1

Proteome analysis of human neutrophil granulocytes from patients with monogenic disease using data-independent acquisition

Running title: DIA proteomics aids clinical genetic diagnostics

Piotr Grabowski[1#], Sebastian Hesse[2#] , Sebastian Hollizeck[2], Meino Rohlfs[2], Uta Behrends[4], Roya Sherkat[5], Hannah Tamary[6], Ekrem Ünal[7], Raz Somech[8], Türkan Patıroğlu[7], Stefan Canzar[9], Jutte van der Werff Ten Bosch[10], Christoph Klein[2#*], Juri Rappsilber[1,3#*]

[1] Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany

[2] Dr. von Hauner Children's Hospital, Department of Pediatrics, University Hospital, LMU Munich, 80337 Munich, Germany

[3] Wellcome Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, UK

[4] Children's Hospital, Hematology-Oncology, Technical University Munich, 80804 Munich, Germany

[5] Acquired Immunodeficiency Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

[6] Schneider Children's Medical Center of Israel, Petah Tikva, Sackler School of Medicine, Tel Aviv University, Israel

[7] Department of Pediatrics, Division of Pediatric Hematology & Oncology, Erciyes University, Kayseri, Turkey

[8] Pediatric Department A and Immunology Service, Jeffrey Modell Foundation Center, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer, Sackler Faculty of Medicine, Tel Aviv University

[9] Gene Center, Ludwig-Maximilians-Universität München, Munich, Germany

[10] Department of Pediatric Hematology Oncology, UZ Brussel, Brussels, Belgium

[*] Communicating authors

[#] Shared first/last authorship

Corresponding authors:
Juri Rappsilber, juri.rappsilber@ed.ac.uk
Christoph Klein, christoph.klein@med.uni-muenchen.de

Summary

Neutrophil granulocytes are critical mediators of innate immunity and tissue regeneration. Rare diseases of neutrophil granulocytes may affect their differentiation and/or functions. However, there are very few validated diagnostic tests assessing the functions of neutrophil granulocytes in these diseases. Here, we set out to probe omics analysis as a novel diagnostic platform for patients with defective differentiation and function of neutrophil granulocytes. We analyzed highly purified neutrophil granulocytes from 68 healthy individuals and 16 patients with rare monogenic diseases. Cells were isolated from fresh venous blood (purity >99%) and used to create a spectral library covering almost 8000 proteins using strong cation exchange fractionation. Patient neutrophil samples were then analyzed by data-independent acquisition proteomics, quantifying 4154 proteins in each sample. Neutrophils with mutations in the neutrophil elastase gene *ELANE*, showed large proteome changes that suggest these mutations may affect maturation of neutrophil granulocytes and initiate misfolded protein response and cellular stress mechanisms. In contrast, only few proteins changed in patients with leukocyte adhesion deficiency (LAD) and chronic granulomatous disease (CGD). Strikingly, neutrophil transcriptome analysis showed no correlation with its proteome. In case of two patients with undetermined genetic causes, proteome analysis guided the targeted genetic diagnostics and uncovered the underlying genomic mutations. Data-independent acquisition proteomics may help to define novel pathomechanisms in neutrophil diseases and provide a clinically useful diagnostic dimension.

Introduction

Neutrophil granulocytes constitute the most abundant population of nucleated cells in human blood. Whereas their role in defense against microbes has been known for more than a century, their sophisticated roles in tissue remodeling, cancer and chronic inflammation has emerged only recently (1, 2).

Rare diseases of neutrophil granulocytes may affect their differentiation and/or function. Severe congenital neutropenia (SCN) comprises a heterogeneous group of monogenic disorders characterized by aberrant premature apoptosis of myeloid progenitor cells (3). Chronic granulomatous disease (CGD) results from mutations in one out of five subunits of the NADP-oxidase complex (4). Leukocyte adhesion deficiency (LAD) causes aberrant transendothelial migration of neutrophil granulocytes due to integrin defects (5).

In striking contrast to their prominent role in health and disease, there are very few validated diagnostic tests assessing the function of neutrophil granulocytes. Whereas quantitative studies (i.e. differential blood counts) are among the most common laboratory tests, validated qualitative studies are virtually absent, with the exception of measurement of NADPH-oxidase activity for CGD and expression of defined cell surface markers to diagnose LAD.

Genetic sequencing assays, based on defined panels, exome, or whole genome sequencing are the gold standard for the diagnosis of monogenic diseases, yielding conclusive results in up to 25-50% of patients (6, 7). To capture the complexities of diseases and to increase the diagnostic yield, additional omics technologies and data integration may be useful. This has recently been shown for strategies associating genome and RNA-sequencing in rare neuromuscular diseases (8, 9). We hypothesized that in-depth proteome-analysis may also provide additional cues for diagnosis of monogenic diseases, such as rare defects of neutrophil granulocytes. In the past, several laboratories have published data on the proteome of healthy neutrophil granulocytes (10–23), however all but one (10) quantified significantly fewer proteins compared to our study. We here systematically analyze proteome changes in neutrophils from patients with different monogenic diseases and demonstrate the usefulness of next-generation proteomics for guiding clinical genetic diagnostics.

Results

Proteome analysis of high-purity neutrophil samples

First, we set up an experimental flow allowing for consistent, robust, and reliable analysis of purified neutrophil granulocytes. Using magnetic-bead based negative selection we isolated neutrophil granulocytes with >99% purity from 16 patients with quantitative or qualitative defects (Supplemental Table 1) and 68 healthy control individuals (Fig. 1a-b). While 14/16 patients had a genetically confirmed molecular diagnosis of neutrophil deficiency (CGD: n = 5, SCN: n = 6, LAD: n = 3), in two patients (one with NADPH-oxidase deficiency, one with congenital neutropenia associated with albinism) routine genetic workup by exome sequencing did not yield a specific defect in any gene known to cause congenital neutropenia or CGD. A set of 4154 proteins was quantified across all samples using data-independent acquisition (DIA) mass spectrometry and neutrophil-specific spectral

libraries generated by us. The estimated protein copy numbers per cell using the proteomic ruler approach (24) and the full list of quantified proteins in each sample are reported in the Supplemental Table 2. To estimate the variability in protein expression measured by our approach, we compared proteomes of the same healthy individual sampled over three different timepoints spanning several months. As expected from DIA proteomics (25), we observed a high level of agreement between replicates with a mean Pearson correlation coefficient of ~0.87 (Fig. 1c). After correcting for batch effects related to sample processing using the ComBat algorithm (26), we looked at the separation of neutrophil proteomes of healthy donors and patients using principal component analysis (PCA) (Fig. 1d). The groups were clearly separated along the two first principal components, explaining in total ~17% of the total variability in the data.

Neutrophil mRNA levels are not well correlated with protein levels

Next, we looked at correlation of protein expression with expression levels of the respective mRNAs in the healthy naive neutrophils. We sequenced the transcriptomes of 22 healthy donor neutrophil samples and compared cumulative protein and protein-coding transcriptome abundances (Fig. 2a). Interestingly, only seven proteins accounted for 50% of the total protein copy number. This suggests that the neutrophil proteome is dominated by a very few highly expressed proteins, with the antimicrobial proteins DEF3A, S100A8, LYZ and CTSG being among the most abundant. A similar dynamic range can be seen for the protein-coding transcriptome (top three abundant transcripts: MYCBP, KIF2C, ATP6V0B). However, the top expressed protein-coding transcripts did not correspond to the top expressed proteins (Fig. 2b-c). The healthy donor neutrophil mRNA abundances were reported in Supplemental Table 3.

We found no correlation between mRNA and protein levels (Pearson correlation coefficient of 0.00; Fig. 2d). This result is in sharp contrast to correlation coefficients obtained in other cells, which typically range from ~0.4-0.8 (27, 28) and agrees with previous observations (13, 16, 29). Naive neutrophils have been shown previously to have reduced transcriptional activity (30). It is therefore possible that the naive neutrophil transcriptome is not reflective of its function, but rather prepared to initialize gene expression programs upon activation, as suggested by (30).

For this reason, we focused on in-depth analysis of proteome rather than transcriptome data.

