical phenotyping. A summary of the clinical phenotype and laboratory characteristics of the patient cohort can be found in Table 1 and a more complete overview can be found in Table S2.

2.2. WGS500

A cohort of 34 CVID patients was selected for WGS as part of the WGS500 project [31]. This is a collaborative project between the University of Oxford and Illumina, which aims to sequence the genomes of 500 individuals with a range of diseases including rare inherited diseases, immunological disorders and cancer. For all variants the frequency of the variant in the non-cancer, non-CVID samples of the WGS500 project (n = 239) is listed in Table 2 and Tables S3–S7 and S15.

2.3. Whole genome sequencing

Genomic DNA was extracted from peripheral blood mononuclear cells (PBMCs) using the FlexiGene DNA kit (Qiagen) according to the manufacturer's instructions. The DNA was quantified by fluorescence using the Qubit Fluorometer (Invitrogen) and the quality assessed by running <500 ng on a 1% TAE agarose gel for 1 h at 70 V. Whole genome sequencing was performed on 3.5–7.5 ng DNA on either the Illumina HiSeq2000 or the HiSeq2500 run in standard mode using v2.5 or v3 sequencing chemistry (Core Genomics, WTCHG). Briefly, the genomic DNA was fragmented, end-paired, A-tailed and adapter-ligated before size selection and amplification for a multiplexed library preparation. The libraries were then paired end sequenced and sequenced to an average coverage between 27 and 40 quality reads per base (100 base pair reads).

2.4. Mapping and variant calling

Reads were mapped to the GRCh37d5/hg19 human reference sequence using Stampy and the SNVs and short InDels were called with Platypus www.well.ox.ac.uk/platypus (Core Bioinformatics, WTCHG) [32].

2.5. Variant filtering

The created vcf files were analysed using QIAGEN's Ingenuity® Variant Analysis™ (IVA) software (www.qiagen.com/Ingenuity, QIAGEN Redwood City). For all analysis a

confidence filter was applied to exclude poor quality calls (Table 3) [33]. Frequency (1000 Genomes and WGS500) and predicted deleterious filters (SIFT and PolyPhen) were applied to prioritise high quality calls for non-synonymous and likely pathogenic variants (Table 3). IVA allowed the use of custom gene lists (Notes S1 and S3) as biological filters with options for ‘1 hop’ analysis both up and downstream of these genes [34]. The ‘1 hop’ function identifies interactions between genes from literature. Here, this function was used to create the list of variants in the genes interacting with described CVID genes.

Genomic regions described in GWAS were transferred from hg18 into hg19 using <http://www.genome.ucsc.edu/cgi-bin/hgLiftOver>. WGS variants were filtered as described in Table 3 and restricted to those in the identified genomic regions [30].

All variants reported (except the results of the pathway analysis) were manually inspected using the Integrative Genome Viewer (IGV) [35]. Heterozygous variants were included if the variant was present in >25% of all reads with >15% in both forward and reverse reads. Homozygous variants were included if present in >80% of all reads and a minimum of 60% of reads in both directions. Variants found in regions with a high density of insertions/deletions were excluded to reduce false positive calls resulting from mapping errors.

2.6. Sanger sequencing

Primers were designed using primer3 software (<http://bioinfo.ut.ee/primer3/>) (Table S14). 25 ng DNA was amplified using a 25 µl PCR reaction containing 2.5 µl 10× Buffer, 1.5 mM MgCl₂, 0.2 mM dNTP, 250 nM primer (each direction) and 0.5 U Platinum Taq DNA polymerase (Invitrogen). Cycling conditions were an initial denaturation step at 94 °C for 2–5 min followed by 30 cycles of 94 °C for 30 s, 56–63 °C (Table S14) for 30 s and 72 °C for 40 s followed by a final elongation step of 5 min at 72 °C. The PCR product was cleaned using either a QIAquick PCR purification kit (Qiagen) or by incubating 0.1 µl ExoI and 0.1 µl CIP with 0.4 µl NEB buffer 3 in an 8 µl reaction at 37 °C for 30 min before inactivation at 95 °C for 5 min. 3.5 µl cleaned-up template DNA was sequenced using the Big Dye Terminator 3.1 according to the manufacturer’s instructions (Applied Biosystems).

2.7. RNA-sequencing

RNA-seq data were generated for enriched B cells from three CVID patients and three healthy controls. PBMCs were isolated using Lymphoprep (Stemcell Technologies) and B-cells were enriched using CD19 microbeads (Miltenyi Biotec) according to the manufacturer’s instructions. After isolation, purity of the B-cells was 61.9–96.1% as determined by FACS analysis. RNA was extracted using a RNeasy Mini Kit (Qiagen) following the manufacturer’s instructions. Spectrophotometry (Nanodrop 2000; Thermo Scientific) was used to quantitate the RNA yield and the quality was determined by on-chip electrophoresis (Biorad Bioanalyzer; Agilent). RNA-seq was performed on 600–1000 ng RNA on the Illumina HiSeq2500 run in standard mode using v1 sequencing chemistry (Core Genomics, WTCHG). Briefly, the mRNA fraction was selected from the total RNA before conversion to cDNA. dUTP was incorporated from the second strand cDNA synthesis. The cDNA was end-repaired, A-tailed and adapter-ligated. Prior to amplification the samples underwent uridine digestion. The prepared libraries were size selected, multiplexed and

QC'ed before paired end sequencing over one lane of a flow cell. Data were aligned to the reference and reads mapped to genes using HTSeq (Core Bioinformatics, WTCHG) [36]. QC to determine the proportion of reads mapping to genes and variation within the dataset was carried out using R (R Core Team, 2012; <http://www.R-project.org>) (see Fig. S4). The genes with mapped reads were filtered so only those with at least 10 reads in at least three samples were retained prior to normalisation. This resulted in a reduction from 32,426 to 16,546 transcripts. The data were normalised using the TMM method and the edgeR R package [37]. Biological pathway enrichment was generated through the use of QIAGEN'S Ingenuity Pathway Analysis (IPA®, www.qiagen.com/ingenuity, QIAGEN Redwood City).

2.8. Protein structural analysis

Variants were mapped to corresponding protein structures in Protein Data Bank (PDB) (LRBA, Protein Data Bank (PDB) ID 1T77; TNFAIP3, PDB ID 3ZJE; SH3BP2, PDB ID 2CR4) (<http://www.rcsb.org/>) [38]. The structures were analysed using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC; <http://pymol.org/>). The effect of a variant on protein stability was assessed using mCSM [39].

3. Results

3.1. Whole genome-sequencing and CVID

As part of the WGS500 project aiming to sequence the genomes of 500 individuals with a range of diseases, we performed WGS for a cohort of 32 sporadic CVID patients and one grandmother–grandson pair [31]. The cohort had an equal sex distribution and the median age of disease onset was 23.5 years. All patients met the European Society for Immunodeficiencies (ESID) diagnostic criteria for CVIDs [2]. We excluded patients with lymphomas, though patient C003 developed a B cell lymphoma four years into the study. Information on the clinical phenotypes and immunotypes of the patients can be found in Table 1 and Table S2. WGS using the Illumina platform with an average coverage between 27 and 40 quality reads per base identified 14,819,871 variants (single nucleotide variants (SNVs) and short InDels) across the 34 patients. Initial analysis confirmed a familial relationship between the grandmother (D232) and grandson (C018) pair and revealed long runs of homozygosity in one of the sporadic patients (D269).

To validate our data we first looked for variants previously described to cause or predispose for CVID-like diseases in our patients (variants are listed in Table S1, results are summarised in Table S3). This analysis revealed five previously reported variants in *TNFRSF13B*, tumour necrosis factor receptor superfamily member 13B (also known as TACI). We identified one patient heterozygous for p.C104R, previously reported to abolish ligand binding in the heterozygous state through haploinsufficiency [24,26,40]. Other *TNFRSF13B* variants (p.P251L, p.V220A, p.R202H and p.R72H) have originally been linked to CVID, but a later study found equal frequencies in CVID patients and healthy controls as we found here (as compared to the frequency in the 1000 genomes project and the non-CVID WGS500 samples) [24–26,41]. Our analysis also identified one patient heterozygous for the p.H304Y variant in the inflammatory modulator gene *NLRP12* [23]. We identified two patients with the p.P21R variant (one homozygous) in tumour necrosis

factor receptor superfamily member 13C gene *TNFRSF13C* (BAFFR) [27] which is reported to have functional consequences through effects on multimerisation [42]. We noted poor coverage of *TNFRSF13C* in our WGS data and subsequent Sanger sequencing of all 34 patients revealed an additional 4 patients heterozygous for this variant (frequency of 0.176 compared to 0.059 in the 1000 Genomes Project) [41]. Details of variants highlighted throughout this report can be found in Table 2.

To identify novel variants of interest we selected high quality, and likely pathogenic variants. Because of the expected polygenic nature of CVID we used the relatively mild selection criteria (1000 Genomes and WGS500 frequency <0.05 and classified as pathogenic or likely pathogenic by Ingenuity Variant Analysis (IVA)), and included both heterozygous and homozygous variants resulting in 4768 variants in 3737 genes (details on the filtering can be found in Table 3). The majority of variants were SNVs (Fig. 1a) in exonic regions together with a number in annotated regulatory regions including promoters, 5' and 3'UTRs, microRNA binding sites, splice sites and noncoding RNAs (Fig. 1b). In our analysis strategy (Fig. 1c) we focused on variants present in regions implicated in GWAS [30], genes involved in described primary immunodeficiency disorders (PIDs) or genes with a direct link with previously identified CVID genes. Moreover, we also looked at pathways in which variants were overrepresented.

3.2. GWAS and PID genes

The CVID GWAS conducted by Orange and colleagues [30] included 29 of the patients described here. Within the regions associated with CVID development we identified 43 potentially pathogenic rare variants (Table S4) of which 39 are located in the HLA region of chromosome 6. We highlight the variants of interest based on a combination of gene function, frequency (1000 Genomes Project and WGS500 compared to our cohort) and predicted deleteriousness (PolyPhen, SIFT and mutation taster prediction score) (Table 2) [41,43–45]. Two variants were identified in *SKIV2L*, a gene associated with trichohepatoenteric syndrome 2 and syndromic diarrhoea, two rare diseases associated with low immunoglobulin levels and the absence of vaccine responses [46]. Four variants (p.R1883Q, p.P1745R, p.R1299Q, p.S1112F) were observed in *MDC1*, a gene involved in DNA damage response [47]. We also identified six patients (17.6%) with predicted deleterious variants in *PRRC2A* (p.R804C, p.R1556W, p.R1397W, p.F2083S) a gene associated with type 1 diabetes and rheumatoid arthritis, suggesting a function in immune regulation [48,49]. We did not find any predicted deleterious variants in *CLEC16A*, a gene recently associated with CVID via GWAS [50].

Next, we hypothesised that a combination of heterozygous variants in genes where homozygous variants cause PIDs (Note S1) may contribute to the pathogenicity of CVID as previously described for rheumatoid arthritis [51]. Analysis resolved 55 variants in described PID genes (Table S5). Again the most likely disease causing variants were identified based on frequency, gene function and predicted deleteriousness (Table 2). Variants of interest include a novel, predicted damaging p.S2221P/p.S2210P variant in *LRBA* and heterozygous predicted deleterious variants in other genes in which loss of function leads to low or absent immunoglobulin levels (*IKBKB*), combined with low or

absent B-cell subtypes (*TCF3*, *BTK*, *CD79A*, *PMS2*, *POLE*), and/or T-cell dysregulation (*IL21R*, *STIM1*, *TAP2*, *DKC1*) [52].

