Cell Host & Microbe, Volume *11*

Supplemental Information

Cofactors Required for TLR7- and TLR9-Dependent Innate Immune Responses

Chih-yuan Chiang, Alex Engel, Amanda M. Opaluch, Irene Ramos, Ana M. Maestre, Ismael Secundino, Paul D. De Jesus, Quy T. Nguyen, Genevieve Welch, Ghislain M.C. Bonamy, Loren J. Miraglia, Anthony P. Orth, Victor Nizet, Ana Fernandez-Sesma, Yingyao Zhou, Gregory M. Barton, and Sumit K. Chanda

Supplemental Inventory

Table S3: Overrepresented functional processes required for TLR7/9 signaling. Related to Figure 1.

Table S4: Confirmed factors involved in autoimmune diseases, inflammatory diseases, and viral infection. Related to Table 1.

Table S5: Binary protein interactions between confirmed factors and known TLR signaling pathway members, and MCODE analysis. Related to Figure 1.

Table S6: Annotations for subcellular localization of confirmed genes, and SVM-predicted genes shown in Figure 2. Related to Figure 2.

Table S7: IL-8 mRNA expression levels, p65 nuclear translocation activities, and ICAM-1 mRNA expression levels in THP-1 cells in response to R848 stimulation. Related to Figure 3.

Supplemental Experimental Procedures

Whole Genome High-throughput siRNA Screens

A 384-well plate-based assay was optimized to identify siRNAs that influence endosomal TLR7 and TLR9 signaling in response to their cognate ligands, R848 and CpG oligonucleotides. Genome-wide libraries comprising 107,734 synthetic siRNAs targeting 19529 unique human genes in total were arrayed in 384-well plates such that each well contained two unique and identifiable siRNAs per gene (7ng siRNA per gene, per well). There were, on average, 3 wells per gene or 6 siRNAs per gene. Each plate also contained positive controls (MyD88 and p65 siRNAs), negative controls (GL3-luciferase siRNA), and scramble siRNA controls. The library was introduced into the HEK293T/TLR7 NF-κB luciferase reporter line or the HEK293T/TLR9 NF-κB luciferase reporter line by a high-throughput transfection process (Konig et al., 2008). For the TLR7 and TLR9 screens, 0.5uM of R848 or 3uM of CpG oligonucleotides (ODN 2006-G5, 5'-TCGTCGTTTTGTCGTTTTGTCGTTGGGGG-3') were added to the wells 72 hours after siRNA transfection. Seven hours post stimulation, Bright-Glo (Promega) was added in equal volumes to each well and the luminescence associated with each sample was analyzed. The TLR7 and TLR9 screens were run in duplicates and statistically analyzed as described below. All steps were performed using a fully integrated high-throughput cellular genomics robotic system (GNF Systems).

siRNA Screening Data Transformation and Primary Hits

Whole-genome siRNA screening data were obtained for 2 assays: TLR7 and TLR9. Each well was measured twice, and readings were normalized and geometrically-averaged using the same data processing pipelines as described previously (Konig et al., 2009; Konig et al., 2008) (also see "siRNA Activity Score" below); briefly, scores were scaled via a non-linear transformation with negative controls set to 1.0 and positive controls set to 0.1. *This process generated a normalized "siRNA activity score" for each well tested.* To aggregate activity scores of multiple siRNAs for the same gene, we applied an RSA (redundant siRNA analysis) algorithm to each screen individually, so that a p-value was obtained for each siRNA in each screen (Konig et al., 2007) (also see "RSA Activity Score" below); *this analysis generated an "RSA activity score" for each gene tested.*

Next, we defined the primary TLR7 and TLR9 hits as those genes with threshold siRNA activity scores less than or equal to 0.4; this resulted in 5660 and 3924 primary hits for the TLR7 and TLR9 screens, respectively. Additionally, whole genome screening data for cytotoxicity (TOX) was obtained previously (Konig et al., 2008). For this dataset, a threshold siRNA activity score was set at less than or equal to 0.7 to identify those genes that were cytotoxic. By this definition, 4958 genes caused significant cellular toxicity.