Differential protein expression in neutrophil granulocytes from patients with monogenic diseases

We asked the question to what extent the composition of the proteome differed between patients with severe congenital neutropenia due to mutations in *ELANE* and healthy individuals. As shown in Fig. 3a, the proteome of *ELANE*-mutated neutrophil granulocytes showed 71 significantly underexpressed and 159 overexpressed proteins. As expected, ELANE was one of the most underexpressed proteins (4.5-fold). In addition, other primary granule antimicrobial proteins, such as azurocidin (AZU1, 4.5-fold) and cAMP receptor protein (CAMP, 5-fold) were also markedly underexpressed. In the group of significantly overexpressed proteins, we identified several endoplasmic reticulum (ER) heat-shock proteins (DNAJB1, HSP90B, HSPA5, HYOU1, SDF2L1, PDIA4, MANF), indicative of

proteostatic stress response. Another prominent group of significantly overexpressed proteins were ribosomal proteins (RPL9, RPL10, RPL11, RPL19, RPL21, RPL22, RPL31, RPL32, RPL38, RPS23) and mitochondrial proteins (SPTLC2, ABCB10, EIF2A. LRPPRC, CYCS, SLC25A1, LONP1, PUS1, LETM1, TOMM22, GRPEL1, HSPA9, SHMT2). Several transcription factors were differentially expressed (decreased levels: TCEAL3, ZNF22, increased levels: LRPPRC, SND1). It is currently not known whether this reflects differences in maturation stages or a specific response to the disturbed subcellular environment in *ELANE*-mutated neutrophil granulocytes. Fig. 3b shows the relative abundance of the top 6 proteins with the most significantly changed expression values. The most enriched Gene Ontology (GO) Biological Process (BP) terms among the upregulated proteins were "translational initiation" (adj. P-value: $9.8*10^{-11}$), "nuclear-transcribed mRNA catabolic process, nonsense-mediated decay" (adj. P-value: $2.8*10^{-5}$) and "protein folding" (adj. P-value: $8.3*10^{-4}$). The most significant BP term enriched among down-regulated proteins was "proteolysis" (adj. P-value: $5.6*10^{-2}$), a term which ecompasses the neutrophil antimicrobial proteins like ELANE and AZU1. In contrast to the proteome of *ELANE*-mutated neutrophil granulocytes, the analysis of neutrophil granulocytes from CGD patients (n=5) revealed less striking differences when compared to healthy individuals (Fig. 3c). Interestingly, the two proteins with the lowest expression levels were CYBA and CYBB (down-regulated 16-fold and 84-fold, respectively). This suggests that mutations in either CYBA or CYBB are destabilizing the membrane-bound heterodimeric complex, whereas the expression of the three cytosolic members of the NADPH-oxidase complex (NCF1, NCF2, NCF4) is not affected (Supplementary Figure 1). A number of secondary granule proteins, such as CAMP, LTF and CRISP3, were also decreased in abundance, albeit to a lower degree. The most enriched GO BP terms among the upregulated proteins in CGD were: "type I interferon signaling pathway" (adj. P-value: $4.4*10^{-11}$), "defense response to virus" (adj. P-value: $1.3*10^{-7}$) and "complement activation, classical pathway" (adj. P-value: $1.1*10^{-2}$). The most significant BP term enriched among downregulated proteins in CGD was: "cell redox homeostasis" (adj. P-value: $2.1*10^{-2}$) which ecompasses the affected proteins CYBA and CYBB together with CAMP and LTF. The proteomes of patients with LAD (2x *ITGB1*, 1x *GFTP/SLC35C1)* were minimally disturbed when compared to healthy individuals and showed ITGB2 and related members of the integrin family (ITGAM, ITGAX, ITGAL) as proteins with strongest underexpression. Moreover, the antimicrobial protein CRISP3 was found to be expressed at lower levels than controls (Supplementary Figure 2). No GO BP terms were enriched among the upregulated proteins in LAD. The most enriched GO BP terms among the downregulated proteins in LAD were: "integrin-mediated signaling pathway" (adj. P-value: $7.6*10^{-4}$), "leukocyte migration" (adj. P-value: $7.1*10^{-4}$), "extracellular matrix organization" (adj. P-value: $2.0*10^{-3}$) and "cell adhesion" (adj. P-value: $1.8*10^{-2}$). All differentially expressed proteins have been reported in the Supplemental Table 4.

Comparison of protein expression profiles of SCN, CGD and LAD
We compared the similarity of differentially expressed proteins in SCN-ELANE, CGD, and LAD (Fig. 3g). As expected, most of the patient samples clustered together according to their clinical phenotypes. There was however one exception: the one LAD2 (*SLC35C1*) sample

was clustered with the SCN group owing to its different genetic background than the two LAD1 cases and protein expression changes more resembling SCN cases than LAD1.

In the heatmap one can observe multiple distinct protein clusters. One prominent cluster of proteins was the interferon 1-response pathway (STAT2, MX1, MX2, TRIM22, IFIT3), overexpressed in four CGD cases, one LAD1 case and one SCN case, but mostly unchanged in other SCN cases and the other two LAD cases. All patient neutrophils showed underexpression of granule proteins (CAMP, CRISP3, AZU1), with SCN cases exhibiting the strongest difference. Finally, a prominently underexpressed CYBA/CYBB (flavocytochrome b558) cluster of four CGD cases can be observed, with the fifth CGD case demonstrating very minimal underexpression of those proteins. Next, we directly compared all significantly over- and under-expressed proteins of the three clinical phenotypes for overlaps (Fig. 3h). Overall, we observe only minimal overlap of significantly dysregulated proteins between the three diseases.

Proteome analysis guiding molecular diagnosis

Two patients in our cohort had clinical signs and symptoms of neutrophil granulocyte deficiency (one patient with CGD confirmed by negative nitroblue tetrazolium (NBT) test and one patient with congenital neutropenia associated with albinism) but exome sequencing analyses did not yield a molecular diagnosis. We therefore analyzed proteomes of those cases and fitted Gaussian models for each protein using the expression data in 68 healthy samples. Next, we calculated the likelihood that a given expression value in a patient sample could come from the healthy distribution. Finally, we plotted the copy number values for the top ten dysregulated proteins for each case (Fig. 4a-b). The lists of significantly affected proteins and their copy numbers have been reported in the Supplemental Table 5.

Fig. 4a shows the ten most anomalous protein expression values of neutrophil granulocytes in the CGD patient, with the most pronounced downregulation of NCF1. Molecular diagnosis of *NCF1* deficiency is challenging, since the human genome contains two *NCF1*-pseudogenes with 99% sequence homology. Using gene- and pseudogene-specific probes (31), two normal NCF1 alleles and four NCF1 pseudogene alleles could be detected in healthy control cells, but only pseudogenes could be detected in patient CGD6 (Fig. 4c). This shows that proteome analysis in neutrophil granulocytes could be used as a pre-screening procedure to classify patients for specific *NCF1* sequencing.

Patient SCN7, a 1-year old boy born to consanguineous parents, had immunodeficiency with recurrent infections of skin and the upper respiratory tract, oral ulcerations, splenomegaly and chronic diarrhea, associated with partial albinism. He had intermittent congenital neutropenia and pancytopenia. Proteome analysis of his neutrophil granulocytes revealed a marked decrease of RAB27A protein expression (Fig. 4b). Sanger sequencing of the *RAB27a* gene confirmed a segregating homozygous mutation (NM_004580.4(*RAB27A*): c.148_149delinsC) in the patient (Fig. 4d) and thus established the diagnosis of Griscelli syndrome (32). Of interest, this patient also showed underexpression of SYTL1, a known binding partner of RAB27A (33).

Discussion

In this study, we looked at neutrophil proteome changes in patients with monogenic diseases. We show that proteomics is helpful in establishing genetic diagnosis in neutrophil deficiencies in cases where routine genetic testing is not conclusive.

Other laboratories have reported data on the proteome of human neutrophil granulocytes (10–23). Rieckmann *et al.* studied the proteomes of leukocyte subpopulations and reported quantitation data for 6007 proteins in naive neutrophil granulocytes using a data-dependent acquisition method optimized for depth and using TrEMBL protein annotation. Our data-independent approach was aimed at rapid and precise quantification of proteins in a large number of samples and relied on manually curated Swiss-Prot annotation. We observed a moderate level of correlation of our estimated protein copy numbers with the Rieckmann *et al.* data (PCC = 0.7). Moreover, we overcame past problems from endogenous proteases in neutrophils causing extensive sample degradation by enhancing the neutrophil sample handling conditions. This resulted in primarily tryptic peptides which improved reliable protein identification and quantitation. The crucial factor for this improvement seems to be the use of high amounts of protease inhibitors shortly after sample collection.

Our RNAseq data showed high agreement (PCC = 0.85) with publicly available data (30). Interestingly, there was no correlation between protein and mRNA levels in our healthy neutrophil samples (PCC = 0.00), which prompted us to focus our analytical efforts on proteome changes rather than transcriptome changes, as the former is closer related to function (34, 35). This also suggests that clinical neutrophil diagnostics should be rather focused on protein expression analysis than transcriptomics. However, we did not analyze the transcriptomes of patients. There is a possibility that in some cases, such as monogenic diseases affecting the maturation of neutrophils, the correlation between the proteome and the transcriptome is higher and in those cases using transcriptomics for clinical diagnostics could be beneficial.

