A complementary approach for genetic diagnosis of inborn errors of immunity using proteogenomic analysis

Fumiaki Sakura[a], Kosuke Noma[a], Takaki Asano[a], Kay Tanita[b], Etsushi Toyofuku[b], Kentaro Kato[c], Miyuki Tsumura[a], Hiroshi Nihira[c], Kazushi Izawa[c], Kanako Mitsui-Sekinaka[d], Ryo Konno[e], Yusuke Kawashima[e], Yoko Mizoguchi[a], Shuhei Karakawa[a], Seiichi Hayakawa[a], Hiroshi Kawaguchi[a], Kohsuke Imai[d], Shigeaki Nonoyama[d], Takahiro Yasumi[c], Hidenori Ohnishi[f], Hirokazu Kanegane[b], Osamu Ohara[e,1], and Satoshi Okada[a,1]

[a]Department of Pediatrics, Hiroshima University Graduate School of Biomedical and Health Sciences, 1-2-3 Kasumi, Minami ward, Hiroshima city, Hiroshima, 734-8551, Japan.; [b]Department of Pediatrics and Developmental Biology, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University (TMDU), 1-5-45 Yushima, Bunkyo city, Tokyo, 113-0034, Japan.; [c]Department of Pediatrics, Kyoto University Graduate School of Medicine, 54 Shogoin Kawaharacho, Sakyo ward, Kyoto city, Kyoto, 606-8507, Japan.; [d]Department of Pediatrics, National Defense Medical College, 3-2 Namiki, Tokorozawa city, Saitama, 359-8513, Japan.; [e]Kazusa DNA Research Institute, 2-6-7 Kazusakamatari, Kisarazu city, Chiba, 292-0818, Japan; [f]Department of Pediatrics, Gifu University Graduate School of Medicine, 1-1 Yanagido, Gifu city, Gifu 501-1112, Japan.

This manuscript was compiled on March 22, 2023

Advances in next-generation sequencing technology have identified many genes responsible for inborn errors of immunity (IEI). However, there is still room for improvement in the efficiency of genetic diagnosis. Recently, RNA sequencing and proteomics using peripheral blood mononuclear cells (PBMCs) have gained attention, but only some studies have integrated these analyses in IEI. Moreover, previous proteomic studies for PBMCs have achieved limited coverage (approximately 3000 proteins). More comprehensive data are needed to gain valuable insights into the molecular mechanisms underlying IEI. Here, we propose a state-of-the-art method for diagnosing IEI using PBMCs proteomics integrated with targeted RNA sequencing (T-RNA-seq), providing unique insights into the pathogenesis of IEI. This study analyzed 70 IEI patients whose genetic etiology had not been identified by genetic analysis. In-depth proteomics identified 6498 proteins, which covered 63% of 527 genes identified in T-RNA-seq, allowing us to examine the molecular cause of IEI and immune cell defects. This integrated analysis identified the disease-causing genes in four cases undiagnosed in previous genetic studies. Three of them could be diagnosed by T-RNA-seq, while the other could only be diagnosed by proteomics. Moreover, this integrated analysis showed high protein-mRNA correlations in B- and T-cell-specific genes, and their expression profiles identified patients with immune cell dysfunction. These results indicate that integrated analysis improves the efficiency of genetic diagnosis and provides a deep understanding of the immune cell dysfunction underlying the etiology of IEI. Our novel approach demonstrates the complementary role of proteogenomic analysis in the genetic diagnosis and characterization of IEI.

Proteomics | Inborn errors of immunity | Targeted RNA sequencing | Genetic diagnosis

Patients with inborn errors of immunity (IEI), previously known as primary immunodeficiency disorders, demonstrate increased susceptibility to infectious diseases, autoimmunity, autoinflammatory diseases, allergies, and malignancies(1). These conditions are generally caused by monogenic germline defects resulting in the dysfunction of encoded proteins. The latest classification of IEI from the International Union of Immunological Societies (IUIS) Expert Committee includes 485 genes as genetic etiologies of IEI, representing an increase of 55 genes since the 2019 IUIS update(2). This breakthrough occurred predominantly due to the application of next-generation sequencing (NGS) technologies, such as targeted gene panel NGS (T-NGS), whole-exome sequencing (WES), or whole-genome sequencing(3–5). Genetic diagnosis plays a pivotal role in the clinical management in IEI patients because elucidating the molecular etiology paves the way for fundamental therapies; 34% of genetically diagnosed cases have distinct therapeutic options(5). However, the diagnostic yield of NGS for IEI is still low and is estimated to be approximately 30 to 40%(5–9). WES and T-NGS have several inherent limitations, explaining these undiagnosed cases. The most challenging of those limitations is the difficulty of interpreting variants of unknown significance(10, 11). Other drawbacks are the inability to detect variants in noncoding regions(12).

RNA sequencing (RNA-seq) has been well employed as one of the most valuable tools to study Mendelian disorders(10, 13), because it provides complementary information about the downstream consequences of genomic variants, such as varia-

Significance Statement

Genetic diagnosis plays a central role in the clinical management of patients with inborn errors of immunity (IEI). However, the diagnostic yield for IEI based on the sequencing of germline DNA is still low and is estimated to be approximately 30%. This study shows the utility of integrated analysis with proteomics and targeted RNA sequencing (T-RNA-seq) of peripheral blood mononuclear cells. We identified the molecular cause and immune cell defects in patients with IEI, increasing the diagnostic yield by 6%. Notably, even in cases missed by T-RNA-seq, proteomics could identify the genetic etiology of the disease, suggesting the pivotal role of proteomic analysis in diagnosing IEI. Our novel approach improves the efficiency of the genetic diagnosis and elucidates the pathogenesis of IEI.

S.O. and O.O. designed research; K.T., E.T., K.K., M.T., H.N., K.I., K.M-S., Y.M., S.K., S.H., K.K., K.I., S.N., T.Y., H.O., and H.K. performed research; R.K., Y.K., and O.O. contributed new reagents/analytic tools; F.S., K.N., R.K., Y.K., and O.O. analyzed data; and F.S., T.A., O.O., and S.O. wrote the paper

The authors declare no competing interest.

[1]To whom correspondence should be addressed. E-mail: sokada@hiroshima-u.ac.jp or ohara@kazusa.or.jp

© The Author(s) 2023. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

tions in RNA abundance, allele-specific expression (ASE) and alternative splicing isoforms(13, 14). Especially the use of targeted RNA-seq (T-RNA-seq) is a well-established approach for investigating low-abundance transcripts or low-input RNA samples(15, 16) and is advantageous in studying IEI, in which the expression of disease-causing genes is often suppressed. Indeed, many studies on IEI have confirmed the effectiveness of RNA-seq or T-RNA-seq(17–21). However, the diagnostic yield of IEI remains in the 7.5-36% range for patients for whom T-NGS or WES is uninformative(10, 22, 23). One of the most significant current discussions regarding RNA-seq is the discordance of RNA and protein expression levels. The controversy about the relationship between protein abundance and its coding mRNA abundance has continued unabated due to the development of high-throughput technologies that simultaneously interrogate the global abundance of protein and mRNA(24–26).

More recently, researchers have shown an increasing interest in proteomics due to technological advances in mass spectrometry (MS)-based protein identification(27, 28). To date, more than 90% of the proteins corresponding to known protein-coding genes have been detected by MS-based proteomics(29). A recent literature review concluded that MS-based proteomics contributed substantially to our understanding of innate immunity(30). This review also pointed out that overcoming problems associated with low abundance of cellular fractions and high abundance of degradative proteases will be required to obtain an unbiased and comprehensive protein profile. Since hematopoietic cells form the basis of the pathogenesis of IEI, expression analysis of peripheral blood mononuclear cells (PBMCs) is useful to determine the molecular pathogenesis. However, previous PBMC proteomics studies using data-independent acquisition (DIA)-MS, which provides higher sensitivity, higher protein coverage, and greater reproducibility than classic data-dependent acquisition, have identified only approximately 3000 proteins(31–33). Considering that patients with IEI have a variety of immune cell defects and disease-causing protein defects, more comprehensive proteomic data are needed to gain rational insights into the molecular mechanisms underlying aberrant immune systems. A few studies have applied proteomics to the genetic diagnosis of IEI(34, 35). However, the current study is the first to examine the utility of in-depth proteomics in integrated analysis in combination with T-RNA-seq.

Here, we propose a state-of-the-art method for diagnosing IEI, providing notable insights into the pathogenesis of IEI. Our single-shot DIA-MS approach, which was high-throughput and cost-effective, enabled proteomic analysis of PBMCs at greater depth. Furthermore, this improved analytical depth achieved protein coverage nearly equivalent to the depth of transcriptome analysis by RNA-seq and allowed integrated analysis with T-RNA-seq. This study aims to highlight the complementary role of integrated analysis of proteomics and T-RNA-seq to canonical genomic analysis in determining the molecular pathogenesis of IEI.

Results

In-depth proteomic data from PBMCs covered many IEI-related genes. The current study encompassed a cohort of 70 patients diagnosed with IEI but without a known genetic etiology. Of these, 48 patients underwent WES, and the remaining 22 underwent T-NGS of a 400 IEI gene panel. Prior genetic analysis was conducted based on criteria established by the American College of Medical Genetics and Genomics (ACMG), along with the patients' phenotypes and the disease's inheritance mode. However, no pathogenic variants were identified that satisfied these criteria. (*SI Appendix*, Table S1). The first set of analyses examined the eligibility of the proteomic data. The initial processing of the proteomic data identified 8857 proteins; after data optimization, 6498 (73% of detected proteins) proteins from 63 IEI patients and six healthy controls (HCs) (91% of all participants) were retained for downstream analysis (Fig 1*A*, *SI Appendix*, Dataset S1). T-RNA-seq provided data for 527 IEI-related genes, almost all of which were highly enriched, in 63 cases (Fig 1*A*, *SI Appendix*, Fig S1*A* and Dataset S2). We then removed the genes with total read counts of less than 1000, leaving T-RNA-seq data for 499 genes in 63 cases (Fig 1*B*). Surprisingly, the refined proteomic data, which excluded nontarget proteins such as plasma and RBCs, identified 8641 proteins from PBMCs, covering 80% of the genes in T-RNA-seq (399 out of 496 genes; three noncoding genes were removed) (Fig 1*C*). Although filtering the data to remove proteins with high missing values (MVs) reduced that coverage to 63% (314 out of 496 genes) (Fig 1*D*), our proteomic data still maintained high coverage. Overall, these results show that our proteomic analysis covered many known IEI genes and allowed us to perform integrated mRNA–protein analysis.

Detailed interpretation of proteomic data enabled optimization of initial processing. We next performed data interpretation to ensure the validity and reproducibility of the proteomic data. We excluded seven samples with a higher proportion of MVs from this study (E1 to E7) based on the PCA for raw data (Fig 1*E*). Regarding the assessment of MVs with linear regression, the refined protein abundance and the proportion of MVs showed a negative correlation, with an R-squared value of 0.61 (Fig 1*F*), which was markedly higher than that of the total protein abundance including nontarget proteins. Moreover, the distribution of mean expression levels was biased toward lower levels for proteins with MVs compared to those without MVs (Fig 1*G*). These results indicated that MVs were abundance-dependent and left-censored. Another significant aspect of this result is that the difference in R-squared values between total proteins and targeted proteins indicates that the dominance of nontarget proteins overwhelmed the abundance of the proteins of interest and increased the number of MVs (Fig 1*F*). Regarding MVs being left-censored data, we adopted small-value imputation methods separately for exploratory and diagnostic analyses. Considering that the MVs were below the detection limit, the zero-value method was adapted for the diagnostic analysis. Meanwhile, the minimum deterministic method was selected for the exploratory analysis because distance-based clustering, such as the k-means method, is not sensitive to zero value, especially in cases with a small k value. Then, based on the results of the NormalyzerDE comparison, we normalized the imputed data with quantile normalization and robust linear regression normalization (*SI Appendix*, Fig S1*B*). Similarly, we normalized the T-RNA-seq data with the variance stabilizing transformation method (*SI Appendix*, Fig S1*C*). In summary, our data interpretation approach revealed the nature of the MVs and allowed data optimization (*SI Appendix*,

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

Dataset S1).

Diagnostic analysis identifies disease-causing protein. Our study allows direct comparison of protein and mRNA expression profiles because the data were generated from the same specimens. Therefore, we examined the utility of proteomic analysis in genetic diagnosis by comparing the protein and mRNA expression levels of 314 overlapping genes (*SI Appendix*, Dataset S3). We identified four cases where a proteomic analysis unveiled the disease-causing protein (Table S1). Bruton tyrosine kinase (BTK) deficiency (B1_P21) and X-linked inhibitor of apoptosis (XIAP) deficiency (B1_P22) exhibited impressive reductions in protein (z-scores; -6.7 and -8.1, respectively) and mRNA (z-scores; -5.3 and -7.8, respectively) (Fig 2*A* and *B*), despite a lack of significant findings in the initial genomic analysis. In contrast, adenosine deaminase 2 (ADA2) deficiency (B1_P29) and LPS-responsive beige-like anchor protein (LRBA) deficiency (B2_P35) presented no reduction in mRNA expression (z-scores; -0.8 and -0.6, respectively) but a considerable reduction in protein expression (z-scores; -5.2 and -6.3, respectively) (Fig 2*C* and *D*). In these cases, only monoallelic variants were identified in genome analysis, and no genetic diagnosis was made. Proteomic analysis thus provided unique information directly related to a definitive diagnosis in these two cases. In addition, the protein expression profiles of these four cases were compared to HCs as a means of making a diagnosis in a single case. Each disease-causing protein was highly expressed in HCs, while its expression was markedly decreased with log2-fold change <-5 in each patient, indicating a decrease of more than 1/32 from the average expression (Fig 2*E*, *F*, *G*, and *H*, Dataset S3).