Correlation Between TLR7 and TLR9 Screens

Altogether, the siRNA activity scores for genes that were profiled in the TLR7 and TLR9 screens have a Pearson correlation coefficient of 0.27 ($p < 10^{-100}$). TLR7 primary hits have a 41% chance to also be considered a hit in the TLR9 screen, while non-TLR7 primary hits only have a 12% chance to be considered TLR9 hits; therefore, the overlap between the two hit lists are highly significant (chi-squared test, $p < 10^{-100}$). Importantly, upon reconfirmation analysis, we did not observe any genes that were TLR7- or TLR9-specific (i.e. all genes impacted both pathways), so we conclude that non-overlapping genes in the primary screen were due to false negative activities.

Evidence-Based Analysis of Screening Data: Integrative evidence-based analysis, gene confirmation, and mechanistic studies (*See Figure S1B for a flow chart of this systematic validation strategy, Figure S1C for a compilation of all evidences, and Table S2 for the metrics and quantification for each gene confirmed by this approach*)

Evidence Analyses Summary

We hypothesized that genes having multiple lines of independent data to support TLR7/TLR9-related functions are less likely to be experimental false positives; therefore, for each gene identified as a primary hit using the threshold siRNA activity scores described above, we compiled various datasets that would provide potential support for association with TLR7/9. In addition to siRNA screening data analyses (siRNA activity scores and RSA activity scores), such types of evidences included: protein network analyses, and gene expression analyses (all evidences are listed below and numbered relative to Figure S1B). Each piece of evidence was transformed into a normalized support score between zero and one, representing a minimum and maximum level of support, respectively. This process of using multiple evidences has been employed in several previous studies (Konig et al., 2009; Konig et al., 2008) and has been shown to effectively prioritize primary hits for confirmation.

Individual Evidences and Analyses Applied:

1. siRNA activity scores and RSA activity scores

To calculate **siRNA activity scores** for each well tested, the raw data was first divided by the plate median, which was centered at 1.0. The duplicate readings of each well were then geometrically averaged into a single siRNA activity score. To further remove plate variations, a non-linear transformation was applied, so that the median activities of positive controls on each plate were scaled to 0.1. The best score for any given gene across multiple siRNAs was kept as the siRNA activity score of the gene (See Table S2, column headings “TLR7” or “TLR9”), and thus the siRNA activity score does not take into account the activities of additional siRNAs targeting the same gene.

To calculate **RSA activity scores**, siRNA wells were ranked based on their normalized activity in each assay, so that the most potent siRNAs were aggregated on the top of the list (Konig et al., 2007). Since our library contained multiple independently arrayed siRNAs for each interrogated gene, we expect gene activities that are not driven by off-target activities to possess multiple siRNAs which to cluster towards the top of the ranked list. Therefore, measuring the degree of bias in the ranked distribution of all siRNAs for a gene provides a quantitative association to whether the gene is likely a true hit (i.e. not an off target effect). Based on this hypothesis, the RSA analysis algorithm calculates the minimum probability that the siRNAs for a given gene have a biased distribution of activities by chance, and then identifies those genes that have multiple siRNAs supporting their activities based on this probability. Specifically, this was done by iteratively applying accumulative hypergeometric probability calculations to multiple siRNA subsets. As a result, each gene was assigned an RSA p-value based on the distribution of its siRNA activities (See Table S2, column headings "TLR7_LogP" and "TLR9_LogP"). Genes with multiple siRNAs clustering towards the top tend to be assigned smaller p-values, and are considered less likely to be active due to off-target activities or experimental noise.