We identified five CVID patients with heterozygous variants in genes which are involved in non-homologous end joining (NHEJ) or V(D)J recombination and associated with T-B-SCID (*DCLRE1C*, *PRKDC*, *RAG2*) [52,53]. We found an additional four variants in other PID genes involved in NHEJ (*NHEJ1*, *MRE11A*, *ATM*, *NLRP2*) [53]. We also identified variants in PID genes involved in B-cell receptor (BCR) signalling (*IKBKB*, *CD79a*, *BTK*, *KRAS*) including a novel hemizygous variant in *BTK* (p.N350fs11/p.N526fs11/p.N560fs11) that leads to a frameshift and a premature stop codon. This male patient was previously screened for known *BTK* mutations and clinically fits the phenotype of X-linked agammaglobulinaemia (XLA) caused by *BTK* deficiency [52]. Functional follow-up is currently being conducted on this finding.

3.3. Genes in sporadic CVID

Next, we further investigated the 31 sporadic CVID patients in our cohort (excluding the grandmother:grandson pair (Note S2) and the possible XLA patient). Filtering as described in Table 3 identified 4422 variants of which 12 were present in five or more patients (Table S6). Of special interest is a likely gain of function variant in *LILRB5*, an inhibitory member of the leukocyte Ig-like receptor (LIR) family, of which little is known though other members of this protein family inhibit the activation of immune cells (Table 2) [54].

Likely pathogenic variants in genes for which there is evidence from the literature of direct interaction with genes involved in CVID-like antibody failure were investigated using the '1 hop' function in IVA (Note S3). We identified 112 variants (Table S7) of which 38 were novel, including predicted damaging variants in *PSMB9*, a gene involved in the production of peptides for MHC class I presentation, and *TNFAIP3* which plays a role in TNF α induced apoptosis and at the protein level is known to interact with *TNIP1* in which we identified two different novel variants [55]. In addition we found variants in seven genes involved in B-cell receptor signalling (*ARID3A*, *INPP5D*, *SH3BP2*, *BANK1*, *BLK*, and *GAB2*), a variant in *CAMLG* encoding a protein involved in both TACI and TCR signalling, and a variant in *FCGR3B* encoding FC γ R3B [56–62]. One patient (C032) was found to have a variant in the 3'UTR regions of two genes (*BCL2L11* and *EBF1*) involved in IL7R α signalling [63,64]. Dysregulation of this pathway leads to low T-cell numbers and low antibody production [52]. Moreover, *EBF1* is a master regulator of early B-cell development in the bone marrow [63,65].

Overall, our analysis strategy to identify variants associated with CVID, variants in GWAS regions, PID genes, the most common variants and variants in genes interacting with described CVID genes (Fig. 1c), identified on average 9.4 variants (range 5–15) in each of the 31 sporadic CVID patients. Two patients do not share any variants with the other patients, while in others, the majority (up to 87%) of variants are also found in one or more other patients (Fig. S1 and S2). A maximum of five variants are shared between two patients. Fig. 2 shows the overlap of variant-containing genes between patients. Visualising this shows that all patients share at least one variant-containing gene with another patient (Fig. 3). Patients did not cluster together based on their clinical phenotype (IgM \pm and

complex vs infections only phenotypes) (Fig S3). We did not have sufficient power to do a comparison analysis between different patient groups.

3.4. Integration with RNA sequencing analysis

As a pilot study, RNA-seq data were generated for B-cell enriched samples from three sporadic CVID patients (C078, C085 and C044, selected on B-cell numbers and sample availability) and three healthy controls (Figs. 4a–b; Fig. S4). We performed differential gene expression analysis between CVID patients and healthy controls for 16,546 transcripts revealing 262 transcripts that were differentially expressed (false discovery rate <0.05 and fold change >2) (Fig. 4c; Table S8). Down-regulated genes included those encoding immunoglobulin heavy and light chains as would be expected. This, together with our ability to reproduce a previous study showing increased *FAS* (a member of the TNF receptor superfamily encoding a receptor inducing programmed cell death) expression in CVID patients suggests that this small pilot study is able to pick up important differences in gene expression (Fig. 4d) [66]. We also find significantly increased expression of four additional PID-associated genes in the CVID patients (Fig. 4d and Fig. S) including *IL10RA*, encoding the receptor for interleukin 10 that mediates immunosuppressive signals. None of the genes located in regions implicated by GWAS or genes highlighted in previous paragraphs (Table 2) were significantly differentially expressed.

The WGS analysis we performed enables the prediction of variants deleterious to protein function. Combining WGS with RNA-seq allowed us to identify non-coding variants with evidence of altered gene expression. For each of the three patients with RNA-seq data, we applied filters to their WGS data as described in Table 3, with an alteration to the predicted deleterious filter to include variants which modify promoter and enhancer regions in addition to those with predicted deleterious effects to protein function. To determine gain of function variants we identified variants linked to genes with higher expression in CVID patients compared to healthy controls (fold change >2 from median expression) ($n = 106$, chi squared p value < 0.0001 , Table S9). Fig. S6 shows the increased expression of *ITGB2*, a described PID gene, and *ICAMI*, which interacts with three CVID related genes (*NLRP12*, *CD81*, *PLCG2*). We also identified variants linked to reduced gene expression ($n = 116$, chi squared p value < 0.0001 , Table S10). These include *SH3BP2* which interacts with and activates other CVID and PID genes and *VPREB3* which is involved in B-cell development (Fig. S6) [67,68]. No variants linked to altered gene expression are found in the regions associated with CVID identified by GWAS [30].

3.5. Pathway analysis using WGS and RNA-seq

Due to the heterogeneous and probable polygenic nature of CVID we hypothesised that different variants or combinations of variants in the same pathway could lead to similar clinical phenotypes. Therefore we focused on the identification of pathways that are possibly dysregulated in CVID. We identified the 100 pathways most enriched for genes with predicted deleterious variants in the 31 sporadic CVID patients based on WGS data (Table S11) and compared this with the 52 pathways enriched for differentially expressed genes in the three CVID patients compared to healthy controls from Ingenuity Pathway Analysis (IPA) (Table S12). We found that 24 of these pathways overlap (Table 4), which is

significantly greater than expected by chance (Fisher's exact test $p < 0.0001$; Fig. 4e). The presence of 'iCOS–iCOSL signalling' in both analyses confirms previous findings [10,11]. Other pathways of interest include those involved in the development of the immune system ('haematopoiesis from pluripotent stem cells') and regulation of apoptosis ('death receptor signalling'); and also pathways involved in cytokine signalling ('IL-6 signalling', 'IL-10 signalling') and differentiation or signalling of T helper cells ('CD28 signalling', 'PKC θ signalling' and 'T helper cell differentiation').

CVID may result from dysregulation of different pathways in different patients and analysing all 31 patients as a group may not reveal this. Therefore we repeated the WGS pathway analysis on a patient by patient basis to identify potentially important pathways at the individual level. This revealed an average of 5.8 (range 1–18) significantly enriched pathways per patient (Table S13). Pathways of particular interest include 'PI3K/AKT signalling' (three patients) and 'the role of BRCA1 in DNA damage response' (three patients) 'Actin Cytoskeleton signalling' (two patients) and 'DNA Double-strand Break repair by non-homologous end joining' (one patient).

3.6. Structural perspective on predicted deleterious variants

To further understand the mechanism by which the predicted deleterious variants (Table 2) could modulate protein structure and function, we mapped them onto available structures. Three selected variants highlighting changes of protein surface properties and interactions between secondary structure elements (helices) are shown in Fig. 5. Changes of protein surface properties could alter protein–protein interactions whereas disruption of bonding within the protein may alter folding, conformation and thus function. All three variants were mapped onto available crystal or solution NMR structures [69,70]. The novel variant p.S2210P that we identified in *LRBA* was found to affect an evolutionarily conserved, partially buried Ser2210 of *LRBA* located within the BEACH domain that is crucial for protein function [69]. The side chain of Ser2210 makes multiple hydrogen bonds with Glu2213 and Arg2397 (Fig. 5a). The Ser2210Pro mutation would abolish these interactions and could thus alter conformation and/or stability of *LRBA*. In contrast, with the variant p.R534W we identified in the B cell receptor signalling protein *SH3BP2*, the side chain of Arg534 does not interact with any residues within the SH2 domain (Fig. 5b). As Arg534 is exposed on the SH2 surface, the Arg534Trp mutation would make *SH3BP2* more hydrophobic and could therefore alter *SH3BP2* interactions with its potential binding partners recognising the surface of *SH3BP2* centred on Arg534. The pR280W mutation of the deubiquitinase *TNFAIP3* (also known as A20) involves Arg280 which is located within the catalytic domain and stabilizes conformation of the domain by binding to Asp279 (Fig. 5c) [70]. The Arg280Trp mutation would abolish these interactions and thus might affect the overall conformation, catalytic activity and/or *TNFAIP3*–protein interactions. We further assessed whether these three variants could affect protein stability using a programme mCSM [39]. All three mutations are predicted to destabilize the proteins (change in Gibbs free energy, ΔG , kcal/mol: *LRBA* Ser2210Pro, -0.236 ; *SH3BP2* Arg534Trp, -0.461 ; *TNFAIP3* Arg280Trp, -0.138).

4. Discussion

Here, we present the first WGS analysis of sporadic CVID patients and show that this does not reveal a unifying genetic aetiology in terms of highly penetrant mutations involving a single gene or pathway but rather that CVID is likely to include a number of different disease aetiologies matching a polygenic model. We applied an analysis strategy considering previously implicated candidate genes, loci implicated by CVID GWAS, together with known primary immunodeficiency disorder genes. Further investigation of genes with experimental evidence of interaction with those involved in CVID-like antibody failure was informative. We adopted an integrated analysis that considered WGS data in the context of differentially expressed genes in CVID patients, allowing interpretation in a biological context. We recognise that significantly larger sample sizes would allow further progress in resolving the heritable basis of CVID and resolving the current classification of disease aetiology. Our finding of a novel BTK mutation illustrates the diagnostic power of the WGS approach.

Among the CVID patients analysed, we found variants in *TNFRSF13B*, *TNFRSF13C* and *NLRP12*, consistent with previous findings [23–27]. We also found a novel heterozygous variant in *LRBA*, which is predicted to influence protein stability. Using our combined analysis strategy, we are able to identify a number of important pathways where we find evidence of pathogenic mutations which together with biological evidence that we and others have generated reveal likely significant pathways in CVID. For example we found significant enrichment in WGS and differential gene expression analysis of ICOS–ICOS ligand signalling confirming the previously found importance of this pathway in antibody failure [10,11].

Other studies have previously shown that deficiency in genes involved in B-cell receptor sequencing (*CD19*, *CD21*, *CD81*, *PLCG2*, *PI3KCD*) can lead to CVID-like antibody deficiencies [4–8,15,16]. In our analysis we found heterozygous coding variants in a variety of genes involved in this pathway (*IKBKB*, *CD79a*, *BTK*, *KRAS*, *ARID3A*, *INPP5D*, *BANK1*, *BLK*, and *GAB2*) including a number of novel variants in *BTK*, *BLK* and *ARID3A*. In one patient we found reduced expression and two variants (one coding and one non-coding) in *SH3BP2*. The coding variant might alter interactions with potential binding partners. Moreover, in our RNA-seq analysis BCR signalling was one of the most significantly enriched pathways in CVID patients.

Previous studies showed that variants in BAFF-R, a receptor important for B-cell survival, can cause or predispose to CVID [12,27,44]. We, and others, show increased expression of *FAS*, an apoptosis inducing receptor, in CVID patients [66]. Moreover, we found a variant which might alter the overall conformation of *TNFAIP3*, a gene inhibiting TNF α induced apoptosis, and novel variants in *TNIP1* which is a known *TNFAIP3* interaction partner. Death receptor signalling was significant in pathway analysis of both the RNA-seq and WGS analyses together suggesting that apoptosis might be dysregulated in a subgroup of CVID patients.

As NHEJ is essential in V(D)J recombination and class-switch recombination, inefficient NHEJ could lead to CVID like symptoms. We find variants in *DCLRE1C*, *PRKDC*, *RAG2*, *NHEJ1*, *MRE11A*, *ATM* and *NLRP2* in our patients, all genes important in the initiation of V(D)J recombination or DSB repair by NHEJ. The hypothesis that NHEJ defects could lead to CVID in a subgroup of patients is supported by the recent finding of a leaky block in B-cell development in a study comparing the bone marrow of CVID patients and healthy controls [71].