Validation analysis links the results of the diagnostic analysis to the clinical diagnosis. Since genetic diagnosis is based on genomic variants, we performed further analysis to validate the results of our diagnostic analysis. The results are summarized in Table 1. In a BTK-deficient case, the intronic variant of c.-196+1G>T was detected by follow-up genomic analysis. This 5'-UTR was not only a splice site but also contained a number of transcriptional regulators that may have explained the results of the diagnostic analysis (*SI Appendix*, Fig. S2), but detailed pathogenicity is currently under analysis. In an XIAP-deficient case, Western blotting and RT–PCR also showed decreased protein and mRNA expression levels. In addition, targeted sequencing covering the entire *XIAP* region identified a large deletion containing a noncoding exon with promoter activity. These results were previously reported by Sbihi et al., and "patient 2" corresponded to this case(36). In an ADA2-deficient case, decreased ADA2 activity was observed in the patient and was a supportive laboratory finding. Some results have already been reported by Nihira et al., and "patient 2" corresponded to this case(37). T-RNA-seq revealed aberrant splicing in this case (*SI Appendix*, Fig S3*A*). The results of LeafCutter show that the aberrant junction is specific to this case (*SI Appendix*, Fig S3*A*). Moreover, variant calling on T-RNA-seq revealed the intronic variant of c.972+102T>G, which generated an abnormal splicing profile, and the known missense variant led to ASE, with unequal expression between the wild-type and mutant alleles (20% and 80%, respectively) (*SI Appendix*, Fig S3*B*). Given that aligned reads harbored missense and intronic variants separately, compound heterozygous variants in *ADA2* were the cause of the disease. In LRBA deficiency cases, the results of the diagnostic analysis are under verification. However, the patient showed various autoimmune abnormalities consistent with the phenotype of LRBA deficiency. In addition, we observed supportive laboratory findings of decreased CTLA4 expression in Tregs and decreased LRBA expression, as determined by Western blotting. These results suggest that our diagnostic analysis can contribute to clinical diagnosis. In summary, although genetic diagnosis was possible in three patients by T-RNA-seq alone, integrated analysis with proteomics enabled genetic diagnosis in one additional patient, increasing the efficiency of genetic diagnosis by 6% in patients who could not be diagnosed by genetic analysis (Table 2).

The protein and mRNA expression levels of B- and T-cell-specific genes show strong correlations. Considering that a discrepancy between protein and mRNA expression of the disease-causing gene was noted in two cases in our diagnostic analysis, we systematically analyzed the correlation between protein and mRNA levels. We first calculated Spearman's correlation coefficients for 314 genes identified by both proteomics and T-RNA-seq among our 63 patients (Fig 3*A* and *SI Appendix*, Dataset S4) and found that the median correlation was 0.29 (interquartile range of 0.07 to 0.52). Furthermore, the distribution of correlation coefficients indicates that more than half of the genes have an absolute correlation coefficient of less than 0.4, that is, weak or no correlation (Fig 3*B*). These results indicate a discrepancy between protein and mRNA expression levels. Because the genes targeted in T-RNA-seq included the immune-cell-specific genes used as cell markers, we also compared protein–mRNA correlations of B-, T-, and NK-cell-specific genes. We identified 10 B-cell- and 13 T-cell-specific genes among the 314 genes but no NK-cell-specific genes. Interestingly, the correlation coefficients for B-cell-specific and T-cell-specific genes were 0.84 and 0.74, respectively, showing a strong correlation (Fig 3*C*).

Exploratory analysis of B-cell-specific proteins enables the identification of B-cell-deficient cases. Based on the strong correlation of proteomic and T-RNA-seq data in B and T cells detected in the current study, we investigated whether proteomic analysis could discriminate the population with immune cell defects, which play a pivotal role in the pathogenesis of IEI. We thus analyzed proteomic data with k-means clustering based on immune cell-specific protein profiles (*SI Appendix*, Dataset S5). First, we extracted 18 B-cell-specific proteins (based on public databases) from our proteomic data (Fig 4*A*) and selected three according to the criteria described in the Methods (see "Exploratory analysis of B- and T-cell deficiency"). We then segregated 12 cases into B-cell-deficient cluster by k-means clustering (Fig 4*B*). Interestingly, eight out of 12 cases categorized as B-cell-deficient cluster were classified in IUIS category 3 as "predominantly antibody deficiencies", and five of them showed apparent B-cell defects in flow cytometry (FCM) analysis (*SI Appendix*, Table S2). To validate the clustering results, we performed GO analysis of significantly downregulated genes (log-fold-change <-1.5 and p-value <0.05) in a two-group comparison (B-cell-deficient clusters vs. others). The results showed that many genes involved in B-cell function were strongly downregulated in the B-cell-deficient group, even in the total protein profile,

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

suggesting that the clustering results were valid (Fig 4*C* and *D*). For further validation of the proteomics results, we compared the results with those of T-RNA-seq (*SI Appendix*, Fig S4*A*). The 14 B-cell-deficient cases identified by T-RNA-seq included all 12 B-cell-deficient cases in the proteomics, indicating the strong protein-mRNA correlation of B-cell-specific genes (Fig 4*E*, *SI Appendix*, Table S3). In summary, PBMC proteomics enabled the identification of cases with B-cell dysfunction based on their quantitative changes.

Comprehensive protein analysis reveals T-cell dysfunction in diverse disease types, and T-RNA-seq reveals diversity in the expression profiles of T-cell-specific genes. Next, we examined T-cell dysfunction, which provides a helpful benchmark for the validity of our study because T-cell function is diverse, and its dysfunction is implicated in the pathogenesis of various forms of IEI. Our proteomic analysis identified 32 T-cell-specific proteins (Fig 5*A* and *SI Appendix*, Dataset S5), and clustering analysis identified 23 cases of T-cell deficiency (Fig 5*B*). The Results show that half of the T-cell-deficient cluster are either combined immunodeficiency or IUIS category 4 as "diseases of immune dysregulation", in which T-cell dysfunction is the predominant pathological feature (*SI Appendix*, Table S4). Most of the remaining cases were suggested to be common variable immune deficiency (CVID), but only three of them were also classified as B-cell deficient. On the other hand, a case of X-linked agammaglobulinemia, which presents as a pure B-cell defect, was not included in the T-cell-deficient cluster, indicating the heterogeneous nature of CVID. GO analysis of the proteins downregulated in the T-cell-deficient cluster vs. others showed that terms involved in ribosome biogenesis and ribosomal RNA were highly enriched (Fig 5*C*), and the protein expression of those involved in T-cell function was also suppressed to the same extent (Fig 5*D*). In contrast to the analysis of B-cell deficiency, only 17 T-cell-deficient cases in T-RNA-seq matched the cluster in the proteomic analysis (Fig 5*E*, and *SI Appendix*, Table S5). This is an unexpected result but is attributed to the fact that clustering based on T-cell-specific genes was highly variable (*SI Appendix*, Fig S4*B*), and the elbow point, which indicates the optimal number of clusters, was uniquely greater than a value of two in T-cell analysis of T-RNA-seq (*SI Appendix*, Fig S5*A* and *B*). These results suggest that T-cell function in IEI is more complex than B-cell function, and in particular, the mRNA expression of T-cell-specific genes exhibits a diverse profile.

Discussion

This study analyzed 63 patients with IEI through in-depth proteomic analysis of PBMCs, identifying 6498 proteins that covered 63% of the genes covered by the T-RNA-seq. The improved comprehensiveness and mRNA coverage allowed an integrated analysis of protein and mRNA and revealed the discrepancies between protein and mRNA expression levels. These findings demonstrate the importance of proteomic analysis and its role as a complement to RNA-seq for IEI. The most important clinically relevant result was that these gene expression analyses enabled genetic diagnosis in four cases, two of which could be diagnosed only by proteomic analysis. In addition, an integrated study with T-RNA-seq elucidated the genomic basis of the disease in one case. Another significant finding was that proteomic data allowed us to classify the cases of immune cell defects based on protein profiles specific to those cells. Exploratory analysis then revealed immune cell dysfunction in terms of comprehensive molecular interactions. These findings suggest that an integrated analysis of proteomics and T-RNA-seq facilitates the understanding of the pathogenesis and underlying immune cell defects in IEI cases.

One fascinating finding was that diagnostic analysis revealed the disease-underlying protein in four cases. Among them, BTK- and XIAP-deficient cases demonstrated a noticeable reduction in both protein and mRNA expression. Further analysis proved that these results were due to genomic variants in the promoter region. In contrast, ADA2- and LRBA-deficient cases exhibited discordance between protein and mRNA expression, where decreased expression was observed only at the protein level. In these cases, the identification of the lack of ADA2 activity and reduced LRBA expression in western blotting aided in the clinical diagnosis. Proteomic analysis thus provides essential information that contributes to clinical diagnosis. Moreover, T-RNA-seq for ADA2 deficiency showed ASE in genomic locations bearing missense variants which may trigger nonsense-mediated decay (NMD). This finding is consistent with previous findings by Rivas et al., who demonstrated that variants generating premature stop codons and predicted to trigger NMD were prone to demonstrate ASE(38). Nevertheless, NMD occurring in the allele of the intronic variant in ADA2 did not significantly affect the mRNA expression levels, and its pathological significance was identified via the decrease in protein expression levels. These findings are consistent with those of Jiang et al., who showed that protein information could explain genetic disease phenotypes that could not be explained by transcript information alone(39). Additionally, reduced expression of disease-causing proteins can be identified through comparison with healthy controls, and the discovery of down-regulated proteins does not necessarily require a cohort. These findings suggest that they can be applied in the clinical setting for diagnosing a single patient.

Another important finding was that target enrichment of RNA-seq allowed us to identify the genomic basis of an ADA2-deficient case. The expression levels of aberrant transcript was very low due to mRNA instability; Leafcutter results show that the number of aberrant splicing reads is only 0.008% of the cluster. However, target enrichment increased the read depth and revealed the aberrant splicing with intronic variant. These results reflect those of Gildea et al. who also found that target RNA-seq method increased the efficiency of identification of rare splice isoforms, which was difficult with standard RNA-seq(40). Given that the guidelines from ACMG state that a null variant in a gene where loss of function (LOF) is a known mechanism of pathogenicity is the strongest evidence of pathogenesis(41), integrated analysis of T-RNA-seq and proteomics provides significant support for genetic diagnosis by detecting an aberrant splicing and reduced protein levels. In addition, integrated analysis can be a useful tool for the diagnosis of IEI because more than 75% of the known IEI variants show autosomal recessive or X-linked recessive inheritance and are considered LOF(2). Taken together, this study contributes to the clinical management of IEI by providing a rationale for essential specific treatment options, such as

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

TNF inhibitors for ADA2 deficiency(42), abatacept for LRBA deficiency(43), and HSCT for XIAP deficiency.

As mentioned in the literature review, lymphocyte subset analysis, which provides the initial evidence of immune system insufficiency, is a fundamental diagnostic approach for IEI, along with genetic testing(44). We classified all cases into two groups based on the profiles of three proteins specific to B- and T cells and performed DEA to explore immune cell defects. The results of GO analysis for B cells are reasonable, with many proteins involved in B-cell function showing decreased expression. Interestingly, patient B1_P17, clinically diagnosed with late-onset combined immunodeficiency, was assigned to the B-cell-deficient cluster, even though the CD19(+)-B-cell abundance in the peripheral blood was 13.1% and no reduction was observed by FCM. These results further support the suitability of proteomics for IEI diagnosis, as its unbiased comprehensiveness provides a quantitative and functional information regarding immune cell status. However, the B-cell-deficient cluster of T-RNA-seq showed no decreased expression in *AICDA*. This rather contradictory result may be due to inadequate target enrichment of *AICDA*; in fact, some cases showed missing values. In contrast to B-cell analysis, T-cell analysis showed that the proteins involved in ribosome biogenesis and ribosomal RNA processing were downregulated to the same extent as those involved in T-cell function. However, paradoxically, these results coincide with those of well-regarded studies indicating that T-cell activation via T-cell receptor signaling enhances ribosome biosynthesis(45, 46); in other words, T-cell dysfunction inhibits ribosome biogenesis. Overall, these findings suggest that comprehensive proteomics provides insight into not only quantitative abnormalities of immune cells but also the functional aspects of immune cells based on quantitative changes in the molecules involved in their cellular function.

Even though the data processing yielded optimized proteome data, the presence of nonnegligible numbers of MVs remains the major limitation of this study. Seven ineligible cases, which were PCA outliers, were excluded to ensure protein coverage of the data, but 2143 proteins (27% of the total) were excluded due to the large number of samples containing MVs for that protein. Moreover, these proteins included 85 genes covered in the T-RNA-seq (decreasing the total from 399 to 314 genes), which may have caused some bias in the results of correlation analysis. Additionally, analyzing only at a one-time point may underestimate the correlation as proteins and mRNAs have different temporal contexts(47, 48). In part, this is why it is important to analyze protein and mRNA in an integrated manner. Another potential weakness of this study is that proteomic analysis cannot be directly linked to genetic diagnosis when disease-causing proteins show no quantitative changes. In such cases, the changes in the molecules associated with the pathogenic protein could provide the initial clues to the pathogenesis of the disease. However, we did not find such results in the current study. Despite these limitations, this study indicates that integrated analysis of PBMCs is a novel and valuable diagnostic tool for IEI to identify immune cell dysfunction that reflects disease pathogenesis and, in several cases, disease-causing proteins. Further improvements in proteomics data analysis and measurement sensitivity, in combination with its use in multilayered expression analysis with RNA-seq, will contribute to increases in diagnostic yield and a deeper understanding of IEI.

Materials and Methods

Clinical samples. Seventy IEI patients were recruited from five institutions in three cohorts, with 34, 28, and 8 patients, respectively. In addition, six HCs participated in another period. Throughout this paper, we refer to the cohorts as Batch1 (B1), Batch2 (B2), Batch3 (B3), or Ctrl (C), and patients are identified by group and a unique ID, for example, B1_P1, B2_P35, or B3_P63. Clinical information, such as classification from IUIS, presumptive diagnosis, and candidate genes, was obtained from clinicians. The primary inclusion criterion for IEI patients was the lack of genetic diagnosis via a canonical diagnostic approach such as WES or T-NGS; that is, patients without pathogenic variants in genes consistent with their clinical features and mode of inheritance, and the interpretation of "pathogenic" was according to the ACMG criteria(41). Therefore, when we identified no pathogenic variants, we designated them as "no candidate." On the other hand, when we identified variants that matched the clinical characteristics but did not meet the ACMG criteria or the mode of inheritance, we designated the gene as a "candidate gene."

The local ethics boards approved this study of Hiroshima University, Tokyo Medical and Dental University, National Defense Medical College, Gifu University, and Kyoto University.

Sample preparation. Methods for sample preparation are described in “SI methods”.

Proteomics and targeted RNA sequencing. Methods for Mass spectrometry-based proteomics and T-RNA-seq are described in “SI methods”.

Integrated proteomics and targeted RNA sequencing analysis. To understand the etiology and pathogenesis of IEI, we carried out three different approaches using R v4.1 and Bioconductor v3.14 packages.

Comparison of proteomics and targeted RNA sequencing in genetic diagnosis for inborn errors of immunity. First, to assess whether proteomic data could contribute to the genetic diagnosis, we examined changes in the abundance of proteins encoded by candidate genes in individual cases and compared these results with those of T-RNA-seq. It was impossible to investigate the DEA by comparing individual cases and HC because statistical significance is not a logical criterion in a single-case situation. Therefore, we analyzed the distribution of the protein abundance and the quantitative differences were calculated using z-scores. The absolute value of the z-score greater than two was defined as significant change. The absolute value of the z-score greater than or equal to 2 was defined as significant change. We also analyzed the quantitative differences between each case and the HCs to obtain further information about the biological significance. We calculated the log fold-change (LFC) and mean expression values using limma(49) and visualized the data using ggplot2 (R package). We also used Integrative Genomics Viewer (IGV) v2.8.7(50) to visualize aligned reads to detect sequence variants and allele-specific expression in T-RNA-seq.

