# Multi-omic profiling of pathogen-stimulated primary immune cells


## Authors

Renee Salz*[1,2], Emil E. Vorsteveld*[2,3], Caspar I. van der Made[2,3,4], Simone Kersten[2,3], Merel Stemerdink[2,5], Tabea V. Riepe[1,2,3], Tsung-han Hsieh[8], Musa Mhlanga[2,8], Mihai G. Netea[3,4], Pieter-Jan Volders[6,7], Alexander Hoischen[#2,3,4], Peter A.C. 't Hoen[#1,2]

## Affiliations

1. Department of Medical BioSciences, Radboud University Medical Center, Nijmegen, the Netherlands
2. RadboudUMC Research Institute for Medical Innovation, Radboud University Medical Center, Nijmegen, the Netherlands
3. Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands
4. Department of Internal Medicine and Radboud Centre for Infectious Diseases (RCI), Radboud University Medical Centre
5. Department of Otorhinolaryngology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands
6. Department of Biomolecular Medicine, Ghent University, Ghent, Belgium
7. Laboratory of Molecular Diagnostics, Department of Clinical Biology, Jessa Hospital, Hasselt, 3500, Belgium
8. Department of Cell Biology, Radboud University, Nijmegen, the Netherlands

* shared first authors

# joint corresponding authors. Contact details: alexander.hoischen@radboudumc.nl; peter-bram.thoen@radboudumc.nl

## Summary

We performed long-read transcriptome and proteome profiling of pathogen-stimulated peripheral blood mononuclear cells (PBMCs) from healthy donors to discover new transcript and protein isoforms expressed during immune responses to diverse pathogens. Long-read transcriptome profiling reveals novel sequences and isoform switching induced upon pathogen stimulation, including transcripts that are difficult to detect using traditional short-read sequencing. Widespread loss of intron retention occurs as a common result of all pathogen stimulations. We highlight novel transcripts of *NFKB1 and CASP1* that may indicate novel immunological mechanisms. RNA expression differences did not result in differences in the amounts of secreted proteins. Clustering analysis of secreted proteins revealed a correlation between chemokine (receptor) expression on the RNA and protein levels in *C. albicans*- and Poly(I:C)-stimulated PBMCs. Isoform aware long-read sequencing of pathogen-stimulated immune cells highlights the potential of these methods to identify novel transcripts, revealing a more complex transcriptome landscape than previously appreciated.

## Introduction

Immune system responses within the context of an infection are shaped by the nature of the infection and by inter-individual variability, contributing to differential susceptibility to infections and to various diseases with an inflammatory component. Dynamic expression of transcripts and proteins in a range of cells responsible for the innate immune response is important to shape the first line of defense against a wide variety of pathogens[1]. Pattern recognition receptors (PRR) initiate acute inflammatory responses, activating signaling cascades that converge on various transcription factors. Multiple levels of regulation orchestrate the dynamic expression of transcripts and proteins, including transcriptional and post-transcriptional checkpoints such as mRNA splicing and protein translation. Examples of this include the regulatory role of alternative splicing of Toll-like receptors (TLR) and their downstream signaling factors[2,3].

Methods for investigating innate immune responses include *in vitro* stimulation of primary immune cells with pathogens or microbial components. These methods allow for specific investigation of host-pathogen interactions that shape the immune response elicited by specific cell types and have been under extensive investigation in research on the innate immune system[4,5]. Stimuli that are commonly used include molecules that stimulate a specific TLR, such as *E. coli* lipopolysaccharide (LPS) for TLR4[6], dsRNA mimicking Poly(I:C) for TLR3[7] and imidazoquinolines for TLR7/8[8]. Stimulation is also often elicited by live or heat-killed pathogens[9,10], which stimulate at the same time a broader range of PRRs[11,12].

Transcriptome characterization is traditionally performed using short read RNA sequencing. These sequencing approaches are limited by their short-read length (approximately 150-300 bp), necessitating the computational reconstruction of whole transcripts and making detection of different transcripts of the same gene inaccurate. This limitation is especially pronounced in immune biology, where tight regulation of isoform expression has previously been described to play a major role in processes, such as the expression of multiple IL-32 transcripts with different inflammatory potency[13] and alternative splicing of *CD45* in T cell activation[14]. Recent long-read sequencing approaches provide a more complete and accurate reflection of the transcriptome. Sequencing technologies provided by PacBio and

Oxford Nanopore allow for the sequencing of mRNA (or cDNA) molecules from the ultimate 3'-end to the ultimate 5'-end, which have given a more comprehensive view into the complexity of the transcriptome. A number of studies have indicated that the isoform landscape is much more complex than previously appreciated[15–17]. Long-read mRNA sequencing has provided insight into regulatory mechanisms of immune responses, for instance in alternative splicing in macrophages[18] , and allows for accurate sequencing of complex transcripts in immune cells[19]. Further work on cell-type specific long-read transcriptomes have shown preferential expression of transcripts in specific cell types[20].

The impact of the newly discovered transcripts as well as post-transcriptional processes can only be fully understood by observing the proteome. Many studies have characterized the transcriptomic landscape of the human immune response, but a multi-omics view of immunity is necessary as mRNA profiles are not enough to understand immune activation[21,22]. Transcript information can be leveraged to study the proteome, including identification of novel proteoforms resulting from alternative splicing. Proteoforms discovered by proteogenomics methodologies have already been found to have a role in immunological processes, for instance in immune-regulating micropeptides[23] and tumor neoantigen production[24].

Here, we stimulated peripheral blood mononuclear cells (PBMCs) with multiple microbial stimuli *in vitro/ex vivo* and performed long- and short-read RNA sequencing and secretome proteomics to gain insight into potential differences in immune response. We aim to provide insight into the immune transcriptome and proteome of immune cells during innate immune responses against a variety of pathogens.

## Material & methods

### Ex vivo PBMC experiments

Venous blood was drawn from five healthy donors[5] and collected in 10mL EDTA tubes. Isolation of peripheral blood mononuclear cells (PBMCs) was conducted as described elsewhere[25]. In brief, PBMCs were obtained from blood by differential density centrifugation over Ficoll gradient (Cytiva, Ficoll-Paque Plus, Sigma-Aldrich) after 1:1 dilution in PBS. Cells were washed twice in saline and re-suspended in serum-free cell culture medium (Roswell Park Memorial Institute (RPMI) 1640, Gibco) supplemented with 50 mg/mL gentamicin, 2 mM L-glutamine and 1 mM pyruvate. Cells were counted using a particle counter (Beckmann Coulter, Woerden, The Netherlands) after which the concentration was adjusted to $5 \times 10^6$/mL. *Ex vivo* PBMC stimulations were performed with $5\times10^5$ cells/well in round-bottom 96-well plates (Greiner Bio-One, Kremsmünster, Austria) for 24 hours at 37°C and 5% carbon dioxide. Cells were treated with lipopolysaccharide (*E. coli* LPS, 10 ng/mL), *Staphylococcus aureus* (ATCC25923 heat-killed, $1\times10^6$/mL), TLR3 ligand Poly I:C (10 µg/mL), *Candida albicans* yeast (UC820 heat-killed, $1\times10^6$/mL), or left untreated in regular RPMI medium as normal control. After the incubation period of 24h and centrifugation, supernatants were collected and stored at -80°C until further processing. For the RNA isolation, cells were stored in 350 µL RNeasy Lysis Buffer (Qiagen, Rneasy Mini Kit, Cat nr. 74104) at −80°C until further processing.

## RNA and protein isolation

RNA was isolated from the samples using the RNeasy RNA isolation kit (Qiagen) according to the protocol supplied by the manufacturer. The RNA integrity of the isolated RNA was examined using the TapeStation HS D1000 (Agilent), and was found to be ≥7.5 for all samples. Accurate determination of the RNA concentration was performed using the Qubit (ThermoFisher).

We extracted the secretome of the 24-hour stimulated PBMCs. To 250 µL of supernatant, 250 µL buffer containing 10% sodium dodecyl sulfate (SDS) and 100 mM triethylammonium bicarbonate (TEAB), pH 8.5 was added. Proteins were reduced by addition of 5 mM dithiothreitol and incubation for 30 minutes at 55˚C and then alkylated by addition of 10 mM iodoacetamide and incubation for 15 minutes at RT in the dark. Phosphoric acid was added to a final concentration of 1.2% and subsequently samples were diluted 7-fold with binding buffer containing 90% methanol in 100 mM TEAB, pH 7.55. The samples were loaded on a 96-well S-Trap™ plate (Protifi) in parts of 400 µL, placed on top of a deepwell plate, and centrifuged for 2 min at 1,500 x g at RT. After protein binding, the S-trap™ plate was washed three times by adding 200 µl binding buffer and centrifugation for 2 min at 1,500 x g at RT. A new deepwell receiver plate was placed below the 96-well S-Trap™ plate and 125 µL 50 mM TEAB containing 1 µg of trypsin was added for digestion overnight at 37°C. Using centrifugation for 2 min at 1,500 x g, peptides were eluted in three times, first with 80 µL 50 mM TEAB, then with 80 µL 0.2% formic acid (FA) in water and finally with 80 µL 0.2% FA in water/acetonitrile (can) (50/50, v/v). Eluted peptides were dried completely by vacuum centrifugation.

## Long-read library preparation and sequencing

Libraries were generated from one donor using the Iso-Seq-Express-Template-Preparation protocol according to the manufacturer's recommendations (PacBio, Menlo Parc, CA, USA). We followed the recommendation for 2-2.5kb libraries, using the 2.0 binding kit, on-plate loading concentrations of final IsoSeq libraries was 90pM (*C. albicans*, *S. aureus*, Poly(I:C), RPMI) and 100pM (LPS) respectively. We used a 30h movie time for sequencing.

The five samples were analyzed using the isoseq3 v3.4.0 pipeline. Each sample underwent the same analysis procedure. First CCS1 v6.3.0 was run with min accuracy set to 0.9. IsoSeq lima v2.5.0 was run in IsoSeq mode as recommended. IsoSeq refine was run with '--require-polya'. The output of IsoSeq refine was used as input for IsoQuant v3.1.2[26] with GRCh38.p13 v43 primary assembly from GENCODE. The settings were set for full length PacBio data, and quantification included ambiguous reads. In IsoQuant, transcripts were considered novel if their intron chains did not match intron chains found in GENCODE annotation version 39. Transcripts with fewer than 5 reads across all samples were excluded from further analyses (Supplemental table 1).

We sought to validate the novel transcripts identified using long-read sequencing using FANTOM5 CAGE data of CD14 monocytes (https://fantom.gsc.riken.jp/5/datafiles/latest/basic/human.primary_cell.CAGEScan/CD14%2b%20monocyte%20derived%20endothelial%20progenitor%20cells%2c%20donor1.NCig10041.11229-116C5.hg19.GCTATA.clusters.bed.gz) that allows for the identification of

171 transcripts with a matching TSS from this 5' sequencing data. Transcripts with novel 5' were
172 considered to be supported with a CAGE peak if within 150 basepairs from the TSS.
173
174 ## Short-read library preparation and sequencing
175 RNA input was normalized to 200 ng for all samples/donors and libraries were generated
176 using the QuantSeq 3' mRNA-Seq Library Prep Kit-FWD from Lexogen (Lexogen) in
177 accordance with the manufacturers' protocol. In order to ensure high quality libraries, two
178 separate preparations were performed, limiting the number of samples to 30 per
179 preparation. End-point PCR was performed with 19 – 22 cycles, as indicated by a
180 quantitative PCR on a 1:10 aliquot of a subset of double stranded cDNA libraries. Accurate
181 quantification and quality assessment of the generated libraries was performed using Qubit
182 dsDNA High Sensitivity assay (Thermo Fisher Scientific) and Agilent 2200 TapeStation (High
183 Sensitivity D1000 ScreenTape, Agilent). Molarity of individual libraries was calculated using
184 the cDNA concentration (Qubit) and average fragment size (TapeStation). Safeguarding
185 sufficient read-depth for each sample, libraries were split in two separate runs. In each run,
186 the baseline RPMI condition across all donors and time-points was included, in turn allowing
187 sequencing bias assessment. The cDNA libraries of 35 samples were pooled equimolarly to
188 100 fmol. After a final dilution of both pools to a concentration of 4 nM, they were
189 sequenced on a NextSeq 500 instrument (Illumina) with a final loading concentration of 1.4
190 pM.
191
192 FastQC v0.11.5 (Babraham Bioinformatics) was used to assess the quality of the obtained
193 sequencing data, followed by removal of adapter sequences and poly(A) tails by Trim
194 Galore! V.0.4.4_dev (Babraham Bioinformatics) and Cutadapt v1.18[27]. Since QuantSeq reads
195 only provide coverage of the 3' end of transcripts, we generated a set of transcripts
196 representative of the full transcriptome by grouping transcripts based on unique 3'
197 sequences. Therefore, we separately mapped the filtered and trimmed reads to the long
198 read transcriptome with Salmon v1.9.0 in mapping-based mode with decoys[28].
199
200 ## Differential expression analyses
201 To measure differential gene expression from long-read RNA sequencing, low abundance
202 genes were filtered using a 10 CPM threshold with the conform package in python.
203 Differentially expressed genes (DEGs) and transcripts were calculated for each condition
204 versus control using the NOISeq R package[29] from the abundances generated with isoquant.
205 TMM normalization was chosen and q-value threshold for DE was set at 0.95.
206
207 DEGs were generated from the salmon-mapped short-read RNA sequencing data using the
208 samples from the same donor using NOISeq[29]. The two control samples (RPMI) per donor
209 were treated as technical replicates. TMM normalization was chosen and q-value threshold
210 for DE was set at 0.95. We validated the DEGs detected from long-read sequencing with
211 those generated with the short-read data by comparing the linear correlation of the
212 log2fold change values for each condition combination between both datasets using the
213 lm() R function.
214
215 The up- and downregulated DEGs per condition-control pair were analyzed for pathway
216 enrichment separately using gProfiler[29]. We used Gene Ontology biological process and

molecular function and TRANSFAC transcription factor motifs gene sets[30,31]. A term size filter of between 100-500 was used to generate the final enrichment profiles.

### Isoform switching

A first-pass isoform switching analysis was performed using swanvis v2.0[32]. For a second-pass isoform switching analysis, the resulting gene-level isoform switch p-values were imported into IsoformSwitchAnalyzeR v1.16.0 package in R[33]. Thresholds for isoform switching were set at 10 DPI (differential percent isoform use) and nominal p-value <0.05. Sequences corresponding to the significant isoform switches were analyzed with CPAT v1.2.4[34], hmmscan v3.3.2 with Pfam[35], and SignalP5[36] as a part of the IsoformSwitchAnalyzeR package.

Pathway analysis and gene network analysis of genes that were found to undergo isoform switching was performed in Cytoscape[37]. Default pathway analysis was performed, filtering for Gene Ontology Biological Process gene sets. An Enrichment Map was built from the enriched gene sets with a Jaccard similarity cutoff of 0.4[38].

Genes found to undergo intron retention gains/losses and genes with domain gains/losses were separately analyzed using gProfiler. We used Gene Ontology Biological Process gene sets with a with a term size filter between 100-500 genes. We separately analyzed genes with domain gains or losses were using dcGOR[39]. We used the gene ontology molecular function gene sets with a term size filter between 100-500 genes.