Our analysis also highlighted the role of T-cells in sporadic, non familial CVID. We found variants in genes associated with T-cell regulation (*IL21R*, *STIM1*, *TAP2*) with WGS. Pathway analysis of both the RNA-seq and WGS data revealed enrichment of a number of T-cell related pathways (T helper cell differentiation; CD28 signalling; ICOS–ICOSL signalling). Dysregulation of T cells could also be the result of pathways found to be significantly enriched in the WGS data related to dendritic cell maturation, interaction between innate and adaptive immune cells and cytokine signalling (IL-18, IL6 and IL10) which would lead to B cell dysfunction through altered B–T cell interactions in CVID patients.

Interestingly, we found a relatively high number of T-cell related pathways in our RNA-seq analysis, having investigated and compared the expression profiles of samples enriched for B-cells. One possible explanation is the presence of residual T-cells in our samples (61.9–96.1% B-cell purity). Alternatively, genes that are important in T-cell signalling pathways could also be expressed in B-cells and therefore be found in our analysis.

In one of our CVID patients we identified variants in the 3'UTR region of two genes involved in IL-7R signalling (*BCL2L1* and *EBF1*) [63,64]. 3' UTRs are important for the determination of translation efficiency, localization and stability of mRNA. In mice IL-7R signalling and EBF-1 expression have been described as central regulators in B cell development where their interplay positively regulates B lineage commitment, survival and proliferation [64]. In humans, EBF-1 is important in B-cell lineage commitment [63]. In contrast, the role of IL-7R signalling in B-cell development seems less important as patients with IL-7R α deficiency have normal B-cell numbers.

Together our data illustrate the strength of a combined approach using WGS and RNA-seq and highlight that CVID is a highly heterogeneous and polygenic disease. Our results corroborate with previous findings in *TNFRSF13B*, *TNFRSF13C* and *NLRP12*. Moreover, our data highlight variants in genes involved in ICOS–ICOSL signalling, regulation of apoptosis, and BCR signalling. We were also able to identify novel pathways which may be important for CVID, including NHEJ, T cell regulation and cytokine signalling. However, functional follow-up studies in individual patients are needed to confirm the role of variable pathways involved in the pathogenesis in CVIDs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

TM is funded by a Long-term Fellowship from the Human Frontier Science Program grant LT000021/2014-L. JCK is supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 281824, and the NIHR Oxford Biomedical Research Centre. The Wellcome Trust grant 090532/Z/09/Z supports the core facilities Wellcome Trust Centre for Human Genetics including the High-throughput Genomics Group. SYP is funded by the NIHR Oxford Biomedical Research Centre and the Jeffrey Model Foundation.

Abbreviations

BCR	(B-cell receptor)
CVIDs	(Common Variable Immunodeficiency Disorders)
ESID	(European Society for Immunodeficiencies)
GWAS	(Genome-wide Association Study)
HLA	(Human Leucocyte Antigen)
IPA	(Ingenuity Pathway Analysis)
IVA	(Ingenuity Variant Analysis)
LIR	(Leucocyte Ig receptor)
PDB	(Protein Data Bank)
PIDs	(Primary Immunodeficiency Disorders)
NHEJ	(Non-homologous End Joining)
RNA-seq	(RNA sequencing)
SNVs	(single nucleotide variants)
WGS	(whole genome sequencing)
XLA	(X-linked agammaglobulinaemia)

References

- [1]. Hammarstrom L, Vorechovsky I, Webster D. Selective IgA deficiency (SIgAD) and common variable immunodeficiency (CVID). *Clin. Exp. Immunol.* 2000; 120:225–231. [PubMed: 10792368]
- [2]. Conley ME, Dobbs AK, Farmer DM, Kilic S, Paris K, Grigoriadou S, Coustan-Smith E, Howard V, Campana D. Primary B cell immunodeficiencies: comparisons and contrasts. *Annu. Rev. Immunol.* 2009; 27:199–227. [PubMed: 19302039]
- [3]. Chapel H, Lucas M, Lee M, Bjorkander J, Webster D, Grimbacher B, Fieschi C, Thon V, Abedi MR, Hammarstrom L. Common variable immunodeficiency disorders: division into distinct clinical phenotypes. *Blood.* 2008; 112:277–286. [PubMed: 18319398]
- [4]. Kanegane H, Agematsu K, Futatani T, Sira MM, Suga K, Sekiguchi T, van Zelm MC, Miyawaki T. Novel mutations in a Japanese patient with CD19 deficiency. *Genes Immun.* 2007; 8:663–670. [PubMed: 17882224]
- [5]. van Zelm MC, Reisli I, van der Burg M, Castano D, van Noesel CJ, van Tol MJ, Woellner C, Grimbacher B, Patino PJ, van Dongen JJ, Franco JL. An antibody-deficiency syndrome due to mutations in the CD19 gene. *N. Engl. J. Med.* 2006; 354:1901–1912. [PubMed: 16672701]

- [6]. van Zelm MC, Smet J, van der Burg M, Ferster A, Le PQ, Schandene L, van Dongen JJ, Mascart F. Antibody deficiency due to a missense mutation in CD19 demonstrates the importance of the conserved tryptophan 41 in immunoglobulin super-family domain formation. *Hum. Mol. Genet.* 2011; 20:1854–1863. [PubMed: 21330302]
- [7]. Thiel J, Kimmig L, Salzer U, Grudzien M, Lebrecht D, Hagen T, Draeger R, Voelxen N, Bergbreiter A, Jennings S, Gutenberger S, Aichem A, Illges H, Hannan JP, Kienzler AK, Rizzi M, Eibel H, Peter HH, Warnatz K, Grimbacher B, Rump JA, Schlesier M. Genetic CD21 deficiency is associated with hypogammaglobulinemia. *J. Allergy Clin. Immunol.* 2012; 129:801–810. e806. [PubMed: 22035880]
- [8]. van Zelm MC, Smet J, Adams B, Mascart F, Schandene L, Janssen F, Ferster A, Kuo CC, Levy S, van Dongen JJ, van der Burg M. CD81 gene defect in humans disrupts CD19 complex formation and leads to antibody deficiency. *J. Clin. Invest.* 2010; 120:1265–1274. [PubMed: 20237408]
- [9]. Kuijpers TW, Bende RJ, Baars PA, Grummels A, Derks IA, Dolman KM, Beaumont T, Tedder TF, van Noesel CJ, Eldering E, van Lier RA. CD20 deficiency in humans results in impaired T cell-independent antibody responses. *J. Clin. Invest.* 2010; 120:214–222. [PubMed: 20038800]
- [10]. Grimbacher B, Hutloff A, Schlesier M, Glocker E, Warnatz K, Drager R, Eibel H, Fischer B, Schaffer AA, Mages HW, Kroczeck RA, Peter HH. Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. *Nat. Immunol.* 2003; 4:261–268. [PubMed: 12577056]
- [11]. Salzer U, Maul-Pavicic A, Cunningham-Rundles C, Urschel S, Belohradsky BH, Litzman J, Holm A, Franco JL, Plebani A, Hammarstrom L, Skrabl A, Schwinger W, Grimbacher B. ICOS deficiency in patients with common variable immunodeficiency. *Clin. Immunol.* 2004; 113:234–240. [PubMed: 15507387]
- [12]. Warnatz K, Salzer U, Rizzi M, Fischer B, Gutenberger S, Bohm J, Kienzler AK, Pan-Hammarstrom Q, Hammarstrom L, Rakhmanov M, Schlesier M, Grimbacher B, Peter HH, Eibel H. B-cell activating factor receptor deficiency is associated with an adult-onset antibody deficiency syndrome in humans. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:13945–13950. [PubMed: 19666484]
- [13]. Salzer E, Daschkey S, Choo S, Gombert M, Santos-Valente E, Ginzel S, Schwendinger M, Haas OA, Fritsch G, Pickl WF, Forster-Waldl E, Borkhardt A, Boztug K, Bienemann K, Seidel MG. Combined immunodeficiency with life-threatening EBV-associated lymphoproliferative disorder in patients lacking functional CD27. *Haematologica.* 2013; 98:473–478. [PubMed: 22801960]
- [14]. van Montfrans JM, Hoepelman AI, Otto S, van Gijn M, van de Corput L, de Weger RA, Monaco-Shawver L, Banerjee PP, Sanders EA, Jol-van der Zijde CM, Betts MR, Orange JS, Bloem AC, Tesselaar K. CD27 deficiency is associated with combined immunodeficiency and persistent symptomatic EBV viremia. *J. Allergy Clin. Immunol.* 2012; 129:787–793. e786. [PubMed: 22197273]
- [15]. Ombrello MJ, Rimmers EF, Sun G, Freeman AF, Datta S, Torabi-Parizi P, Subramanian N, Bunney TD, Baxendale RW, Martins MS, Romberg N, Komarow H, Aksentijevich I, Kim HS, Ho J, Cruse G, Jung MY, Gilfillan AM, Metcalfe DD, Nelson C, O'Brien M, Wisch L, Stone K, Douek DC, Gandhi C, Wanderer AA, Lee H, Nelson SF, Shianna KV, Cirulli ET, Goldstein DB, Long EO, Moir S, Meffre E, Holland SM, Kastner DL, Katan M, Hoffman HM, Milner JD. Cold urticaria, immunodeficiency, and autoimmunity related to PLCG2 deletions. *N. Engl. J. Med.* 2012; 366:330–338. [PubMed: 22236196]
- [16]. Zhou Q, Lee G-S, Brady J, Datta S, Katan M, Sheikh A, Martins MS, Bunney TD, Santich BH, Moir S, Kuhns DB, Priel DAL, Ombrello A, Stone D, Ombrello MJ, Khan J, Milner JD, Kastner DL, Aksentijevich I. A hypermorphic missense mutation in PLCG2, encoding phospholipase C γ 2, causes a dominantly inherited autoinflammatory disease with immunodeficiency. *Am. J. Hum. Genet.* 2012; 91:713–720. [PubMed: 23000145]
- [17]. Alangari A, Alsultan A, Adly N, Massaad MJ, Kiani IS, Aljebreen A, Raddaoui E, Almomen AK, Al-Muhsen S, Geha RS, Alkuraya FS. LPS-responsive beige-like anchor (LRBA) gene mutation in a family with inflammatory bowel disease and combined immunodeficiency. *J. Allergy Clin. Immunol.* 2012; 130:481–488. e482. [PubMed: 22721650]
- [18]. Lopez-Herrera G, Tampella G, Pan-Hammarstrom Q, Herholz P, Trujillo-Vargas CM, Phadwal K, Simon AK, Moutschen M, Etzioni A, Mory A, Srugo I, Melamed D, Hultenby K, Liu C,