Correlation analysis of proteomics and targeted RNA sequencing. Second, we examined the discrepancy between protein and mRNA expression levels. Based on the gene profiles identified by both proteomics and T-RNA-seq in 63 of the cases analyzed, the protein–mRNA correlation for each gene was analyzed using Spearman’s correlation coefficient. In addition, the correlation coefficients of genes specific to B, T, and NK cells were compared for later exploratory analysis. Cell-specific proteins were obtained from the database of Immune Cells(51) in The Human Protein Atlas(52). The degree of correlation was set as follows based on the absolute value of the correlation coefficient: 0.7 or higher is strong, 0.4 to 0.7 is moderate, 0.2 to 0.4 is weak, and 0.2 or lower is no correlation.

Exploratory analysis of B- and T-cell deficiency. Finally, we conducted an exploratory process to identify B-cell- or T-cell-deficient populations. In proteomic analysis, three cell-specific proteins were

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

selected according to the following criteria: (i) proteins with higher specificity and (ii) proteins without MVs or with fewer MVs. In T-RNA-seq, on the other hand, the analysis was based on gene profiles selected based on the criteria described in (i), since T-RNA-seq data are already target-enriched and contain no MVs. We then normalized the data with the z score using Genefilter(53), and performed a heatmap analysis of k-means clustering using ComplexHeatmaps(54). The k value was set to two to discriminate the data points into cell deficiency clusters and others, and the results of proteomics and T-RNA-seq were compared. The validity of the k-value was examined by PCA and the elbow method, which determines the optimal number of clusters. We performed differential expression analysis (DEA) on the comprehensive proteomic data to further validate the clustering results. DEA was compared in the cell-deficient cluster vs. others and was performed using DEP(55), which borrows its statistical models from limma(49). In the DEP results, P values of <0.05 and LFC of <-1.5 were set as the thresholds for significant differential expression. We then performed Gene Ontology (GO) enrichment analysis of significantly suppressed proteins using ClusterProfiler(56). GO terms related to biological processes were selected, and those with adjusted P values below 0.01 were considered significant.

ACKNOWLEDGMENTS. We thank all patients, their families, and attending physicians for participating in this study; Masao Kobayashi and members of the Pediatrics Laboratory at Hiroshima University Graduate School of Biomedical and Health Sciences for meaningful discussions; Yoshinori Hasegawa for conducting RNA sequencing; Masaki Takazawa for technical assistance. This study was supported in part by the Japan Agency for Medical Research and Development (AMED) (grant number: JP20ek0109480 to S.O.) and the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant numbers JP18KK0228, 19H03620 and 22H03041 to S.O.)

Data availability

The proteomic data underlying this article are available in [ProteomeXchange] at [http://www.proteomexchange.org], and can be accessed with [PXD038352]. And other data and codes used in this study are described in *SI Appendix*.

Author contributions statement

S.O. and O.O. designed research; K.T., E.T., K.K., M.T., H.N., K.I., K.M-S., Y.M., S.K., S.H., K.K., K.I., S.N., T.Y., H.O., and H.K. performed research; R.K., Y.K., and O.O. contributed new reagents/analytic tools; F.S., K.N., R.K., Y.K., and O.O. analyzed data; and F.S., T.A., O.O., and S.O. wrote the paper

1. SG Tangye, et al., Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* **40**, 24–64 (2020).
2. SG Tangye, et al., Human Inborn Errors of Immunity: 2022 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* (2022).
3. IJ Nijman, et al., Targeted next-generation sequencing: A novel diagnostic tool for primary immunodeficiencies. *J. Allergy Clin. Immunol.* **133**, 529–534.e1 (2014).
4. E Karimi, et al., Approach to genetic diagnosis of inborn errors of immunity through next-generation sequencing. *Mol. Immunol.* **137**, 57–66 (2021).
5. P Arts, et al., Exome sequencing in routine diagnostics: A generic test for 254 patients with primary immunodeficiencies. *Genome Medicine* **11**, 38 (2019).
6. EE Vorsteveld, A Hoischen, CI van der Made, Next-Generation Sequencing in the Field of Primary Immunodeficiencies: Current Yield, Challenges, and Future Perspectives. *Clin. Rev. Allergy & Immunol.* **61**, 212–225 (2021).
7. T Okano, et al., Whole-Exome Sequencing-Based Approach for Germline Mutations in Patients with Inborn Errors of Immunity. *J. Clin. Immunol.* **40**, 729–740 (2020).
8. A Stray-Pedersen, et al., Primary immunodeficiency diseases: Genomic approaches delineate heterogeneous Mendelian disorders. *The J. Allergy Clin. Immunol.* **139**, 232–245 (2017).
9. CD Platt, et al., Efficacy and economics of targeted panel versus whole-exome sequencing in 878 patients with suspected primary immunodeficiency. *The J. Allergy Clin. Immunol.* **147**, 723–726 (2021).
10. HD Gonorazky, et al., Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *The Am. J. Hum. Genet.* **104**, 466–483 (2019).
11. DF Vears, E Niemiec, HC Howard, P Borry, Analysis of VUS reporting, variant reinterpretation and recontact policies in clinical genomic sequencing consent forms. *Eur. J. Hum. Genet.* **26**, 1743–1751 (2018).
12. M Ma, et al., Disease-associated variants in different categories of disease located in distinct regulatory elements. *BMC Genomics* **16**, S3 (2015).
13. LS Kremer, et al., Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* **8**, 15824 (2017).
14. Z Wang, M Gerstein, M Snyder, RNA-Seq: A revolutionary tool for transcriptomics. *Nat. reviews. Genet.* **10**, 57–63 (2009).
15. F Curion, et al., Targeted RNA sequencing enhances gene expression profiling of ultra-low input samples. *RNA Biol.* **17**, 1741–1753 (2020).
16. TR Mercer, et al., Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat. Protoc.* **9**, 989–1009 (2014).
17. S Sakata, et al., Autosomal recessive complete STAT1 deficiency caused by compound heterozygous intronic mutations. *Int. Immunol.* **32**, 663–671 (2020).
18. D Merico, et al., Compound heterozygous mutations in the noncoding RNU4ATAC cause Roifman Syndrome by disrupting minor intron splicing. *Nat. Commun.* **6**, 8718 (2015).
19. S Ichikawa, et al., Reticular dysgenesis caused by an intronic pathogenic variant in AK2. *Cold Spring Harb. Mol. Case Stud.* **6**, a005017 (2020).
20. YH Chen, et al., Functional and structural analysis of cytokine-selective IL6ST defects that cause recessive hyper-IgE syndrome. *J. Allergy Clin. Immunol.* **148**, 585–598 (2021).
21. SG Tangye, et al., Hyper-IgE Syndrome due to an Elusive Novel Intronic Homozygous Variant in DOCK8. *J. Clin. Immunol.* **42**, 119–129 (2022).
22. BB Cummings, et al., Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Medicine* **9**, eaal5209 (2017).
23. L Frésard, et al., Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat. Medicine* **25**, 911–919 (2019).
24. Y Liu, A Beyer, R Aebersold, On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
25. C Vogel, EM Marcotte, Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (year?).
26. M Jovanovic, et al., Dynamic profiling of the protein life cycle in response to pathogens. *Science* (2015).
27. AFM Altelaar, J Munoz, AJR Heck, Next-generation proteomics: Towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14**, 35–48 (2013).
28. Y Kawashima, et al., Optimization of Data-Independent Acquisition Mass Spectrometry for Deep and Highly Sensitive Proteomic Analysis. *Int. J. Mol. Sci.* **20**, 5932 (2019).
29. S Adhikari, et al., A high-stringency blueprint of the human proteome. *Nat. Commun.* **11**, 5301 (2020).
30. A Sukumaran, et al., Decoding communication patterns of the innate immune system by quantitative proteomics. *J. leukocyte biology* **106**, 1221–1232 (2019).
31. Y Tong, et al., Data-independent acquisition-based quantitative proteomic analysis reveals differences in host immune response of peripheral blood mononuclear cells to sepsis. *Scand. J. Immunol.* **89**, e12748 (2019).
32. H Li, et al., A Comprehensive Proteome Analysis of Peripheral Blood Mononuclear Cells (PBMCs) to Identify Candidate Biomarkers of Pancreatic Cancer. *Cancer Genomics & Proteomics* **16**, 81–89 (2019).
33. N Jiang, et al., A data-independent acquisition (DIA)-based quantification workflow for proteome analysis of 5000 cells. *J. Pharm. Biomed. Analysis* **216**, 114795 (2022).
34. TW Kuijpers, et al., Combined immunodeficiency with severe inflammation and allergy caused by ARPC1B deficiency. *J. Allergy Clin. Immunol.* **140**, 273–277.e10 (2017).
35. P Grabowski, et al., Proteome Analysis of Human Neutrophil Granulocytes From Patients With Monogenic Disease Using Data-independent Acquisition. *Mol. & cellular proteomics: MCP* **18**, 760–772 (2019).
36. Z Sbihi, et al., Identification of Germline Non-coding Deletions in XIAP Gene Causing XIAP Deficiency Reveals a Key Promoter Sequence. *J. Clin. Immunol.* (2022).
37. H Nihira, et al., Detailed analysis of Japanese patients with adenosine deaminase 2 deficiency reveals characteristic elevation of type II interferon signature and STAT1 hyperactivation. *J. Allergy Clin. Immunol.* **148**, 550–562 (2021).
38. MA Rivas, et al., Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Sci. (New York, N.Y.)* **348**, 666–669 (2015).
39. L Jiang, et al., A Quantitative Proteome Map of the Human Body. *Cell* **183**, 269–283.e19 (2020).
40. MA Gildea, ZW Dwyer, JA Pleiss, Multiplexed primer extension sequencing: A targeted RNA-seq method that enables high-precision quantitation of mRNA splicing isoforms and rare pre-mRNA splicing intermediates. *Methods* **176**, 34–45 (2020).
41. S Richards, et al., Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. medicine : official journal Am. Coll. Med. Genet.* **17**, 405–424 (2015).
42. NT Deuitch, et al., TNF inhibition in vasculitis management in adenosine deaminase 2 deficiency (DADA2). *J. Allergy Clin. Immunol.* (2021).
43. B Lo, et al., AUTOIMMUNE DISEASE. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. *Sci. (New York, N.Y.)* **349**, 436–440 (2015).
44. CS Ma, SG Tangye, Flow Cytometric-Based Analysis of Defects in Lymphocyte Differentiation and Function Due to Inborn Errors of Immunity. *Front. Immunol.* **10**, 2108 (2019).
45. T Wolf, et al., Dynamics in protein translation sustaining T cell preparedness. *Nat. immunology* **21**, 927–937 (2020).
46. A Galloway, et al., Upregulation of RNA cap methyltransferase RNMT drives ribosome biogenesis during T cell activation. *Nucleic Acids Res.* **49**, 6722–6738 (2021).
47. L Peshkin, et al., On the Relationship of Protein and mRNA Dynamics in Vertebrate Embry-

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

onic Development. *Dev. Cell* **35**, 383–394 (2015).
48. K Becker, et al., Quantifying post-transcriptional regulation in the development of Drosophila melanogaster. *Nat. Commun.* **9**, 4970 (2018).
49. ME Ritchie, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
50. JT Robinson, et al., Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
51. M Uhlen, et al., A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* (2019).
52. M Uhlén, et al., Tissue-based map of the human proteome. *Science* (2015).
53. R Gentleman, VJ Carey, W Huber, F Hahne, Genefilter: Genefilter: Methods for filtering genes from high-throughput experiments (Bioconductor version: Release (3.14)) (2022).
54. Z Gu, R Eils, M Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
55. X Zhang, et al., Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat. Protoc.* **13**, 530–550 (2018).
56. G Yu, LG Wang, Y Han, QY He, clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* **16**, 284–287 (2012).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

Table 1. Summary of the results of the diagnostic analysis and its validation analysis

Patient ID	Pathogenic gene	Variants detected in prior genetic analysis	Results of diagnostic analysis	Genomic variants detected in follow-up analysis	Supportive laboratory findings
B1_P21	*BTK*	No pathogenic variants	Decreased protein and mRNA expression levels	c.-196+1G>T (variant in splice-site and cis-regulatory region)	B-cell defects via flow cytometry (0.1% of total lymphocytes)
B1_P22	*XIAP*	No pathogenic variants	Decreased protein and mRNA expression levels	Large deletion in promoter region(36)	Decreased XIAP expression in RT–PCR and WB(36)
B1_P29	*ADA2*	c.982G>A:p.Glu328Lys (heterozygous)	Decreased expression only at the protein level	• Aberrant splicing with intoronic variant of c.972+102T>G • Allele specific expression	Decreased ADA2 activity(37)
B2_P35	*LRBA*	c.1219_1220del: p.Leu408Valfs*7 (heterozygous)	Decreased expression only at the protein level	Being analyzed	• Decreased CTLA4 expression in Tregs • Decreased LRBA expression via WB

RT–PCR, reverse transcription PCR; WB, Western blotting; CTLA4, cytotoxic T-lymphocyte associated protein 4; Tregs, regulatory T cells

Table 2. Diagnostic efficiency in undiagnosed patients using WES or T-NGS

Method	Number of diagnosed patients	The increase in diagnostic efficiency	Notes
T-RNA-seq	3 (BTK deficiency, XIAP deficiency, ADA2 deficiency)	4%	ADA2 deficiency could possibly be diagnosed via T-RNA-seq alone by identifying aberrant splicing
Proteomics	1 (LRBA deficiency)	2%	• Proteomics was the only diagnostic evidence of LRBA deficiency • Proteomics provided supportive findings at the protein level in BTK, XIAP, and ADA2 deficiency
T-RNA-seq + Proteomics	4	6%	