We looked at protein expression profiles of three rare monogenic diseases of neutrophil granulocytes. The largest differences could be observed for the SCN cases. This disease affects the maturation of neutrophils by the mislocalized protease ELANE, leading to increased proteostatic stress within the cell (36). Indeed, we observe an overexpression of the heat-shock response systems which might be an effect of a large number of misfolded proteins within the cell. Increased expression of ribosomal and mitochondrial proteins may be related to the immature phenotype of SCN-affected neutrophils observed in our cytospin controls. We observed decreased expression of proteins in all granule subsets in SCN neutrophils, indicating antimicrobial deficiency beyond reduced cell numbers. This further expands the results of a previous study that analyzed neutrophils of genetically undetermined SCN patients and showed loss of alpha defensins and LL37 (37).

Reduced LL37 (CAMP) level in serum was proposed as a marker for SCN patients, potentially differentiating it from autoimmune neutropenia by indicating faulty granulopoiesis (38). We confirmed the observation of CAMP underexpression in neutrophils of ELANE

patients but also showed CAMP underexpression in CGD neutrophils. This warrants caution to the claim that low CAMP levels always indicate disturbance of granulopoiesis.

The proteome changes in the CGD cases were more localized, affecting mostly the protein complex affected by the mutations (flavocytochrome b558) and other proteins localized in secondary granules. In the Supplementary Figure 1 we plotted expression of all neutrophil NADPH oxidase members. Patients with *CYBA* and *CYBB* mutations have low expression of both proteins as has been observed before (39) with no loss of other complex members. It is known that CGD patients suffer from a state of hyperinflammation but the molecular mechanism has not yet been identified. We did not observe any overexpression of inflammatory mediators and inducers in CGD nor did we see dysregulation of TLRs as mentioned in a previous study (40). One of the LAD cases with *ITGB2* mutation also shows an overexpression of the IFN1 response network. Clinically the patient was very sick which points at a possibility of systemic hyperinflammation. We therefore assumed that the IGN1 response network activation in the analyzed neutrophil proteomes could be related to infection rather than the underlying genetic causes. Analyzing the LAD cases as single samples, we note that proteins with lowest expression levels in LAD1 (*ITGB2* mutation) were the integrins ITGB2, ITGAM and ITGAX while in LAD2 (*SLC35C1* mutation) integrin expression was not affected.

Of interest, one of the CGD patients (CGD1) with a splice site mutation showed a protein expression pattern similar to both CGD and SCN, with milder underexpression of CYBA and CYBB but with substantial underexpression of NCF1 (Supplementary Figure 1). Interestingly, the patient has the same splice site mutation as another CYBB case in our cohort (CGD3) but in addition to immunodeficiency and lack of respiratory burst reaction, shows low absolute neutrophil counts (ANC 500 - 1400/µl) and a severe neurological phenotype with mental retardation and facial dysmorphia. A polygenic etiology must be suspected, potentially altering his neutrophil proteome in additional ways. An *NCF1* mutation in the patient was excluded by the Gene-scan method (communicated by the treating physician, data not shown). We also excluded Williams Beuren Syndrome, very rarely associated with CGD (41) and found no mutation in XL, thereby excluding McLeod syndrome. The elucidation of causes of the underlying disease warrants a follow-up study.

Finally, data-independent acquisition proteomics helped us guide genetic diagnostics of cases for which standard exome sequencing did not provide clear answers. The observed NCF1 underexpression was a clear hint for the underlying molecular mechanism of the disease as 20% of CGD patient cases are caused by NCF1 mutations. Sequencing efforts for *NCF1* are known to be be complicated by two pseudogenes sharing 99% homology. A commercially available solution using a published Gene-scan method (42) is able to circumvent this problem. However, this is a costly method if applied to all patients. Our proteomics results allowed us to narrow-down the most likely cause of the disease and could be confirmed using the Gene-scan method.

The second case with inconclusive exome sequencing was a patient with neutropenia and partial albinism. Proteomics allowed us to guide genetic analysis and suggested RAB27A protein as the possible cause. Mutations in *RAB27a* are known to cause Griscelli 2 syndrome, matching well with the patient's phenotype. Sanger sequencing of *RAB27a* exons demonstrated a homozygous mutation in exon 2. Indeed, Sanger sequencing and whole

exome sequencing confirmed that the patient's mother was heterozygous for this mutation. When we compared the exome result of patient and mother, it became apparent, that in the patient the last part of exon 2 was not covered by the sequencing reads, explaining why we could not detect his mutation initially.

Mass spectrometry is a powerful technique for parallel quantitation of thousands of proteins. Its true potential is only being revealed in the clinical setting. Neutrophils are cells that are hard to work with for molecular diagnostics as they contain numerous highly abundant proteases and RNAses leading to quick degradation of their contents. We presented here a workflow for high purity and high quality neutrophil proteome samples acquired by state-of-the-art DIA mass spectrometry which allows for quick profiling of neutrophils (1.5 days of sample preparation plus 130 minutes per sample acquisition time). We believe that this approach can be further used to aid in genetic and clinical diagnosis of neutrophil diseases and expand our knowledge of the neutrophil biology itself.

Experimental Procedures

Experimental Design and Statistical Rationale

The total number of analyzed neutrophil samples was 84 (68 healthy controls and 16 patients). The patient samples were measured without replication. The size of the healthy control group allowed to average out biological variation. The rationale for choosing the large healthy control group was that it allowed to better estimate parameters of the Gaussian curves fitted to protein expression profiles for outlier detection in the two unclear clinical cases. Differential protein expression analysis was performed with limma R package (43), while blocking for batch effects.

Patient cohort

Patients were recruited from pediatric centers in Germany (LMU University, Dr. von Hauner Children's Hospital; and TU University, Department of Pediatrics), Turkey (Erciyes University, Fevzi Mercan Children's Hospital, Kayseri), Iran (Isfahan University, Imam Hossein Children's Hospital, Isfahan) and Israel (Schneider Children's Medical Center of Israel, Tel Aviv). Informed consent was given by the parents/legal guardians in accordance with the Declaration of Helsinki and European legislation. Children were asked for their informed assent. The study was approved by the LMU Munich ethics committee as well as ethics boards of local clinical institutions.

Extraction and purification of neutrophils

Blood was drawn from patients and healthy donors into EDTA-containing collection tubes and immediately processed within a 4 hour time window. Neutrophils were isolated with the MACSexpress human neutrophil isolation kit (Miltenyi, 130-104-434) according to the vendor's protocol. For total erythrocyte depletion the MACSxpress Erythrocyte Depletion Kit (Miltenyi, 130-098-196) was used according to the vendor's protocol. After isolation, neutrophils were twice washed in PBS (Gibco, 14190250), microscopically counted in a hemocytometer and divided into aliquots of $1x10^6$ or $2.5x10^5$ cells. A cytospin stained with May-Grunwald Giemsa for cell purity control was prepared when possible. After pelleting, the supernatant was removed and 5 µl of 25x protease inhibitor was added (Roche,

04693159001). Cells were then frozen in a -80℃ freezer before being transferred to storage in liquid nitrogen until final proteome preparation.

Neutrophil sample preparation for mass spectrometry

Purified neutrophil samples were processed with the Filter-Aided Sample Preparation (FASP) method as follows: ~10^6 cells were lysed using 50 µl of SDS lysis buffer (0.5% SDS, 0.1 M DTT in 0.1 M Tris-HCl pH 7.6) and sonicated for 30 sec using a Branson Ultrasonics 250A analog sonifier (10% duty cycle, energy level 1). Samples were then heated for 5 min at 95°C in a heating block. Subsequently, 150 µl of UA buffer (8 M urea in 0.1 M Tris-HCl pH 8.5) were added to the samples to a total volume of 200 µl, loaded onto 0.5 ml Microcon 30 kDa-cutoff Ultracel membrane filters (Merck, catalog number MRCF0R030) and spun down in an Eppendorf 5418 centrifuge for 20 min at 14.000 rcf. Next, 200 µl of UA buffer was added and the centrifugation repeated. Subsequently, 50 µl of IAA buffer (0.05 M iodoacetamide in UA buffer) were added to the filters and incubated in darkness for 20 min at room temperature. Next, two washing steps using 150 µl and 200 µl of UA buffer were performed, each time spinning down the samples for 20 min at 14.000 rcf. As a final washing step, the samples were washed twice with 200 µl of ABC buffer (50 mM ammonium bicarbonate in ddH2O) and spun down as described above.

The filters with washed samples were then transported to new collector tubes and MS-grade trypsin (Thermo Fisher, catalog number 90057) in digestion buffer (1 M urea in 0.1 M Tris-HCl 8.5) was added in 1:100 ratio. The filters were wrapped in parafilm to prevent drying and placed in a wet chamber for overnight digestion at 37°C.