- Baronio M, Vitali M, Philippet P, Dideberg V, Aghamohammadi A, Rezaei N, Enright V, Du L, Salzer U, Eibel H, Pfeifer D, Veelken H, Stauss H, Lougaris V, Plebani A, Gertz EM, Schaffer AA, Hammarstrom L, Grimbacher B. Deleterious mutations in LRBA are associated with a syndrome of immune deficiency and autoimmunity. *Am. J. Hum. Genet.* 2012; 90:986–1001. [PubMed: 22608502]
- [19]. Chen K, Coonrod EM, Kumanovics A, Franks ZF, Durtschi JD, Margraf RL, Wu W, Heikal NM, Augustine NH, Ridge PG, Hill HR, Jorde LB, Weyrich AS, Zimmerman GA, Gundlapalli AV, Bohnsack JF, Voelkerding KV. Germline mutations in NFKB2 implicate the noncanonical NF-kappaB pathway in the pathogenesis of common variable immunodeficiency. *Am. J. Hum. Genet.* 2013; 93:812–824. [PubMed: 24140114]
- [20]. Liu Y, Hanson S, Gurugama P, Jones A, Clark B, Ibrahim MA. Novel NFKB2 mutation in early-onset COVID. *J. Clin. Immunol.* 2014; 24:686–690. [PubMed: 24888602]
- [21]. Angulo I, Vadas O, Garcon F, Banham-Hall E, Plagnol V, Leahy TR, Baxendale H, Coulter T, Curtis J, Wu C, Blake-Palmer K, Perisic O, Smyth D, Maes M, Fiddler C, Juss J, Cilliers D, Markelj G, Chandra A, Farmer G, Kielkowska A, Clark J, Kracker S, Debre M, Picard C, Pellier I, Jabado N, Morris JA, Barcenas-Morales G, Fischer A, Stephens L, Hawkins P, Barrett JC, Abinun M, Clatworthy M, Durandy A, Doffinger R, Chilvers ER, Cant AJ, Kumararatne D, Okkenhaug K, Williams RL, Condliffe A, Nejentsev S. Phosphoinositide 3-kinase delta gene mutation predisposes to respiratory infection and airway damage. *Science.* 2013; 342:866–871. [PubMed: 24136356]
- [22]. Lucas CL, Kuehn HS, Zhao F, Niemela JE, Deenick EK, Palendira U, Avery DT, Moens L, Cannons JL, Biancalana M, Stoddard J, Ouyang W, Frucht DM, Rao VK, Atkinson TP, Agharahami A, Hussey AA, Folio LR, Olivier KN, Fleisher TA, Pittaluga S, Holland SM, Cohen JI, Oliveira JB, Tangye SG, Schwartzberg PL, Lenardo MJ, Uzel G. Dominant-activating germline mutations in the gene encoding the PI(3)K catalytic subunit p110delta result in T cell senescence and human immunodeficiency. *Nat. Immunol.* 2014; 15:88–97. [PubMed: 24165795]
- [23]. Borte S, Celiksoy MH, Menzel V, Ozkaya O, Ozen FZ, Hammarstrom L, Yildiran A. Novel NLRP12 mutations associated with intestinal amyloidosis in a patient diagnosed with common variable immunodeficiency. *Clin. Immunol.* 2014; 154:105–111. [PubMed: 25064839]
- [24]. Castigli E, Wilson SA, Garibyan L, Rachid R, Bonilla F, Schneider L, Geha RS. TACI is mutant in common variable immunodeficiency and IgA deficiency. *Nat. Genet.* 2005; 37:829–834. [PubMed: 16007086]
- [25]. Pan-Hammarstrom Q, Salzer U, Du L, Bjorkander J, Cunningham-Rundles C, Nelson DL, Bacchelli C, Gaspar HB, Offer S, Behrens TW, Grimbacher B, Hammarstrom L. Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. *Nat. Genet.* 2007; 39:429–430. [PubMed: 17392797]
- [26]. Salzer U, Chapel HM, Webster AD, Pan-Hammarstrom Q, Schmitt-Graeff A, Schlesier M, Peter HH, Rockstroh JK, Schneider P, Schaffer AA, Hammarstrom L, Grimbacher B. Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat. Genet.* 2005; 37:820–828. [PubMed: 16007087]
- [27]. Losi CG, Silini A, Fiorini C, Soresina A, Meini A, Ferrari S, Notarangelo LD, Lougaris V, Plebani A. Mutational analysis of human BAFF receptor TNFRSF13C (BAFF-R) in patients with common variable immunodeficiency. *J. Clin. Immunol.* 2005; 25:496–502. [PubMed: 16160919]
- [28]. Flinsenberg TW, Janssen WJ, Herczenik E, Boross P, Nederend M, Jongeneel LH, Scholman RC, Boelens JJ, Maas C, van Gijn ME, van Montfrans JM, Leusen JH, Boes M. A novel FcgammaRIIa Q27W gene variant is associated with common variable immune deficiency through defective FcgammaRIIa downstream signaling. *Clin. Immunol.* 2014; 155:108–117. [PubMed: 25242138]
- [29]. Olerup O, Smith CI, Bjorkander J, Hammarstrom L. Shared HLA class II-associated genetic susceptibility and resistance, related to the HLA-DQB1 gene, in IgA deficiency and common variable immunodeficiency. *Proc. Natl. Acad. Sci. U. S. A.* 1992; 89:10653–10657. [PubMed: 1438261]
- [30]. Orange JS, Glessner JT, Resnick E, Sullivan KE, Lucas M, Ferry B, Kim CE, Hou C, Wang F, Chiavacci R, Kugathasan S, Sleasman JW, Baldassano R, Perez EE, Chapel H, Cunningham-Rundles C, Hakonarson H. Genome-wide association identifies diverse causes of common

- variable immunodeficiency. *J. Allergy Clin. Immunol.* 2011; 127:1360–1367. e1366. [PubMed: 21497890]
- [31]. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier JB, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C, Aricescu AR, Attar M, Babbs C, Becq J, Beeson D, Bento C, Bignell P, Blair E, Buckle VJ, Bull K, Cais O, Cario H, Chapel H, Copley RR, Cornall R, Craft J, Dahan K, Davenport EE, Dendrou C, Devuyst O, Fenwick AL, Flint J, Fugger L, Gilbert RD, Goriely A, Green A, Greger IH, Grocock R, Gruszczyk AV, Hastings R, Hatton E, Higgs D, Hill A, Holmes C, Howard M, Hughes L, Humburg P, Johnson D, Karpe F, Kingsbury Z, Kini U, Knight JC, Krohn J, Lambie S, Langman C, Lonie L, Luck J, McCarthy D, McGowan SJ, McMullin MF, Miller KA, Murray L, Nemeth AH, Nesbit MA, Nutt D, Ormondroyd E, Oturai AB, Pagnamenta A, Patel SY, Percy M, Petousi N, Piazza P, Piret SE, Polanco-Echeverry G, Popitsch N, Powrie F, Pugh C, Quek L, Robbins PA, Robson K, Russo A, Sahgal N, van Schouwenburg PA, Schuh A, Silverman E, Simmons A, Sorensen PS, Sweeney E, Taylor J, Thakker RV, Tomlinson I, Trebes A, Twigg SR, Uhlig HH, Vyas P, Vyse T, Wall SA, Watkins H, Whyte MP, Witty L, Wright B, Yau C, Buck D, Humphray S, Ratcliffe PJ, Bell JI, Wilkie AO, Bentley D, Donnelly P, McVean G. *Nature genetics.* 2015; 47:717–726. [PubMed: 25985138]
- [32]. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 2011; 21:936–939. [PubMed: 20980556]
- [33]. Howard MF, Murakami Y, Pagnamenta AT, Daumer-Haas C, Fischer B, Hecht J, Keays DA, Knight SJ, Kolsch U, Kruger U, Leiz S, Maeda Y, Mitchell D, Mundlos S, Phillips JA III, Robinson PN, Kini U, Taylor JC, Horn D, Kinoshita T, Krawitz PM. Mutations in PGAP3 impair GPI-anchor maturation, causing a subtype of hyperphosphatasia with mental retardation. *Am. J. Hum. Genet.* 2014; 94:278–287. [PubMed: 24439110]
- [34]. Scott AF, Mohr DW, Ling H, Scharpf RB, Zhang P, Liptak GS. Characterization of the genomic architecture and mutational spectrum of a small cell prostate carcinoma. *Genes (Basel).* 2014; 5:366–384. [PubMed: 24823478]
- [35]. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 2013; 14:178–192. [PubMed: 22517427]
- [36]. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015; 31:166–169. [PubMed: 25260700]
- [37]. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26:139–140. [PubMed: 19910308]
- [38]. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
- [39]. Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics.* 2014; 30:335–342. [PubMed: 24281696]
- [40]. Lee JJ, Jabara HH, Garibyan L, Rauter I, Sannikova T, Dillon SR, Bram R, Geha RS. The C104R mutant impairs the function of transmembrane activator and calcium modulator and cyclophilin ligand interactor (TACI) through haploinsufficiency. *J. Allergy Clin. Immunol.* 2010; 126:1234–1241. e1232. [PubMed: 20889194]
- [41]. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012; 491:56–65. [PubMed: 23128226]
- [42]. Pieper K, Rizzi M, Speletas M, Smulski CR, Sic H, Kraus H, Salzer U, Fiala GJ, Schamel WW, Lougaris V, Plebani A, Hammarstrom L, Recher M, Gemenis AE, Grimbacher B, Warnatz K, Rolink AG, Schneider P, Notarangelo LD, Eibel H. A common single nucleotide polymorphism impairs B-cell activating factor receptor’s multimerization, contributing to common variable immunodeficiency. *J. Allergy Clin. Immunol.* 2014; 133:1222–1225. [PubMed: 24406071]
- [43]. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat. Methods.* 2010; 7:248–249. [PubMed: 20354512]

- [44]. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 2009; 4:1073–1081. [PubMed: 19561590]
- [45]. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods.* 2014; 11:361–362. [PubMed: 24681721]
- [46]. Fabre A, Charroux B, Martinez-Vinson C, Roquelaure B, Odul E, Sayar E, Smith H, Colomb V, Andre N, Hugot JP, Goulet O, Lacoste C, Sarles J, Royet J, Levy N, Badens C. SKIV2L mutations cause syndromic diarrhea, or trichohepatoenteric syndrome. *Am. J. Hum. Genet.* 2012; 90:689–692. [PubMed: 22444670]
- [47]. Coster G, Goldberg M. The cellular response to DNA damage: a focus on MDC1 and its interacting proteins. *Nucleus.* 2010; 1:166–178. [PubMed: 21326949]
- [48]. McKinnon E, Morahan G, Nolan D, James I. Association of MHC SNP genotype with susceptibility to type 1 diabetes: a modified survival approach. *Diabetes Obes. Metab.* 2009; 11(Suppl. 1):92–100. [PubMed: 19143821]
- [49]. Singal DP, Li J, Lei K. Genetics of rheumatoid arthritis (RA): two separate regions in the major histocompatibility complex contribute to susceptibility to RA. *Immunol. Lett.* 1999; 69:301–306. [PubMed: 10528792]
- [50]. Li J, Jorgensen SF, Maggadottir SM, Bakay M, Warnatz K, Glessner J, Pandey R, Salzer U, Schmidt RE, Perez E, Resnick E, Goldacker S, Buchta M, Witte T, Padyukov L, Videm V, Folseraas T, Atschekzei F, Elder JT, Nair RP, Winkelmann J, Gieger C, Nothen MM, Buning C, Brand S, Sullivan KE, Orange JS, Fevang B, Schreiber S, Lieb W, Aukrust P, Chapel H, Cunningham-Rundles C, Franke A, Karlsen TH, Grimbacher B, Hakonarson H, Hammarstrom L, Ellinghaus E. Association of CLEC16A with human common variable immunodeficiency disorder and role in murine B cells. *Nat. Commun.* 2015; 6:6804. [PubMed: 25891430]
- [51]. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A, Ortmann W, Bhangale T, Denny JC, Carroll RJ, Eyler AE, Greenberg JD, Kremer JM, Pappas DA, Jiang L, Yin J, Ye L, Su DF, Yang J, Xie G, Keystone E, Westra HJ, Esko T, Metspalu A, Zhou X, Gupta N, Mirel D, Stahl EA, Diogo D, Cui J, Liao K, Guo MH, Myouzen K, Kawaguchi T, Coenen MJ, van Riel PL, van de Laar MA, Guchelaar HJ, Huizinga TW, Dieude P, Mariette X, Bridges SL Jr, Zhernakova A, Toes RE, Tak PP, Miceli-Richard C, Bang SY, Lee HS, Martin J, Gonzalez-Gay MA, Rodriguez-Rodriguez L, Rantapaa-Dahlqvist S, Arlestig L, Choi HK, Kamatani Y, Galan P, Lathrop M, Eyre S, Bowes J, Barton A, de Vries N, Moreland LW, Criswell LA, Karlson EW, Taniguchi A, Yamada R, Kubo M, Liu JS, Bae SC, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Raychaudhuri S, Stranger BE, De Jager PL, Franke L, Visscher PM, Brown MA, Yamanaka H, Mimori T, Takahashi A, Xu H, Behrens TW, Siminovitch KA, Momohara S, Matsuda F, Yamamoto K, Plenge RM. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014; 506:376–381. [PubMed: 24390342]
- [52]. Al-Herz W, Bousfiha A, Casanova JL, Chatila T, Conley ME, Cunningham-Rundles C, Etzioni A, Franco JL, Gaspar HB, Holland SM, Klein C, Nonoyama S, Ochs HD, Oksenhendler E, Picard C, Puck JM, Sullivan K, Tang ML. Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency. *Front. Immunol.* 2014; 5:162. [PubMed: 24795713]
- [53]. Malu S, Malshetty V, Francis D, Cortes P. Role of non-homologous end joining in V(D)J recombination. *Immunol. Res.* 2012; 54:233–246. [PubMed: 22569912]
- [54]. Tedla N, Lee CW, Borges L, Geczy CL, Arm JP. Differential expression of leukocyte immunoglobulin-like receptors on cord-blood-derived human mast cell progenitors and mature mast cells. *J. Leukoc. Biol.* 2008; 83:334–343. [PubMed: 17998301]
- [55]. Ramirez VP, Gurevich I, Aneskievich BJ. Emerging roles for TNIP1 in regulating post-receptor signaling. *Cytokine Growth Factor Rev.* 2012; 23:109–118. [PubMed: 22542476]
- [56]. Bram RJ, Crabtree GR. Calcium signalling in T cells stimulated by a cyclophilin B-binding protein. *Nature.* 1994; 371:355–358. [PubMed: 7522304]
- [57]. de la Fuente MA, Kumar L, Lu B, Geha RS. 3BP2 deficiency impairs the response of B cells, but not T cells, to antigen receptor ligation. *Mol. Cell. Biol.* 2006; 26:5214–5225. [PubMed: 16809760]