2. Protein Network Data Analyses –TLR7 and TLR9

Binary human protein-protein interaction data used in this study was derived from Reactome (http://reactome.org), BIND (http://www.bind.ca), MINT (http://mint.bio.uniroma2.it/mint/Welcome.do), HPRD (http://www.hprd.org), CORUM (http://mips.helmholtz-muenchen.de/genre/proj/corum/index.html), Hynet (Prolexys, Inc.) (Bader et al., 2003; Ceol et al., 2010; Croft et al., 2011; Peri et al., 2003; Ruepp et al., 2010), and a previously reported functional NF-κB/TNF-α interaction map (Bouwmeester et al., 2004). For each gene list, all direct interactions among the encoded proteins were extracted from the databases described above, and three networks were constructed: a TLR7 network, a TLR9 network, and a TOX network (p-values <0.001). Each network was generated based on those genes that had siRNA activity scores below the thresholds indicated above (see siRNA Screening Data Transformation and Primary Hits) for the TLR7, TLR9 and TOX screens. We hypothesized that candidate innate factors that form highly interconnected and statistically significant networks with each other and canonical TLR pathway members are more likely to be relevant regulators of innate signaling. Conversely, if the candidate innate factors form statistically significant networks with cytotoxic proteins, these factors are less likely to be relevant regulators of innate signaling and more likely to be toxic themselves. Therefore, we prioritized those genes within TLR7 and TLR9 networks, and used the TOX network to exclude 2907 genes that were considered to be toxic (See Table S2, column headings "TLR7_UnfilteredNet_Direct," "TLR9_UnfilteredNet_Direct," and "TLR9_UnfilteredNet_Indirect").

3. Protein Network Data Analyses - InnateDB Protein Interaction Enrichment

We collected 822 genes that could be linked to the TLR7/TLR9 pathway components, based on three different sources: *1)* Approximately 40 genes were extracted from Bouwmeester et. al., 2004; *2)* All the 1-hop binary protein-protein interactions involving canonical known TLR7/9 pathway members were extracted from Oda and Kitano et. al., 2006; *3)* All the 1-hop binary protein-protein interactions involving canonical known TLR7/9 pathway members were extracted from InnateDB (Lynn et al., 2008). After removing redundant genes, the final list contained 822 genes, which we collectively refer to as InnateDB genes in this study. Next, we determined whether any of the genes identified by our own study were previously annotated as InnateDB genes. This data was kept as a first line of InnateDB evidence (See Table S2, column heading "TLRPathwayInnate"). Subsequently, we used several of the protein interaction databases described above (BIND, MINT, Reactome, HPRD, CORUM, and HYNET) to generate a list of total proteins predicted to interact with genes in the TLR7 and TLR9 networks. The overall number of direct interactions between each newly identified TLR7/9 regulatory co-factor and InnateDB genes was kept as a second line of InnateDB evidence (See Table S2, column heading "TLRPathwayInnate_Direct").

4. Human Tissue Gene Expression Analyses

GNF Tissue Atlas consists of 79 human tissue samples measured on an U133A_v2 gene chip in replicates (Su et al., 2004). Genes expected to function in immune surveillance pathways were expected to be present in the following eleven blood-derived samples, including "CD33 positive bone marrow derived myeloid cells", "CD34 positive stem cells", "BDCA4 positive dendritic cells", "CD14 positive monocytes", "CD19 positive B cells", "CD56 positive natural killer cells", "CD71 positive early erythroid cells", "CD105 positive endothelial cells", "CD8 positive T cells", CD4 positive T cells" and "Whole blood". For each gene, an absent/present score was calculated based on expression in the above samples using the Affymetrix MAS5 algorithm (max score of 1). If a gene was present in only one of the two replicates, it was counted as a 0.5 occurrence (See Table S2, column heading "Blood Present").

The tissue expression profiles were also log-transformed, and the Pearson correlation coefficient between genes and that of TLR7 and MyD88 were calculated. Candidate innate regulators with higher correlation scores were weighted favorably (See Table S2, column headings "CorrTLR7_Positive" and "CorrMYD88_Positive"). Expression profiles of both TLR7 and MyD88 showed differential blood-biased expression; therefore, the two correlation coefficients are well-correlated (r^2 = 0.94). TLR9 was not on the array and its expression profile was not available.