Finally, the samples were spun down for 15 min at 14.000 rcf, 50 µl of ABC buffer were added to the filters and the centrifugation step repeated. The samples were then acidified to pH ~2.5 using 20% TFA (trifluoroacetic acid). The peptide yield was estimated using Thermo Fisher NanoDrop 2000c.

The samples were cleaned-up and desalted using a C18-StageTip approach as described in (44) and stored at -20°C.

Generation of spectral library: SCX chromatography

To create a comprehensive neutrophil spectral library, selected healthy donor and patient peptide samples were eluted from StageTips using 80% acetonitrile (ACN) in 0.1% formic acid, dried in Eppendorf Concentrator plus 5305, reconstituted in SCX buffer A (5 mM K2HPO4 in 10% ACN) and pooled. 100 µg of pooled peptides were separated on a Shimadzu LC-20AD HPLC system using a PolyLC PolySULFOETHYL-A SCX column (100x2.1 mm, 3 µm beads, 300 Å pores) with a 12 min non-linear gradient of SCX buffer B (1 M KCl, 5 mM K2HPO4 in 10% ACN) while collecting fractions every 15 sec. The fractions were then concentrated, reconstituted in 0.1% TFA and desalted using C18-StageTips. Lower complexity fractions were pooled together prior to mass spectrometric analysis.

Generation of spectral library: data-dependent acquisition

The spectral library SCX fractions were analyzed on Thermo Fisher QExactive HF mass spectrometer coupled to a Dionex UltiMate 3000 HPLC system using a 50 cm C-18 Thermo Fisher EasySpray column (catalog number ES803), heated to 50°C. Roughly 2 µg of peptides were loaded onto the column for each run. All samples contained spiked-in iRT peptides (Biognosys, catalog number Ki-3002-2) for retention-time alignment.

A 135 min gradient was used as follows: the flow rate was set to 300 nl/min, start at 2% buffer B (80% ACN in 0.1% FA) with a linear increase to 35% B for 90 min followed by a linear increase to 41% until 102 min and to 99% B until 104 min with a hold at 99% B until 120 min for washing. The gradient was then reduced to 2% B at 120 min and held at 2% for 15 minutes for column re-equilibration.

A Top10 DDA method was used as follows: 1 full MS1 scan between 350 - 1300 m/z at resolution of 120.000, with AGC target of 3e6, maximum injection time (IT) of 50 ms and a default charge state of 2. Ten most intense peaks were selected for MS2 fragmentation using resolution 15.000, AGC target of 1e5 and a maximum IT of 80 ms. The isolation window was set to 1.6 m/z, fixed first mass to 100 m/z and a normalized collision energy (NCE) to 30. Additionally, a dynamic exclusion of 30 sec was set.

Generation of spectral library: raw data processing

Peak lists obtained from DDA MS/MS spectra were identified using X! Tandem Vengeance (2015.12.15.2) (45), Andromeda version 1.5.3.4 (46), MS Amanda version 1.0.0.7501 (47), MS-GF+ version Beta (v10282) (48) and Comet version 2016.01 rev. 3 (49). The search was conducted using SearchGUI version 3.2.19 (50).

Protein identification was conducted against a complete Uniprot SwissProt human protein sequence database (state on 27.06.2016). The decoy sequences were created by reversing the target sequences in SearchGUI. The identification settings were as follows: trypsin, specific, with a maximum of 2 missed cleavages. 10.0 ppm as MS1 and 20 ppm as MS2 tolerances; fixed modifications: carbamidomethylation of C (+57.021464 Da); variable modifications: acetylation of protein N-term (+42.010565 Da), oxidation of M (+15.994915 Da); fixed modifications during refinement procedure: carbamidomethylation of C (+57.021464 Da); variable modifications during refinement procedure: pyrolidone from E (--18.010565 Da), pyrolidone from Q (--17.026549 Da), pyrolidone from carbamidomethylated C (--17.026549 Da). Peptides and proteins were inferred from the spectrum identification results using PeptideShaker version 1.16.11 (51). Peptide Spectrum Matches (PSMs), peptides and proteins were validated at a 1% False Discovery Rate (FDR) estimated using the decoy hit distribution. All engine-specific settings were set kept as default. The list of peptides identified in this study can be found in the Supplemental Table 6.

Data-independent acquisition of neutrophil proteomes

The data-independent acquisitions were performed on the same equipment as the spectral library DDA measurements using the exact same chromatography conditions. Each patient sample was measured once due to the size of the cohort. All the samples contained spiked-in iRT peptides (Biognosys, catalog number Ki-3002-2) for retention-time alignment. The mass spectrometry settings were as follows: one MS1 scan was performed between 350 - 1300 m/z at resolution of 120.000, AGC target of 5e6 and a maximum IT of 100 ms, followed by ten 12.5 m/z MS2 windows, ten 37.5 m/z MS2 windows and a final single 450 m/z MS2 window (21 MS2 windows total). In case of all the MS2 windows the resolution was set to 30.000, AGC target to 1e6, maximum IT to "auto" and the collision energy to 30. Both MS1 and MS2 scans were recorded in profile mode. The DIA method resulted in a median of 7 data points per peak. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (52) partner repository with the dataset identifier PXD010701.

Protein identification and quantification

Biognosys Spectronaut 11 was used for DIA search and protein quantification. The DIA raw files were converted into HTRMS format using Biognosys HTRMS converter.

Our sample-specific SCX spectral library containing 119193 precursors, 87757 peptides and 7977 proteins, was imported. Minimum of 3 up to 6 best fragments per peptide were used. The DIA search and quantification were performed with the following settings: using precision iRT and non-linear iRT calibration, MS1 and MS2 mass tolerance strategy were set to "Dynamic", XIC RT extraction window was set to "Dynamic". Precursor FDR was set to 1% and protein FDR was set to 5% (using decoy method set to "inverse"). Data filtering was set to "Qvalue percentile 0.5", cross-run normalization was set to "Qvalue complete". Peptide quantification was performed using mean precursor quantity (using up to 3 top precursors) and the area under the MS2 signal. Protein quantification was performed using mean peptide quantity (using up to 3 top peptides per protein). Protein inference was set to automatic. All settings for our Spectronaut analyses can be found in the Spectronaut experiment file (.sne) uploaded to the PRIDE archive with raw files.

As a final filtering step, known contaminants according to MaxQuant (53) were removed from the list of quantified proteins.

Differential protein expression analysis

Limma R package (43) was used to perform differential protein expression analysis using empirical Bayes moderation. The log2-transformed expression values were normally distributed. In order to increase the power of the analysis, the extra term (sample processing date) was added to the model for blocking as a mean of batch-effect control. The expression matrix used in this analysis was not processed by the ComBat algorithm. Only proteins identified by two or more peptides were used for differential abundance testing. Hits with Benjamini-Hochberg P-adjusted value <0.01 were considered statistically significant.

Gaussian model fitting for protein expression anomaly detection

Log2-transformed protein expression values in healthy donors were used to fit a Gaussian model for each protein using R package MASS (54). Proteins quantified with at least three peptides were used for increased stringency. Protein expression values of each of the two patients were used to calculate probability that a given protein is expressed similarly to healthy donors. Finally, proteins were sorted by increasing probability of having a healthy expression value. Proteins with assigned probability <0.01 were used for more detailed manual inspection.

Sanger sequencing

Patient DNA was extracted from full blood samples with Qiagen DNeasy blood and tissue Kit (catalog number 69504). PCR reaction was performed using OneTaq DNA polymerase kit (catalog number M0480X), all with the same cycler conditions (5': 95°, 30'': 95°, 30'': 55°, 45'': 72°, 5': 72°). Before sequencing samples were treated with Illustra ExoProStar (catalog number US78220) for removal of primers and dNTP. Sequencing was performed by Eurofins Genomics Germany, data were analyzed using SeqMan Pro of the DNAstar software suite. The primers for *ELANE* were designed in-house, while the primers for *RAB27a* were kindly

provided by Meeths *et al.* (55). The complete list of sequencing primers can be found in the Supplemental Table 7.

Whole exome sequencing

High throughput sequencing was performed by NextSeq500 with 2x150 bp reads. The exomes were enriched using SureSelect Human All Exon Kit Agilent V6+UTR. BWA (56) (version 0.7.15) algorithm for short Illumina reads was used to align the short reads to the human reference genome GRCh37. The Genome Analysis Tool Kit (57) (GATK version 3.6) was used for analysis of the whole exome sequencing data. All algorithms and parameters were chosen according to the GATK best practice pipeline. Functional Annotation was performed with the Variant Effect Predictor (58) (VEP version 85). Final variants were filtered with a custom database encompassing gnomAD (59) and GME (60) following a rare homozygous and compound heterozygous inheritance. SIFT (61) and PolyPhen-2 (62) were used as provided by VEP.