- [58]. Gold MR. To make antibodies or not: signaling by the B-cell antigen receptor. *Trends Pharmacol. Sci.* 2002; 23:316–324. [PubMed: 12119152]
- [59]. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 2014; 42:D199–D205. [PubMed: 24214961]
- [60]. Tidwell JA, Schmidt C, Heaton P, Wilson V, Tucker PW. Characterization of a new ARID family transcription factor (Brightlike/ARID3C) that co-activates Bright/ARID3A-mediated immunoglobulin gene transcription. *Mol. Immunol.* 2011; 49:260–272. [PubMed: 21955986]
- [61]. von Bulow GU, Bram RJ. NF-AT activation induced by a CAML-interacting member of the tumor necrosis factor receptor superfamily. *Science.* 1997; 278:138–141. [PubMed: 9311921]
- [62]. Yokoyama K, Ih I.H.Su. Tezuka T, Yasuda T, Mikoshiba K, Tarakhovsky A, Yamamoto T. BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP(3) receptor. *EMBO J.* 2002; 21:83–92. [PubMed: 11782428]
- [63]. Bohle V, Doring C, Hansmann ML, Kuppers R. Role of early B-cell factor 1 (EBF1) in Hodgkin lymphoma. *Leukemia.* 2013; 27:671–679. [PubMed: 23174882]
- [64]. Clark MR, Mandal M, Ochiai K, Singh H. Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. *Nat. Rev. Immunol.* 2014; 14:69–80. [PubMed: 24378843]
- [65]. Nechanitzky R, Akbas D, Scherer S, Gyory I, Hoyle T, Ramamoorthy S, Diefenbach A, Grosschedl R. Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. *Nat. Immunol.* 2013; 14:867–875. [PubMed: 23812095]
- [66]. Artac H, Kara R, Reisli I. In vivo modulation of the expressions of Fas and CD25 by intravenous immunoglobulin in common variable immunodeficiency. *Clin. Exp. Med.* 2010; 10:27–31. [PubMed: 19730984]
- [67]. Deckert M, Rottapel R. The adapter 3BP2: how it plugs into leukocyte signaling. *Adv. Exp. Med. Biol.* 2006; 584:107–114. [PubMed: 16802602]
- [68]. Shirasawa T, Ohnishi K, Hagiwara S, Shigemoto K, Takebe Y, Rajewsky K, Takemori T. A novel gene product associated with mu chains in immature B cells. *EMBO J.* 1993; 12:1827–1834. [PubMed: 8491176]
- [69]. Gebauer D, Li J, Jogl G, Shen Y, Myszka DG, Tong L. Crystal structure of the PH-BEACH domains of human LRBA/BGL. *Biochemistry.* 2004; 43:14873–14880. [PubMed: 15554694]
- [70]. Kulathu Y, Garcia FJ, Mevissen TE, Busch M, Arnaudo N, Carroll KS, Barford D, Komander D. Regulation of A20 and other OTU deubiquitinases by reversible oxidation. *Nat. Commun.* 2013; 4:1569. [PubMed: 23463012]
- [71]. Anzilotti C, Kienzler AK, Lopez-Granados E, Gooding S, Davies B, Pandit H, Lucas M, Price A, Littlewood T, van der Burg M, Patel SY, Chapel H. Key stage of bone marrow B-cell maturation are defective in patients with common variable immunodeficiency disorders. *J. Clin. Immunol.* 2015 <http://dx.doi.org/1.1016/j.jaci.2014.12.1943> (Epub ahead of print) (accepted for publication).

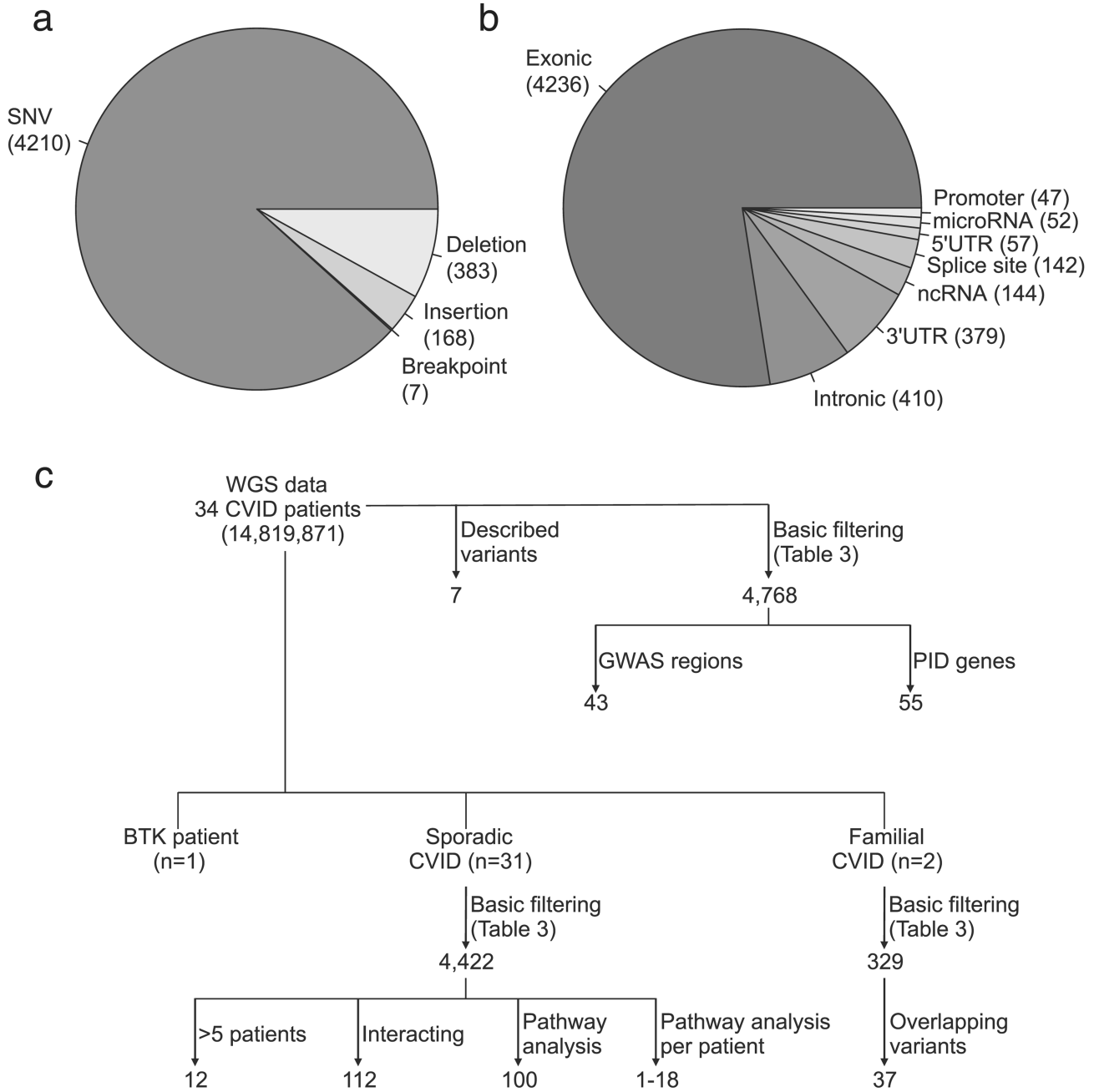


Fig. 1. Overview of analysis strategy and distribution of variant type and genetic location of variants. (a) Distribution of variant type for variants that passed the filtering steps as described in Table 3. (b) Number of variants in different genetic locations in the 4768 variants that passed the basic filtering described in Table 3. (c) Overview of the analysis strategies applied to the WGS data.

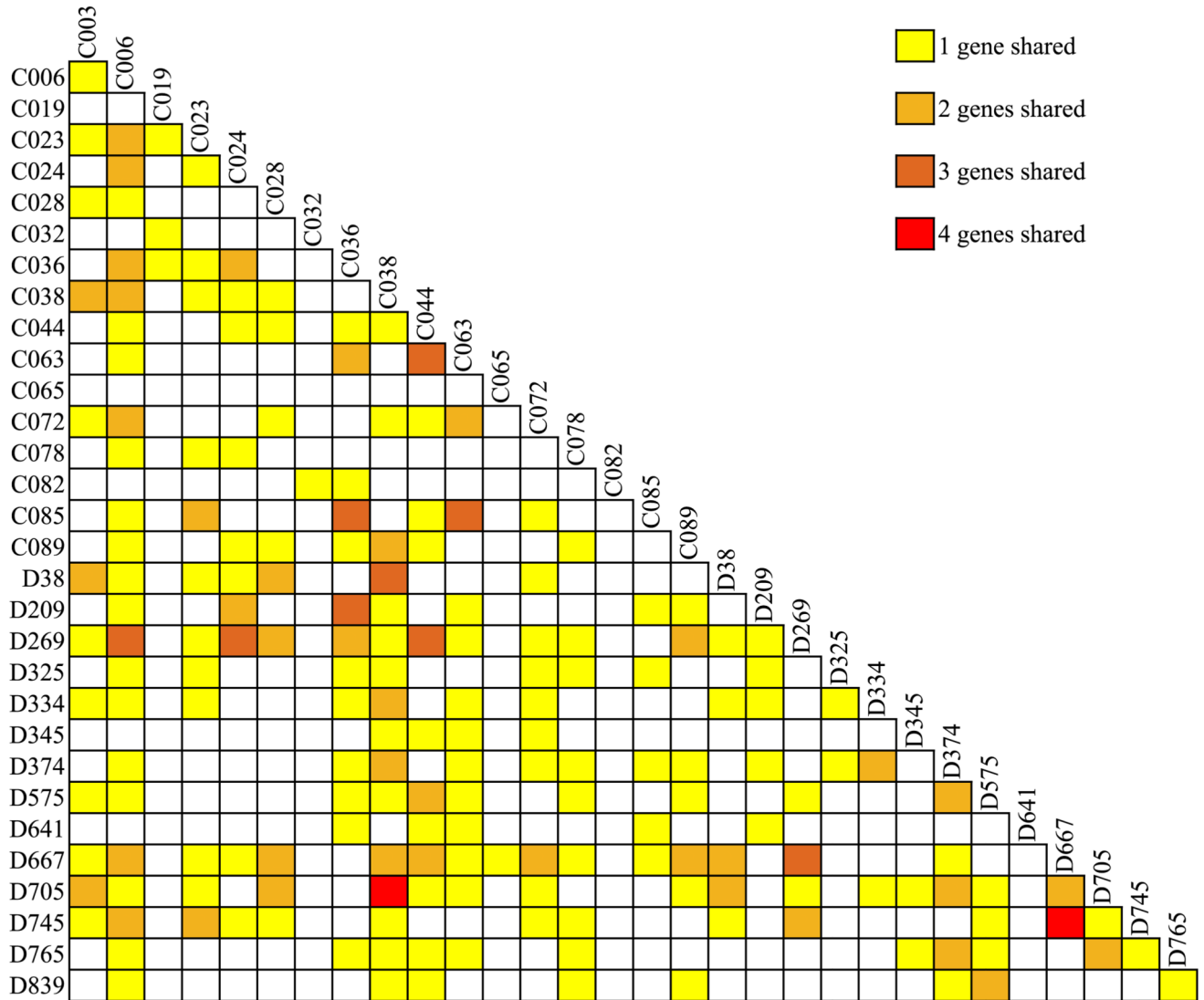


Fig. 2. Distribution of variant-containing genes between patients. The number of variant-containing genes from Tables S3–S7 shared between patients.