5. Gene Expression Analyses in Response to TLR7/TLR8/TLR9 Stimulation

Previously, HEK293/TLR7 and HEK293/TLR8 stable cell lines were profiled as three groups, untreated, stimulated through TLR7, or stimulated through TLR8 (unpublished data). For each treatment, the gene expression by microarray analysis was profiled at 2-hour, 4-hour and 8-hour time points. Fold changes were first calculated within the TLR7 and TLR8 group as Max/Min among the three time points; fold change values were then subtracted by the fold change observed from the untreated group, resulting in a delta fold change (dFC), and the larger dFC value was kept as an evidence (See Table S2, column heading "TLR_DeltaFoldChange"). This data set of nine experiments was also analyzed by an Ontology-based Pattern Identification (OPI) algorithm in order to identify genes that share both similar expression profiles and similar biological functions (See Table S2, column headings "OPI_Support" and "OPI_MicroarraySupport") (Zhou et al., 2005). Additionally, previously generated microarray data from CpG ODN-treated RAW264.7 cells was analyzed (Tross et al., 2009). Gene expression microarray data was available for cells stimulated at 4-hour or 12-hour time points, and dFC was calculated as described above (See Table S2, column heading "FC_CpGODN").

Evidence Compilation and Gene Hitpick Criteria

For each gene, we collected a series of supporting evidences using the criteria described above. All individual evidences were transformed into a normalized score between zero and one (where zero indicates no support and one indicates strong support), compiled into an evidence table, and the scores were then used for candidate hit selection.

To determine which experimental and orthogonal datasets would be most informative towards identification of relevant innate co-factors, we initially evaluated each dataset independently for their enrichment of canonical TLR pathway members. Specifically, the predictive value of a given line of evidence was quantified by calculating statistical enrichment of known TLR pathway members utilizing the previously described RSA-like log p-value calculation; when an evidence was found to have a log p value < 0.2, the evidence was deemed not predictive, and was eliminated from further analyses. Subsequently, we evaluated the pairwise correlation coefficients among evidences; when a line of evidence was highly correlated with another (coefficient > 0.8), the less predictive evidence was eliminated to reduce redundancy in the evidence pool. For instance, the siRNA activity scores for the TLR9 and TLR7 screens were highly correlated, so only the TLR9 siRNA activity score was kept as an informative line of evidence. Based on these criteria, there were eight remaining highly informative evidence datasets: direct protein-protein interactions derived from TLR9 screen data, indirect protein-protein interaction derived from TLR9 screen data, gene expression in myeloid or lymphoid cells or tissues, tissue expression correlation with TLR7, gene expression fold change analysis in

response to TLR7/8 stimulation, two quantitative measures of the microarray OPI clusters, and siRNA activity score in the TLR9 screen (See Table S2, column headings "TLR9_UnfilteredNet_Direct," "TLR9_UnfilteredNet_Direct," "BloodPresent," "CorrTLR7_Positive," "TLR_DeltaFoldChange," "OPI_Support," "OPI_MicroarraySupport," and "TLR9," respectively).

Next, a genetic algorithm was applied to assign weights to these informative lines of evidence (Goldberg, 1989). Briefly, twenty independent search processes, each with 1000 iterations, were initiated by initially assigning random weights to each line of evidence. Each iteration aimed to optimize the weights by improving upon the best weight obtained in the previous iteration, so that the calculated linear combination scores could produce ranked gene lists where more known (canonical) pathway members were included in the top 250 positions, out of more than 1500 positions. Genes that were ranked consistently highly (ie. within the top 250 positions) by at least five out of the twenty search processes tended to be the ones reliably supported by multiple informative evidences, though we found that many top genes were actually supported by twenty out of twenty search processes. All genes that were ranked highly were considered favorable candidates for further prioritization based on their siRNA activity scores, RSA p-values, or number of protein-protein interactions with TLR pathway members. The above described automated hit selection process was particularly valuable because it relies on multiple orthogonal lines of evidence to select genes, and eliminates user bias from the selection process.

Complementary to this automated hit picking approach, additional genes were selected for confirmation based solely on their siRNA or their RSA activity scores, in either the TLR7 screen, the TLR9 screen, or both. By using both the automated and activity-based ranking criteria, some genes were selected in more than one list. By removing redundant genes, we obtained a final list of 546 primary hits for gene confirmation studies. Ultimately, 38% of the 546 primary hits were selected based on the automated hit pick strategy, and 62% were selected based on siRNA or RSA activity score ranking.