### LC-MS/MS analysis

Peptides were re-dissolved in 20 µL loading solvent A (0.1% trifluoroacetic acid in water/acetonitrile) (98:2, v/v)) of which 4 µL was injected for LC-MS/MS analysis on an Ultimate 3000 RSLCnano system in-line connected to a Q Exactive HF mass spectrometer (Thermo). Trapping was performed at 10 µL/min for 4 min in loading solvent A on a 20 mm trapping column (made in-house, 100 µm internal diameter (I.D.), 5 µm beads, C18 Reprosil-HD, Dr. Maisch, Germany). The peptides were separated on a 250 mm Waters nanoEase M/Z HSS T3 Column, 100Å, 1.8 µm, 75 µm inner diameter (Waters Corporation) kept at a constant temperature of 45°C. Peptides were eluted by a non-linear gradient starting at 1% MS solvent B reaching 33% MS solvent B (0.1% formic acid (FA) in water/acetonitrile (2:8, v/v)) in 100 min, 55% MS solvent B (0.1% FA in water/acetonitrile (2:8, v/v)) in 135 min, 97% MS solvent B in 145 minutes followed by a 5-minute wash at 97% MS solvent B and re-equilibration with MS solvent A (0.1% FA in water).

The mass spectrometer was operated in data-dependent acquisition mode, automatically switching between MS and MS/MS acquisition for the 16 most abundant ion peaks per MS spectrum. Full-scan MS spectra (375-1500 m/z) were acquired at a resolution of 60,000 in the Orbitrap analyzer after accumulation to a target value of 3,000,000. The 16 most intense ions above a threshold value of 15,000 were isolated with a width of 1.5 m/z for fragmentation at a normalized collision energy of 28% after filling the trap at a target value of 100,000 for maximum 80 ms. MS/MS spectra (200-2000 m/z) were acquired at a resolution of 15,000 in the Orbitrap analyzer.

263 ## Protein identification and quantification
264 Two search databases were constructed; one database for proteoform detection and one
265 database for quantification. The database used for sensitive detection of proteoforms was
266 generated using a slightly adapted version of the Long Read Proteogenomics pipeline by
267 Miller *et al*[40]. Since the pipeline uses a different long-read transcriptomics tool, small syntax
268 adjustments were made to accommodate the use of Isoquant output. Additionally, a custom
269 script was written to have Isoquant output mimic the required input format. The pipeline
270 generated a GENCODE-PacBio hybrid database. The proteome from *C. albicans* (taxon ID
271 5476) and *S. aureus* (taxon ID 1280) were downloaded from UniProt and added to the
272 search database. The search database used for quantification was created by downloading
273 the proteome from *H. sapiens* (taxon ID 9609), *C. albicans* (taxon ID 5476) and *S. aureus*
274 (taxon ID 1280) from UniProt. Metamorpheus default contaminants were added to both
275 search databases.
276
277 Mass spectra were identified using Metamorpheus v1.0.0[41]. The Human Proteome Project
278 Mass Spectrometry Data Interpretation Guidelines version 3.0 were applied[42].
279 Quantification was performed using FlashLFQ v 1.2.4.294[43] with all five individuals set as
280 biological replicates and the two control (RPMI) samples per individual set as technical
281 replicates. The following options enabled: normalization, shared peptide quantification,
282 Bayesian fold change analysis, and match between runs (Supplemental table 2). An adapted
283 version of SQANTI protein was used to search for novel peptides in the Metamorpheus
284 identifications. Enrichment of secreted proteins was determined using the predicted
285 secreted proteins from Human protein atlas[44] as reference.
286
287 ## Protein clustering
288 FlashLFQ raw protein expression values originating from the quantification database search
289 were first square root transformed. To normalize for donor effects, the mean protein
290 expression value per gene/individual was subtracted from all the expression values from the
291 same gene/individual. Then z-score normalization was performed across all individuals per
292 gene. K-means clustering was then performed using the kmeans() function in R with seed
293 #82 and default parameters. We found four clusters to optimally represent the data
294 according to the elbow plots (Supplemental figure 1). A heatmap was constructed with
295 those clusters using the ComplexHeatmap package[45]. The proteins identifiers assigned to
296 cluster #4 were converted to gene names and analyzed using gProfiler for enrichment
297 analysis using both Gene Ontology Biological Process and Molecular Function gene sets. We
298 further analyzed the protein found to form cluster 4 through a protein network analysis in
299 Cytoscape[37].
300
301 # Results
302 We stimulated PBMCs from five donors with four different microbial stimuli, mimicking
303 bacterial (*E. coli* LPS, *S. aureus*), viral (Poly(I:C)) and fungal (*C. albicans*) infections. PBMCs
304 were stimulated for 24 hours. RPMI incubation was used as a negative control (Figure 1A).
305 To characterize full-length transcript structures, we performed long-read sequencing on
306 PBMCs from one donor (Figure 1B). Additionally, shotgun proteomics data was generated
307 from supernatants of the samples from all five donors. The proteomics data serves to
308 corroborate differential gene/transcript expression and provide evidence of the protein-

309 coding potential of novel transcripts identified through long-read RNA sequencing (Figure
310 1C). Short-read 3' sequencing data of all five donors was generated to validate differential
311 gene expression data generated from long-read RNA sequencing (Figure 1D).
312

313 ## Long read transcriptomes of both control and pathogen-stimulated conditions show novelty

314 Sequences detected using long-read sequencing were categorized in terms of novelty
315 according to their intron chains. Transcripts are divided into three categories that
316 encompass reference transcripts (GENCODE), novel in catalog (transcripts that contain
317 annotated introns) and transcripts that are novel not in catalog (containing unannotated
318 introns) (Figure 2A). We identified a total of 37,312 unique transcript sequences from
319 11,872 genes across all samples. The majority of transcripts were in protein coding genes
320 (Supplemental figure 2A) including ~10% immune-related genes (Supplemental figure 2B).
321 We found 47.4% of detected transcripts to be novel, while these accounted for only 20.3%
322 of the total reads (Figure 2B). The distribution of reads per novel transcript was similar to
323 that of known transcripts with a slight skew towards lower abundance (Supplemental figure
324 3A). Exon elongations were the most observed feature distinguishing novel from known
325 transcripts, occurring in nearly a third of the novel transcripts found in RPMI. This was
326 similar for the stimulated conditions (Figure 2C, Supplemental figure 3B). The percentage of
327 novel transcripts and transcript deviations were similar for all conditions (Figure 2D). To
328 corroborate the existence of novel transcripts, we analyzed FANTOM5 CAGE peaks in the
329 vicinity of the transcription start sites for novel transcripts with novel 5' ends. We found
330 8,233 (51.3%) novel 5' end transcripts across all conditions to be supported by a CAGE peaks
331 from unstimulated human monocytes (within 150 nucleotides)[46].
332

333 Principal component analysis of the expression levels for each transcript indicated that
334 stimulated conditions were more similar to each other than to RPMI. *S. aureus* and *C.*
335 *albicans* were most similar to each other (Figure 2F). Genes and transcripts expressed were
336 similar in the stimulated conditions with average Jaccard similarity indices of 0.9 and 0.82
337 for genes and transcripts, respectively (Figure 2G). Novel transcripts had similar Jaccard
338 indices to each other than for known transcripts (not shown). Differential expression
339 analysis yielded an average of 949 DEGs and 2,076 differentially expressed transcripts per
340 condition (Supplemental figure 4, Supplemental table 3-4).
341

342 We validated the DEGs through 3' transcript counting (QuantSeq)[47]. We gathered a set of
343 representative transcripts based on sequence differences at the 3' end of transcripts
344 (29,760 transcripts, 79.8% of total) and investigated the correlation of differential
345 expression in the long-read sequencing data with the separately generated short read
346 dataset of the same donor. The DEGs that overlapped between both datasets correlate well
347 ($R^2$ 0.62-0.81). Best matching pairs of stimulated conditions between the short- and long-
348 read confirmed the concordance of both sequencing approaches (Supplemental figure 5,
349 Supplemental table 5).
350

351 ## Pathogen stimuli display upregulation of different pathways

352 Differential gene expression analysis using the long-read sequencing data resulted in a total
353 of 1,733 genes that were differentially expressed in stimulated conditions compared to
354 control. We performed pathway analysis for each condition using gProfiler[48] (Supplemental
355 table 6). By overlapping the gene sets enriched in each of the four conditions, we discerned

356 biological processes/functions specific to certain pathogen-stimulated conditions. There are
357 a lot of constants in host response regardless of the pathogen, and indeed the largest set of
358 pathways was in the overlap between all stimulated conditions (211 pathways, Figure 3A).
359 This set has an enrichment of genes involved in type II interferon (IFN-γ) responses. Genes
360 involved in tertiary and specific granules, which play a role in the defense against pathogens
361 were found to be enriched among upregulated genes in all conditions. Surprisingly, we also
362 find these and related gene sets to be enriched among downregulated genes as a result of *S.*
363 *aureus* and Poly(I:C) stimulation, potentially a result of the regulation of the inflammatory
364 response. Further gene sets included the response to molecules of bacterial origin (including
365 LPS), innate immune response signaling such as PRR signaling, antigen processing and
366 presentation and IL-1 production (Figure 3B).
367
368 Some pathogen-stimulated conditions had more enriched pathways in common than others.
369 There was a notable overlap of 131 gene sets enriched in *C. albicans-, S. aureus*- and
370 Poly(I:C)-stimulated conditions. Some of these were common to the set overlapping
371 between all conditions, such as interferon responses. The LPS-excluding set showed
372 particular enrichment related to viral processes such as the defense against viruses,
373 regulation of the viral lifecycle, likely due to interferon-stimulated gene expression, such as
374 *STAT1, OAS1/3, OASL and IFIH1*. Also, transcription factor binding matches (TRANSFAC) such
375 as *IRF-2, 5, 8* and *9* were enriched, reflecting downstream signaling through various
376 signaling pathways leading to the regulation of the production of interferons and immune
377 cell development (Figure 3C)[49].
378
379 LPS and Poly(I:C) were the 2 stimuli with the most enriched pathways unique to a single
380 stimulus. For 55 gene sets unique to LPS, there was a downregulation of T cell receptor
381 signaling, in part due to the downregulation of *CD4* expression, which has previously been
382 described as a result of endogenous production of TNF-α and IL-1β as a result of LPS
383 stimulation[50]. We further found an upregulation of gene sets involved in metabolic
384 processes such as oxidoreductase complexes and cellular responses to oxygen, possibly
385 reflecting metabolic changes previously described to occur in immune cells such as
386 monocytes upon LPS stimulation[51]. Furthermore, there was an upregulation of genes
387 involved in humoral immune responses (Figure 3D). For 53 gene sets enriched uniquely in
388 Poly(I:C), we found functions including viral gene expression, apoptosis related signaling
389 (regulation of cysteine-type endopeptidase activity) and B-cell related gene sets such as
390 increased antibody levels and BCR signaling. Finally, there was an enrichment of MHC class II
391 antigen presentation (Figure 3E).
392
393 ### Isoform switches highlight transcriptome differences between conditions and control
394 Isoform switching (IS) genes are defined by a change (increase/decrease) of expression of a
395 particular transcript isoform as measured by percent of total reads for a gene. In different
396 samples/conditions, a particular transcript isoform may comprise a different isoform
397 fraction (dIF) value for a given gene. Here, a change of at least 10% (0.10 dIF) in control and
398 the opposite change (decrease/increase) of expression of a different transcript isoform in
399 the same gene of at least 10% in the pathogen-stimulated condition is considered an IS.
400
401 A total of 999 IS were detected in 398 genes. Nearly half (N=192, 48.2%) of these IS genes
402 were unique to their respective stimulus conditions, while 10.3% were found in all

403 conditions (N=41) (Figure 4A, Supplemental table 7-8). The majority of genes demonstrating
404 IS were not differentially expressed in their respective conditions (327 genes; 77%). Most
405 genes that were found to undergo IS displayed only one IS instance (Supplemental figure
406 6A). Pathway analysis of genes undergoing IS were enriched for gene sets involved in
407 metabolic processes, mRNA splicing, protein transport and catabolism. Furthermore,
408 immune and stress-related pathways such as MHC type I antigen processing and transport
409 through vesicles, inflammasomes, oxidative stress and apoptosis were found to be
410 represented in genes undergoing IS (Figure 4B, Supplemental tables 9-13).
411
412 We sought to understand the molecular consequences of IS upon pathogen stimulation by
413 categorizing the differing features of the isoform pairs involved in the switch. Each of the IS
414 was annotated with one or more of the following predicted protein characteristics: change
415 in ORF length, ORF gain/loss, domain gain/loss, NMD sensitivity, intron retention (IR)
416 gain/loss, coding probability (ORF presence), and signal peptide gain/loss. These
417 consequences are not independent and often multiple consequences could be attributed to
418 one IS (Supplemental Figure 6B). We observed general IS trends on a genome-wide scale
419 (Supplemental figure 6C, Supplemental table 14). Strikingly, we found IR loss to be the most
420 common consequence of IS in this dataset. Isoforms with retained introns comprised a
421 higher isoform fraction for genes in the control condition, while their respective intron-
422 excluding counterparts had a higher isoform fraction for genes in the pathogen-stimulated
423 conditions. Genes displaying loss of IR were enriched for pathways involved in mRNA
424 processing, including spliceosome-related gene sets, antigen processing and IL-1 production
425 (Supplemental figure 7). IR has previously been described as a regulatory mechanism of RNA
426 processing, splicing, vesicle transport and type I interferon production in the development
427 of various immune cell types, including macrophages[52,53], granulocytes[54] and B cells[55,56]. Our
428 findings support previously described associations of IR losses in immune-related processes,
429 and adds new genes regulated by IR loss during immune responses (Supplemental figure 7,
430 Supplemental table 15).
431
432 In addition, we found a higher proportion of transcripts to have domain gains than domain
433 losses. This could indicate that stimulation by a pathogen causes a gene to switch expression
434 to a transcript isoform that codes for a protein with an extra function. Other observed
435 trends included longer ORFs and NMD insensitivity in transcript isoforms induced by
436 pathogen stimulation (Figure 4C).
437
438 Since the addition or loss of domains could directly reveal protein function changes, we
439 explored the IS that had this consequence type. We found that genes with domain gain/loss
440 (N=158, Supplemental table 16) were enriched for involvement in various catabolic
441 processes. We also found enrichment of T cell activation genes, an effect previously
442 described as a functional consequence of CD8+ T cell co-stimulation[57]. Other enriched gene
443 sets include leukocyte cell-cell adhesion and activation and general innate immune response
444 genes (Supplemental table 17). When looking more specifically at the molecular functions of
445 the gained domains themselves, we found an enrichment of domains with potassium
446 channel regulator activity, kinase- and transferase activity concerning phosphorus-
447 containing groups and nucleic acid binding. These results potentially indicate functional and
448 cell-type specific effects of domain gains as a result of IS in immune responses
449 (Supplemental figure 8, Supplemental table 18).