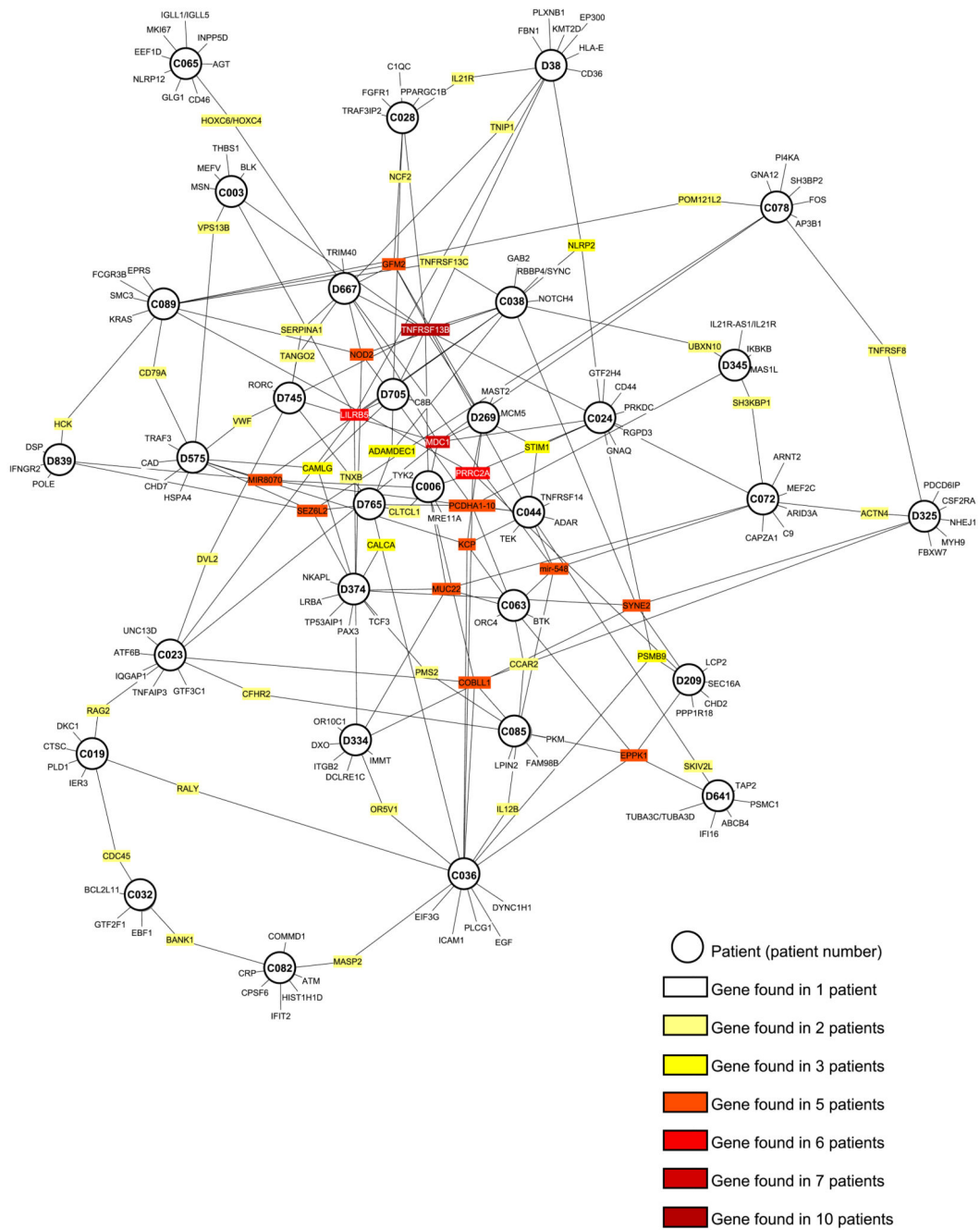


Fig. 3. Visualisation of the overlap of variant containing genes between patients. The distribution of variant-containing genes (small nodes) listed in Tables S3–S7 between patients (big nodes).

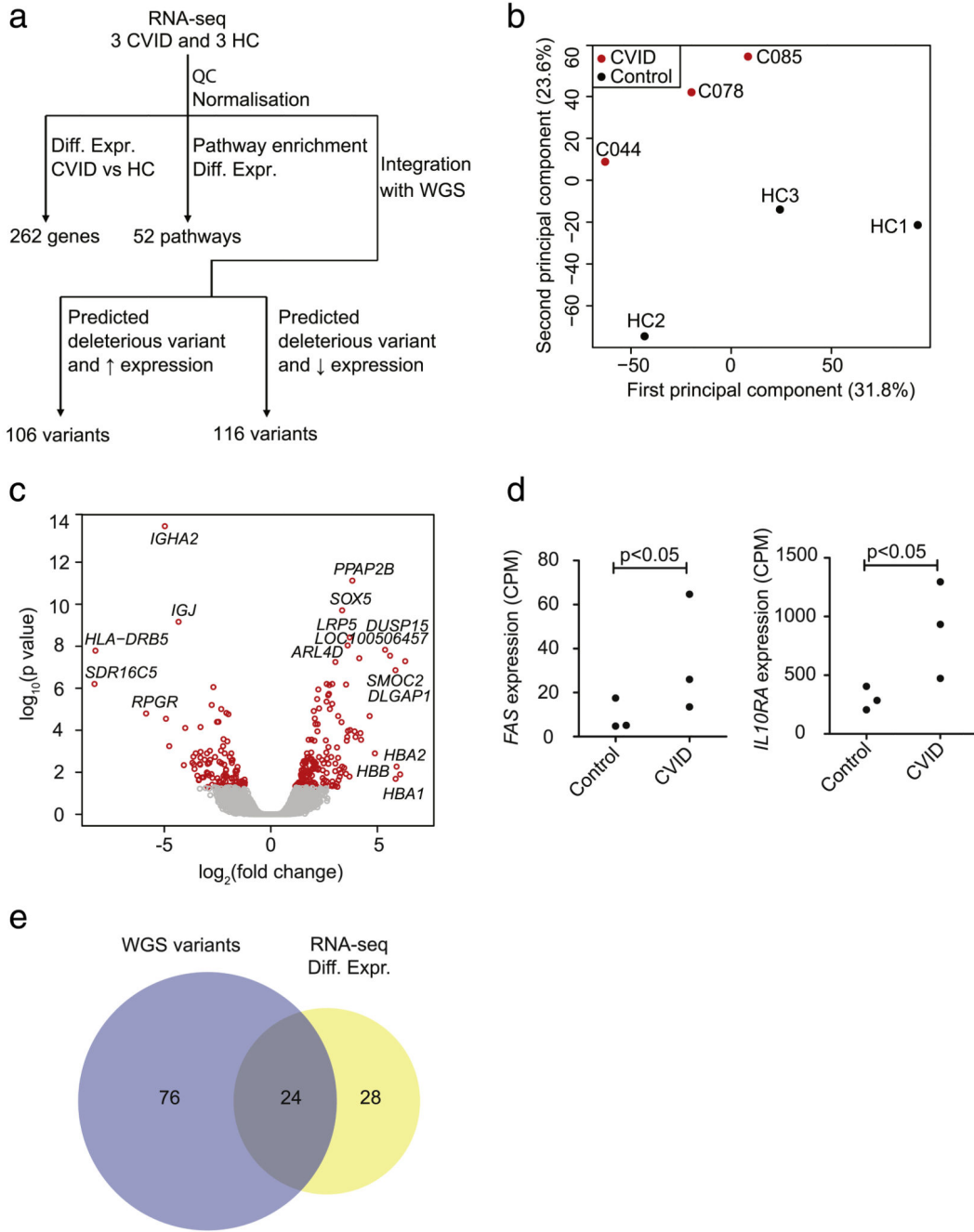


Fig. 4. Overview of differentially expressed genes and overlap of pathway analysis of WGS and RNA-seq data. (a) Overview of the analysis strategy applied to the RNA-seq data. QC, quality control; Diff. Expr., differentially expressed; HC, healthy controls. (b) The first two principal components are plotted with the proportion of variance explained by each component (based on reads which map to genes). (c) Volcano plot of differentially expressed genes. The significance of differential expression between CVID patients and healthy controls is plotted against fold change. Each point represents a single gene. Red

points highlight genes with an adjusted p value < 0.05 and fold change > 2 . A positive fold change indicates upregulation in the CVID patients. Some of the most differentially expressed genes are named. (d) Differential expression of *FAS* and *IL10RA* between CVID patients and healthy controls. (e) The 100 most significantly enriched pathways for WGS predicted deleterious variants were reported by IVA. Fifty-two pathways were enriched for differentially expressed genes from the RNA-seq analysis. The overlap of enriched pathways between these two datasets is greater than expected by chance (Fisher's exact test p value < 0.0001).