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

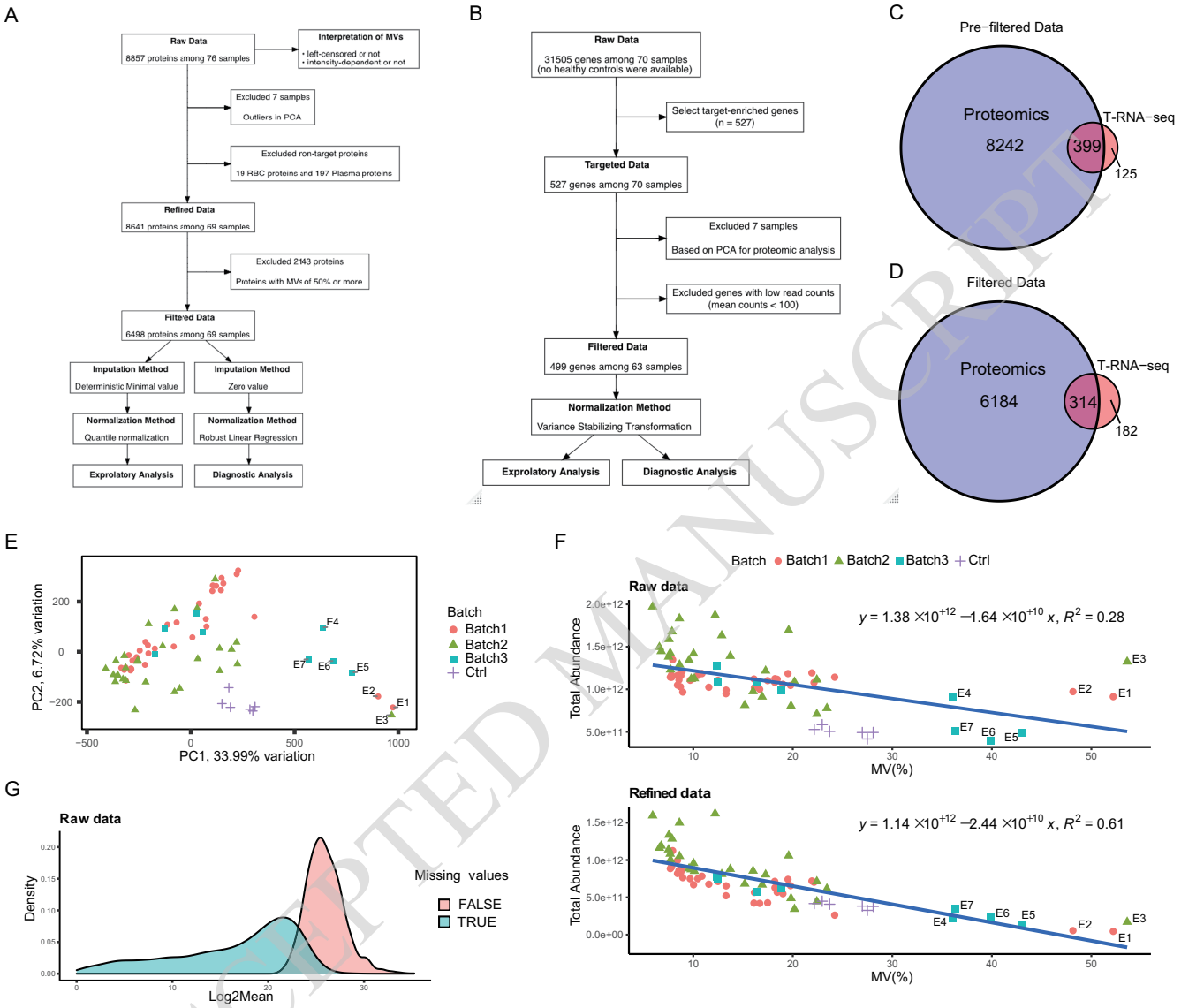


Fig. 1. Overview of the initial processing of proteomics and RNA-seq data. (A) Schematic diagram of proteomic analysis. DIA-MS yielded 8857 proteins from 70 IEI patients and six healthy donors. Interpretation for MVs was performed with raw data. MVs filtering was performed on 8641 protein data points among 69 patients, resulting in a filtered dataset of 6498 proteins. Downstream analysis was performed using two methods, each with optimal MVs imputation and normalization. (B) Schematic diagram of targeted RNA-seq (T-RNA-seq). No RNA-seq data from healthy controls were available. Quality control was performed with data from 527 target-enriched genes, yielding filtered data for 499 genes among 63 cases. (C) Venn diagrams of genes identified by proteomics and T-RNA-seq (8641 vs. 524). The blue circle reflects the proteomics data excluding RBC and plasma proteins, and the red circle reflects the targeted genes. Among the 527 targeted genes in T-RNA-seq, four noncoding RNAs were excluded. D, Venn diagram for filtered data (6498 vs. 496). Three noncoding RNAs were excluded from 499 genes. (E) PCA of raw proteomic data showing the eligibility of the data. Batches are indicated by shape and color. The x-axis shows the first principal component (PC1), and the y-axis shows the second principal component (PC2). Only excluded samples are labeled (E1 to E7). (F) Correlation of total protein abundance and MVs proportion. Protein intensity excluding RBCs and plasma proteins (targeted protein) is shown in the top figure, and the raw protein abundance is shown in the bottom figure. The linear regression, its formula, and R-squared values are shown in the figure. The shape and color coding are the same as in Figure E. (G) Density plot represents the distribution of protein abundance with or without MVs. The blue area contains proteins with MVs, and the red area contains proteins without MVs.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

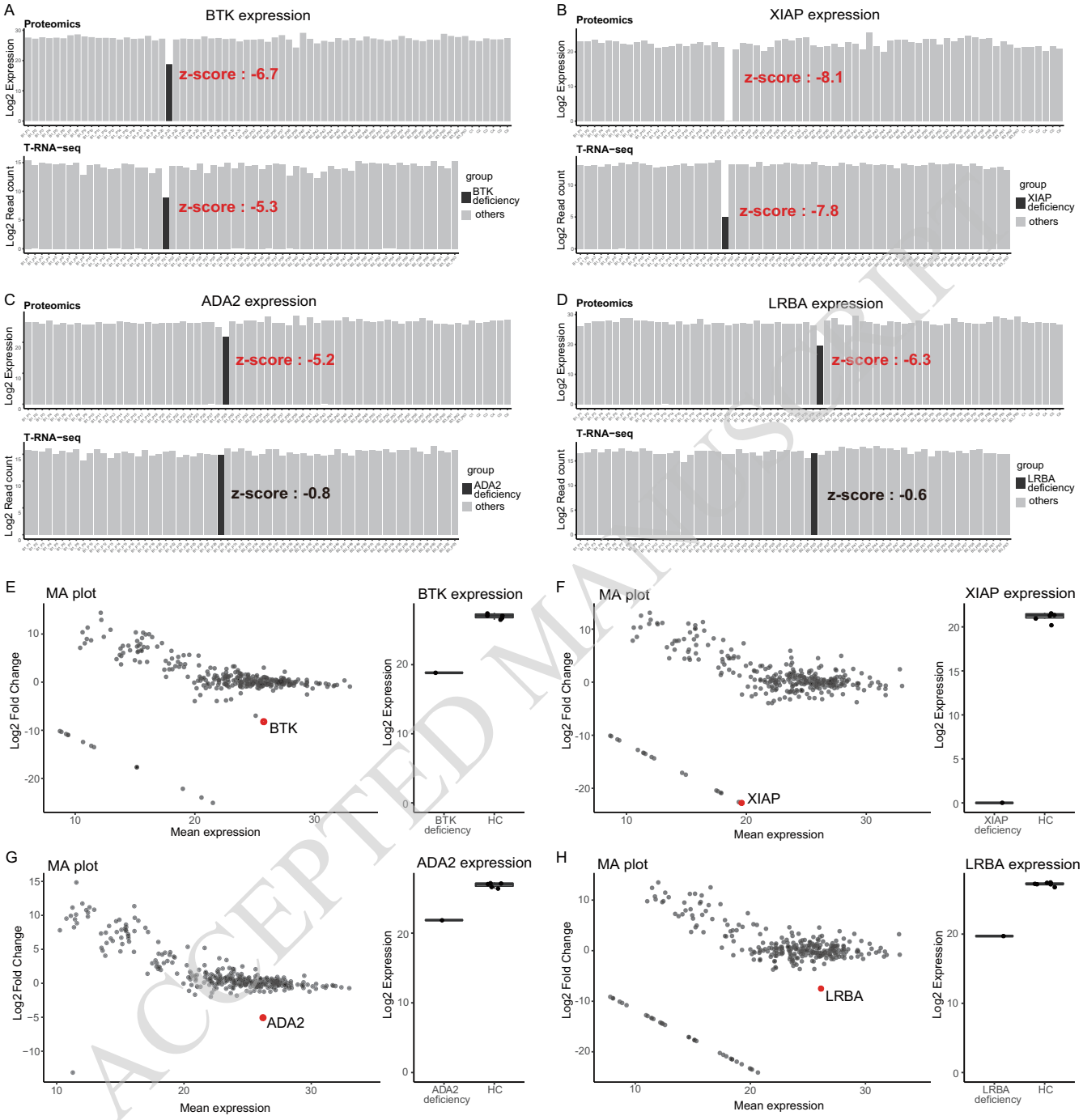


Fig. 2. Diagnostic analysis of disease-causing genes and Correlation analysis of proteomics and T-RNA-seq. (A), (B), (C), and (D) The bar plot shows the distribution of disease-causing protein and mRNA expression levels per sample. Throughout the figure, protein expression is shown at the top and mRNA expression is shown at the bottom; the case is shown in black, and other samples are shown in gray. (A); BTK deficiency, (B); XIAP deficiency, (C); ADA2 deficiency, (D); LRBA deficiency. (E), (F), (G), and (H) Decreased expression of disease-causing proteins compared to healthy controls (HCs). MA plot shows that the disease-causing protein is prominently downregulated (left panel). The x-axis shows the log mean expression of each protein, and the y-axis shows the log fold change of protein expression between the patient and HCs. The plots shown on a straight line in the lower left of the figure are proteins showing the MVs in the patients. The right panel shows the distribution of disease-causing protein expression in the patient and HCs. (E); BTK deficiency, (F); XIAP deficiency, (G); ADA2 deficiency, (H); LRBA deficiency.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

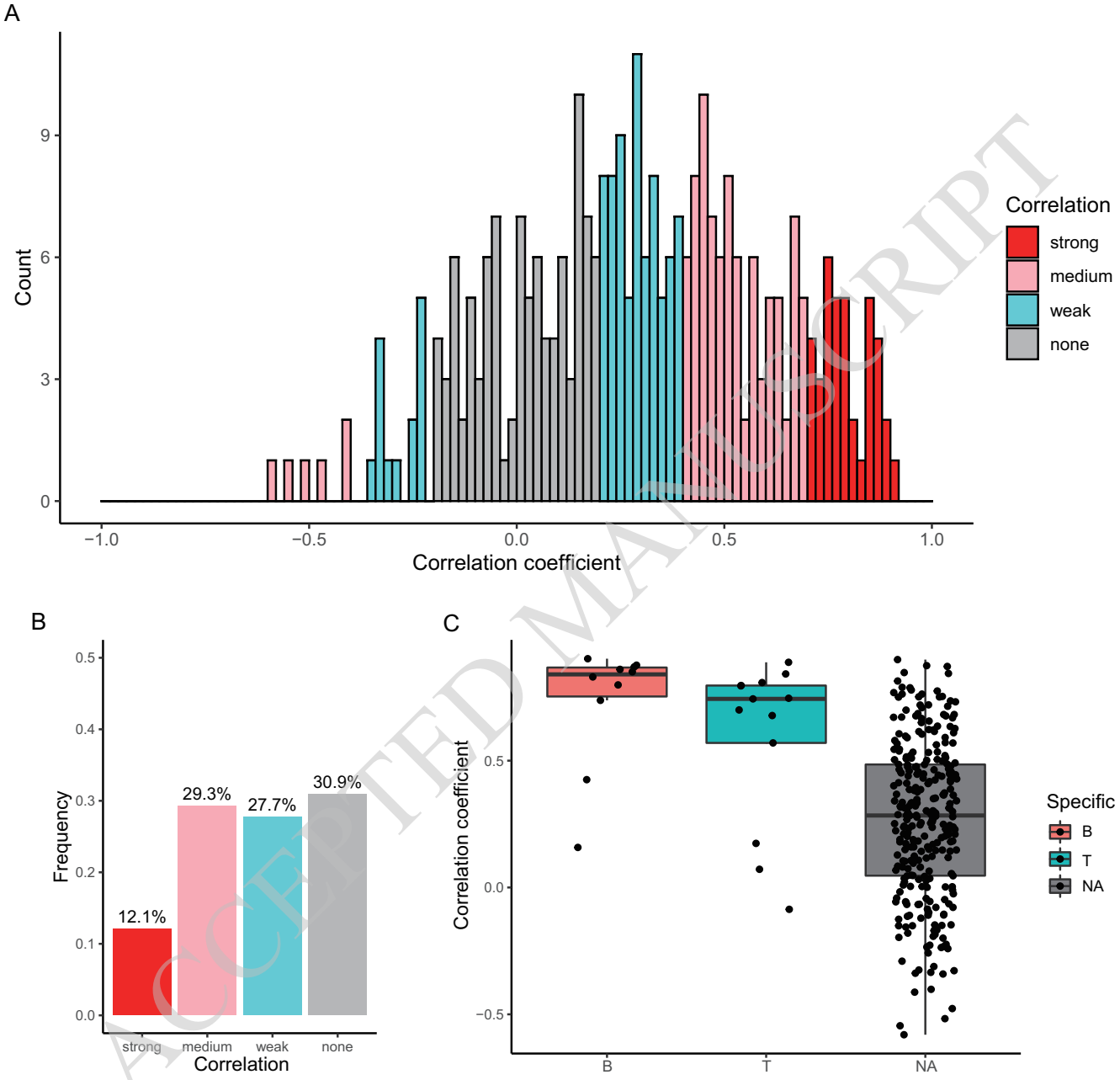


Fig. 3. Correlation analysis of IEI-related genes. (A) Spearman correlation coefficients of protein and mRNA levels for genes identified by proteomics and T-RNA-seq. The color scale reflects the degree of correlation, with red bars indicating a strong correlation, pink bars indicating a moderate correlation, blue bars indicating a weak correlation, and gray bars indicating no correlation. (B) The bar chart indicates the frequency of each degree of correlation among the 314 genes. The color coding is the same as in Figure E. (C) Jitter boxplot showing the distribution of the correlation coefficients for cell-specific genes. The color scale reflects specific cell types, with red indicating B cells, blue indicating T cells, and gray indicating genes that do not correspond to a specific cell type (NA).