450
451 Novel transcripts play an important role in IS. Of the 999 IS, more than half (N=592) had at
452 least one novel transcript involved in the IS. In most cases (N=438), the switch was from a
453 novel transcript to a known transcript (Figure 4D, Supplemental table 19). Compared to IS
454 cases where only known transcripts were involved, the IS consequences were more often
455 NMD insensitivity and IR loss (Figure 4E). Conversely, shorter ORFs, domain losses and NMD
456 sensitivity were more common effects when the IS was from a known to a novel transcript
457 isoform. In conclusion, the unstimulated condition is characterized by the presence of many
458 novel transcripts with retained introns, which are difficult to detect with short read
459 sequencing. IR is likely a mechanism to prepare a cell for fast action after an immune
460 stimulus, when splicing of the retained intron could quickly generate a functional transcript
461 with coding potential, which has been for instance been described in CD4+ T cells[58].
462

463 ### A novel read-through transcript including *CARD16* and *CASP1*

464 As an example of a remarkable finding with possible biological impact once validated, we
465 identified a read-through transcript that includes both *CARD16* and *CASP1* (Figure 5A). Read-
466 through transcripts involve transcription that extends beyond the normal polyadenylation
467 site (PAS), terminating at the PAS of an adjacent gene or other nearby locus[59]. These
468 transcripts have been found to be expressed in specific circumstances, including malignancy
469 and infection[59,60]. This particular novel transcript encompasses the coding region of *CASP1*
470 and has an extended 5' UTR which spans *CARD16,* and thus contains two ORFs. This IS was
471 annotated as an intron retention loss, as the novel transcript loses an intronic region in its 3'
472 UTR (Figure 5B). Both the known and novel transcripts in this IS are predicted to be coding
473 (both 100%). *CASP1* was found to be differentially expressed upon Poly(I:C) stimulation
474 (log2FC 1.73, p=0.049; Figure 5C). The isoform expression of the known transcript was found
475 to decrease upon Poly(I:C) stimulation, while the novel transcript was found to increase
476 (Figure 5D). This is further reflected in the isoform fraction, increasing from 8.3% to 24.8%,
477 while the known transcript decreased from 85.5% to 74.2% (Figure 5E).
478
479 *CARD16* and *CASP1* both have a function in proinflammatory IL-1β signaling, where *CARD16*
480 has been shown to play a role in *CASP1* assembly, although there remains discussion on the
481 exact regulatory effect of *CARD16* on this process [61,62]. We have identified an IS specifically
482 for Poly(I:C) stimulation, where a novel transcript of *CASP1* was found to harbor *CARD16* in
483 its 5' UTR was upregulated upon stimulation. This finding could suggest a novel molecular
484 mechanism in IL-1β signaling, potentially through the regulation of *CASP1* by its regulator
485 *CARD16*.
486

487 ### A novel coding transcript of *NFKB1*

488 We identified a novel *NFKB1* transcript that demonstrated IS in all four conditions. This
489 novel transcript was shorter than the canonical transcripts (Figure 6A). Further analysis
490 revealed that the novel transcript start site was supported by multiple nearby CAGE peaks
491 (Figure 6B). Strikingly, this novel transcript lacks a part of its Rel homology domain, a
492 conserved domain responsible for functions such as dimerization and DNA binding (Figure
493 6C)[63]. *NFKB1* was not found to be significantly differentially expressed, although gene
494 expression was found to be higher in pathogen-stimulated condition compared to
495 unstimulated condition (only *C. albicans* shown, Figure 6D). The expression of the novel
496 transcript was found to increase upon pathogen stimulation (Figure 6E). This is reflected in

497 the isoform fraction, which increases from 23.5% to 50.7%, while the known transcript
498 decreases from 39.0% to 21.2% (Figure 5F).
499
500 *NFKB1* plays a central role in immune responses, regulating the response to infections
501 through transcriptional activation[57]. Furthermore, the Rel homology domain region is known
502 to harbor disease-causing variants responsible for common viable immunodeficiency
503 (CVID)[64], highlighting the importance of this domain in normal B cell function. This finding
504 could suggest a novel regulatory mechanism of *NFKB1*.
505

506 ### Isoform switching in *CLEC7A* and *OAS1* in a stimulus-specific manner

507 We sought to identify genes with stimulus-specific IS patterns. We identified an IS in
508 *CLEC7A,* which codes for Dectin-1, a receptor that recognizes fungal glucans, triggering the
509 immune response[65]. While this gene was not differentially expressed upon pathogen
510 stimulation, we did identify an IS specific to *C. albicans* stimulation in this gene, involving a
511 decrease in expression of an NMD sensitive transcript, with the increase in expression of the
512 canonical coding transcript and a non-coding transcript (Supplemental figure 9A). In
513 contrast, IR loss in *CLEC7A* was previously identified as a result of stimulation with multiple
514 pathogen stimuli in monocytes. Additionally, no difference in gene expression levels was
515 found between IAV-stimulated and resting cells, where the change in splicing was most
516 pronounced[66]. While we find IR loss in *CLEC7A* specifically upon *C. albicans* stimulation, this
517 could therefore also indicate a shared transcriptional response to pathogen stimuli.
518
519 Additionally, we identified an IS involving *OAS1*, which is involved in antiviral immunity. This
520 gene is differentially expressed in response to *C. albicans, S. aureus* and Poly(I:C)
521 (Supplemental figure 9B). We identified an IS in this gene resulting in an intron loss and a
522 domain gain. This IS was only observed for Poly(I:C) stimulation, potentially indicating this
523 transcript is necessary for antiviral immune responses (Supplemental figure 9C). Previous
524 work has identified common *OAS1* haplotypes responsible for a decrease in protein
525 abundance through the expression of NMD sensitive transcript p42, which contributes to
526 COVID-19 severity[67]. We find this transcript to be downregulated upon Poly(I:C) stimulation.
527 However, the transcript we find to be upregulated lacks the prenylation site needed for
528 antiviral function, as shown for p46[68].
529

530 ### Detecting secreted peptides

531 We sought to obtain evidence of the protein-coding potential of novel transcripts found
532 through long-read RNA sequencing. Mass spectrometry was performed for 30 secretome
533 samples from five donors' stimulated PBMCs, which includes the samples from the
534 individual for which long read RNA sequencing was performed (see methods). These include
535 2 control samples and 1 of each 24-hour pathogen stimulation condition for each individual.
536
537 We designed a search database comprising all proteins that we suspected could be in the
538 sample. This includes the GENCODE human proteome, the proteomes of the pathogens
539 used, as well as ORFs derived from novel transcripts found using long-read RNA sequencing.
540 Novel transcripts do not always correspond to novel ORFs; 32% of the novel transcripts had
541 an ORF that was present in the GENCODE reference database (Supplemental Figure 10). In
542 the collection of 30 samples, a total of 38,703 peptides from 15,964 proteins were
543 identified. We found 404 (7.37%) of identified proteins were known to be secreted

544 according to the human protein atlas, which constitutes a significant enrichment (OR=2.12,
545 $p=3.88x10^{-21}$, Fisher's exact test). We did not detect microbial proteins in the samples. Many
546 of the novel ORFs predicted from the transcriptome have high similarity to GENCODE ORFs,
547 resulting in a small number of novel peptides that could uniquely identify these. After
548 rigorous filtering, we were unable to confidently identify peptides that mapped uniquely to
549 the predicted novel ORFs.
550
551 ### Wider deviations in expression in the secretome
552 To assess whether differences in transcript expression resulted in differences in the
553 amounts of secreted proteins, we performed a label-free quantification of the proteins in
554 the cells' supernatants. Using PCA, we found that a large portion of variation in the
555 proteome was explained by inter-individual differences and that these differences were
556 larger than the differences induced by the immune stimuli (Supplemental figure 11).
557
558 We found a total of 418 differentially expressed proteins (DEPs) between the stimuli and
559 control when controlling for individual variation. Differential protein expression was not
560 equally distributed between stimuli with over a third (N=131) of the DEGs unique to
561 Poly(I:C) stimulation (Supplemental figure 12A). With the exception of the *S. aureus*
562 condition, more proteins were significantly downregulated than significantly upregulated in
563 the secretome (Supplemental figure 12, Supplemental table 20). We found few overlapping
564 proteins per condition, which could indicate either a high specificity in response to different
565 pathogens or a lack of protein secretion in a subset of samples.
566
567 To determine which explanation is more likely, we visualized the specific (groups of)
568 proteins associated with response stemming from the stimuli. We clustered protein
569 expression values normalized by individual and stimulus (Figure 7A, Supplemental table 21).
570 The clustering revealed a separation between poly(I:C) samples and the rest of the stimuli.
571 *C. albicans* showed a large overlap with poly(I:C) in the protein expression profiles. Some *C.*
572 *albicans* samples were grouped with poly(I:C) samples, which confirms the results from the
573 differential protein expression analysis (34 common DEPs, Supplemental figure 12A). Other
574 stimulus conditions could not reliably be separated from RPMI.
575
576 We identified a cluster of proteins that are highly expressed in Poly(I:C) and *C. albicans*
577 (cluster 4, Figure 7A). This group of proteins is enriched for genes with functions in
578 leukocyte migration and chemotaxis, exemplified by neutrophil migration. We identified
579 further enrichments of gene sets involved in the response to IL-1, humoral antimicrobial
580 response, and cellular responses to LPS and type II interferons. Analysis of the molecular
581 functions of these genes indicated an enrichment of cytokine activity and receptor binding,
582 GPCR receptor binding and various catalytic functions, likely due to immune cell
583 differentiation and immune responses involving the degradation of extracellular matrix
584 proteins during immune cell migration[69] (Figure 7B, Supplemental table 22). We further
585 assessed the proteins in cluster 4 through a gene network analysis (Figure 7C, Supplemental
586 table 23-24). Of the 84 proteins in this network, 61 were differentially expressed on the
587 protein level (72.6%, any condition). Of these DEPs, 18 are involved in cytokine signaling
588 (29.5%), of which 13 genes are chemokines (71.2%). A high proportion of proteins are found
589 in the extracellular region (n=47, 77.0%), for instance through secretion in granules. The
590 biological functions of the DEPs in cluster 4 reflect those found for the complete set of

591 proteins in cluster 4, mainly corresponding to pathways associated with functions in
592 neutrophil migration and chemotaxis (Supplemental table 25). As these pathways are not
593 necessarily specific to these two stimuli, this may indicate Poly(I:C) and *C. albicans* may be
594 more effective at eliciting differential protein secretion or have less delay in secretion
595 compared to the other stimuli.
596

597 ## Comparison with RNA expression

598 As established earlier, a multi-omics approach is currently the best way to understand the
599 human immune response. Correlation between the RNA and protein levels, or lack thereof,
600 can provide important clues about the host response to pathogens. To assess the
601 correlation of differential gene and protein expression levels, we assessed the concordance
602 of differential expression on the RNA and protein level. This metric corresponds to the
603 percentage of genes for which differential expression on both levels matched in
604 directionality (out of all genes where DE was observed on both levels) (Figure 7D,
605 Supplemental table 26).
606
607 We observed an overall poor concordance of directionality and fold change of expression
608 levels at the RNA and protein levels in the different stimulus conditions, with the exception
609 of *C. albicans* with 73% overall concordance. We overlaid the genes in group 4 from our
610 clustering analysis with the genes found to be DE on both RNA and protein levels. There was
611 an overrepresentation of the genes in this cluster in the total group of dual-level DE genes
612 (OR=6.99, p=4.813e-16). Further analysis of concordant differential expression matches
613 arising from proteins in cluster 4 (triangles in Figure 7D), we observed high concordance in
614 the genes induced by *C. albicans* and/or Poly(I:C). Directionality concordance for Poly(I:C)
615 and *C. albicans* for genes in in cluster 4 was significantly higher than overall directionality
616 concordance (p=0.0313 Poly(I:C), p=0.0003 *C. albicans*, Fisher's test one-tailed). The cluster
617 4 proteins in the LPS and *S. aureus* conditions are in the lower right quadrant, indicating that
618 the increase of RNA translated into a decrease of secreted proteins for these genes (Figure
619 7D).
620
621 We hypothesized in the IS analysis that a major regulatory mechanism in the host response
622 to pathogens was the loss of intron retention for rapid protein generation. We cross-
623 referenced the secreted proteins to support this conjecture. By overlapping upregulated
624 isoforms from intron retention loss events, we found 20 cases from 7 genes (Supplemental
625 table 27). Of these genes, 2 were upregulated on the protein level, supporting our
626 hypothesis. The genes were *GZMB* and *B2M*, which are important immune-regulatory genes
627 that are both secreted[70,71]. Considering the remaining 5 genes that were *downregulated* on
628 the protein level, however, this is not convincing evidence that intron retention loss in
629 general provides a rapid increase of protein production.
630

631 ## Discussion

632 The identification of novel transcripts and subsequent production of additional protein
633 isoforms could help identify molecular mechanisms that play a role in various biological
634 processes, including immune responses. Various immune system processes have previously
635 been found to be regulated by alternative splicing [72,73 74 75]. Immune responses display
636 significant inter-individual differences. Donor-specific effects such as sex and ancestry have
637 been shown to significantly influence the transcriptome. Previous studies have further

638 shown the impact of QTLs in the heritability of cytokine production capacity[5,76–78]. However,
639 the effect of these processes on host defense mechanisms against pathogens, together with
640 the large inter-individual differences in transcription and protein expression, remain to be
641 elucidated.
642
643 We have generated a long-read transcriptome of pathogen-challenged primary immune
644 cells (PBMCs) together with the secreted proteome to investigate mechanisms underlying
645 immune responses during infection. We described the accurate identification of known and
646 novel transcripts in both control and pathogen-challenged conditions. Of these transcripts,
647 we identified a subset that is differentially expressed as a result of pathogen stimulation,
648 which we validated by short read RNA sequencing data (including 4 additional individuals)
649 and publicly available CAGE data from neutrophils.
650
651 We examined isoform switching that occurred as a result of pathogen stimulation, insight
652 into transcripts that may play a role in pathogen responses. On a genome-wide level,
653 widespread intron retention losses were observed. Retained introns that rendered the
654 transcript unusable in the control condition were spliced out as a result of microbial
655 stimulation; a trend we observed in all conditions regardless of microbe. We postulate that
656 these are examples of unproductive splicing in unstimulated cells switching to productive
657 splicing after stimulation enabling fast production of proteins relevant for the immune
658 response. Genes that undergo intron retention loss mainly have functions in mRNA splicing
659 and processing and in immunity. Tissue- or cell-type specific unproductive splicing has been
660 widely observed as an autoregulatory process for mRNA splicing factors[79], which is
661 supported by our data in immune cells. We were however not able to confirm changes in
662 protein expression of genes that underwent IR losses using our secretome proteomics data,
663 likely because these proteins are not generally secreted. A couple of pertinent examples
664 have been illustrated in greater detail. We identified an IS specific to the viral stimulus that
665 involves a novel read-through transcript of *CASP1* and *CARD16*. We found an instance of IS
666 to a novel *NFKB1* transcript with a shortened DNA binding domain that was found in all four
667 conditions. Additionally, we describe IS in *CLEC7A* and *OAS1* for *C. albicans* and Poly(I:C),
668 which highlight stimulus-specific alternative splicing. Taken together, these results highlight
669 the potential for long-read sequencing to accurately resolve novel transcripts with potential
670 relevance in immune responses, including intron retention loss events that are generally
671 difficult to detect using short-read sequencing.
672
673 The extent to which conclusions can be drawn about immune response mechanisms is
674 limited by the low sample size for long-read sequencing. In this explorative study meant to
675 provide insights into the novel technical possibilities utilizing latest sequencing approaches,
676 we generated long-read sequencing data for only a single individual because of the
677 expensive nature of this technology in combination with the required sequencing depth and
678 the number of conditions studied. This design did not allow us to investigate the inter-
679 individual differences in the transcriptome. Novel transcripts that were detected could thus
680 be specific to this individual. Future follow-up including the sequencing of more individuals
681 using accurate long-read sequencing methods and functional studies could provide
682 additional insight into the more general relevance of these transcripts in immune responses.
683 This study focused on the appraisal of the transcriptome and proteome in PBMCs, which
684 consist of multiple cell types. Use of freshly isolated PBMCs accurately represents the

complete immune cell population in the peripheral blood and allows for communication between cell types during pathogen stimulation, thereby potentially giving an accurate representation of this cell population *in vivo*. However, no information on cell type specificity of transcripts is available. This could be resolved by recent developments in single cell long-read sequencing[80].