Gene Confirmation

We selected 546 genes based on the above-described evidence optimization results, and additional genes for the purpose of completeness and comparison. These genes were further profiled in confirmation studies, where the individual siRNAs from the original wells, plus additional siRNAs for each gene, were arrayed and examined by monitoring NF-κB luciferase reporter activity in response to R848 and CpG. 190 genes containing at least two independent siRNAs showing an siRNA activity score of 0.4 and below were considered confirmed. The activity score was based on a non-linear scale where the positive controls were set to 0.1 and the median of the plate was set to 1.0 (See also siRNA activity scores and RSA activity scores). These genes were also determined to be non-toxic, using the previously described criteria (See siRNA Screening Data Transformation and Primary Hits).

Protein-Protein Interaction Networks, MCODE analysis, and Functional Enrichment Analyses

To generate the protein-protein interaction networks depicted in Figures S1F and S1G (Figure 1C), we initially compiled a list of 190 confirmed TLR7/9 co-factors and 20 known canonical members of the TLR7/9 pathway. These 210 genes were then used to construct a network (Figure S1F) as described previously (Konig et al., 2008). Next, additional primary hits were added to the original list of 210 genes in order to expand the network. The additional primary hits were selected based on the following criteria: genes that are in TLR7 and TLR9 Network, but not in the TOX network; genes that have TLR7 and TLR9 siRNA activity scores < 0.5; genes that have hematopoeitic tissue gene expression absent/present scores > 50% and have Pearson Correlation coefficients > 0.5 for either TLR7 or MYD88; and genes that directly interact with at least one known canonical member of the TLR7/9 pathway as well as one confirmed hit. By applying these criteria, we added another 62 genes (gray nodes) to the original network (Figure S1F) to obtained the expanded network (Figure S1G, also Figure 1C).

If a network was too complex for visual interpretation, Molecular Complex Detection (MCODE) analysis was applied to identify densely connected network components (Bader and Hogue, 2003). All network visualization was based on Cytoscape (version 2.8.0) (Shannon et al., 2003). Additionally, significance of these networks (p-value <0.001) was determined by comparing the complexity of iterative randomized (1000X) networks of the same size.

Functional enrichment analyses were applied to provide biological context for confirmed genes or other genes that were original screening hits, protein networks, or MCODE subnetworks. For these analyses, gene ontology data were collected from multiple sources, including Gene Ontology, GeneGo process, CORUM, and in-house accumulated manual curations. Statistical significance of each functional category was scored using the standard accumulative hypergeometric probability function.

Ligand-profiling of Confirmed Hits

The 190 confirmed hits from the RNAi analysis were cross-profiled in TLR7, TLR9, TNFR, TLR5, and IL-1R assays. For the TLR7 assay, siRNAs were introduced into the HEK293T/TLR7/NF-κB luciferase reporter line and stimulated with 0.5uM of R848. For the TLR9, TNFR, TLR5, and IL-1R assays, siRNAs were introduced into HEK293T/TLR9/NF-κB luciferase reporter line and stimulated with 3uM of CpG oligonucleotides, 10ng/ml of TNF-α, 100ng/ml of flagellin, and 10ng/ml of IL-1β, respectively. For each assay, activity data were normalized and scaled using the same process as described before. Slightly different activity thresholds were determined for each assay to capture the variations in signal distributions; we chose 0.6 (TLR7), 0.6 (TLR9), 0.45 (TNFR), 0.5 (TLR5), and 0.65 (IL-1R), respectively, to be the thresholds. Genes with activity scores below the individual thresholds were considered active, and genes

with activity scores above the thresholds were considered inactive. The activity profiles were then hierarchically clustered and mechanisms of hits were assigned based on the heat map visualization and on the known downstream pathway circuitry for each receptor interrogated. siRNAs that were active only in both TLR7 and TLR9 and inactive in TLR5, IL1R, and TNFR were categorized as "TLR7/9 Specific"; siRNAs that were active in TLR7, TLR9, TLR5 and IL-1R but inactive in TNFR but were categorized as "MyD88-dependent Signaling"; siRNAs that were active in TLR7, TLR9, TLR5, IL-1R, TNFR, were categorized as "General NF-κB Activity". Using this approach, 80 hits were assigned mechanisms. For 110 genes, we were unable to assign a categorization, as these specific genes did not display activities that were discernable between ligands and/or not consistent with currently known pathway circuitries.