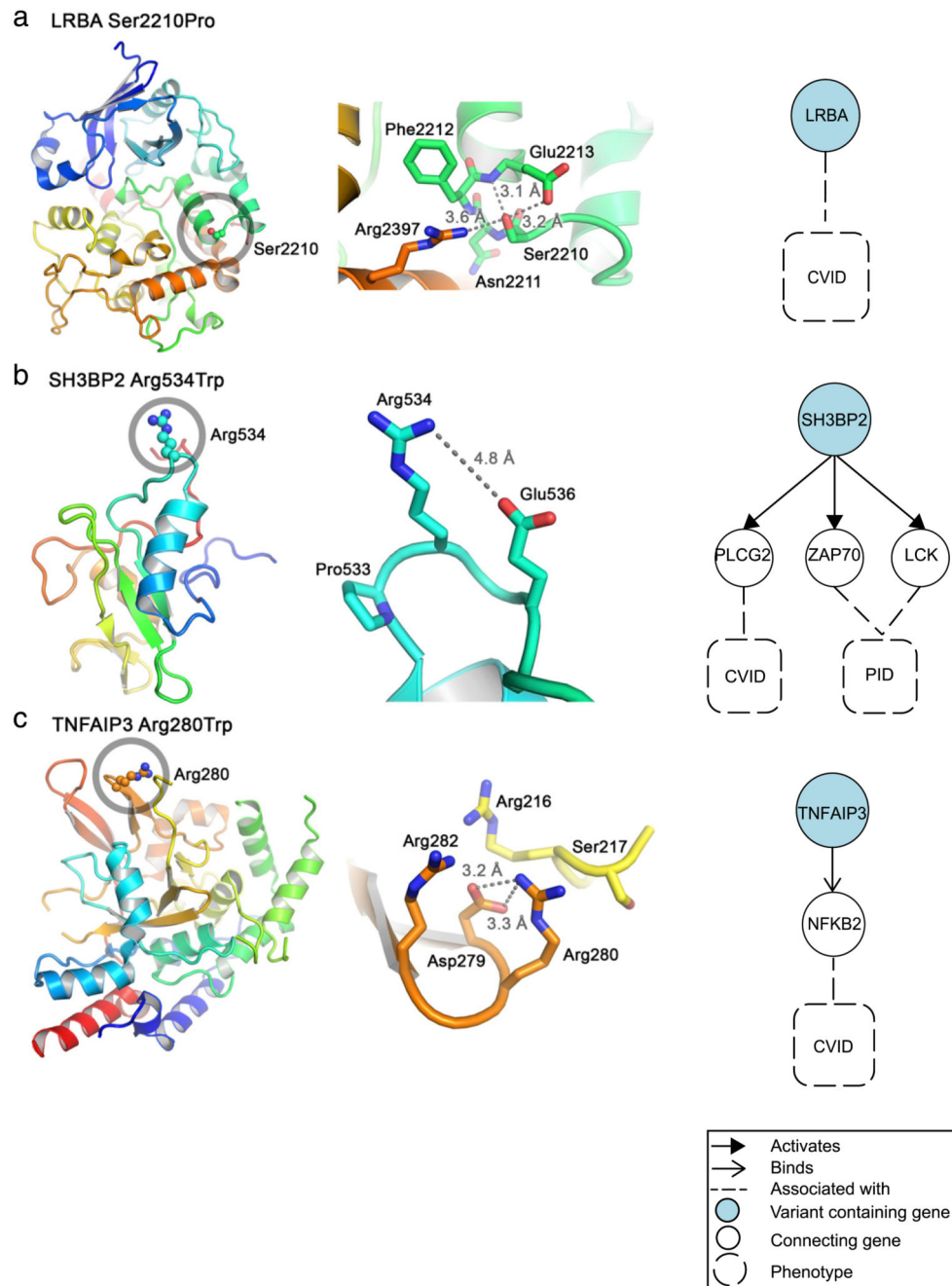


Fig. 5. Mapping of predicted deleterious variants onto protein structures and linking proteins of interest to CVID. Proteins are shown as ribbon diagrams and coloured in blue-to-red transitions (N-termini, blue; C-termini, red). Side chains of selected variants are shown as connected spheres (oxygen, red; nitrogen, blue) and are circled in the panels on the left. Zoom-in views are presented in the panels on the middle with side chains shown as sticks. Distances between selected atoms are indicated with dashed lines. The right panel shows how these proteins could be linked to CVID. (a) The crystal structure of the C-terminal

fragment of human LRBA encompassing the pleckstrin homology (PH, blue) and the beige and Chediak–Higashi syndrome (BEACH, green-to-red) domains (PDB code 1T77) [69]. The side chain of Ser2210 makes hydrogen bonds to Glu2213 (both main chain and side chain) and the side chain of Arg2397. Amino acid numbering corresponds to UniProt entry P50851, isoform 2. (b) Solution NMR structure of the SH2 domain of SH3BP2 (PDB 2CR4; UniProt P78314, isoform 4). The side chain of Arg534 is exposed on the protein surface and does not make any bonds with other residues. (c) The crystal structure of the catalytic domain of TNFAIP3 (PDB 3ZJE; UniProt P21580) [70]. The side chain of Arg280 is exposed on the surface and makes hydrogen bonds to the side chain of Asp279.

Table 1

Overview of clinical information on the 34 COVID patients.

Patient no	Sex	Age at onset of symptoms	CD19 count ^d 0.1–0.5	CD3 count ^d 0.7–2.1	CD4 count ^d 0.3–1.4	CD8 count ^d 0.2–0.9	IgG levels ^d 6.0–16.0	IgM levels ^d 0.4–2.5	IgA levels ^d 0.8–3.0	Clinical phenotype ³
Normal										
C003	F	54	↓↓	↓	Normal	Normal	↓↓	Normal	↓↓	Polyclonal lymphoproliferation
C006	M	34	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	Polyclonal lymphoproliferation
C018	M	8	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	No disease-related complications
C019	M	18	Normal	Normal	Normal	↑	↓↓	↓↓	↓↓	No disease-related complications
C023	M	25	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	No disease-related complications
C024	F	48	Normal	↓	Normal	↓	↓↓	↓↓	↓↓	No disease-related complications
C028	F	47	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	Polyclonal lymphoproliferation
C032	M	26	↓	Normal	↓	↓	Normal	Normal	↓↓	Polyclonal lymphoproliferation
C036	F	19	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	No disease-related complications
C038	M	5	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	No disease-related complications
C044	F	31	Normal	Normal	Normal	Normal	↓	↓↓	↓	Polyclonal lymphoproliferation; autoimmune cytopenias
C063	F	48	Normal	Normal	Normal	Normal	↓↓	Normal	↓↓	No disease-related complications
C065	M	5	Normal	Normal	Normal	Normal	↓	↓	↓↓	No disease-related complications
C072	F	34	↓↓	Normal	Normal	Normal	↓	↓↓↓	↓↓	No disease-related complications
C078	M	13	Normal	Normal	Normal	Normal	↓↓	Normal	↓↓	No disease-related complications
C082	F	11	↓↓	Normal	Normal	Normal	↓↓	↓↓	↓↓ ^b	Autoimmune cytopenias
C085	F	37	Normal	Normal	Normal	Normal	↓	↓↓	↓↓	No disease-related complications
C089	F	17	↓↓	Normal	Normal	Normal	↓	↓	↓↓	No disease-related complications
D038	M	22	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	No disease-related complications
D209	M	41	Normal	Normal	Normal	Normal	↓↓	↓	↓↓	No disease-related complications
D232	F	44	↑	↑	↑	Normal	Normal	Normal	↓↓	No disease-related complications
D269	F	44	Normal	↓	↓	Normal	Normal	Normal	↓↓	Polyclonal lymphoproliferation; autoimmune cytopenias
D276	M	1	↓↓↓	Normal	Normal	Normal	↓	↓	↓↓	No disease-related complications
D325	F	35	Normal	↓	Normal	Normal	↓	↓	↓↓	Polyclonal lymphoproliferation; autoimmune cytopenias
D334	M	15	Normal	Normal	Normal	Normal	↓	↓	↓↓	No disease-related complications

Patient no	Sex	Age at onset of symptoms	CD19 count ^a	CD3 count ^a	CD4 count ^a	CD8 count ^a	IgG levels ^a	IgM levels ^a	IgA levels ^a	Clinical phenotype ^b
Normal			0.1–0.5	0.7–2.1	0.3–1.4	0.2–0.9	6.0–16.0	0.4–2.5	0.8–3.0	
D345	F	20	Normal	↑	Normal	↑	↓↓	↓	↓↓	No disease-related complications
D374	M	3	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	No disease-related complications
D575	M	2	Normal	Normal	Normal	Normal	↓↓	↓↓	↓↓	No disease-related complications
D641	F	20	Normal	↓	↓	Normal	↓↓	↓↓↓	↓↓	No disease-related complications
D667	M	13	Normal	Normal	Normal	Normal	↓↓	↓	↓↓	No disease-related complications
D705	F	49	Normal	Normal	Normal	Normal	↓↓↓	↓↓↓	Normal/↓	No disease-related complications
D745	M	5	Normal	↑	↑	↑	↓↓	Normal	↓↓	No disease-related complications
D765	F	30	Normal	Normal	Normal	Normal	↓↓	Normal	↓↓	No disease-related complications
D839	M	25	Normal	↓	Normal	Normal	↓↓	↓	↓↓	Polyclonal lymphoproliferation

↓ = between lower normal and half lower normal values; ↓↓ = less than half lower normal value; ↓↓↓ = undetectable; ↑ = between upper normal value and twice upper normal value; ↑↑ = more than twice upper normal value.

^aValues at diagnosis.

^bValues at diagnosis not available; values listed are post-infusion of intravenous immunoglobulin, which does not alter IgA levels.

Table 2

Additional details on variants highlighted in the text.

Chromosome: position	Gene symbol	Transcript variant	Protein variant	dbSNP ID	Cases samples with variant (freq)	1000 Genomes frequency	No. WGS50 (freq)	SIFT function prediction	SIFT score	Pol-Phen-2 function prediction	Conservation annot p-value	Samples
<i>Described variants</i>												
17:16842991	<i>TNFRSF13B</i>	c.752C>T	p.P251L	34562254	7 (20.59)	18.07	47.0 (19.67)	Tolerated	0.69	Benign		C006, C028, C038, C072, D667, D705, D745 D269 (hom)
17:16843084	<i>TNFRSF13B</i>	c.659T>C	p.V220A	56063729	1 (2.94)	1.03	10.0 (4.18)	Tolerated	0.86	Benign		D38
17:16843666	<i>TNFRSF13B</i>	c.605G>A	p.R202H	104894649	1 (2.94)	0.32	2.0 (0.84)	Damaging	0.02	Benign	2.90E-00	D38
17:16852187	<i>TNFRSF13B</i>	c.310T>C	p.C104R	34857412	1 (2.94)	0.32	5.0 (2.09)	Damaging	0	Probably damaging		D38
17:16852282	<i>TNFRSF13B</i>	c.215G>A	p.R72H	55916807	1 (2.94)	0.32	3.0 (1.26)	Tolerated	0.29	Benign		C003
19:54314003	<i>NLRP12</i>	c.910C>T	p.H304Y	141245482	1 (2.94)	0.27	1.0 (0.42)	Damaging	0.04	Probably damaging		C065
22:42322716	<i>TNFRSF13C</i>	c.62C>G	p.P21R	77874543	2 (5.88)	5.59	22.0 (9.21)	Tolerated	0.64	Benign		C038, C089 (Hom)
<i>GWAS</i>												
6:30671229	<i>MDC1</i>	c.5648G>A	p.R188Q	28994875	2 (5.88)	1.14	5 (2.09)	Damaging	0.03	Probably damaging	5.25E-04	C023, C078
6:30671726	<i>MDC1</i>	c.5234C>G	p.P1745R	28994871	5 (14.71)	1.37	9 (3.77)	Damaging	0.04	Probably damaging		C006, C024, D269 (hom), D667, D745
6:30673064	<i>MDC1</i>	c.3896G>A	p.R1299Q	144657716	5 (14.71)	3.60	9 (3.77)	Activating	1	Benign		C006, C024, D269 (hom), D667, D745
6:30673625	<i>MDC1</i>	c.3335C>T	p.S112F	28987085	5 (14.71)	1.37	9 (3.77)	Damaging	0.01	Possibly damaging		C006, C024, D269 (hom), D667, D745
6:31598523	<i>PRRC2A</i>	c.2410C>T	p.R804C	11538262	1 (2.94)	0.74	7 (2.93)	Damaging	0.01	Probably damaging	3.89E-03	C089
6:31599916	<i>PRRC2A</i>	c.3466C>T	p.R1156W	20057972	1 (2.94)	0.74	7 (2.93)	Damaging	0.01	Probably damaging	6.38E-04	C036
6:31600639	<i>PRRC2A</i>	c.4189C>T	p.R1397W	201074309	1 (2.94)	0.74	7 (2.93)	Damaging	0.01	Probably damaging	3.05E-03	D209
6:31605016	<i>PRRC2A</i>	c.6248T>C	p.F2083S	35595439	3 (8.82)	0.51	7 (2.93)	Damaging	0.01	Possibly damaging	8.13E-04	C006, C024, D269 (hom)
6:31929737	<i>SKA2L</i>	c.970C>T	p.R324W	36038685	1 (2.94)	1.10	4 (1.67)	Damaging	0.01	Possibly damaging	5.46E-06	D641
6:31931903	<i>SKA2L</i>	c.1860+L1860+10delGTGGCTCTGT			1 (2.94)	1.10	4 (1.67)	Damaging	0.01	Possibly damaging	5.46E-06	C044
<i>PID genes</i>												
2:22022915	<i>NHEJ1</i>	c.170G>A	p.R57Q	61755339	1 (2.94)	1.03	1.0 (0.42)	Damaging	0	Probably damaging	4.54E-02	D325
4:15192815*	<i>LRBA</i>	c.666T>C; c.662KT>C	p.S221F; p.S221Q		1 (2.94)	0.43	9.0 (3.77)	Damaging	0	Probably damaging	7.55E-04	D374
6:32803458	<i>TAP2</i>	c.701T>A	p.L234Q	138708621	1 (2.94)	0.43	9.0 (3.77)	Damaging	0.02	Probably damaging		D641
7:6045634	<i>PM52</i>	c.52A>G	p.I18V	63750123	2 (5.88)	0.43	9.0 (3.77)	Damaging	0	Probably damaging	8.00E-03	C085, D374
8:42176165	<i>IKZF8</i>	c.1330C>T; c.1336C>T; c.1159C>T	p.R444W; p.R387W; p.R446W	202136671	1 (2.94)	0.43	9.0 (3.77)	Damaging	0.01	Possibly damaging	7.91E-01	D345

van Schouwenburg et al.