The proteome, in contrast, was generated for all samples from all 5 individuals and highlighted significant differences between the secretome of individual donors, before and after response to immune stimuli. Concordance between the transcriptome and proteome levels was high in Poly(I:C) and *C. albicans*, and lower in LPS and *S. aureus*. We found that genes with high correlation on the RNA- and protein levels form a cluster of protein expression, separating the former two stimuli from the latter. These proteins are enriched for secreted immune-related proteins, indicating that pathogen stimulation successfully led to secretion of relevant proteins. This would indicate that cells have responded faster to the Poly(I:C) and *C. albicans* stimuli than to the LPS and *S. aureus* stimuli, because RNA and protein were isolated simultaneously from our samples. Delay in protein production after expression of an mRNA may partially explain the lack of correlation of differential expression on RNA and protein level. This delay is presumably even longer in the secretome as proteins need to be first produced and subsequently secreted[81].

We focused our study on the secretome to reduce the complexity of the protein mixture analyzed, and to obtain better peptide coverage of the secreted proteins that play an important role in immune signaling. However, this limited our view on the complete proteome affected by immune stimuli. Also, there is the added complication that only a small number of peptides exist that could discriminate between proteoforms. To detect the proteoforms derived from our long-read sequencing data, much deeper shotgun proteomics must be performed[82]. These limitations are reasons why no evidence of novel transcripts could be validated with the proteome.

Multi-omics approaches are a promising method to further our understanding of immune responses. Our study scratches the surface of biological insight to be reaped from a combination of multi-omics and long-read sequencing data and was hindered only by the aforementioned limitations in the samples themselves. Removing these limitations will undoubtedly result in deeper mechanistic understanding and will translate into better outcomes for patients. Insights gained from this methodology can be used immediately in rare disease diagnostics applications, such as the reannotation of variants using more accurate reference transcriptomes for specific tissues[83], contributing to the development of more personalized medicine.

## Acknowledgements

### Funding

MGN was supported by an ERC Advanced Grant (#833247) and a Spinoza Grant of the Netherlands Organization for Scientific Research. AH was supported by the Solve-RD project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 779257.

## Acknowledgement
None

## Author contributions
Conceptualization: RS, PJV, AH, PACH; data curation: RS; formal analysis: RS; investigation: RS, EEV, CIM, TVR, TH; resources: SK, MS; software: RS, TVR; supervision: MM, MGN, PJV, AH, PACH; visualization: RS, EEV; writing—original draft: RS, EEV; writing—review and editing: All authors. All authors read and approved the final manuscript.

## Keywords
Long-read RNA sequencing, transcriptomics, proteomics, innate immune response, alternative splicing, isoform switching, protein secretion

## Declaration of interests
### Competing interests
The authors declare no competing interests.

### Consent for publication
Not applicable

## Ethics approval and consent to participate
PBMCs were retrieved form healthy, anonymized donors, as part of the human functional genomics project (HFGP). The HFGP study was approved by the Ethical Committee of Radboud University Nijmegen, the Netherlands (no. 42561.091.12). Experiments were conducted according to the principles expressed in the Declaration of Helsinki. Samples of venous blood were drawn after informed consent was obtained.

## Availability of data and material
Raw PacBio sequencing data and transcriptome is available on EGA under accession number EGAS00001006779. Raw QuantSeq sequencing data is available on EGA under accession number EGAS50000000007. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[84] partner repository with the dataset identifier PXD045237 and 10.6019/PXD045237. Scripts used to generate the results described in this paper can be found at https://github.com/cmbi/hpi_isoseq_paper.

## References
1 Medzhitov R, Horng T. Transcriptional control of the inflammatory response. *Nature Reviews Immunology 2009 9:10* 2009; **9**: 692–703.
2 Carpenter S, Ricci EP, Mercier BC, Moore MJ, Fitzgerald KA. Post-transcriptional regulation of gene expression in innate immunity. *Nature Reviews Immunology 2014 14:6* 2014; **14**: 361–376.
3 Wells CA, Chalk AM, Forrest A *et al.* Alternate transcription of the Toll-like receptor signaling cascade. *Genome Biol* 2006; **7**: 1–17.

775 4 Oosting M, Kerstholt M, ter Horst R *et al.* Functional and Genomic Architecture of
776 Borrelia burgdorferi-Induced Cytokine Responses in Humans. *Cell Host Microbe* 2016;
777 **20**: 822–833.
778 5 Li Y, Oosting M, Smeekens SP *et al.* A Functional Genomics Approach to Understand
779 Variation in Cytokine Production in Humans. *Cell* 2016; **167**: 1099-1110.e14.
780 6 Lu YC, Yeh WC, Ohashi PS. LPS/TLR4 signal transduction pathway. *Cytokine* 2008; **42**:
781 145–151.
782 7 Alexopoulou L, Holt AC, Medzhitov R, Flavell RA. Recognition of double-stranded RNA
783 and activation of NF-kappaB by Toll-like receptor 3. *Nature* 2001; **413**: 732–738.
784 8 Van Der Made CI, Simons A, Schuurs-Hoeijmakers J *et al.* Presence of Genetic Variants
785 Among Young Men With Severe COVID-19. *JAMA* 2020; **324**: 663–673.
786 9 Smeekens SP, Ng A, Kumar V *et al.* Functional genomics identifies type I interferon
787 pathway as central for host defense against Candida albicans. *Nat Commun* 2013; **4**.
788 doi:10.1038/NCOMMS2343.
789 10 Bruno M, Kersten S, Bain JM *et al.* Transcriptional and functional insights into the host
790 immune response against the emerging fungal pathogen Candida auris. *Nat Microbiol*
791 2020; **5**: 1516–1531.
792 11 Askarian F, Wagner T, Johannessen M, Nizet V. Staphylococcus aureus modulation of
793 innate immune responses through Toll-like (TLR), (NOD)-like (NLR) and C-type lectin
794 (CLR) receptors. *FEMS Microbiol Rev* 2018; **42**: 656–671.
795 12 Netea MG, Gow NAR, Munro CA *et al.* Immune sensing of Candida albicans requires
796 cooperative recognition of mannans and glucans by lectin and Toll-like receptors. *J*
797 *Clin Invest* 2006; **116**: 1642–1650.
798 13 Heinhuis B, Koenders MI, Van De Loo FA, Netea MG, Van Den Berg WB, Joosten LAB.
799 Inflammation-dependent secretion and splicing of IL-32γ in rheumatoid arthritis. *Proc*
800 *Natl Acad Sci U S A* 2011; **108**: 4962–4967.
801 14 Oberdoerffer S, Moita LF, Neems D, Freitas RP, Hacohen N, Rao A. Regulation of CD45
802 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science* 2008; **321**:
803 686–691.
804 15 Sharon D, Tilgner H, Grubert F, Snyder M. A single-molecule long-read survey of the
805 human transcriptome. *Nat Biotechnol* 2013; **31**: 1009–1014.
806 16 Glinos DA, Garborcauskas G, Hoffman P *et al.* Transcriptome variation in human
807 tissues revealed by long-read sequencing. *Nature 2022 608:7922* 2022; **608**: 353–359.
808 17 Al'Khafaji AM, Smith JT, Garimella K V. *et al.* High-throughput RNA isoform
809 sequencing using programmed cDNA concatenation. *Nature Biotechnology 2023*
810 2023; : 1–5.
811 18 Vollmers AC, Mekonen HE, Campos S, Carpenter S, Vollmers C. Generation of an
812 isoform-level transcriptome atlas of macrophage activation. *J Biol Chem* 2021; **296**.
813 doi:10.1016/J.JBC.2021.100784.
814 19 Cole C, Byrne A, Adams M, Volden R, Vollmers C. Complete characterization of the
815 human immune cell transcriptome using accurate full-length cDNA sequencing.
816 *Genome Res* 2020; **30**: 589–601.
817 20 Inamo J, Suzuki A, Ueda M *et al.* Immune Isoform Atlas: Landscape of alternative
818 splicing in human immune cells. *bioRxiv* 2022; : 2022.09.13.507708.
819 21 Kanno T, Konno R, Miyako K *et al.* Characterization of proteogenomic signatures of
820 differentiation of CD4+ T cell subsets. *DNA Research* 2023; **30**: 1–11.

821 22 Shi ZR, Duan YX, Cui F *et al.* Integrated proteogenomic characterization reveals an
822 imbalanced hepatocellular carcinoma microenvironment after incomplete
823 radiofrequency ablation. *Journal of Experimental and Clinical Cancer Research* 2023;
824 **42**: 1–18.
825 23 Proteogenomics R, Bautista De Sanctis J, Subbannayya Y *et al.* Proteogenomics
826 Analysis Reveals Novel Micropeptides in Primary Human Immune Cells. *Immuno*
827 *2022, Vol 2, Pages 283-292* 2022; **2**: 283–292.
828 24 Rivero-Hinojosa S, Grant M, Panigrahi A *et al.* Proteogenomic discovery of
829 neoantigens facilitates personalized multi-antigen targeted T cell immunotherapy for
830 brain tumors. *Nature Communications 2021 12:1* 2021; **12**: 1–15.
831 25 Oosting M, Kerstholt M, ter Horst R *et al.* Functional and Genomic Architecture of
832 Borrelia burgdorferi-Induced Cytokine Responses in Humans. *Cell Host Microbe* 2016;
833 **20**: 822–833.
834 26 Prjibelski A, Mikheenko A, Joglekar A, Smetanin A, Lapidus A, Tilgner H. IsoQuant: a
835 tool for accurate novel isoform discovery with long reads. 2022.
836 doi:10.21203/RS.3.RS-1571850/V1.
837 27 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing
838 reads. *EMBnet J* 2011; **17**: 10–12.
839 28 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-
840 aware quantification of transcript expression. *Nat Methods* 2017; **14**: 417–419.
841 29 Tarazona S, Furió-Tarí P, Turrà D *et al.* Data quality aware analysis of differential
842 expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res* 2015; **43**: e140.
843 30 Ashburner M, Ball CA, Blake JA *et al.* Gene Ontology: tool for the unification of
844 biology. *Nature Genetics 2000 25:1* 2000; **25**: 25–29.
845 31 Wingender E, Dietze P, Karas H, Knüppel R. TRANSFAC: a database on transcription
846 factors and their DNA binding sites. *Nucleic Acids Res* 1996; **24**: 238–241.
847 32 Reese F, Mortazavi A. Swan: a library for the analysis and visualization of long-read
848 transcriptomes. *Bioinformatics* 2021; **37**: 1322–1323.
849 33 Vitting-Seerup K, Sandelin A, Berger B. IsoformSwitchAnalyzeR: analysis of changes in
850 genome-wide patterns of alternative splicing and its functional consequences.
851 *Bioinformatics* 2019; **35**: 4469–4471.
852 34 Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: coding-Potential
853 Assessment Tool using an alignment-free logistic regression model. *Nucl Acids Res*
854 2013; **41**: e74.
855 35 Eddy SR. HMMER User's Guide Biological sequence analysis using profile hidden
856 Markov models. 2020.http://hmmer.org (accessed 2 Feb2022).
857 36 Almagro Armenteros JJ, Tsirigos KD, Sønderby CK *et al.* SignalP 5.0 improves signal
858 peptide predictions using deep neural networks. *Nature Biotechnology 2019 37:4*
859 2019; **37**: 420–423.
860 37 Shannon P, Markiel A, Ozier O *et al.* Cytoscape: a software environment for
861 integrated models of biomolecular interaction networks. *Genome Res* 2003; **13**:
862 2498–2504.
863 38 Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment Map: A Network-Based
864 Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* 2010; **5**:
865 e13984.
866 39 Fang H. dcGOR: an R package for analysing ontologies and protein domain
867 annotations. *PLoS Comput Biol* 2014; **10**. doi:10.1371/JOURNAL.PCBI.1003929.

868 40 Miller RM, Jordan BT, Mehlferber MM *et al.* Enhanced protein isoform
869 characterization through long-read proteogenomics. *Genome Biol* 2022; **23**.
870 doi:10.1186/s13059-022-02624-y.
871 41 Miller RM, Millikin RJ, Rolfs Z, Shortreed MR, Smith LM. Enhanced Proteomic Data
872 Analysis with MetaMorpheus. *Methods Mol Biol* 2023; **2426**: 35–66.
873 42 Deutsch EW, Lane L, Overall CM *et al.* Human Proteome Project Mass Spectrometry
874 Data Interpretation Guidelines 3.0. *J Proteome Res* 2019; **18**: 4108–4116.
875 43 Millikin RJ, Shortreed MR, Scalf M, Smith LM. Fast, Free, and Flexible Peptide and
876 Protein Quantification with FlashLFQ. *Methods Mol Biol* 2023; **2426**: 303–313.
877 44 Digre A, Lindskog C. The Human Protein Atlas-Spatial localization of the human
878 proteome in health and disease. *Protein Sci* 2021; **30**: 218–233.
879 45 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
880 multidimensional genomic data. *Bioinformatics* 2016; **32**: 2847–2849.
881 46 Noguchi S, Arakawa T, Fukuda S *et al.* FANTOM5 CAGE profiles of human and mouse
882 samples. *Sci Data* 2017; **4**. doi:10.1038/SDATA.2017.112.
883 47 Moll P, Ante M, Seitz A, Reda T. QuantSeq 3' mRNA sequencing for RNA
884 quantification. *Nature Methods 2014 11:12* 2014; **11**: i–iii.
885 48 Raudvere U, Kolberg L, Kuzmin I *et al.* g:Profiler: a web server for functional
886 enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*
887 2019; **47**: W191–W198.
888 49 Paun A, Pitha PM. The IRF family, revisited. *Biochimie* 2007; **89**: 744–753.
889 50 Herbein G, Doyle AG, Montaner LJ, Gordon S. Lipopolysaccharide (LPS) down-
890 regulates CD4 expression in primary human macrophages through induction of
891 endogenous tumour necrosis factor (TNF) and IL-1 beta. *Clin Exp Immunol* 1995; **102**:
892 430–437.
893 51 Lachmandas E, Boutens L, Ratter JM *et al.* Microbial stimulation of different Toll-like
894 receptor signalling pathways induces diverse metabolic programmes in human
895 monocytes. *Nat Microbiol* 2016; **2**. doi:10.1038/NMICROBIOL.2016.246.
896 52 Green ID, Pinello N, Song R *et al.* Macrophage development and activation involve
897 coordinated intron retention in key inflammatory regulators. *Nucleic Acids Res* 2020;
898 **48**: 6513–6529.
899 53 Song R, Tikoo S, Jain R *et al.* Dynamic intron retention modulates gene expression in
900 the monocytic differentiation pathway. *Immunology* 2022; **165**: 274–286.
901 54 Wong JJL, Ritchie W, Ebner OA *et al.* Orchestrated intron retention regulates normal
902 granulocyte differentiation. *Cell* 2013; **154**. doi:10.1016/J.CELL.2013.06.052.
903 55 Ullrich S, Guigó R. Dynamic changes in intron retention are tightly associated with
904 regulation of splicing factors and proliferative activity during B-cell development.
905 *Nucleic Acids Res* 2020; **48**: 1327–1340.
906 56 O'Grady TM, Baddoo M, Flemington SA, Ishaq EY, Ungerleider NA, Flemington EK.
907 Reversal of splicing infidelity is a pre-activation step in B cell differentiation. *Front*
908 *Immunol* 2022; **13**. doi:10.3389/FIMMU.2022.1060114.
909 57 Karginov TA, Ménoret A, Vella AT. Optimal CD8+ T cell effector function requires
910 costimulation-induced RNA-binding proteins that reprogram the transcript isoform
911 landscape. *Nat Commun* 2022; **13**. doi:10.1038/S41467-022-31228-0.
912 58 Ni T, Yang W, Han M *et al.* Global intron retention mediated gene regulation during
913 CD4+ T cell activation. *Nucleic Acids Res* 2016; **44**: 6817–6829.