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

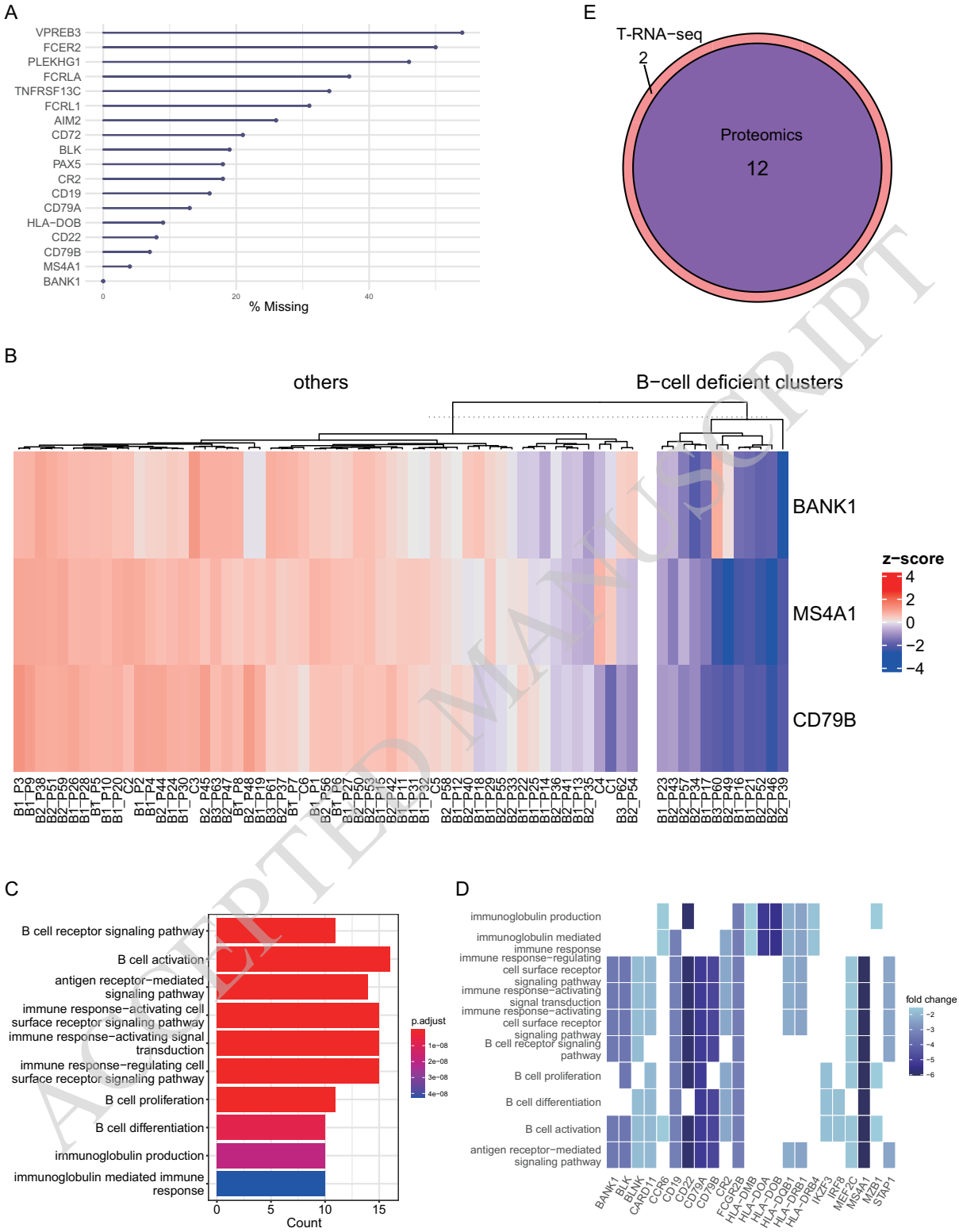


Fig. 4. Exploratory analysis of B-cell dysfunction. (A) Missing value analysis of B-cell-specific proteins. The x-axis shows the percentage of missing values among 69 samples. (B) The heatmap of k-means clustering shows cluster segregation with decreased expression of B-cell-specific proteins. The color scale reflects the z score, with red indicating a positive value and blue a negative value. (C) Top 10 enriched GO terms for proteins downregulated in the comparison of the B-cell-deficient cluster and others. The color scale reflects the adjusted P value of each GO term. (D) Heatmap of proteins associated with the top 10 enriched GO terms. The color scale shows the log fold change; the darker the color tone is, the lower the expression. (E) Venn diagram showing that all B-cell-deficient clusters identified by proteomics are included in the clusters identified by RNA-seq. The blue area indicates proteomics, and the red area indicates RNA-seq.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

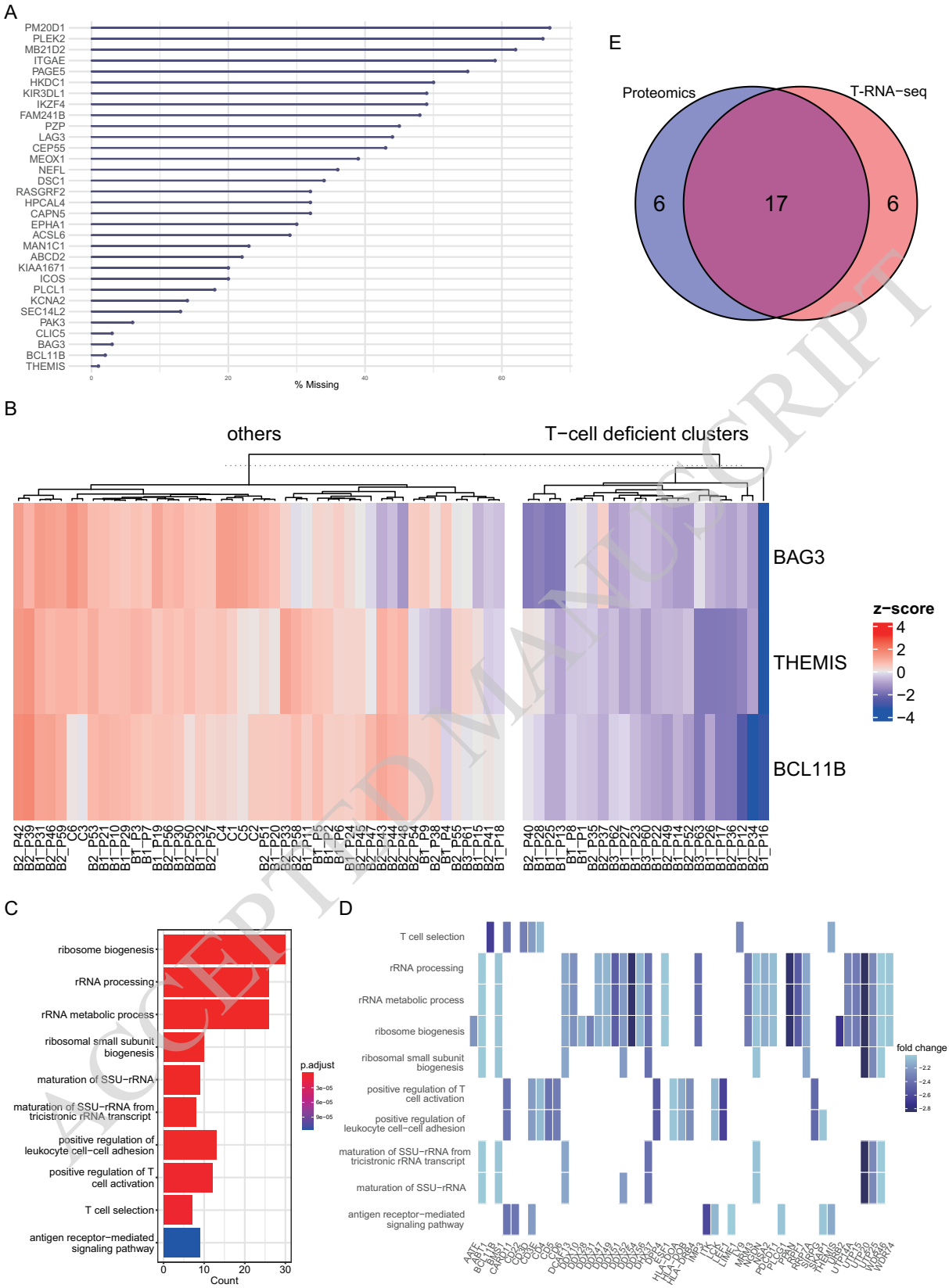


Fig. 5. Exploratory analysis of T-cell dysfunction. (A) Missing value analysis of T-cell-specific proteins. The x-axis indicates the percentage of missing values among the 69 samples. (B) The heatmap of k-means clustering shows cluster segregation with decreased expression of T-cell-specific proteins. The color scale reflects the z score, with red indicating a positive value and blue a negative value. (C) Top 10 enriched GO terms for proteins downregulated in the comparison of the T-cell-deficient cluster and others. The color scale reflects the adjusted P value of each GO term. (D) Heatmap of proteins associated with the top 10 enriched GO terms. The color scale reflects the log fold change; the darker the color tone is, the lower the expression. (E) Venn diagram of T-cell-deficient clusters identified by proteomics and RNA-seq. The blue area indicates proteomics, and the red area indicates RNA-seq.

ACCEPTED MANUSCRIPT

Downloaded from https://academic.oup.com/pnasnexus/advance-article/doi/10.1093/pnasnexus/pgad104/7091737 by USACO HANSHIN FREIGHT user on 30 March 2023

Supporting Information for

Demonstration of a complementary approach for genetic diagnosis of inborn errors of immunity using proteogenomic analysis

Fumiaki Sakura, Kosuke Noma, Takaki Asano, Kay Tanita, Etsushi Toyofuku, Kentaro Kato, Miyuki Tsumura, Hiroshi Nihira, Kazushi Izawa, Kanako Mitsui-Sekinaka, Ryo Konno, Yusuke Kawashima, Yoko Mizoguchi, Shuhei Karakawa, Seiichi Hayakawa, Hiroshi Kawaguchi, Kohsuke Imai, Shigeaki Nonoyama, Takahiro Yasumi, Hidenori Ohnishi, Hirokazu Kanegane, Osamu Ohara, and Satoshi Okada

Satoshi Okada
E-mail: sokada@hiroshima-u.ac.jp
Osamu Ohara
E-mail: ohara@kazusa.or.jp

This PDF file includes:

Supporting text
Figs. S1 to S6
Tables S1 to S5
Legends for Dataset S1 to S6
SI References

Other supporting materials for this manuscript include the following:

Datasets S1 to S6

Supporting Information Text

Methods

Sample preparation. PBMCs were isolated from EDTA-coated fresh peripheral blood using density-gradient centrifugation, and red blood cells (RBCs) were lysed with an erythrocyte lysis reagent. The cells were washed in PBS and centrifuged at 300 ×g three times. Then the cells were resuspended with TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and stored at -80 °C for further analysis.

Mass spectrometry-based proteomic. Proteins and RNAs were isolated from samples containing TRIzol according to a previously described method(1). The protein fraction was washed twice with 0.8 mL of acetonitrile and then dissolved in 0.5% sodium dodecanoate and 100 mM Tris-HCl, pH 8.5, using a water bath-type sonicator (Bioruptor II, Cosmo Bio, Tokyo, Japan). Pretreatment for shotgun proteomic analysis was performed as previously reported(2). For LC separation, mobile phases consisted of 0.1% (v/v) formic acid as solvent A and 0.1% (v/v) formic acid/80% (v/v) acetonitrile as solvent B. Each peptide sample (200 ng) was directly injected onto a 75 μm × 12 cm-nanoLC nano-capillary column (Nikkyo Technos Co., Ltd., Tokyo, Japan) at 40 °C and then separated with an 80 min gradient at a flow rate of 200 nl/min using an UltiMate 3000 RSLCnano LC system (Thermo Fisher Scientific). Peptides eluting from the column were analyzed on a Q Exactive HFX (Thermo Fisher Scientific) for overlapping window DIA-MS(2, 3). MS1 spectra were collected in the range of 495-785 m/z at 30,000 resolution to set an automatic gain control (AGC) target of 3×10^6 and maximum injection time of 55. MS2 spectra were collected in the range of more than 200 m/z at 30,000 resolution to set an AGC target of 3×10^6, maximum injection time of auto, and stepped normalized collision energy of 22%, 26%, and 30%. The isolation width for MS2 was set to 4 m/z, and overlapping window patterns in 500-780 m/z were used window placements optimized by Skyline v4.1(4). MS files were searched against a human spectral library using Scaffold DIA (Proteome Software, Inc., Portland, OR) as previously reported(5).

Initial processing of the proteomic data. To assess the eligibility of the population, we carried out principal component analysis (PCA) using PCAtools(6). Highly abundant and specific proteins of RBCs and plasma (Dataset S1), which were exemplified in a report by Byrk(7) and Lan(8), were excluded from the dataset because they impaired the analysis of the abundance of residual proteins of interest. Prior to handling missing values (MVs), we assessed the relationship between the protein abundance and MVs with the Pearson correlation coefficient to interpret the nature of the MVs. We then excluded proteins that showed MVs in more than 50% of cases to validate the eligibility of the data, and the remaining MVs were replaced with numerical values using NAguideR(9). To suppress batch effects, we performed normalization using NormalyzerDE(10), which considers several normalization methods and suggests the most effective method for controlling sample dispersion.

Procedures for targeted RNA sequencing. RNA was extracted and recovered from TRIzol using conventional methods. After the quantification and quality control of RNA on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), an adjusted amount of total RNA with RNA integrity numbers of 3 to 10 was treated with the NEBNext rRNA Depletion Kit (New England Biolabs, Ipswich, MA, USA) to deplete ribosomal RNA. Then, the NEBNext Ultra Ⅱ Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) was used for library preparation. In this procedure, cDNA libraries were captured and amplified for nine cycles with 527 targeted genes, including known IEI genes and PBMC markers (Dataset S6). After recovering the enriched T-RNA-seq libraries, the samples were run on a HiSeq2500 (Illumina, San Diego, CA, USA) in 50-nucleotide single-end mode. A mean read depth of approximately 20 million reads per sample was generated and stored in FASTQ format.

Preprocessing for targeted RNA sequencing data analysis. The FASTQ files were quality controlled and trimmed by sickle(11), then aligned to the GENCODE human reference genome GRCh38.p13 with STAR aligner v2.6(12). Reads were quantified with RSEM v1.3.3(13) to obtain the feature read counts per gene, and the identified genes were refined to the targeted 527 genes. After filtering the targeted genes based on the total read counts, we normalized the data for downstream analysis (Dataset S1).

Detection of splicing outliers in targeted RNA sequencing. Splicing outliers were disclosed using LeafCutter(14) with four inconclusive cases as controls and then visualized with LeafViz(14). LeafCutter utilizes a Dirichlet-multinomial generalized linear model to identify differential splicing based on read counts in an intron cluster. The percentage spliced index (PSI) was used to measure the relative expression of transcript isoforms within intron clusters. The change in PSI (ΔPSI) was used to quantify the relative differential expression of each transcript isoform between the case and controls.

Flow cytometry analysis of CTLA4. PBMCs were isolated from the patient with LRBA deficiency and the healthy control and resuspended at 1×10^6 /ml in RPMI medium with 10% fetal bovine serum. Cells were then incubated with anti-CD3/CD28 antibodies (Miltenyi Biotec, Bergisch Gladbach, Germany) for 16 hrs. After incubation, cells were washed and stained with anti-CD4-FITC (BioLegend, San Diego, CA, USA). Following fixation and permeabilization, cells were stained with anti-FOXP3-Alexa Fluor 647 (BioLegend) and anti-CTLA4-PerCP (BioLegend). Data were collected with a BD FACSVerse (BD Biosciences, Franklin Lakes, NJ, USA) and analyzed with FlowJo software (BD Life Sciences).