RNA sequencing

Neutrophils from 22 healthy donors were isolated using the same approach as for the proteomics samples. RNA was isolated with the RNAeasy plus mini-kit from Qiagen (catalog number 74134) according to the vendor protocol. Magnetic oligo-dT beads (NEB) were used for mRNA enrichment starting with 1 to 10 ng total RNA with RIN values above 9. The NEBNext Ultra II directional RNA Library prep Kit (NEB) was used to prepare RNA-seq libraries according to the manufacturer's protocol. Paired-end sequencing was performed with 2x75 cycles on an Illumina NextSeq 500 (Care-for-Rare Genomics Facility at the Dr. von Hauner Children's Hospital). The short reads were aligned with STAR (63) version 2.5.0a to the human reference genome GRCh37.p13 with basic two-pass method. The counts were then generated using the featureCounts program from the subread toolkit (64) version 1.5.1. The counts were normalized for sequencing depth and RNA type using DESeq2 (65). Median transcripts-per-million (TPM) values were calculated by normalizing the median gene read counts to the length of the union of all possible exons coded by a given gene. The raw RNAseq data have been deposited into Gene Expression Omnibus (66) (GEO) under accession number GSE118644.

Gene Ontology term enrichment analysis

The GO term enrichment was performed using DAVID (67) online tool using proteins quantified in the healthy neutrophils as background. The GO BP DIRECT terms were used for analysis.

Data analysis and plotting

All data were analyzed using R 3.4 and plotted using ggplot2 R package (68). Differential protein expression analysis was performed using the limma R package (43). Single-patient protein expression anomaly detection was performed using R MASS package (54).

Author contributions

C.K. designed and coordinated the study, supervised S.He. and provided laboratory resources. S.He. conceived experiments, obtained and prepared the samples, P.G.

generated and processed proteomics data, J.R. supervised P.G. and provided laboratory resources, S.Ho., M.R. and S.C. processed and analyzed the sequencing data, U.B., Ro.Sh., H.T., J.v.WTB, E.Ü., Ra.So. and T.P. care for patients, provided patient data and helped in obtaining the samples, P.G., S.He. C.K. and J.R. analyzed data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

References:

1. Borregaard, N. (2010) Neutrophils, from marrow to microbes. *Immunity* 33, 657–670

2. Nauseef, W. M., and Borregaard, N. (2014) Neutrophils at work. *Nat. Immunol.* 15, 602–611

3. Klein, C. (2011) Genetic defects in severe congenital neutropenia: emerging insights into life and death of human neutrophil granulocytes. *Annu. Rev. Immunol.* 29, 399–413

4. Arnold, D. E., and Heimall, J. R. (2017) A Review of Chronic Granulomatous Disease. *Adv. Ther.* 34, 2543–2557

5. Harris, E. S., Weyrich, A. S., and Zimmerman, G. A. (2013) Lessons from rare maladies: leukocyte adhesion deficiency syndromes. *Curr. Opin. Hematol.* 20, 16–25

6. Taylor, J. C., Martin, H. C., Lise, S., Broxholme, J., Cazier, J.-B., Rimmer, A., Kanapin, A., Lunter, G., Fiddy, S., Allan, C., Aricescu, A. R., Attar, M., Babbs, C., Becq, J., Beeson, D., Bento, C., Bignell, P., Blair, E., Buckle, V. J., Bull, K., Cais, O., Cario, H., Chapel, H., Copley, R. R., Cornall, R., Craft, J., Dahan, K., Davenport, E. E., Dendrou, C., Devuyst, O., Fenwick, A. L., Flint, J., Fugger, L., Gilbert, R. D., Goriely, A., Green, A., Greger, I. H., Grocock, R., Gruszczyk, A. V., 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, J. C., Krohn, J., Lamble, S., Langman, C., Lonie, L., Luck, J., McCarthy, D., McGowan, S. J., McMullin, M. F., Miller, K. A., Murray, L., Németh, A. H., Nesbit, M. A., Nutt, D., Ormondroyd, E., Oturai, A. B., Pagnamenta, A.,

Patel, S. Y., Percy, M., Petousi, N., Piazza, P., Piret, S. E., Polanco-Echeverry, G., Popitsch, N., Powrie, F., Pugh, C., Quek, L., Robbins, P. A., Robson, K., Russo, A., Sahgal, N., van Schouwenburg, P. A., Schuh, A., Silverman, E., Simmons, A., Sørensen, P. S., Sweeney, E., Taylor, J., Thakker, R. V., Tomlinson, I., Trebes, A., Twigg, S. R., Uhlig, H. H., Vyas, P., Vyse, T., Wall, S. A., Watkins, H., Whyte, M. P., Witty, L., Wright, B., Yau, C., Buck, D., Humphray, S., Ratcliffe, P. J., Bell, J. I., Wilkie, A. O., Bentley, D., Donnelly, P., and McVean, G. (2015) Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat. Genet.* 47, 717–726

7. Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M. R., Leduc, M. S., Kirby, A., Pham, P., Scull, J., Wang, M., Ding, Y., Plon, S. E., Lupski, J. R., Beaudet, A. L., Gibbs, R. A., and Eng, C. M. (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* 369, 1502–1511

8. Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., Haack, T. B., Graf, E., Schwarzmayr, T., Terrile, C., Koňaříková, E., Repp, B., Kastenmüller, G., Adamski, J., Lichtner, P., Leonhardt, C., Funalot, B., Donati, A., Tiranti, V., Lombes, A., Jardel, C., Gläser, D., Taylor, R. W., Ghezzi, D., Mayr, J. A., Rötig, A., Freisinger, P., Distelmaier, F., Strom, T. M., Meitinger, T., Gagneur, J., and Prokisch, H. (2017) Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8, 15824

9. Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., Bolduc, V., Waddell, L. B., Sandaradura, S. A., O'Grady, G. L., Estrella, E., Reddy, H. M., Zhao, F., Weisburd, B., Karczewski, K. J., O'Donnell-Luria, A. H., Birnbaum, D., Sarkozy, A., Hu, Y., Gonorazky, H., Claeys, K., Joshi, H., Bournazos, A., Oates, E. C., Ghaoui, R., Davis, M. R., Laing, N. G., Topf, A., Genotype-Tissue Expression Consortium, Kang, P. B., Beggs, A. H., North, K. N., Straub, V., Dowling, J. J., Muntoni, F., Clarke, N. F., Cooper, S. T., Bönnemann, C. G., and MacArthur, D. G. (2017) Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* 9,

10. Rieckmann, J. C., Geiger, R., Hornburg, D., Wolf, T., Kveler, K., Jarrossay, D., Sallusto, F., Shen-Orr, S. S., Lanzavecchia, A., Mann, M., and Meissner, F. (2017) Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat. Immunol.* 18, 583–593

11. Serwas, N. K., Huemer, J., Dieckmann, R., Mejstrikova, E., Garncarz, W., Litzman, J., Hoeger, B., Zapletal, O., Janda, A., Bennett, K. L., Kain, R., Kerjaschky, D., and Boztug, K. (2018) CEBPE-Mutant Specific Granule Deficiency Correlates With Aberrant Granule Organization and Substantial Proteome Alterations in Neutrophils. *Front. Immunol.* 9, 588

12. Pedersen, C. C., Refsgaard, J. C., Østergaard, O., Jensen, L. J., Heegaard, N. H. H., Borregaard, N., and Cowland, J. B. (2015) Impact of microRNA-130a on the neutrophil proteome. *BMC Immunol.* 16, 70

13. Rørvig, S., Østergaard, O., Heegaard, N. H. H., and Borregaard, N. (2013) Proteome profiling of human neutrophil granule subsets, secretory vesicles, and cell membrane: correlation with transcriptome profiling of neutrophil precursors. *J. Leukoc. Biol.* 94, 711–721

14. Uriarte, S. M., Powell, D. W., Luerman, G. C., Merchant, M. L., Cummins, T. D., Jog, N. R., Ward, R. A., and McLeish, K. R. (2008) Comparison of Proteins Expressed on Secretory Vesicle Membranes and Plasma Membranes of Human Neutrophils. *The Journal of Immunology* 180, 5575–5581

15. Loi, A. L. T., Hoonhorst, S., van Aalst, C., Langereis, J., Kamp, V., Sluis-Eising, S., Ten Hacken, N., Lammers, J.-W., and Koenderman, L. (2017) Proteomic profiling of peripheral blood neutrophils identifies two inflammatory phenotypes in stable COPD patients. *Respir. Res.* 18, 100