914 59 Brunet MA, Levesque SA, Hunting DJ, Cohen AA, Roucou X. Recognition of the
915 polycistronic nature of human genes is critical to understanding the genotype-
916 phenotype relationship. *Genome Res* 2018; **28**: 609–624.
917 60 Heinz S, Texari L, Hayes MGB *et al.* Transcription Elongation Can Affect Genome 3D
918 Structure. *Cell* 2018; **174**: 1522-1536.e22.
919 61 Karasawa T, Kawashima A, Usui F *et al.* Oligomerized CARD16 promotes caspase-1
920 assembly and IL-1β processing. *FEBS Open Bio* 2015; **5**: 348–356.
921 62 Devi S, Indramohan M, Jäger E *et al.* CARD-only proteins regulate in vivo
922 inflammasome responses and ameliorate gout. *Cell Rep* 2023; **42**.
923 doi:10.1016/J.CELREP.2023.112265.
924 63 Oeckinghaus A, Ghosh S. The NF-kappaB family of transcription factors and its
925 regulation. *Cold Spring Harb Perspect Biol* 2009; **1**.
926 doi:10.1101/CSHPERSPECT.A000034.
927 64 Fliegauf M, Kinnunen M, Posadas-Cantera S *et al.* Detrimental NFKB1 missense
928 variants affecting the Rel-homology domain of p105/p50. *Front Immunol* 2022; **13**.
929 doi:10.3389/FIMMU.2022.965326.
930 65 Mata-Martínez P, Bergón-Gutiérrez M, del Fresno C. Dectin-1 Signaling Update: New
931 Perspectives for Trained Immunity. *Front Immunol* 2022; **13**: 812148.
932 66 Rotival M, Quach H, Quintana-Murci L. Defining the genetic and evolutionary
933 architecture of alternative splicing in response to infection. *Nat Commun* 2019; **10**: 1–
934 15.
935 67 Banday AR, Stanifer ML, Florez-Vargas O *et al.* Genetic regulation of OAS1 nonsense-
936 mediated decay underlies association with COVID-19 hospitalization in patients of
937 European and African ancestries. *Nature Genetics 2022 54:8* 2022; **54**: 1103–1116.
938 68 Wickenhagen A, Sugrue E, Lytras S *et al.* A prenylated dsRNA sensor protects against
939 severe COVID-19. *Science (1979)* 2021; **374**.
940 doi:10.1126/SCIENCE.ABJ3624/SUPPL_FILE/SCIENCE.ABJ3624_MDAR_REPRODUCIBILI
941 TY_CHECKLIST.PDF.
942 69 Perišić Nanut M, Pečar Fonović U, Jakoš T, Kos J. The Role of Cysteine Peptidases in
943 Hematopoietic Stem Cell Differentiation and Modulation of Immune System Function.
944 *Front Immunol* 2021; **12**. doi:10.3389/FIMMU.2021.680279.
945 70 Thiery J, Keefe D, Saffarian S *et al.* Perforin activates clathrin- and dynamin-
946 dependent endocytosis, which is required for plasma membrane repair and delivery
947 of granzyme B for granzyme-mediated apoptosis. *Blood* 2010; **115**: 1582–1593.
948 71 Momoi T, Suzuki M, Titani K, Hisanaga S, Ogawa H, Saito A. Amino acid sequence of a
949 modified β2-microglobulin in renal failure patient urine and long-term dialysis patient
950 blood. *Clinica Chimica Acta* 1995; **236**: 135–144.
951 72 Fukuhara K, Okumura M, Shiono H *et al.* A study on CD45 isoform expression during
952 T-cell development and selection events in the human thymus. *Hum Immunol* 2002;
953 **63**: 394–404.
954 73 Orta-Mascaró M, Consuegra-Fernández M, Carreras E *et al.* CD6 modulates
955 thymocyte selection and peripheral T cell homeostasis. *J Exp Med* 2016; **213**: 1387–
956 1397.
957 74 De Arras L, Alper S. Limiting of the innate immune response by SF3A-dependent
958 control of MyD88 alternative mRNA splicing. *PLoS Genet* 2013; **9**.
959 doi:10.1371/JOURNAL.PGEN.1003855.

960 75 Pozzi B, Bragado L, Mammi P *et al.* Dengue virus targets RBM10 deregulating host cell
961 splicing and innate immune response. *Nucleic Acids Res* 2020; **48**: 6824–6838.
962 76 Stein MM, Conery M, Magnaye KM *et al.* Sex-specific differences in peripheral blood
963 leukocyte transcriptional response to LPS are enriched for HLA region and X
964 chromosome genes. *Sci Rep* 2021; **11**. doi:10.1038/S41598-020-80145-Z.
965 77 Piasecka B, Duffy D, Urrutia A *et al.* Distinctive roles of age, sex, and genetics in
966 shaping transcriptional variation of human immune responses to microbial
967 challenges. *Proc Natl Acad Sci U S A* 2018; **115**: E488–E497.
968 78 Li Y, Oosting M, Deelen P *et al.* Inter-individual variability and genetic influences on
969 cytokine responses to bacteria and fungi. *Nature Medicine 2016 22:8* 2016; **22**: 952–
970 960.
971 79 Mironov A, Petrova M, Margasyuk S *et al.* Tissue-specific regulation of gene
972 expression via unproductive splicing. *Nucleic Acids Res* 2023; **51**: 3055–3066.
973 80 Hardwick SA, Hu W, Joglekar A *et al.* Single-nuclei isoform RNA sequencing unlocks
974 barcoded exon connectivity in frozen brain tissue. *Nat Biotechnol* 2022; **40**: 1082–
975 1092.
976 81 Meissner F, Scheltema RA, Mollenkopf HJ, Mann M. Direct proteomic quantification
977 of the secretome of activated immune cells. *Science (1979)* 2013; **340**: 475–478.
978 82 Sinitcyn P, Richards AL, Weatheritt RJ *et al.* Global detection of human variants and
979 isoforms by deep proteome sequencing. *Nature Biotechnology 2023* 2023; : 1–11.
980 83 Salz R, Saraiva-Agostinho N, Vorsteveld E *et al.* SUsPECT: a pipeline for variant effect
981 prediction based on custom long-read transcriptomes for improved clinical variant
982 annotation. *BMC Genomics* 2023; **24**: 1–10.
983 84 Perez-Riverol Y, Bai J, Bandla C *et al.* The PRIDE database resources in 2022: a hub for
984 mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 2022; **50**: D543–
985 D552.
986

987 ## Figure legends

988 Figure 1: Experimental setup. A) Human peripheral blood mononuclear cells (PBMCs)
989 isolated from five donors were exposed to four different pathogenic stimuli and analyzed
990 after 24 hours. B) PacBio long read RNA-sequencing was performed on samples from one of
991 the five donors. Long-reads were used to estimate differential transcript expression and
992 isoform switching. C) The supernatant from all samples (all donors) was collected and
993 peptides were detected to quantify protein levels in the secretome. D) Short read RNA
994 sequencing (QuantSeq) was performed on all samples (all donors) and differential
995 expression estimates were compared to those measured in long-read sequencing.
996

997 Figure 2: Transcriptome novelty in the control condition and comparison between stimuli
998 transcriptomes in the five long-read samples. A) Transcript novelty categories. GENCODE
999 (blue) is the set of all known reference transcripts. Novel in catalog (orange) contains a
1000 novel combination of annotated introns. Novel not in catalog (green) contains one or more
1001 unannotated introns. B) Reads (top) and unique transcripts (bottom) of events in each pre-
1002 defined transcript novelty category in RPMI. C) Novelty-inducing events occurring in the
1003 RPMI transcriptome. D) Unique transcripts by novelty category for each of the stimulus
1004 conditions. E) Unique transcripts by novelty category that remain at various transcript
1005 abundance thresholds in the C*. albicans* condition. F) Transcriptomes of the samples plotted

1006 on the first two principal components of PCA. G) Jaccard distances of genes (left) and known
1007 transcripts (right) of the transcriptomes, not considering transcript abundance.
1008
1009 Figure 3: Differential pathway analysis originating from differentially expressed genes on the
1010 RNA level. A) Overlap between enriched pathways generated from the differentially
1011 expressed genes from the four conditions. B) Selected pathways found to be enriched for all
1012 conditions, C) three of the four conditions (Poly(I:C), *C. albicans* and *S. aureus*), D)
1013 specifically for LPS and E) specifically for Poly(I:C).
1014
1015 Figure 4: Isoform switching induced by pathogen stimulation. A) Overlap of isoform
1016 switching genes between the four stimulus conditions. B) Pathway network analysis derived
1017 from genes found to undergo isoform switching (IS) upon pathogen stimulation. Each
1018 pathway is colored by p value, where a darker red indicates a lower p value. C) Proportions
1019 of total IS events in each stimulated condition per IS consequence. D) Number of IS by
1020 category of switch pairs. Categories are defined by involvement of novel transcripts in a
1021 given IS. "Novel down" indicates that the isoform switched from a higher proportion of the
1022 novel transcript in control to a higher proportion of a known transcript in the stimulus
1023 condition. "Both known" indicates that the IS occurs between 2 reference transcripts. E)
1024 Fraction of each transcript novelty combination per IS consequence. Normalized by total
1025 number of IS events per novelty category.
1026
1027 Figure 5: *A novel readthrough transcript of CASP1.* A) USCS genome browser track of the
1028 transcripts detected in the control condition (RPMI) and stimulated condition (Poly(I:C)). The
1029 novel readthrough transcript containing both CASP1 and CARD16 is presented in light blue.
1030 Known transcripts in in GENCODE are presented below. B) Representation of the domains in
1031 the novel *CASP1* transcript, indicating that *CARD16* is entirely included in the 5' UTR of the
1032 transcript. C) Gene and transcript expression and isoform fraction of the *CASP1* transcripts
1033 that were detected.
1034
1035 Figure 6: A novel transcript of *NFKB1.* A) UCSC genome browser track of the transcripts
1036 detected in the control condition (RPMI) and stimulated conditions. The novel transcript is
1037 presented in light blue. Known transcripts in GENCODE are presented below. B) Zoomed
1038 view of the transcription start site of the novel transcript with CAGE peaks (monocyte) in
1039 this region. C) Representation of the domains in the known and novel *NFKB1* transcripts that
1040 were detected. D) Gene and transcript expression and isoform fraction of the *NFKB1*
1041 transcripts that were detected.
1042
1043 Figure 7: Protein expression in the secretome. A) Normalized protein expression detected
1044 from five donors in the five conditions. Donors are denoted with numbers 1 through 5.
1045 Clusters originating from kmeans clustering are shown in the heatmap. B) Gene ontology
1046 Biological process (top) and molecular function (bottom) pathways of proteins found in
1047 cluster 4. C) Clustering of genes found in cluster 4. Genes are found to be differentially
1048 expressed on the protein level are in color, others greyed. D) Fold change of gene
1049 differential expression versus fold change of protein DE for genes that were differentially
1050 expressed on both levels, colored by stimulus. Triangle-shaped points correspond with
1051 cluster 4 genes from A. Genes with concordant protein and RNA expression are in the upper
1052 right and lower left quadrants.

1053
1054 ## Supplemental figures
1055 Supplemental figure 1: *Elbow plots using kmeans clustering*. Elbow plots to determine the
1056 optimal number of kmeans clusters to use for clustering of proteomics data.
1057
1058 Supplementary figure 2: *Gene types where novel transcripts are found*. A) Most common
1059 categories of genes. B) Categorization by genes that are related to immunity
1060
1061 Supplemental figure 3: *Transcriptome trends in all conditions.* A) TPM distribution for each
1062 of the conditions measured. B) Ratios of novelty events in the transcriptome by condition.
1063
1064 Supplemental figure 4: *Differential expression on the RNA level.* A) Differentially expressed
1065 genes. B) Differentially expressed transcripts.
1066
1067 Supplemental figure 5: *Differential gene expression validation using short read sequencing*
1068 *(QuantSeq).* A) Correlation of DEGs of each condition between long- and short-read
1069 sequencing. B) Correlation between long-and short-read sequencing when comparing each
1070 condition to each other.
1071
1072 Supplemental figure 6: *Overview of isoform switching.* D) Number of isoform switches
1073 occurring per gene. E) Max number of measured consequences per isoform switch. F) Total
1074 number of each type of isoform switching consequence
1075
1076 Supplemental figure 7: *Pathway analysis of genes affected by intron retention loss.*
1077 Enrichment was calculated using gProfiler using 304 intron retention loss isoform switches
1078 from 145 genes.
1079
1080 Supplemental figure 8: *Domains that were most frequently gained/lost.* A) Domains gained.
1081 B) Domains lost
1082
1083 Supplemental figure 9: *Examples of stimulus-specific isoform switching.* A) isoform switching
1084 in *CLEC7A* specific to *C. albicans* stimulation. B) Differential expression of *OAS1* for three
1085 stimuli. C) isoform switching of *OAS1* specific to Poly(I:C) stimulation.
1086
1087 Supplemental figure 10: *Open reading frames resulting from novel transcripts and their*
1088 *detectability*. A) RNA vs predicted protein novelty. Color indicates RNA novelty category. X-
1089 axis corresponds to unique transcripts. B) (Predicted) protein sequences that comprise the
1090 full search database. C) Unique peptide sequences in the peptide search database by origin.
1091 Multi-mapping indicates 2 or more possible origin categories.
1092
1093 Supplemental figure 11: *Principal component analysis of peptide identifications on the*
1094 *secretome of all samples.* A) Colored by individuals 1-5. B) Colored by stimulus condition.
1095
1096 Supplemental figure 12: *Differential expression on the secretome.* A) Differentially expressed
1097 proteins by condition. B) Differentially expressed protein volcano plots per condition.
1098

## Supplemental tables

Supplemental table 1: *Transcript counts.*

Supplemental table 2: *Protein quantification.*

Supplemental table 3: *Differentially expressed genes.*

Supplemental table 4: *Differentially expressed transcripts.*

Supplemental table 5: *Correlation long-read and short-read Differentially expressed genes.*

Supplemental table 6: *Pathway enrichment of differentially expressed genes.*

Supplemental table 7: *Consequences per isoform switch.*

Supplemental table 8: *Features per isoform switch.*

Supplemental table 9: *Nodes from the network analysis of genes undergoing isoform switching.* Generated using Cytoscape.

Supplemental table 10: *Edges from the network analysis of genes undergoing isoform switching.* Generated using Cytoscape.

Supplemental table 11: *Pathway analysis of genes undergoing isoform switching.* Generated using Cytoscape.

Supplemental table 12: *Nodes from the enrichment map from genes undergoing isoform switching.* Generated using Cytoscape.

Supplemental table 13: *Edges from the enrichment map from genes undergoing isoform switching.* Generated using Cytoscape.