Western blotting of LRBA. Proteins lysates from PBMCs of the patient and healthy control were separated on 10% acrylamide gel, followed by transfer onto the PVDF membrane. LRBA was detected using the primary antibody of polyclonal rabbit anti-LRBA/BGL (Abcam, Kenbridge, UK) and the secondary antibody of ECL anti-rabbit IgG (GE Healthcare, Uppsala, Sweden). Beta-actin was used as a loading control.

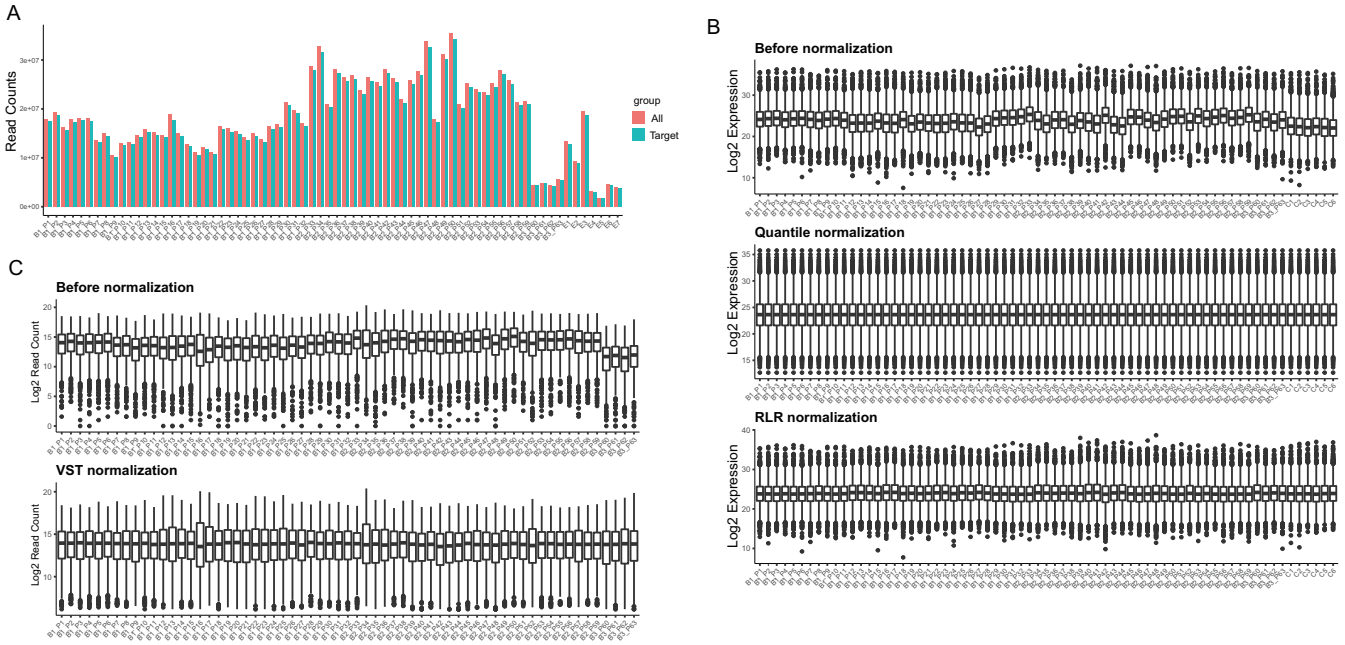


Fig. S1. Interpretation and optimization of the data. (A) The efficiency of target enrichment in RNA-seq. Red bars show total read counts, and blue bars show targeted read counts. (B) Comparison of the distribution of expression profiles in raw and normalized proteomic data (top; Raw data, middle; Quantile normalization, bottom; Robust Linear Regression normalization). (C) Comparison of the distribution of expression profiles in raw and normalized T-RNA-seq data (top; Raw data, bottom; variance stabilizing normalization).

Fumiaki Sakura, Kosuke Noma, Takaki Asano, Kay Tanita, Etsushi Toyofuku, Kentaro Kato, Miyuki Tsumura, Hiroshi Nihira, Kazushi Izawa, Kanako Mitsui-Sekinaka, Ryo Konno, Yusuke Kawashima, Yoko Mizoguchi, Shuhei Karakawa, Seiichi Hayakawa, Hiroshi Kawaguchi, Kohsuke Imai, Shigeaki Nonoyama, Takahiro Yasumi, Hidenori Ohnishi, Hirokazu Kanegane, Osamu Ohara, and Satoshi Okada

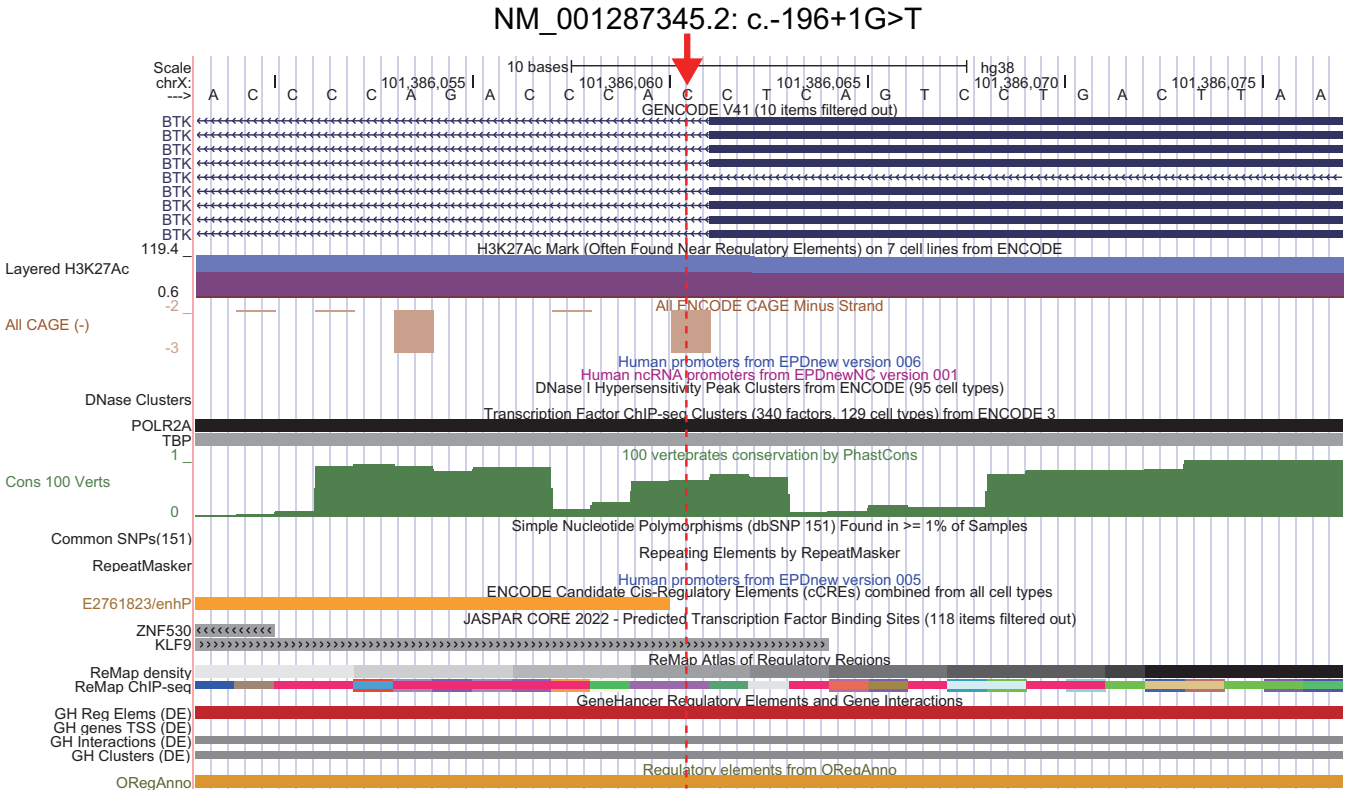
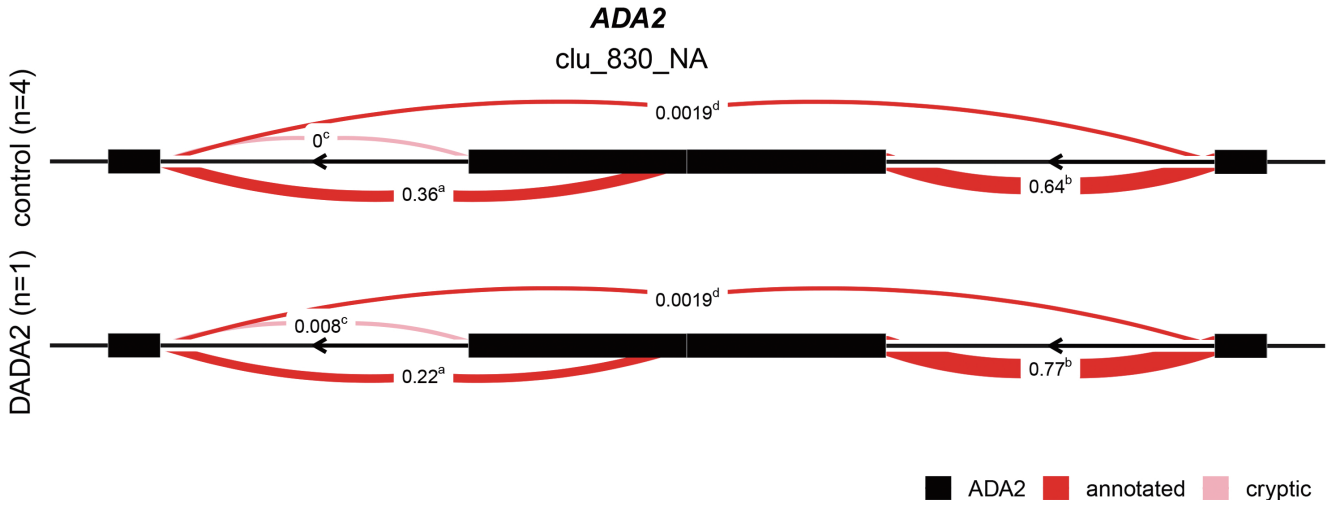


Fig. S2. UCSC Genome Browser view for the splice-site variant in *BTK* (GRCh38/hg38). The red arrow and dashed line indicate the position of c.-196+1. The GENCODE Genes track of *BTK* is shown in the top panel. Below that, transcriptional regulatory elements from the database of ENCODE, EPDnew, GeneHancer, JASPARCORE 2022, ORegAnno, and ReMap Atras are shown. Green bars in the center track indicate evolutionary conservation in 100 vertebrate species.

A



	chr	start	end	verdict	dPSI
a	chr22	17188447	17189942	annotated	−0.133
b	chr22	17190032	17191683	annotated	0.125
c	chr22	17188447	17189840	cryptic_fiveprime	0.008
d	chr22	17188447	17191683	annotated	−0.001

B

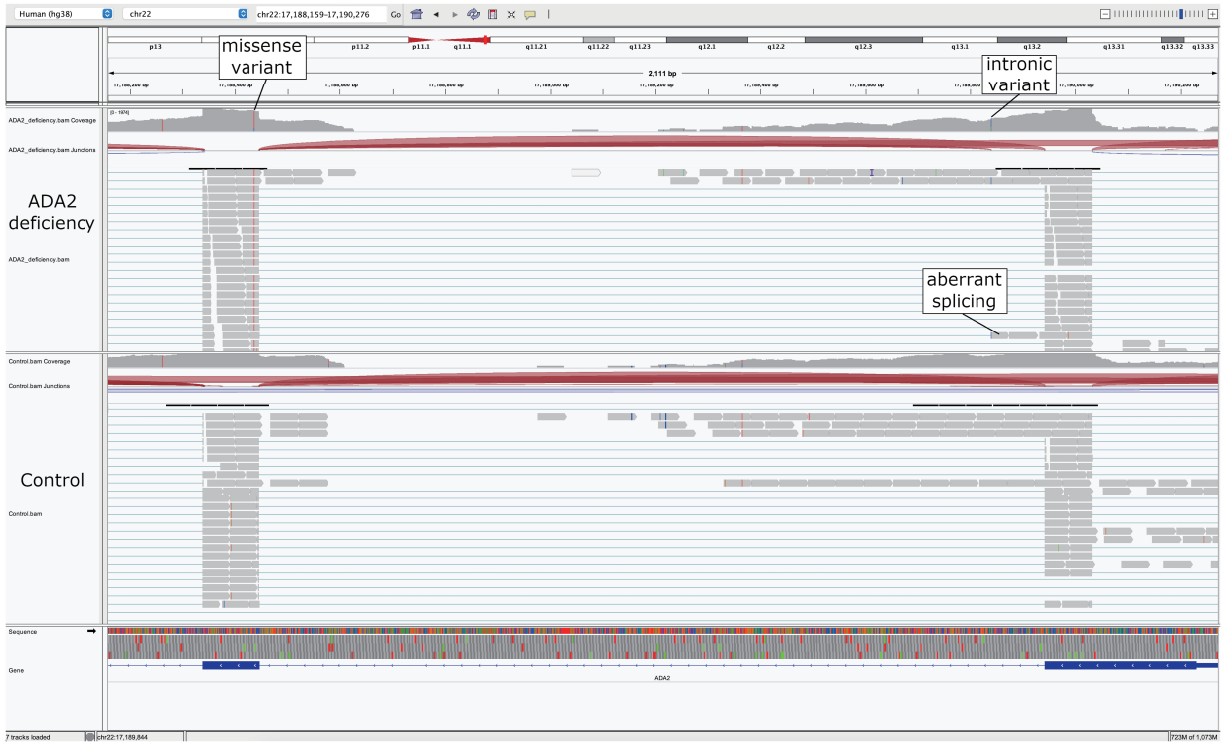


Fig. S3. Abnormal findings of T-RNA-seq in the patient wiht ADA2 deficiency. (A) Visualization of differential splicing in comparison with ADA2 deficiency and four controls. The red lines indicate splicing junctions connecting to the exons (shown in black). All junctions are annotated with percentage spliced index (PSI), reflecting the ratio of junction reads within the cluster. The PSI value of controls is 0 in the cryptic junction (shown in pink), which means that the junction is specific to the patient with ADA2 deficiency. (B) Visualization of the abnormal findings in T-RNA-seq by IGV. Heterozygous missense variant shows allele-specific expression in T-RNA-seq, with the variant allele accounting for 80% of all reads and the wild type for 20%. Abnormal splicing harbors a heterozygous intron variant at the 5' end, generating an inappropriate donor site. Intronic and missense variants are carried by different reads, indicating compound heterozygosity.