16. Teles, L. M. B., Aquino, E. N., Neves, A. C. D., Garcia, C. H. S., Roepstorff, P., Fontes, B., Castro, M. S., and Fontes, W. (2012) Comparison of the neutrophil proteome in trauma patients and normal controls. *Protein Pept. Lett.* 19, 663–672

17. Kotz, K. T., Xiao, W., Miller-Graziano, C., Qian, W.-J., Russom, A., Warner, E. A., Moldawer, L. L., De, A., Bankey, P. E., Petritis, B. O., Camp, D. G., 2nd, Rosenbach, A. E., Goverman, J., Fagan, S. P., Brownstein, B. H., Irimia, D., Xu, W., Wilhelmy, J., Mindrinos, M. N., Smith, R. D., Davis, R. W., Tompkins, R. G., Toner, M., and Inflammation and the Host Response to Injury Collaborative Research Program (2010) Clinical microfluidics for neutrophil genomics and proteomics. *Nat. Med.* 16, 1042–1047

18. Lominadze, G., Powell, D. W., Luerman, G. C., Link, A. J., Ward, R. A., and McLeish, K. R. (2005) Proteomic analysis of human neutrophil granules. *Mol. Cell. Proteomics* 4, 1503–1521

19. Tomazella, G. G., da Silva, I., Laure, H. J., Rosa, J. C., Chammas, R., Wiker, H. G., de Souza, G. A., and Greene, L. J. (2009) Proteomic analysis of total cellular proteins of human neutrophils. *Proteome Sci.* 7, 32

20. Jethwaney, D., Islam, M. R., Leidal, K. G., de Bernabe, D. B.-V., Campbell, K. P., Nauseef, W. M., and Gibson, B. W. (2007) Proteomic analysis of plasma membrane and secretory vesicles from human neutrophils. *Proteome Sci.* 5, 12

21. de Souza Castro, M., de Sá, N. M., Gadelha, R. P., de Sousa, M. V., Ricart, C. A. O., Fontes, B., and Fontes, W. (2006) Proteome analysis of resting human neutrophils. *Protein Pept. Lett.* 13, 481–487

22. Tak, T., Wijten, P., Heeres, M., Pickkers, P., Scholten, A., Heck, A. J. R., Vrisekoop, N., Leenen, L. P., Borghans, J. A. M., Tesselaar, K., and Koenderman, L. (2017) Human CD62Ldim neutrophils identified as a separate subset by proteome profiling and in vivo pulse-chase labeling. *Blood* 129, 3476–3485

23. Ramos-Mozo, P., Madrigal-Matute, J., Martinez-Pinna, R., Blanco-Colio, L. M., Lopez, J. A., Camafeita, E., Meilhac, O., Michel, J.-B., Aparicio, C., Vega de Ceniga, M., Egido, J., and Martín-Ventura, J. L. (2011) Proteomic analysis of polymorphonuclear neutrophils identifies catalase as a novel biomarker of abdominal aortic aneurysm: potential implication of oxidative stress in abdominal aortic aneurysm progression. *Arterioscler. Thromb. Vasc. Biol.* 31, 3011–3019

24. Wiśniewski, J. R., Hein, M. Y., Cox, J., and Mann, M. (2014) A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* 13, 3497–3506

25. Collins, B. C., Hunter, C. L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S. L., Chan,

D. W., Gibson, B. W., Gingras, A.-C., Held, J. M., Hirayama-Kurogi, M., Hou, G., Krisp, C., Larsen, B., Lin, L., Liu, S., Molloy, M. P., Moritz, R. L., Ohtsuki, S., Schlapbach, R., Selevsek, N., Thomas, S. N., Tzeng, S.-C., Zhang, H., and Aebersold, R. (2017) Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* 8, 291

26. Johnson, W. E., Li, C., and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127

27. Liu, Y., Beyer, A., and Aebersold, R. (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* 165, 535–550

28. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342

29. Fessler, M. B., Malcolm, K. C., Duncan, M. W., and Worthen, G. S. (2002) A genomic and proteomic analysis of activation of the human neutrophil by lipopolysaccharide and its mediation by p38 mitogen-activated protein kinase. *J. Biol. Chem.* 277, 31291–31302

30. Ecker, S., Chen, L., Pancaldi, V., Bagger, F. O., Fernández, J. M., Carrillo de Santa Pau, E., Juan, D., Mann, A. L., Watt, S., Casale, F. P., Sidiropoulos, N., Rapin, N., Merkel, A., BLUEPRINT Consortium, Stunnenberg, H. G., Stegle, O., Frontini, M., Downes, K., Pastinen, T., Kuijpers, T. W., Rico, D., Valencia, A., Beck, S., Soranzo, N., and Paul, D. S. (2017) Genome-wide analysis of differential transcriptional and epigenetic variability across human immune cell types. *Genome Biol.* 18, 18

31. Roos, D., de Boer, M., Köker, M. Y., Dekker, J., Singh-Gupta, V., Ahlin, A., Palmblad, J., Sanal, O., Kurenko-Deptuch, M., Jolles, S., and Wolach, B. (2006) Chronic granulomatous disease caused by mutations other than the common GT deletion in NCF1, the gene encoding the p47phox component of the phagocyte NADPH oxidase. *Hum. Mutat.* 27, 1218–1229

32. Klein, C., Philippe, N., Le Deist, F., Fraitag, S., Prost, C., Durandy, A., Fischer, A., and Griscelli, C. (1994) Partial albinism with immunodeficiency (Griscelli syndrome). *J. Pediatr.* 125, 886–895

33. Strom, M., Hume, A. N., Tarafder, A. K., Barkagianni, E., and Seabra, M. C. (2002) A family of Rab27-binding proteins. Melanophilin links Rab27a and myosin Va function in melanosome transport. *J. Biol. Chem.* 277, 25423–25430

34. Grabowski, P., Kustatscher, G., and Rappsilber, J. (2018) Epigenetic Variability Confounds Transcriptome but not Proteome Profiling for Coexpression-based Gene Function Prediction. *Mol. Cell. Proteomics*,

35. Kustatscher, G., Grabowski, P., and Rappsilber, J. (2017) Pervasive coexpression of spatially proximal genes is buffered at the protein level. *Mol. Syst. Biol.* 13, 937

36. Nayak, R. C., Trump, L. R., Aronow, B. J., Myers, K., Mehta, P., Kalfa, T., Wellendorf, A. M., Valencia, C. A., Paddison, P. J., Horwitz, M. S., Grimes, H. L., Lutzko, C., and Cancelas, J. A. (2015) Pathogenesis of ELANE-mutant severe neutropenia revealed by induced pluripotent stem cells. *J. Clin. Invest.* 125, 3103–3116

37. Pütsep, K., Carlsson, G., Boman, H. G., and Andersson, M. (2002) Deficiency of

antibacterial peptides in patients with morbus Kostmann: an observation study. *Lancet* 360, 1144–1149

38. Ye, Y., Carlsson, G., Karlsson-Sjöberg, J. M. T., Borregaard, N., Modéer, T. U., Andersson, M. L., and Pütsep, K. L.-A. (2015) The antimicrobial propeptide hCAP-18 plasma levels in neutropenia of various aetiologies: a prospective study. *Sci. Rep.* 5, 11685

39. Parkos, C. A., Dinauer, M. C., Jesaitis, A. J., Orkin, S. H., and Curnutte, J. T. (1989) Absence of both the 91kD and 22kD subunits of human neutrophil cytochrome b in two genetic forms of chronic granulomatous disease. *Blood* 73, 1416–1420

40. Hartl, D., Lehmann, N., Hoffmann, F., Jansson, A., Hector, A., Notheis, G., Roos, D., Belohradsky, B. H., and Wintergerst, U. (2008) Dysregulation of innate immune receptors on neutrophils in chronic granulomatous disease. *J. Allergy Clin. Immunol.* 121, 375–382.e9

41. Stasia, M. J., Mollin, M., Martel, C., Satre, V., Coutton, C., Amblard, F., Vieville, G., van Montfrans, J. M., Boelens, J. J., Veenstra-Knol, H. E., van Leeuwen, K., de Boer, M., Brion, J.-P., and Roos, D. (2013) Functional and genetic characterization of two extremely rare cases of Williams-Beuren syndrome associated with chronic granulomatous disease. *Eur. J. Hum. Genet.* 21, 1079–1084

42. Dekker, J., de Boer, M., and Roos, D. (2001) Gene-scan method for the recognition of carriers and patients with p47phox-deficient autosomal recessive chronic granulomatous disease. *Exp. Hematol.* 29, 1319–1325

43. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47

44. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896–1906

45. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466–1467

46. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805

47. Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., and Mechtler, K. (2014) MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.* 13, 3679–3684