Supplemental table 14: *Enrichment of consequences of isoform switching.*

Supplemental table 15: *Pathway enrichment of genes affected by intron retention loss/gain*

Supplemental table 16: *Domain gains and losses.*

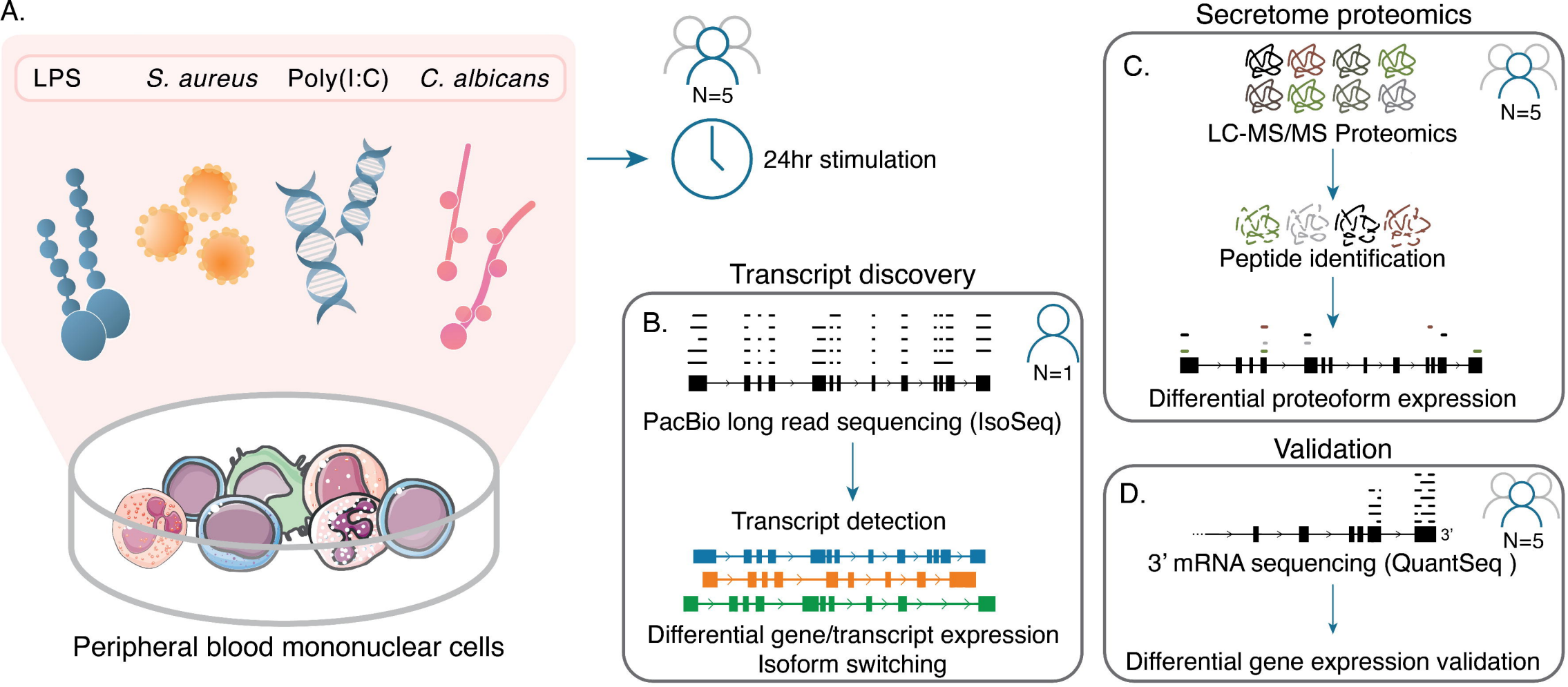
Supplemental table 17: *Pathway analysis of genes found to undergo domain gains or losses.*

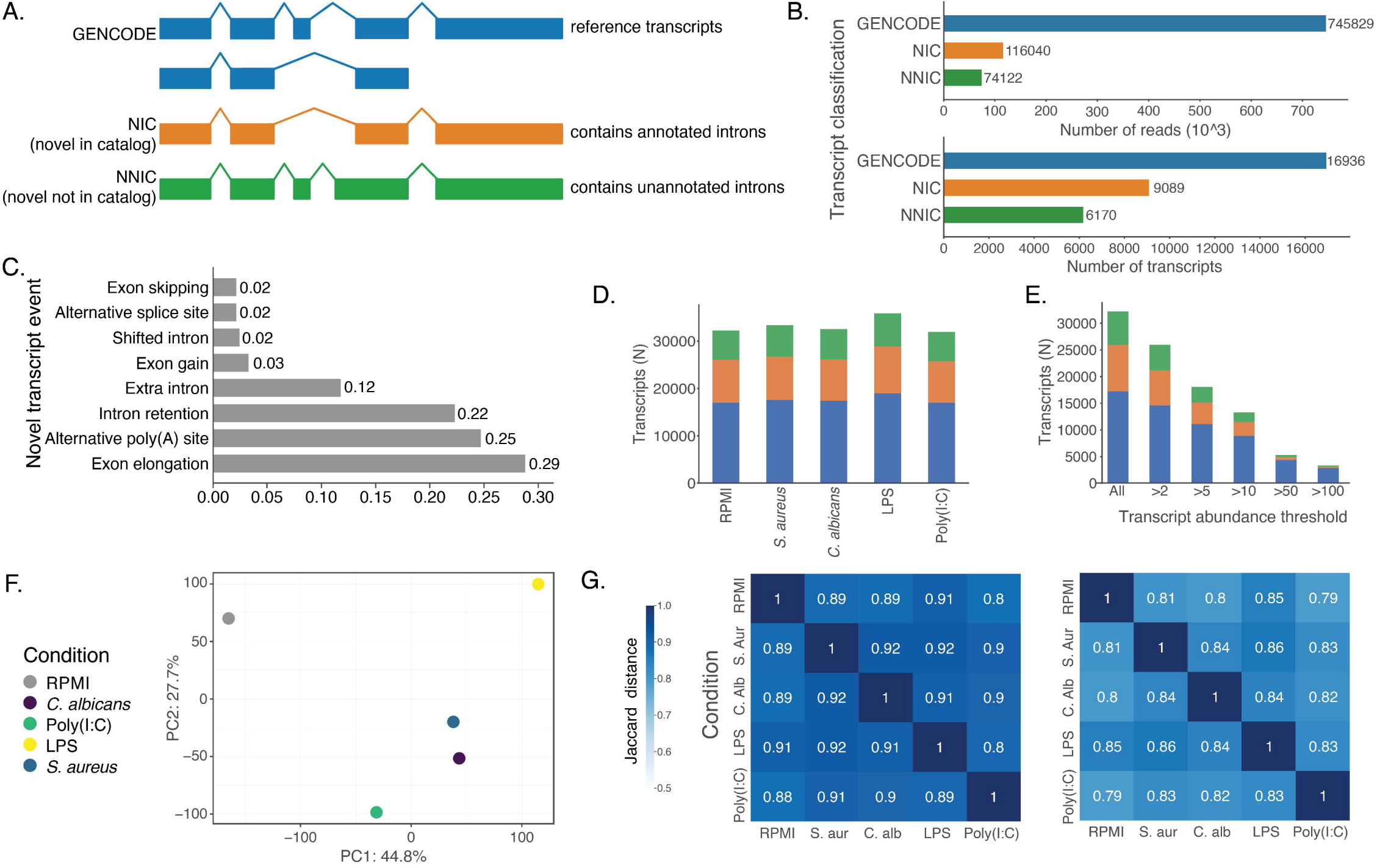
Supplemental table 18: *Analysis of Gene Ontology molecular functions of domains gained or lost using dcGOR.*

Supplemental table 19: *Transcript novelty combination counts per isoform switch consequence.*

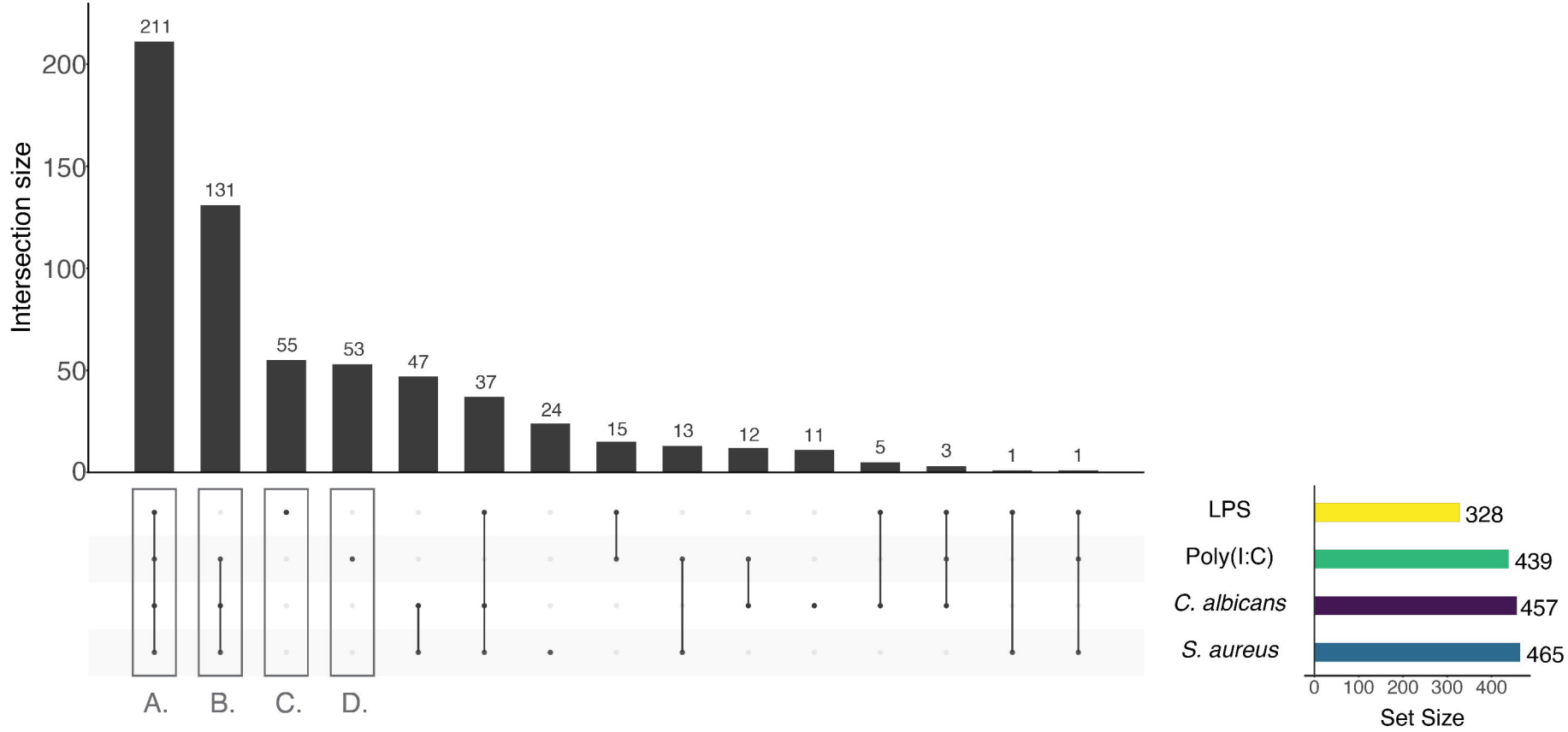
Supplemental table 20: *Differentially expressed proteins.*

1146
1147 Supplemental table 21: *Proteome heatmap expression values.*
1148
1149 Supplemental table 22: *Pathway analysis of proteins in cluster 4.*
1150
1151 Supplemental table 23: *Nodes from the network analysis of proteins found in cluster 4.*
1152 Generated using Cytoscape.
1153
1154 Supplemental table 24: *Edges from the network analysis of proteins found in cluster 4.*
1155 Generated using Cytoscape.
1156
1157 Supplemental table 25: *Pathway analysis of differentially expressed proteins found in cluster*
1158 *4.*
1159
1160 Supplemental table 26: *Correlation between differentially expressed genes and proteins.*
1161
1162 Supplemental table 27: *Proteome evidence for intron retentions.*

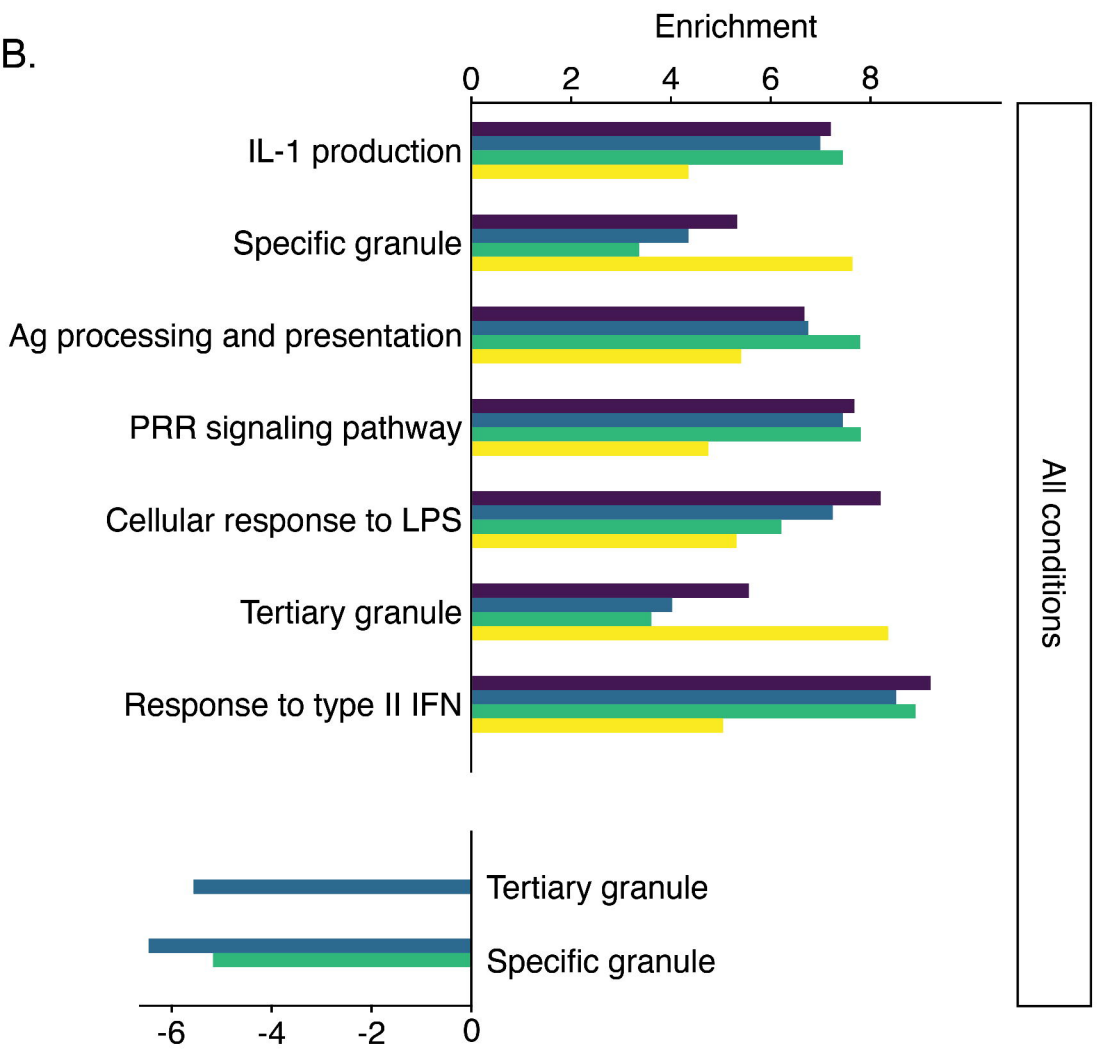

A.
LPS
S. aureus
Poly(I:C)
C. albicans
Peripheral blood mononuclear cells
N=5
24hr stimulation
Transcript discovery
B.
N=1
PacBio long read sequencing (IsoSeq)
Transcript detection
Differential gene/transcript expression
Isoform switching
Secretome proteomics
C.
N=5
LC-MS/MS Proteomics
Peptide identification
Differential proteoform expression
Validation
D.
3'
N=5
3' mRNA sequencing (QuantSeq )
Differential gene expression validation