Chromosome: position	Gene symbol	Transcript variant	Protein variant	dbSNP ID	Case samples with variant (freq)	1000 Genomes frequency	No. WGS500 (freq)	SIFT function prediction	SIFT score	PolyPhen-2 function prediction	Conservation phyloP p-value	Samples
8:48691175	<i>PRKDC</i>	c.11698C>G; c.11695C>G	p.L3869Y; p.L3900V	201214138	1 (2.94)			Probably damaging		Probably damaging	3.27E+00	C024
8:48775055*	<i>PRKDC</i>	c.5818T>C	p.Y1940H		1 (2.94)			Probably damaging		Probably damaging	1.31E-02	D232
10:14965056	<i>DCLRE1C</i>	c.640T>A; c.985T>A; c.625T>A	p.L214M; p.L329M; p.L209M	41299658	1 (2.94)	0.05	1.0 (0.42)	Damaging	0.01	Probably damaging	3.91E-03	D334 C024, C063, D232, D334
11:4108992	<i>STIM1</i>	c.1859+1G>A; c.1541+319G>A		118128831	4 (11.76)	0.51	9.0 (3.77)	Damaging	0.02	Probably damaging		C019, C082, C089
11:56614561	<i>RAG2</i>	c.1158C>A	p.F368L	34629171	2 (5.88)	0.37	8.0 (3.35)	Damaging		Probably damaging	9.66E-01	C006
11:94178974*	<i>MRE11A</i>	c.1783+1411T>C; c.1867+2T>C		1800057	1 (2.94)	1.51	11.0 (4.6)	Damaging	0.01	Probably damaging	1.33E-02	C082
11:108143456	<i>ATM</i>	c.3161C>G	p.P1054R		1 (2.94)			Damaging		Probably damaging		C089
12:25368405	<i>KRAS</i>	c.451-5560T>A; c.540T>A	p.C180*	373169526	1 (2.94)			Damaging	0	Probably damaging	8.51E+00	D839
12:133252342*	<i>POLE</i>	c.1085A>G	p.Y362C		1 (2.94)			Damaging		Probably damaging	7.41E-03	C038, D38
16:27414502	<i>IL21R</i>	c.116G>A; c.-17+565G>A	p.R41H	117535117	2 (5.88)	0.65	4.0 (1.67)	Damaging	0	Possibly damaging		D345
16:27460561	<i>IL21R-AS1; IL21R</i>	c.1574T>A; c.1640T>A	p.V525D; p.V547D	200674281	1 (2.94)	0.63	2.0 (0.84)	Damaging	0.01	Possibly damaging	5.18E+00	D374
19:1627422	<i>TCF3</i>	c.302A>G	p.K101R	41275842	1 (2.94)	0.31	1.0 (0.42)	Damaging	0.02	Probably damaging	3.30E+00	C089, D575
19:42383644	<i>CD79A</i>	c.308C>A; c.419C>A	p.T102N; p.T140N	148797987	2 (5.88)			Damaging		Probably damaging		C024, C088, D38
19:55494283	<i>NLRP2</i>	c.1151C>G; c.1148C>G; c.1217C>G	p.T383R; p.T406R; p.T384R	139905347	3 (8.82)	0.34	2.0 (0.84)	Damaging	0.01	Probably damaging	3.32E-01	D276 (hem1)
X:100609671*	<i>BTX</i>	c.1049dupA; c.1577dupA; c.1679dupA	p.N530I6*11; p.N526I8*11; p.N560I8*11		1 (2.94)							C063
X:100611132*	<i>BTX</i>	c.1474C>T; c.1039-1430C>T; c.1576C>T	p.R492C; p.R526C		1 (2.94)			Damaging	0.05	Possibly damaging		C019 (hem1)
X:154005089*	<i>DKC1</i>	c.1495_1497delAAG; c.*718_720delAAG; c.1510_1512delAAG	p.K505del; p.K500del		1 (2.94)		2.0 (0.84)					
Most common variants												
19:54756246	<i>LILRB5</i>	c.1556T>G; c.1259T>C; c.1559T>C	p.H420T; p.L519T; p.L520T	117421142	6 (19.35)	0.87	9 (3.77)	Activating	1	Benign		C003, C023, C038, D38, D334, D705
<i>I hop genes</i>												
1:161599778	<i>FCGR3B</i>	c.214G>A; c.217G>A; c.109G>A; c.58G>A	p.V73M; p.V37M; p.V72M; p.V20M	375557751	1 (3.23)		0	Damaging	0	Probably damaging		C089
2:11923737	<i>BCL2L1</i>	c.*2124C>T; c.*2139C>T; c.*2014C>T; c.*2151C>T; c.*1920C>T; c.*2141C>T; c.*2195C>T; c.*2144C>T		144847549	1 (3.23)	0.61	(0)				9.10E-03	C032
2:234113242	<i>INPP5D</i>	c.3443C>G; c.3446C>G	p.A1149G; p.A1148G	375825105	1 (3.23)		0			Possibly damaging	2.90E-04	C065
4:2834080	<i>SH3BP2</i>	c.1513C>T; c.1600C>T; c.1429C>T	p.R534W; p.R477W; p.R505W	148761331	1 (3.23)	0.28	3 (1.26)	Damaging	0.01	Possibly damaging		C078
4:102751014	<i>BANK1</i>	c.1206>C; c.306C>C; c.71-25178G>C	p.W40C; p.W10C	35978636	2 (6.45)	0.55	8 (3.35)	Damaging	0	Probably damaging	4.55E-05	C032, C082
5:134086744	<i>CAMLG</i>	c.*104C>A		11552196	3 (9.68)	0.8	5 (2.09)					D374, D575, D705
5:150422217*	<i>TNPI1</i>	c.1018C>G; c.839C>G	p.Q238T; p.Q340E		1 (3.23)		0	Damaging	0.05	Probably damaging	3.30E-06	D38
5:150441774*	<i>TNPI1</i>	c.113-1G>T; c.272-1G>T			1 (3.23)		0					D667
5:158124754*	<i>EBF1</i>	c.*1364_1365insT			1 (3.23)		1 (0.42)					C032
6:32823948	<i>PSMB9</i>	c.94G>A	p.V32I	241419	1 (3.23)	2.19	10 (4.18)	Damaging	0		7.98E-06	D209
6:32826267	<i>PSMB9</i>	c.517C>T	p.R173C	17213861	2 (6.45)	0.14	4 (1.67)	Damaging	0.01			C024, C063

Chromosome: position	Gene symbol	Transcript variant	Protein variant	dbSNP ID	Case samples with variant (freq)	1000 Genomes frequency	No. WGS500 (freq)	SIFT function prediction	SIFT score	PolyPhen-2 function prediction	Conservation phyloP p-value	Samples
6:138198245	<i>TNFAIP3</i>	c.838C>T	p.R280W	150198888	1 (3.23)	0.04	0	Damaging	0	Possibly damaging		C023
8:11418861*	<i>BLK</i>	c.1080C>A	p.D360E		1 (3.23)		0	Damaging	0	Possibly damaging	2.13E-04	C003
11:77937972	<i>GAB2</i>	c.746A>G; c.632A>G	p.Y249C; p.Y211C	146509778	1 (3.23)		1 (0.42)	Damaging	0.03	Possibly damaging	1.69E-05	C038
19:932497*	<i>ARID3A</i>	c.460_462delGAG	p.E154del		1 (3.23)		0					C072

The complete results of the different analyses can be found in Tables S3–S7. Variants are heterozygous unless the sample number is followed by (hom) which indicates a homozygous variant or (hemi) in which case the variant is hemizygous. Novel variants are marked with *. No. patients WGS500 indicates the number of patients in the WGS500 cohort (excluding COVID) with the variant. All variants have been confirmed in IGV.

Table 3

Overview of basic filtering applied in the majority of the analyses.

Filter	Thresholds	Rationale	Number of variants in 34 patients (number of genes)	Number of variants in 31 patients (number of genes)
Confidence	Call quality is at least 20 in any sample Variant passed upstream pipeline filtering Read depth is at least 10 in any sample Variant is outside top 0.2% most exomically variable 100 base windows in healthy public genomes	Exclude poor quality variant calls	14,819,871 (20,891)	14,457,884 (20,627)
No. reverse reads	1 reads containing variant	Reduce strand bias	11,118,902 (20,583)	10,859,616 (20,564)
No. forward reads	1 reads containing variant	Reduce strand bias	10,940,142 (20,571)	10,687,781 (20,554)
Common variants	Variant has an allele frequency less than 5% in 1000 Genomes Project	Exclude common variants	4,808,176 (20,132)	4,582,532 (20,080)
WGS500 MAF	Variant has an allele frequency less than 5% in non-CVID, non-cancer WGS500 samples	Exclude technical artefacts	2,721,014 (19,678)	2,537,263 (19,584)
Predicted deleterious	Pathogenic/likely pathogenic or associated with any type of gain of function or associated with loss of function (frameshift, insertions/deletions, changes in start of stop codon, missense mutations which are predicted to not be tolerated by SIFT or PolyPhen, splice site loss up to 2 bp into the intron, deleterious to a micro RNA or variants leading to a structural variants)	Exclude variants which are unlikely to have effect on gene function	4768 (3737)	4422 (3524)

Details on the settings used and the rationale behind using these filters and the number of variants left after filtering in either 31 or 34 patients are indicated.

Table 4

Overlap of enriched pathways between WGS and RNA-seq. Overview of the 24 pathways found in both the 100 most significantly enriched pathways for WGS and the pathways significantly enriched for differentially expressed genes from the RNA-seq analysis.

Pathway	p value RNA-seq	p value WGS
Allograft rejection signalling	2.14E-03	4.74E-08
CD28 signalling in T helper cells	8.91E-03	1.01E-06
Haematopoiesis from pluripotent stem cells	1.82E-02	1.45E-06
Type I diabetes mellitus signalling	6.61E-03	3.90E-06
Graft-versus-host disease signalling	1.38E-02	7.66E-06
Autoimmune thyroid disease signalling	1.55E-04	8.86E-06
Dendritic cell maturation	2.69E-05	2.31E-05
Role of osteoblasts, osteoclasts and chondrocytes in rheumatoid arthritis	2.95E-03	3.83E-05
Altered T cell and B cell signalling in rheumatoid arthritis	1.45E-02	5.36E-05
iCOS-iCOSL signalling in T helper cells	2.95E-02	5.61E-05
Communication between innate and adaptive immune cells	3.24E-03	5.88E-05
Systemic lupus erythematosus signalling	1.12E-02	6.39E-05
Role of NFAT in regulation of the immune response	5.89E-04	8.62E-05
T helper cell differentiation	9.33E-04	2.47E-04
PKC θ , signalling in T lymphocytes	8.91E-03	8.02E-04
IL-6 signalling	3.80E-02	8.83E-04
Role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis	1.91E-02	1.06E-03
IL-10 signalling	8.71E-04	1.31E-03
Tec kinase signalling	7.41E-03	1.89E-03
Type II diabetes mellitus signalling	3.80E-02	2.28E-03
Death receptor signalling	2.57E-02	3.78E-03
LPS-stimulated MAPK signalling	4.57E-02	4.84E-03
PEDF signalling	4.27E-02	1.34E-02
RANK signalling in osteoclasts	1.51E-02	1.80E-02