Fumiaki Sakura, Kosuke Noma, Takaki Asano, Kay Tanita, Etsushi Toyofuku, Kentaro Kato, Miyuki Tsumura, Hiroshi Nihira, Kazushi Izawa, Kanako Mitsui-Sekinaka, Ryo Konno, Yusuke Kawashima, Yoko Mizoguchi, Shuhei Karakawa, Seiichi Hayakawa, Hiroshi Kawaguchi, Kohsuke Imai, Shigeaki Nonoyama, Takahiro Yasumi, Hidenori Ohnishi, Hirokazu Kanegane, Osamu Ohara, and Satoshi Okada

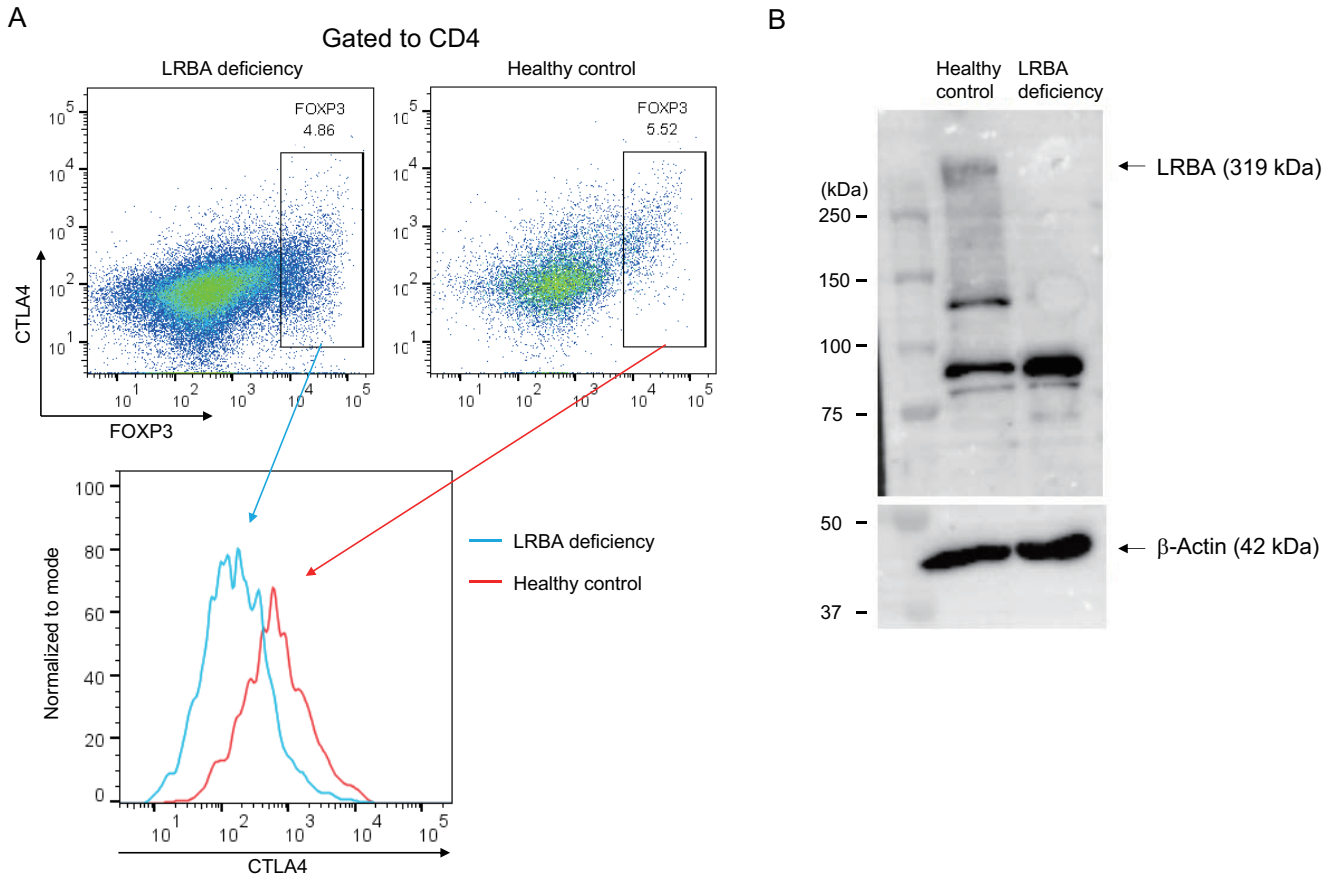


Fig. S4. Laboratory findings of an LRBA-deficient case. (A) CTLA4 expression on CD4+FOXP3+ Tregs in flow cytometric analysis. The upper panel shows the expression of CTLA4 and FOXP3 in PBMCs stimulated with anti-CD3/CD28 antibodies. The bottom panel shows the reduced CTLA4 expression of Tregs in the patient compared to the healthy control. (B) LRBA expression in Western blotting. LRBA protein (319 kDa) is not detected on PBMCs in the patient compared to the healthy control.

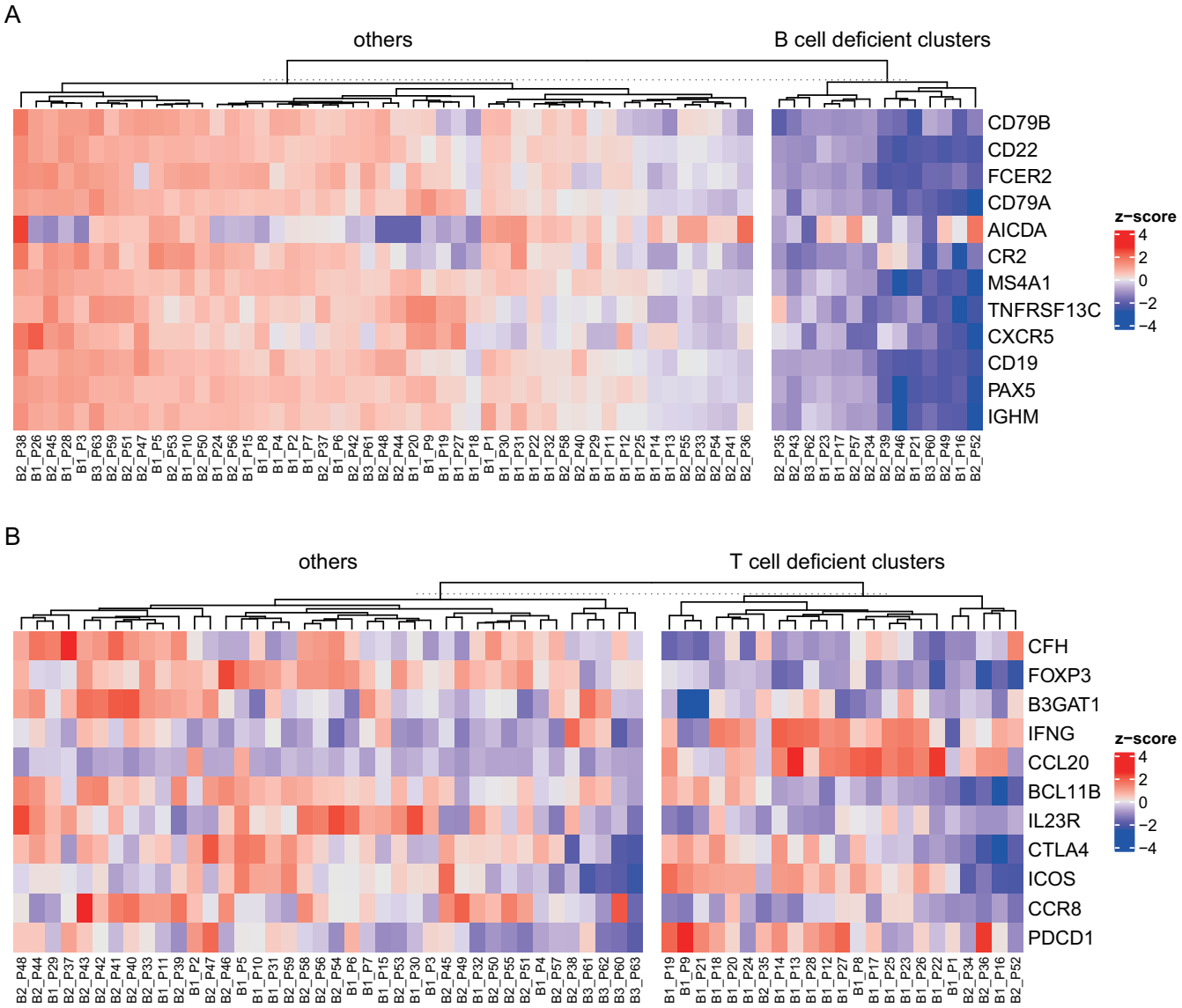


Fig. S5. Hierarchical clustering of B- and T-cell deficiency in T-RNA-seq. (A) The heatmap of k-means clustering shows cluster segregation with decreased expression of B-cell-specific genes. The color scale reflects the z score, with red indicating a positive value and blue a negative value. (B) The heatmap of k-means clustering of T-cell-sepecific genes in T-RNA-seq.

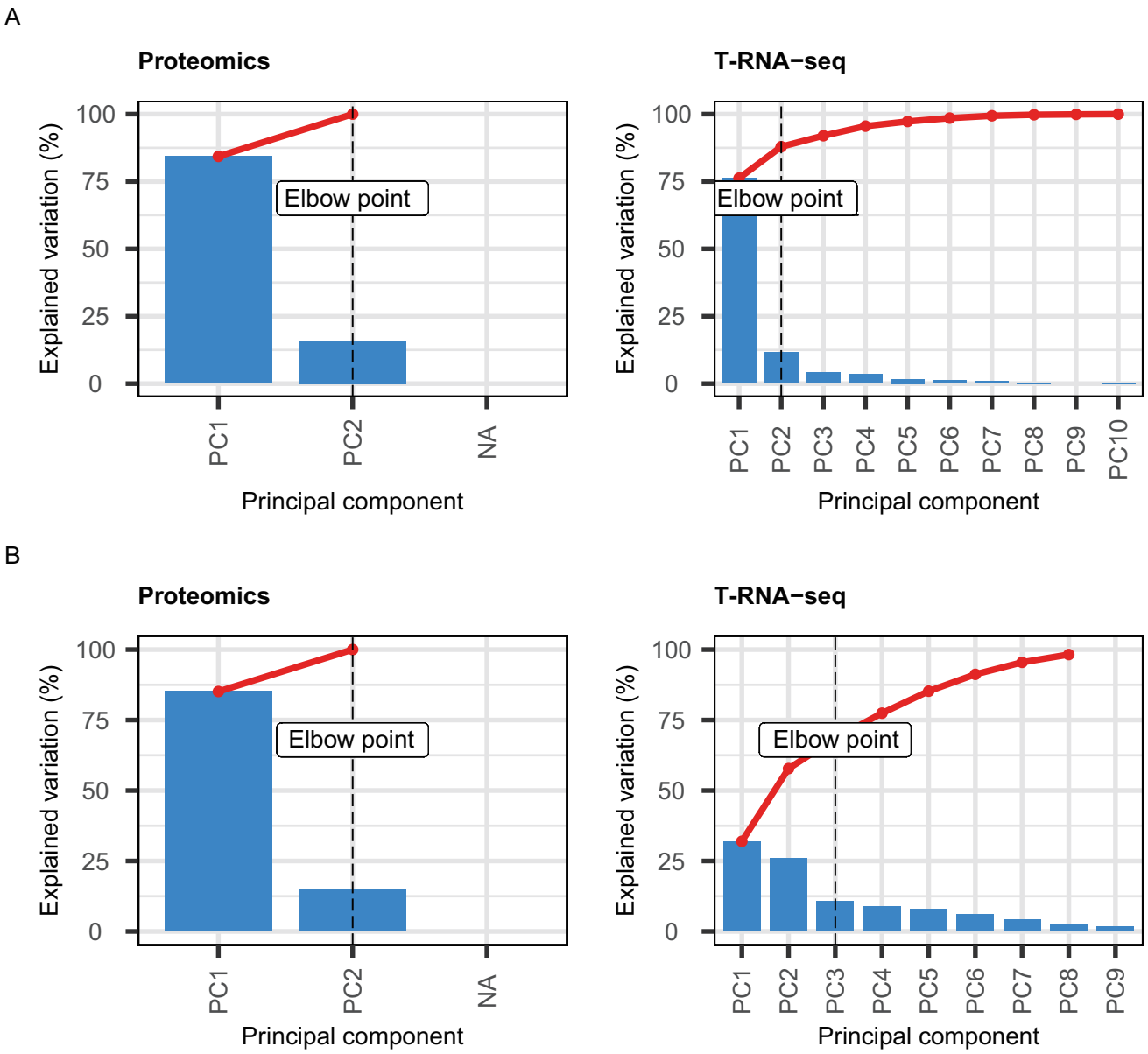


Fig. S6. Comparison of clustering analysis of B-cell and T-cell in proteomics and T-RNA-seq. (A) Scree plot of principal components as criteria for clustering in B-cell. Throughout all figures, the x-axis shows all principal components, and the y-axis shows the proportion of the variable explained by each principal component. The broken red line indicates the cumulative proportion of the variables explained by the principal components. The elbow point suggests the optimal number of clusters, which means the optimal k-value in the k-means method. (B) Scree plot of principal components as criteria for clustering in T-cell. The elbow point in RNA-seq is the fourth principal component (all other elbow points are second principal components)