A.
GENCODE
reference transcripts
NIC
(novel in catalog)
contains annotated introns
NNIC
(novel not in catalog)
contains unannotated introns
B.
Transcript classification
GENCODE
NIC
NNIC
745829
116040
74122
0 100 200 300 400 500 600 700
Number of reads (10^3)
GENCODE
NIC
NNIC
16936
9089
6170
0 2000 4000 6000 8000 10000 12000 14000 16000
Number of transcripts
C.
Novel transcript event
Exon skipping 0.02
Alternative splice site 0.02
Shifted intron 0.02
Exon gain 0.03
Extra intron 0.12
Intron retention 0.22
Alternative poly(A) site 0.25
Exon elongation 0.29
0.00 0.05 0.10 0.15 0.20 0.25 0.30
D.
Transcripts (N)
0 10000 20000 30000
RPMI
S. aureus
C. albicans
LPS
Poly(I:C)
E.
Transcripts (N)
0 5000 10000 15000 20000 25000 30000
All >2 >5 >10 >50 >100
Transcript abundance threshold
F.
Condition
RPMI
C. albicans
Poly(I:C)
LPS
S. aureus
PC2: 27.7%
PC1: 44.8%
100 50 0 −50 −100
−100 0 100
G.
Jaccard distance
1.0 0.9 0.8 0.7 0.6 0.5
Condition
RPMI 1 0.89 0.89 0.91 0.8
S. Aur 0.89 1 0.92 0.92 0.9
C. Alb 0.89 0.92 1 0.91 0.9
LPS 0.91 0.92 0.91 1 0.8
Poly(I:C) 0.88 0.91 0.9 0.89 1
RPMI S. aur C. alb LPS Poly(I:C)
RPMI 1 0.81 0.8 0.85 0.79
S. Aur 0.81 1 0.84 0.86 0.83
C. Alb 0.8 0.84 1 0.84 0.82
LPS 0.85 0.86 0.84 1 0.83
Poly(I:C) 0.79 0.83 0.82 0.83 1
RPMI S. aur C. alb LPS Poly(I:C)

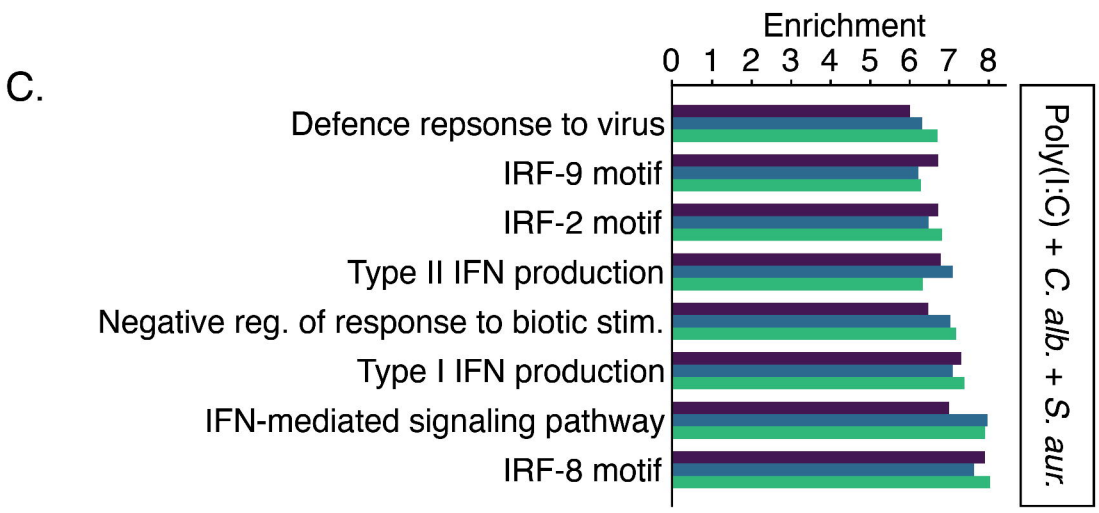
A.


Intersection size
211
131
55
53
47
37
24
15
13
12
11
5
3
1
1
A.
B.
C.
D.
LPS
Poly(I:C)
C. albicans
S. aureus
328
439
457
465
0 100 200 300 400
Set Size

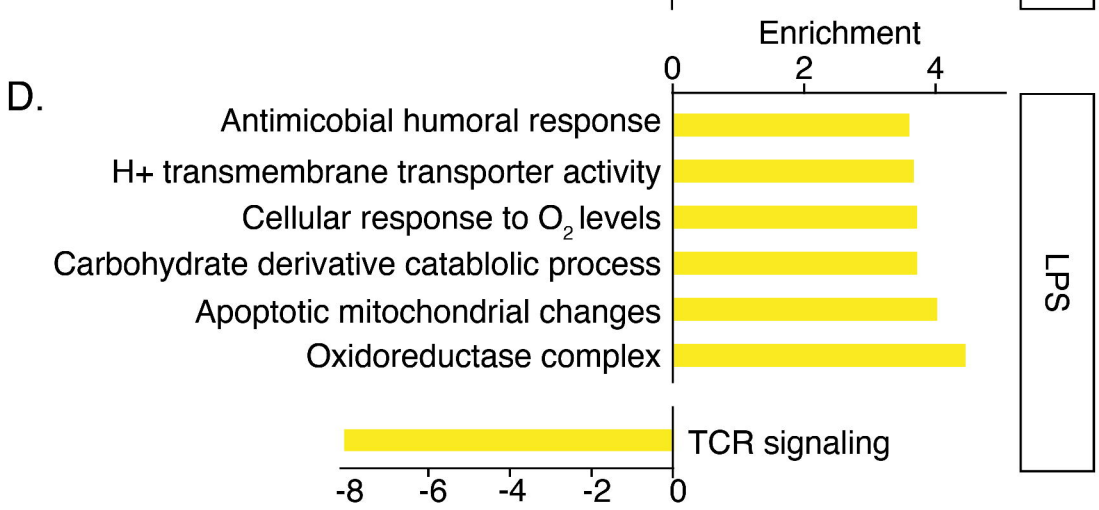
B.


Enrichment
0 2 4 6 8
IL-1 production
Specific granule
Ag processing and presentation
PRR signaling pathway
Cellular response to LPS
Tertiary granule
Response to type II IFN
All conditions
Tertiary granule
Specific granule
-6 -4 -2 0

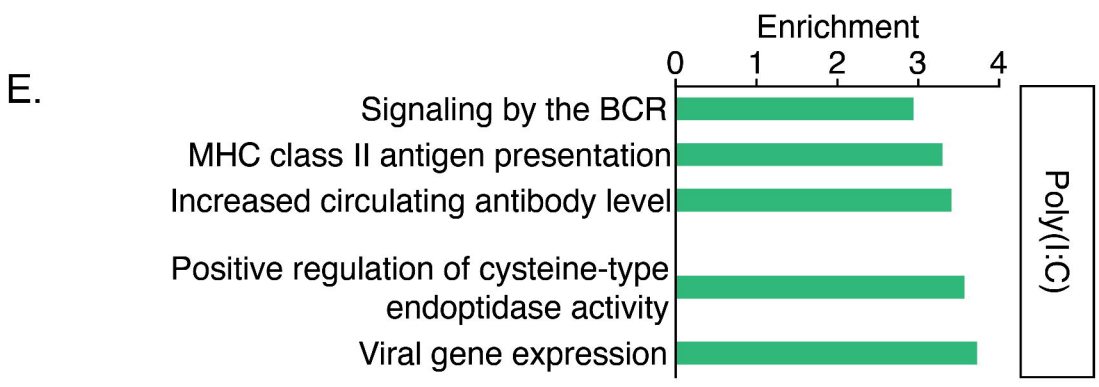
C.

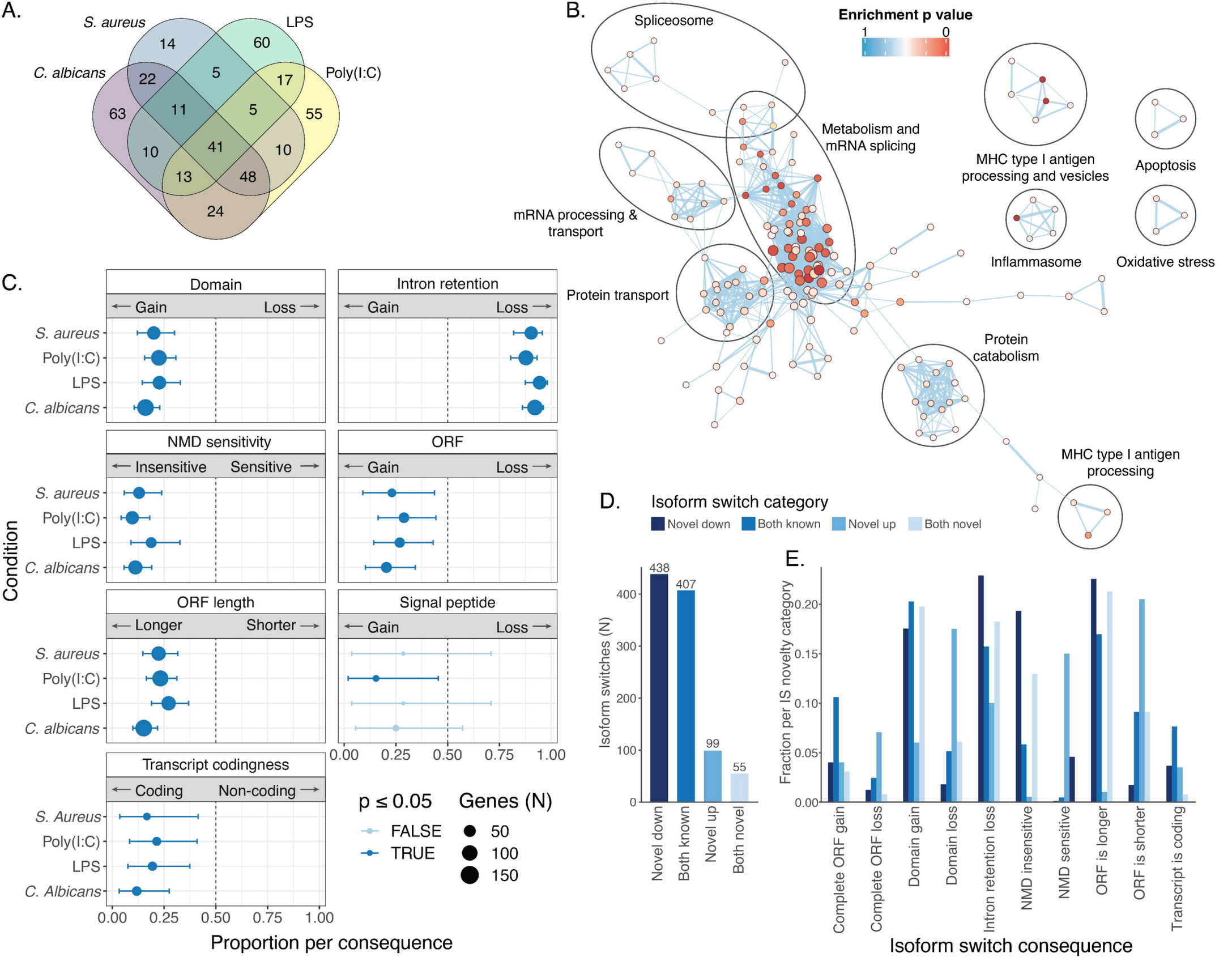

Enrichment
0 1 2 3 4 5 6 7 8
Defence repsonse to virus
IRF-9 motif
IRF-2 motif
Type II IFN production
Negative reg. of response to biotic stim.
Type I IFN production
IFN-mediated signaling pathway
IRF-8 motif
Poly(I:C) + C. alb. + S. aur.

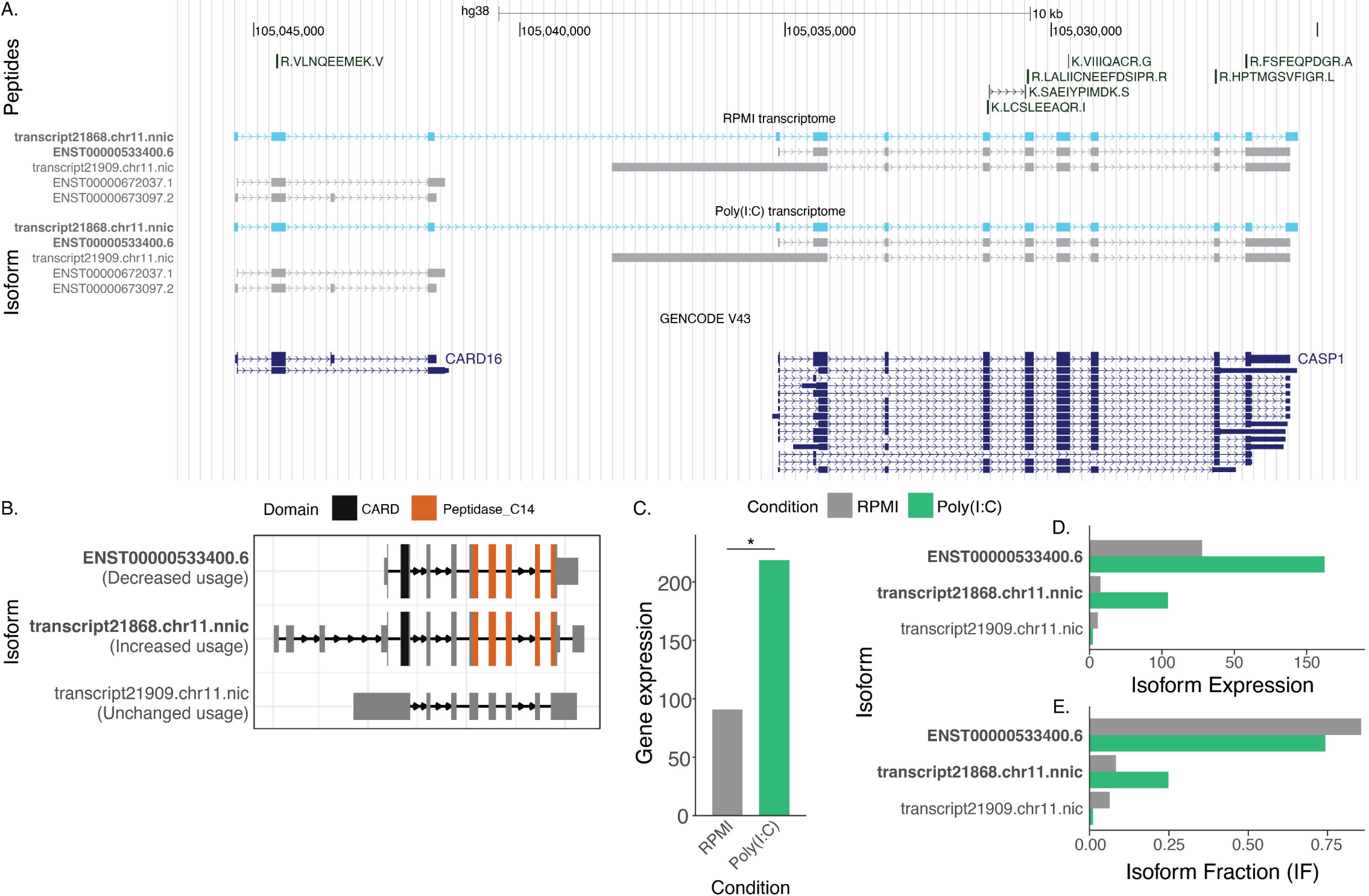
D.