48. Kim, S., and Pevzner, P. A. (2014) MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* 5, 5277

49. Eng, J. K., Jahan, T. A., and Hoopmann, M. R. (2013) Comet: an open-source MS/MS sequence database search tool. *Proteomics* 13, 22–24

50. Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A., and Martens, L. (2011) SearchGUI: An open-source graphical user interface for simultaneous OMSSA and

X!Tandem searches. *Proteomics* 11, 996–999

51. Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., Martens, L., and Barsnes, H. (2015) PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* 33, 22–24

52. Vizcaíno, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, 11033

53. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301

54. Venables, W. N., and Ripley, B. D. (2002) *Modern Applied Statistics with S*

55. Meeths, M., Bryceson, Y. T., Rudd, E., Zheng, C., Wood, S. M., Ramme, K., Beutel, K., Hasle, H., Heilmann, C., Hultenby, K., Ljunggren, H.-G., Fadeel, B., Nordenskjöld, M., and Henter, J.-I. (2010) Clinical presentation of Griscelli syndrome type 2 and spectrum of RAB27A mutations. *Pediatr. Blood Cancer* 54, 563–572

56. Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760

57. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303

58. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and Cunningham, F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26, 2069–2070

59. Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M. I., Moonshine, A. L., Natarajan, P., Orozco, L., Peloso, G. M., Poplin, R., Rivas, M. A., Ruano-Rubio, V., Rose, S. A., Ruderfer, D. M., Shakir, K., Stenson, P. D., Stevens, C., Thomas, B. P., Tiao, G., Tusie-Luna, M. T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D. M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J. C., Gabriel, S. B., Getz, G., Glatt, S. J., Hultman, C. M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Neale, B. M., Palotie, A., Purcell, S. M., Saleheen, D., Scharf, J. M., Sklar, P., Sullivan, P. F., Tuomilehto, J., Tsuang, M. T., Watkins, H. C., Wilson, J. G., Daly, M. J., MacArthur, D. G., and Exome Aggregation Consortium (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291

60. Scott, E. M., Halees, A., Itan, Y., Spencer, E. G., He, Y., Azab, M. A., Gabriel, S. B., Belkadi, A., Boisson, B., Abel, L., Clark, A. G., Greater Middle East Variome Consortium, Alkuraya, F. S., Casanova, J.-L., and Gleeson, J. G. (2016) Characterization of Greater Middle Eastern genetic variation for enhanced disease gene

discovery. *Nat. Genet.* 48, 1071–1076

61. Kumar, P., Henikoff, S., and Ng, P. C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081

62. Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* Chapter 7, Unit7.20

63. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21

64. Liao, Y., Smyth, G. K., and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930

65. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550

66. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., and Soboleva, A. (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995

67. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57

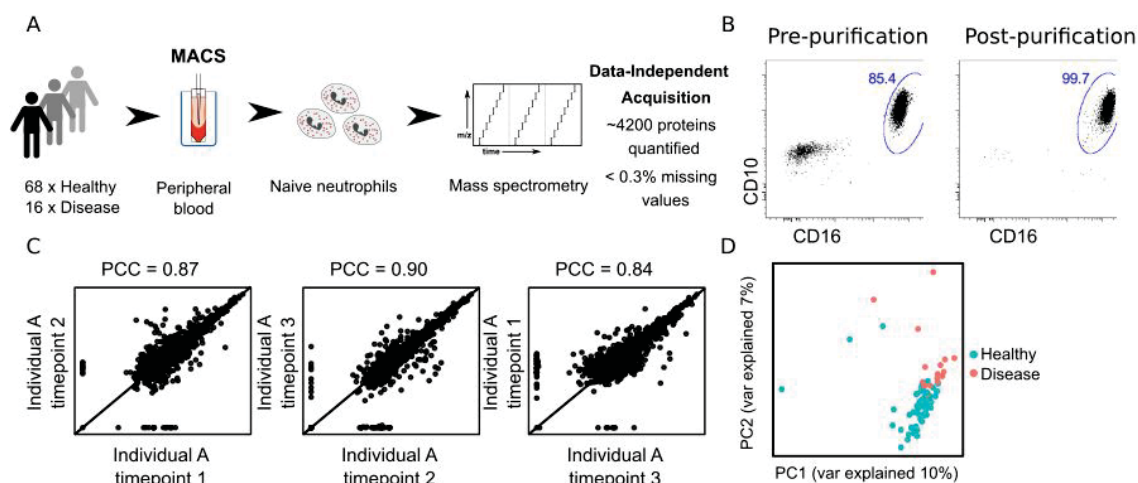68. Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis* (Springer)

Fig. 1 Mass spectrometry-based workflow for neutrophil proteome profiling **a** Neutrophil granulocyte samples were extracted from peripheral blood of healthy donors and patients, purified and subjected to data-independent acquisition mass spectrometry. **b** Flow cytometry showed high neutrophil purity (>99%) after isolation. **c** Pearson correlation analysis of samples from the same individual obtained at three different timepoints shows high level of

correlation between biological replicates. PCC, Pearson correlation coefficient. **d** Principal component analysis demonstrates good separation of healthy and diseased individuals using full protein expression profiles.
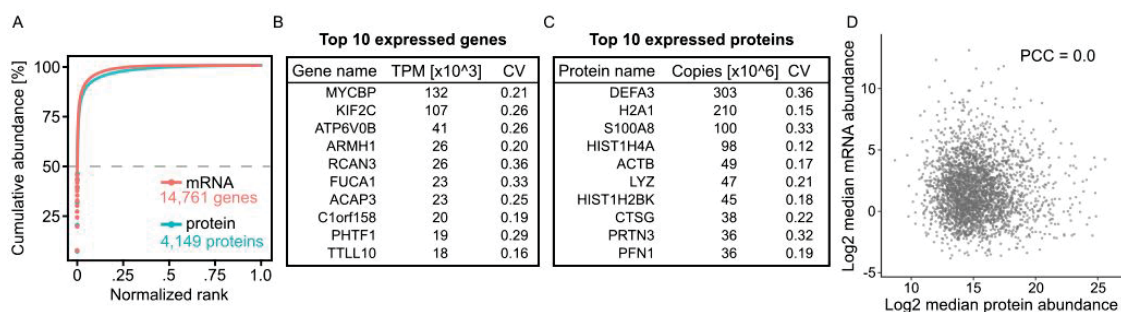