Table S1. Characteristics of patients and the result of prior genetic studies

Patient ID	IUIS classification	Clinical diagnosis	Clinical phenotypes	Genomic analysis	Candidate genes (inheritance mode)	Variants (ACMG Classification)
B1_P1	4	Immune dysregulation	Thrombocytopenia, Splenomegaly	WES	No candidate	N.A.
B1_P2	3	CVID	Hypogammaglobulinemia, Atopic dermatitis, asthma	WES	No candidate	N.A.
B1_P3	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P4	3	CVID	Hypogammaglobulinemia, ADEM, Recurrent infection	WES	No candidate	N.A.
B1_P5	7	Autoinflammation	Recurrent fever, Abdominal pain	WES	No candidate	N.A.
B1_P6	7	Autoinflammation	Recurrent fever, Hepatitis	WES	No candidate	N.A.
B1_P7	4	Lymphomatoid granulomatosis	Chronic EBV infection, Polyneuropathy	WES	No candidate	N.A.
B1_P8	1	LOCID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P9	4	Immune dysregulation	B-LPD	WES	No candidate	N.A.
B1_P10	3	CID	Recurrent infection, T-cell dysfunction	WES	No candidate	N.A.
B1_P11	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P12	3	CVID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B1_P13	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P14	1	LOCID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B1_P15	3	CVID	Hypogammaglobulinemiaa	WES	No candidate	N.A.
B1_P16	3	CVID	Idiopathic enteritis	T-NGS	No candidate	N.A.
B1_P17	1	LOCID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B1_P18	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P19	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P20	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P21	3	XLA (BTK deficiency)	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P22	4	XIAP deficiency	Recurrent HLH	T-NGS	No candidate	N.A.
B1_P23	3	CVID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B1_P24	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P25	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P26	2	EDA-ID	Ectodermal dysplasia	WES	No candidate	N.A.
B1_P27	3	CVID	Hypogammaglobulinemia	WES	No candidate	N.A.
B1_P28	1	CID	EBV-lymphoma	WES	No candidate	N.A.
B1_P29	7	Autoinflammation	Cerebral infraction	T-NGS	*ADA2* (AR)	c.982G>A hetero (VUS, but known pathogenic(15))
B1_P30	7	Juvenile Behçet's disease	Aphthous stomatitis	WES	No candidate	N.A.
B1_P31	4	Immune dysregulation	No data	WES	No candidate	N.A.
B1_P32	7	Autoinflammation	Recurrent fever, Aseptic meningitis	WES	No candidate	N.A.
B2_P33	3	CVID	Hypogammaglobulinemiaa, Recurrent infection	T-NGS	No candidate	N.A.
B2_P34	3	APDS	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B2_P35	4	LRBA deficiency	AIHA, AIN, ITP	WES	*LRBA* (AR)	c.1219_1220delTT hetero (Likely Pathogenic)
B2_P36	4	IBD	Abdominal pain, Recurrent diarrhea	T-NGS	No candidate	N.A.
B2_P37	1	SCID	Recurrent infection	WES	No candidate	N.A.
B2_P38	6	MSMD	BCG myelitis	WES	No candidate	N.A.
B2_P39	3	CVID	Hyper IgM, ITP, SLE	WES	*TCF3* (AD, AR)	c.319G>A hetero (VUS) / c.1592C>T hetero (Benign)
B2_P40	3	CVID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B2_P41	3	CVID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B2_P42	3	CVID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B2_P43	3	CVID	Hypogammaglobulinemia	T-NGS	*TRAF3* (AD)	c.1415C>T hetero (VUS)
B2_P44	3	CVID	Hypogammaglobulinemia	T-NGS	*MSH2* (AD)	c.118G>A hetero (VUS)
B2_P45	2	MOPD1	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B2_P46	3	CVID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B2_P47	3	CVID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B2_P48	3	CVID	Hypogammaglobulinemia	T-NGS	No candidate	N.A.
B2_P49	4	CID	Hashimoto thyroiditis, Type 1 diabetes mellitus	WES	No candidate	N.A.
B2_P50	7	CINCA syndrome	Vasculitis, Sensorineural hearing impairment	WES	No candidate	N.A.
B2_P51	7	Interferonopathy	Periodic fever, Asthma	WES	No candidate	N.A.
B2_P52	4	CID	Evans syndrome, B-cell deficiency	WES	No candidate	N.A.
B2_P53	7	Interferonopathy	Periodic fever, Failure to thrive	WES	No candidate	N.A.
B2_P54	7	Autoinflammation	Systemic vasculitis	WES	*PIK3CD* (AD, AR)	c.2689G>A hetero (VUS)
B2_P55	7	Autoinflammation	Unexplained fever	WES	No candidate	N.A.
B2_P56	7	Interferonopathy	Unexplained fever, Spondylitis	WES	No candidate	N.A.
B2_P57	7	JIA	Macrophage activation syndrome	WES	No candidate	N.A.
B2_P58	7	Interferonopathy	Recurrent fever, Aphthous stomatitis	WES	No candidate	N.A.
B2_P59	4	APDS	Hyper IgE syndrome	WES	No candidate	N.A.
B3_P60	3	CVID	Hypogammaglobulinemia, Skeletal dysplasia	WES	No candidate	N.A.
B3_P61	3	CVID	Hypogammaglobulinemia	T-NGS	No cadidate	N.A.
B3_P62	3	CVID	Hypogammaglobulinemia, Abnormal pigmentation	WES	No candidate	N.A.
B3_P63	7	IBD	Multiple intestinal stricture, Interstitial pneumonia	T-NGS	No candidate	N.A.
E1	3	CVID	Hypogammaglobulinemiaa	WES	No candidate	N.A.
E2	3	CVID	Hypogammaglobulinemia	WES	No cadidate	N.A.
E3	3	CVID	Hypogammaglobulinemia, focal epilepsy, ASD	WES	*RTEL1* (AD, AR)	c.3064C>G hetero (VUS)
E4	4	APS	No data	WES	*PEPD* (AR)	c.410C>T hetero (VUS) / c.1291C>T hetero (VUS)
E5	3	LOCID	Reccurent pneumonia	WES	No candidate	N.A.
E6	4	Immune dysregulation	ALPS like	WES	No candidate	N.A.
E7	7	IBD	Recurrent fever	T-NGS	No candidate	N.A.

AR, autosomal recessive; AD, autosomal dominant; VUS, variant of unknown significance

Table S2. IUIS classification, Clinical diagnosis, and Lymphocyte subset in the B-cell-deficient clusters in proteomics

Patient ID	IUIS classification	Clinical diagnosis	B-cell %/lymphocytes	T-cell %/lymphocytes
B1_P16	3	CVID	0	57.2
B1_P17	1	LOCID	13.1	67.4
B1_P21	3	XLA	0.1	65.5
B1_P23	3	CVID	2.2	77.3
B2_P34	3	APDS	2.2	42.5
B2_P39	3	CVID	N.A.	N.A.
B2_P43	3	CVID	0.8	95
B2_P46	3	CVID	N.A.	N.A.
B2_P49	4	CID	N.A.	N.A.
B2_P52	4	CID	N.A.	N.A.
B2_P57	7	JIA	N.A.	N.A.
B3_P60	3	CVID	N.A.	N.A.

Table S3. Characteristics of the B-cell-deficient clusters in T-RNA-seq

Patient ID	IUIS classification	Clinical diagnosis	B-cell %/lymphocytes	T-cell %/lymphocytes
B1_P16	3	CVID	0	57.2
B1_P17	1	LOCID	13.1	67.4
B1_P21	3	XLA (BTK deficiency)	0.1	65.5
B1_P23	3	CVID	2.2	77.3
B2_P34	3	APDS	2.2	42.5
B2_P35	4	LRBA deficiency	7.7	86.1
B2_P39	3	CVID	N.A.	N.A.
B2_P43	3	CVID	0.8	95
B2_P46	3	CVID	N.A.	N.A.
B2_P49	4	CID	N.A.	N.A.
B2_P52	4	CID	N.A.	N.A.
B2_P57	7	JIA	N.A.	N.A.
B3_P60	3	CVID	N.A.	N.A.
B3_P62	3	CVID	N.A.	N.A.

Fumiaki Sakura, Kosuke Noma, Takaki Asano, Kay Tanita, Etsushi Toyofuku, Kentaro Kato, Miyuki Tsumura, Hiroshi Nihira, Kazushi Izawa, Kanako Mitsui-Sekinaka, Ryo Konno, Yusuke Kawashima, Yoko Mizoguchi, Shuhei Karakawa, Seiichi Hayakawa, Hiroshi Kawaguchi, Kohsuke Imai, Shigeaki Nonoyama, Takahiro Yasumi, Hidenori Ohnishi, Hirokazu Kanegane, Osamu Ohara, and Satoshi Okada

Table S4. IUIS classification, clinical diagnosis, and lymphocyte subset in the T-cell-deficient clusters in proteomics

Patient ID	IUIS classification	Clinical diagnosis	T-cell %/lymphocytes	CD4+ T-cell %/CD3+ T-cell	CD8+ T-cell %/CD3+ T-cell	B-cell %/lymphocytes
B1_P1	4	Immune dysregulation	53.9	58.6	38.0	20.9
B1_P8	1	LOCID	90.4	37.6	54.1	1.1
B1_P12	3	CVID	68.9	57.1	36.3	4.1
B1_P13	3	CVID	73.4	38.6	52.7	9.4
B1_P14	1	LOCID	71.7	78.3	11.4	2.4
B1_P16	3	CVID	57.2	34.1	54.2	0.0
B1_P17	1	LOCID	67.4	61.4	29.9	13.1
B1_P22	4	XIAP deficiency	N.A.	N.A.	N.A.	N.A.
B1_P23	3	CVID	77.3	29.3	66.3	2.2
B1_P25	3	CVID	77.4	38.7	54.0	10.6
B1_P26	2	EDA-ID	33.4	70.3	29.7	32.4
B1_P27	3	CVID	71.4	37.4	42.8	2.0
B1_P28	1	CID	56.8	40.9	56.6	35.6
B2_P34	3	APDS	42.5	42.1	45.8	2.2
B2_P35	4	LRBA deficiency	86.1	37.3	57.1	7.7
B2_P36	4	IBD	83.4	47.8	48.6	6.8
B2_P37	1	SCID	46.8	48.8	41.4	6.3
B2_P40	3	CVID	N.A.	N.A.	N.A.	N.A.
B2_P49	4	CID	N.A.	N.A.	N.A.	N.A.
B2_P52	4	CID	N.A.	N.A.	N.A.	N.A.
B3_P60	3	CVID	N.A.	N.A.	N.A.	N.A.
B3_P62	3	CVID	N.A.	N.A.	N.A.	N.A.
B3_P63	7	IBD	N.A.	N.A.	N.A.	N.A.

Table S5. Characteristics of the T-cell-deficient cluster in T-RNA-seq

Patient ID	IUIS classification	Clinical diagnosis	T-cell %/lymphocytes	CD4+ T-cell %/CD3+ T-cell	CD8+ T-cell %/CD3+ T-cell	B-cell %/lymphocytes
B1_P1	4	Immune dysregulation	53.9	58.6	38.0	20.9
B1_P4	3	CVID	91.6	83.8	11.0	2.7
B1_P8	1	LOCID	90.4	37.6	54.1	1.1
B1_P9	4	Immune dysregulation	N.A.	N.A.	N.A.	N.A.
B1_P12	3	CVID	68.9	57.1	36.3	4.1
B1_P13	3	CVID	73.4	38.6	52.7	9.4
B1_P14	1	LOCID	71.7	78.3	11.4	2.4
B1_P16	3	CVID	57.2	34.1	54.2	0.0
B1_P17	1	LOCID	67.4	61.4	29.9	13.1
B1_P18	3	CVID	84.2	68.9	21.1	0.47
B1_P19	3	CVID	79.0	52.0	42.0	13.6
B1_P20	3	CVID	73.4	62.9	31.5	10.5
B1_P21	3	XLA (BTK deficiency)	65.6	51.8	30.2	0.1
B1_P22	4	XIAP deficiency	N.A.	N.A.	N.A.	N.A.
B1_P23	3	CVID	77.3	29.3	66.3	2.2
B1_P24	3	CVID	73.8	69.1	21.2	20
B1_P25	3	CVID	77.4	38.7	54.0	10.6
B1_P26	2	EDA-ID	33.4	70.3	29.7	32.4
B1_P27	3	CVID	71.4	37.4	42.8	2.0
B1_P28	1	CID	56.8	40.9	56.6	35.6
B2_P34	3	APDS	42.5	42.1	45.8	2.2
B2_P35	4	LRBA deficiency	86.1	37.3	57.1	7.7
B2_P36	4	IBD	83.4	47.8	48.6	6.8
B2_P38	6	MSMD	N.A.	N.A.	N.A.	N.A.
B2_P52	4	CID	N.A.	N.A.	N.A.	N.A.
B3_P62	3	CVID	N.A.	N.A.	N.A.	N.A.
B3_P63	7	IBD	N.A.	N.A.	N.A.	N.A.

Fumiaki Sakura, Kosuke Noma, Takaki Asano, Kay Tanita, Etsushi Toyofuku, Kentaro Kato, Miyuki Tsumura, Hiroshi Nihira, Kazushi Izawa, Kanako Mitsui-Sekinaka, Ryo Konno, Yusuke Kawashima, Yoko Mizoguchi, Shuhei Karakawa, Seiichi Hayakawa, Hiroshi Kawaguchi, Kohsuke Imai, Shigeaki Nonoyama, Takahiro Yasumi, Hidenori Ohnishi, Hirokazu Kanegane, Osamu Ohara, and Satoshi Okada

SI Dataset S1 (proteome_data.xlsx)
Data and codes of proteomic analysis

SI Dataset S2 (RNA-seq_data.xlsx)
Data and codes of targeted RNA sequencing

SI Dataset S3 (diagnostic_analysis.xlsx)
Data and codes of diagnostic analysis

SI Dataset S4 (protein_RNA_correlation.xlsx)
Data for the correlation analysis of protein and RNA expression levels

SI Dataset S5 (exploratory_analysis.xlsx)
Data and codes of exploratory analysis

SI Dataset S6 (RNAtarget_genes.xlsx)
Targeted genes in RNA sequencing

References

1. Y Kawashima, et al., Proteogenomic Analyses of Cellular Lysates Using a Phenol-Guanidinium Thiocyanate Reagent. *J. Proteome Res.* **18**, 301–308 (2019).
2. Y Kawashima, et al., Optimization of Data-Independent Acquisition Mass Spectrometry for Deep and Highly Sensitive Proteomic Analysis. *Int. J. Mol. Sci.* **20**, 5932 (2019).
3. D Amodei, et al., Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. *J. Am. Soc. for Mass Spectrom.* **30**, 669–684 (2019).
4. B MacLean, et al., Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinforma. (Oxford, England)* **26**, 966–968 (2010).
5. H Sato, et al., In-Depth Serum Proteomics by DIA-MS with In Silico Spectral Libraries Reveals Dynamics during the Active Phase of Systemic Juvenile Idiopathic Arthritis. *ACS omega* **7**, 7012–7023 (2022).
6. K Blighe, PCAtools: Everything Principal Component Analysis (2022).
7. AH Bryk, JR Wiśniewski, Quantitative Analysis of Human Red Blood Cell Proteome. *J. Proteome Res.* **16**, 2752–2761 (2017).
8. J Lan, et al., Systematic Evaluation of the Use of Human Plasma and Serum for Mass-Spectrometry-Based Shotgun Proteomics. *J. Proteome Res.* **17**, 1426–1435 (2018).
9. S Wang, et al., NAguideR: Performing and prioritizing missing value imputations for consistent bottom-up proteomic analyses. *Nucleic Acids Res.* **48**, e83 (2020).
10. J Willforss, A Chawade, F Levander, NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *J. Proteome Res.* **18**, 732–740 (2019).
11. najoshi, Sickle - A windowed adaptive trimming tool for FASTQ files using quality (2021).
12. A Dobin, et al., STAR: Ultrafast universal RNA-seq aligner. *Bioinforma. (Oxford, England)* **29**, 15–21 (2013).
13. B Li, CN Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinforma.* **12**, 323 (2011).
14. YI Li, et al., Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).
15. N Keer, M Hershfield, T Caskey, S Unizony, Novel compound heterozygous variants in CECR1 gene associated with childhood onset polyarteritis nodosa and deficiency of ADA2. *Rheumatology* **55**, 1145–1147 (2016).