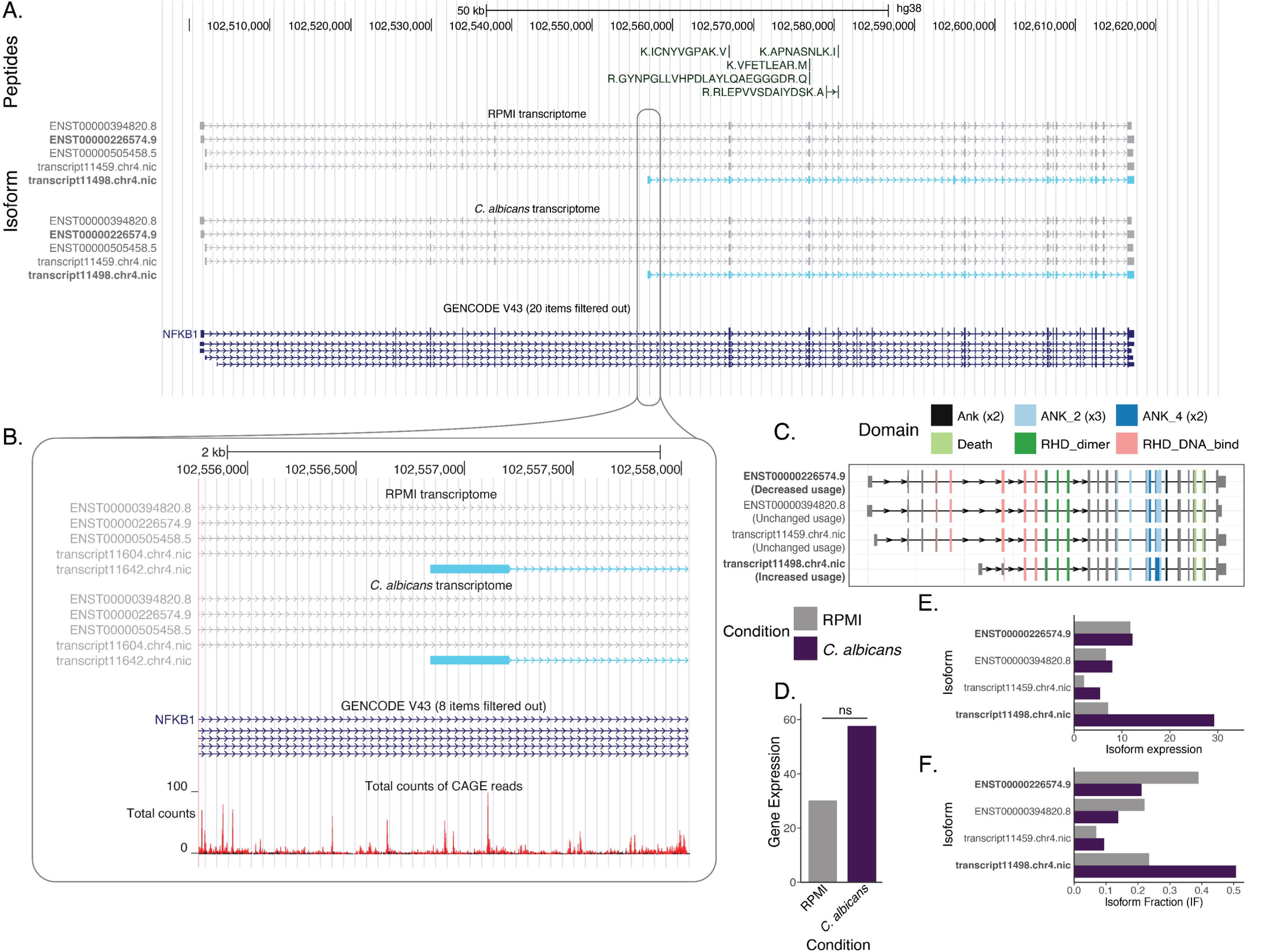

Enrichment
0 2 4
Antimicobial humoral response
H+ transmembrane transporter activity
Cellular response to O2 levels
Carbohydrate derivative catablolic process
Apoptotic mitochondrial changes
Oxidoreductase complex
TCR signaling
-8 -6 -4 -2 0
LPS

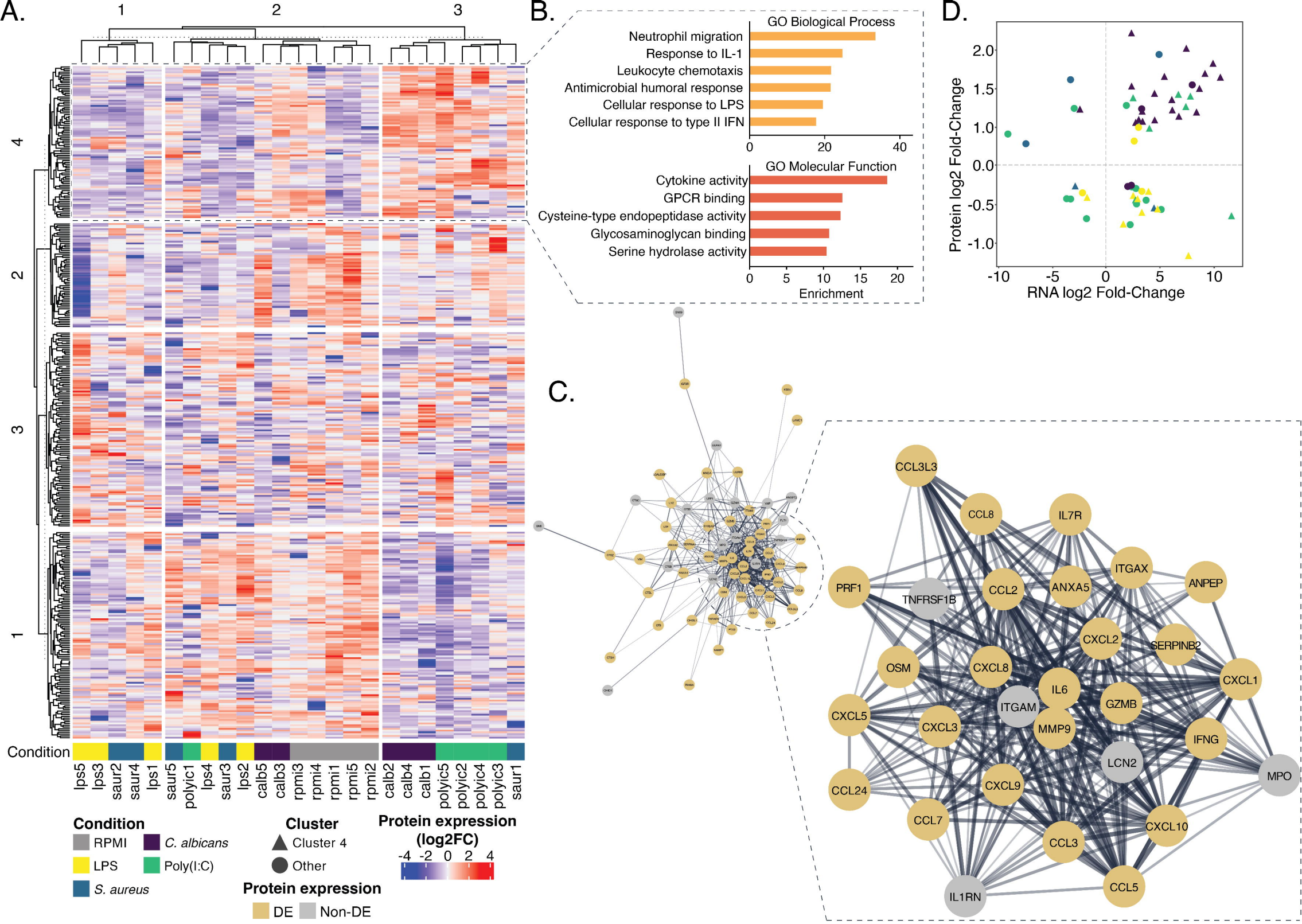
E.


Enrichment
0 1 2 3 4
Signaling by the BCR
MHC class II antigen presentation
Increased circulating antibody level
Positive regulation of cysteine-type endoptidase activity
Viral gene expression
Poly(I:C)


A.
S. aureus
LPS
C. albicans
Poly(I:C)
14
60
22
5
17
63
11
5
55
10
41
10
13
48
24
B.
Enrichment p value
1
0
Spliceosome
Metabolism and mRNA splicing
mRNA processing & transport
Protein transport
Protein catabolism
MHC type I antigen processing and vesicles
Apoptosis
Inflammasome
Oxidative stress
MHC type I antigen processing
C.
Domain
Intron retention
NMD sensitivity
ORF
ORF length
Signal peptide
Transcript codingness
← Gain
Loss →
← Insensitive
Sensitive →
← Longer
Shorter →
← Coding
Non-coding →
S. aureus
Poly(I:C)
LPS
C. albicans
S. Aureus
C. Albicans
Condition
0.00
0.25
0.50
0.75
1.00
Proportion per consequence
p ≤ 0.05
FALSE
TRUE
Genes (N)
50
100
150
D.
Isoform switch category
Novel down
Both known
Novel up
Both novel
Isoform switches (N)
438
407
99
55
0
100
200
300
400
E.
Fraction per IS novelty category
0.00
0.05
0.10
0.15
0.20
Complete ORF gain
Complete ORF loss
Domain gain
Domain loss
Intron retention loss
NMD insensitive
NMD sensitive
ORF is longer
ORF is shorter
Transcript is coding
Isoform switch consequence


A.
hg38
10 kb
105,045,000
105,040,000
105,035,000
105,030,000
Peptides
R.VLNQEEMEK.V
K.VIIIQACR.G
R.FSFEQPDGR.A
R.LALIICNEEFDSIPR.R
R.HPTMGSVFIGR.L
K.SAEIYPIMDK.S
K.LCSLEEAQR.I
RPMI transcriptome
transcript21868.chr11.nnic
ENST00000533400.6
transcript21909.chr11.nic
ENST00000672037.1
ENST00000673097.2
Poly(I:C) transcriptome
transcript21868.chr11.nnic
ENST00000533400.6
transcript21909.chr11.nic
ENST00000672037.1
ENST00000673097.2
Isoform
GENCODE V43
CARD16
CASP1
B.
Domain
CARD
Peptidase_C14
ENST00000533400.6
(Decreased usage)
transcript21868.chr11.nnic
(Increased usage)
transcript21909.chr11.nic
(Unchanged usage)
Isoform
C.
Condition
RPMI
Poly(I:C)
*
Gene expression
200
150
100
50
0
RPMI
Poly(I:C)
Condition
D.
ENST00000533400.6
transcript21868.chr11.nnic
transcript21909.chr11.nic
0
100
50
150
Isoform Expression
Isoform
E.
ENST00000533400.6
transcript21868.chr11.nnic
transcript21909.chr11.nic
0.00
0.25
0.50
0.75
Isoform Fraction (IF)


A.
50 kb
hg38
102,510,000
102,520,000
102,530,000
102,540,000
102,550,000
102,560,000
102,570,000
102,580,000
102,590,000
102,600,000
102,610,000
102,620,000
Peptides
K.ICNYVGPAK.V
K.APNASNLK.I
K.VFETLEAR.M
R.GYNPGLLVHPDLAYLQAEGGGDR.Q
R.RLEPVVSDAIYDSK.A
Isoform
RPMI transcriptome
ENST00000394820.8
ENST00000226574.9
ENST00000505458.5
transcript11459.chr4.nic
transcript11498.chr4.nic
C. albicans transcriptome
ENST00000394820.8
ENST00000226574.9
ENST00000505458.5
transcript11459.chr4.nic
transcript11498.chr4.nic
GENCODE V43 (20 items filtered out)
NFKB1
B.
2 kb
102,556,000
102,556,500
102,557,000
102,557,500
102,558,000
RPMI transcriptome
ENST00000394820.8
ENST00000226574.9
ENST00000505458.5
transcript11604.chr4.nic
transcript11642.chr4.nic
C. albicans transcriptome
ENST00000394820.8
ENST00000226574.9
ENST00000505458.5
transcript11604.chr4.nic
transcript11642.chr4.nic
GENCODE V43 (8 items filtered out)
NFKB1
Total counts of CAGE reads
100
Total counts
0
C.
Domain
Ank (x2)
ANK_2 (x3)
ANK_4 (x2)
Death
RHD_dimer
RHD_DNA_bind
ENST00000226574.9
(Decreased usage)
ENST00000394820.8
(Unchanged usage)
transcript11459.chr4.nic
(Unchanged usage)
transcript11498.chr4.nic
(Increased usage)
Condition
RPMI
C. albicans
D.
ns
Gene Expression
60
40
20
0
RPMI
C. albicans
Condition
E.
Isoform
ENST00000226574.9
ENST00000394820.8
transcript11459.chr4.nic
transcript11498.chr4.nic
0
10
20
30
Isoform expression
F.
Isoform
ENST00000226574.9
ENST00000394820.8
transcript11459.chr4.nic
transcript11498.chr4.nic
0.0
0.1
0.2
0.3
0.4
0.5
Isoform Fraction (IF)


A.
1
2
3
4
2
3
1
Condition
lps5
lps3
saur2
saur4
lps1
saur5
polyic1
lps4
saur3
lps2
calb5
calb3
rpmi3
rpmi4
rpmi1
rpmi5
rpmi2
calb2
calb4
calb1
polyic5
polyic2
polyic4
polyic3
saur1
Condition
RPMI
C. albicans
LPS
Poly(I:C)
S. aureus
Cluster
Cluster 4
Other
Protein expression
DE
Non-DE
Protein expression (log2FC)
-4 -2 0 2 4
B.
GO Biological Process
Neutrophil migration
Response to IL-1
Leukocyte chemotaxis
Antimicrobial humoral response
Cellular response to LPS
Cellular response to type II IFN
0
20
40
GO Molecular Function
Cytokine activity
GPCR binding
Cysteine-type endopeptidase activity
Glycosaminoglycan binding
Serine hydrolase activity
0
5
10
15
20
Enrichment
C.
CCL3L3
CCL8
IL7R
ITGAX
ANPEP
PRF1
TNFRSF1B
CCL2
ANXA5
CXCL2
SERPINB2
OSM
CXCL8
IL6
CXCL1
ITGAM
GZMB
CXCL5
CXCL3
MMP9
IFNG
LCN2
MPO
CCL24
CXCL9
CCL7
CXCL10
CCL3
IL1RN
CCL5
D.
Protein log2 Fold-Change
2.0
1.5
1.0
0.0
-0.5
-1.0
-10
-5
0
5
10
RNA log2 Fold-Change