Fig. 2 Protein and transcript expression profiles in healthy neutrophils. **a** Cumulative abundances of proteins and protein-coding transcripts follow a similar pattern. Only seven proteins account for ~50% of the total protein copy number in neutrophils. **b** Top 10 expressed protein-coding transcripts in healthy donor neutrophils. TPM, transcripts per million; CV, coefficient of variation. **c** Top 10 expressed proteins in healthy donor neutrophils. CV, coefficient of variation **d** We observed no correlation (PCC = 0.00) of transcriptome and proteome in healthy naive neutrophil granulocytes. PCC, Pearson correlation coefficient.
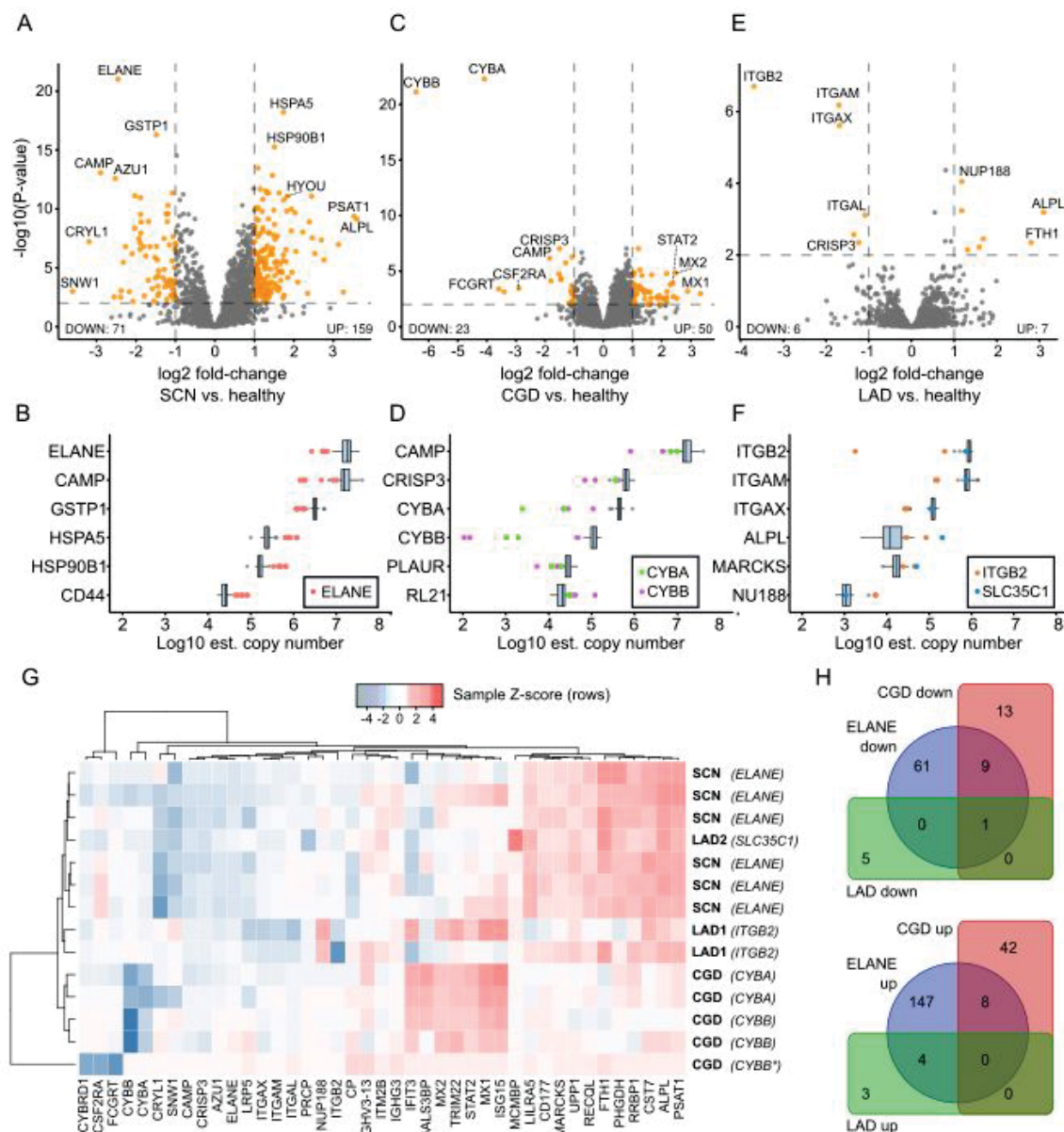
Fig. 3 Differential protein expression in SCN, CGD and LAD. **a** Volcano plot illustrating changes in neutrophil granulocyte protein expression between SCN-ELANE (n=6) and healthy individuals (n=68). X-axis showing log2 fold change in comparison to healthy control, Y-axis showing -log10(P-value). A total of 71 proteins were significantly underexpressed in *ELANE*-mutant cells, 159 were significantly overexpressed. Points marked in orange have adjusted P-values <0.01 and absolute fold-change >2. **b** Top six significantly differentially expressed proteins in SCN, ordered by decreasing mean estimated log10 copy number. The differentially expressed proteins span four orders of magnitude of abundance. **c** Differential protein analysis of CGD cases vs. healthy individuals. The most down-regulated proteins are members of the flavocytochrome b558 complex affected in CGD, namely CYBA and CYBB. Interferon-1-responsive network members (MX1, MX2, STAT2) were significantly upregulated in CGD. Points marked in orange as in Fig. 3a. **d** Top six significantly differentially expressed proteins in CGD, ordered by decreasing mean estimated log10 copy

number. **e** Differential protein analysis of LAD cases vs. healthy control. Only modest proteome-level effects could be detected. The most down-regulated proteins were integrins which are members of the affected protein family (ITGB2, ITGAM, ITGAX). **f** Top six significantly differentially expressed proteins in LAD, ordered by decreasing mean estimated log10 copy number. **g** Heatmap of protein expression levels in all analyzed cases normalized to expression in healthy individuals and z-transformed within each sample. Samples and proteins were ordered using hierarchical clustering. Dendrograms show relative Euclidean distances between samples or proteins. Samples tend to be clustered close together as expected by genetic analysis, with the exception of the one LAD2 case. Very homogenous expression profiles can be seen for the six SCN samples and four CGD samples. *CYBB**, *CYBB* splice donor variant. **h** Venn diagram comparing differentially expressed proteins in SCN-ELANE, CGD, and LAD (note the low degree of overlap between diseases).
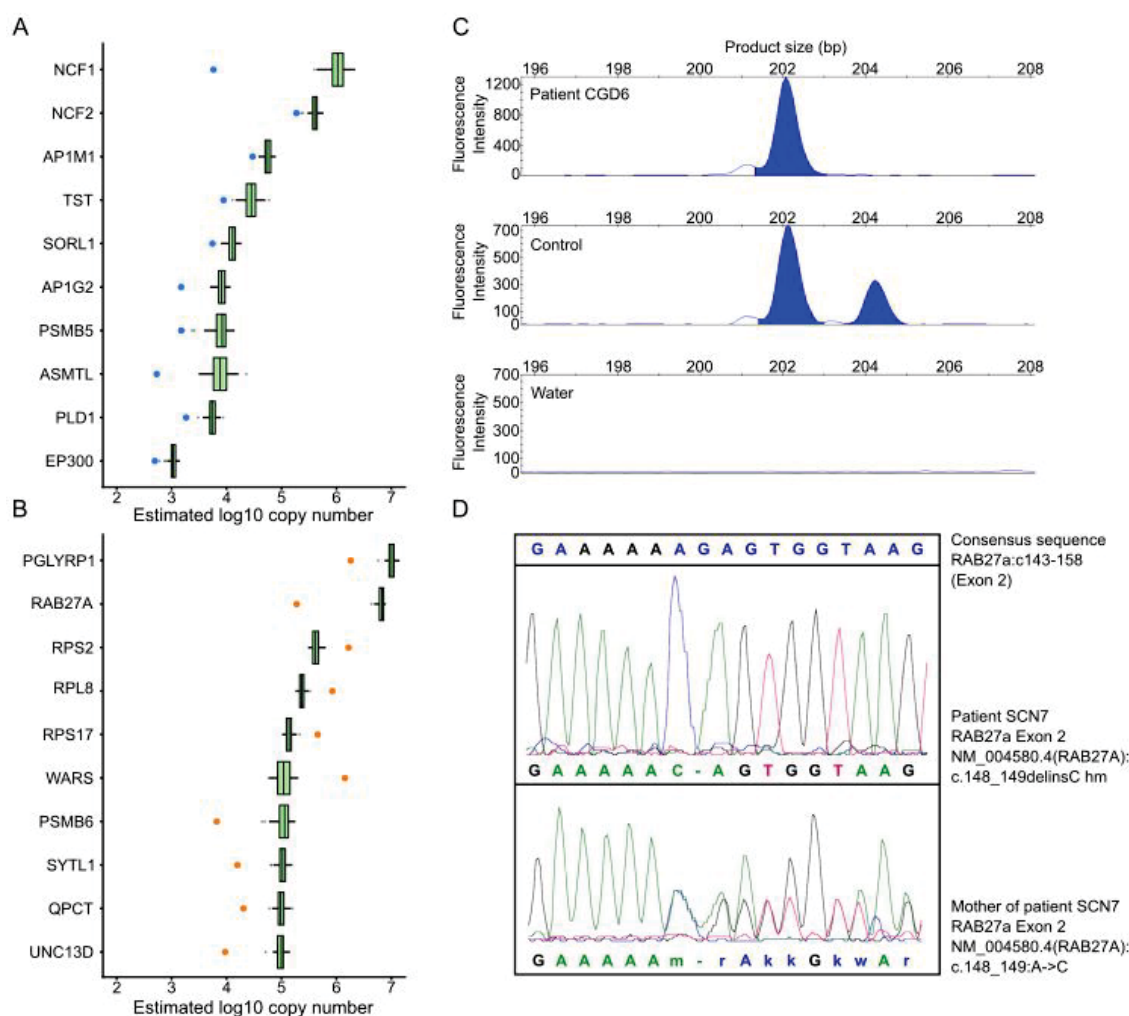
Fig. 4 Proteome analysis guiding molecular diagnosis. **a** Top 10 deregulated proteins in patient CGD6. **b** Proteome deregulation analysis for the other neutropenia case without known genetic mutation. The protein-level analysis found a significant underexpression of

RAB27A protein (by two orders of magnitude) and its interaction partner, SYTL1. **c** Gene-scan analysis using gene- and pseudogene-specific fluorescent primers in patient CGD6. Note two normal NCF1 alleles (product size 204 bp) and 4 NCF1 pseudogene alleles (product size 202 bp) in a healthy control which result in the physiologically normal 2:1 signal ratio. In patient CGD6, only 4 pseudogene alleles but no normal NCF1 alleles are present. DNA-free water served as a negative control. **d** Sanger sequencing of the *RAB27a* gene in patient SCN7 and his mother revealed a homozygous mutation in exon 2 (NM_004580.4(RAB27A):c.148_149delinsC) in the patient. The patient's mother was a heterozygous carrier of the mutation.