Support-Vector Machine (SVM) for Functional Activity Prediction

Out of 190 confirmed hits, 85 confirmed co-factors and 20 known pathway members formed a statistically significant protein network (See Figure S1F). Among these 85 co-factors, our ligand cross-profiling studies (see above) assigned 18 proteins as "TLR7/9 Specific" genes, 5 proteins as "MyD88-dependent Signaling" genes, and 19 proteins as "General NF-κB Activity" genes. Forty-three out of these 85 confirmed genes were not assigned a mechanism; nevertheless, they are still connected to known TLR pathway members within the network (Oda and Kitano, 2006); therefore, the interaction of these genes within the network was further assessed. Specifically, these genes were placed within a matrix, where each row represents a confirmed gene, each column represents a pathway member, and each cell represents the closeness of interactions between the two proteins. The closeness of the confirmed gene and the pathway member can be calculated by using the formula: closeness score = 0.5^{n-1}, where *n* equals to the number of connections between the confirmed gene and the known pathway member. For example, if the confirmed gene and the pathway member directly interact (one connection) the closeness score is 1.0. When we compared to genes assigned "General NF-κB Activity", genes assigned as "TLR7/9 Specific" are statistically significantly closer to upstream pathway components UBE2N (p-value = 0.09) and also to MAP3K7IP2 (TAB2) (p-value = 0.05), indicating that we could reliably use the network architecture to predict possible molecular intersections between identified innate signaling regulators and known components of the TLR pathway. The complete closeness matrix was then used as the featured matrix for two-class support-vector machine (SVM) training. A three-fold cross validation test indicated a prediction accuracy of 71% for "TLR7/9 Specific" genes and 65% for "General NF-κB Activity" genes using SVM. We then applied the trained classifier to assign predictive activity labels to the 43 confirmed genes that were not previously assigned. Using this approach, 17 of the 43 genes were classified as "TLR7/9 Specific", 21 of the 43 genes were classified as "General NF-κB Activity", and 5 remaining genes were unable to be classified. The predictions were then incorporated into the cellular map of TLR7/9 signaling (Figure 2) based on literature curation and also on predicted protein

interactions derived from the networks in Figures S1F and S1G (see also Table S6).

Luciferase and Cytotoxicity Assays

All luciferase reporter activities were quantified with Britelite Plus (PerkinElmer). For the cytotoxic assay, HEK293T/TLR9/NF-κB luciferase reporter cells were transfected with siRNA. Three days post transfection, viability of the cells was quantified with ATP Lite (PerkinElmer). Both the luciferase assay and cytotoxic assay were quantified by using the PHERAstar luminometer (BMG Labtech).

Evaluation of TLR Mutant Signaling in HEK293T

HEK293T cells were transfected with WT-TLR9-HA or mutant TLR2, TLR7, or TLR9 MSCV2.2 retroviral vectors, as well as an NF-κB luciferase reporter plasmid where luciferase expression was driven by an endothelial leukocyte adhesion molecule 1 (ELAM-1) promoter (Whelan et al., 1991). Twenty-four hours after transfection, cells were stimulated for 14h with TLR or reference ligands, as described. Cell lysates were obtained using passive lysis buffer (Promega) and assayed for luciferase activity.

RNA interference

Double stranded RNA duplexes were purchased from Qiagen. The targeting sequences were as follows: MyD88, 5'-CATCCTGAGTTTATAATAATA-3'; p65, 5'-CGGATTGAGGAGAAACGTAAA-3'; UNC93B1, CTGCCCGACATCGACAGCAAA; RIG-I, AACGTTTACAACCAGAATTTA; GL-3 luciferase, 5'-AACTTACGCTGAGTACTTCGA-3'; PDK2-1, TAGGTCTGTGATGGTCCCTAA; PDK2-2, CACAGTAAAGAGGAGACTGAA; PDK2-3, CCACGTACCGCGTCAGCTA; PDK2-4, CAACGTCTCTGAGGTGGTC; HRS-1, CACGTCCGGAGTAACACTACA; HRS-2, GCACGTCTTTCCAGAATTCAA; HRS-3, CCGGAACGAGCCCAAGTACAA; HRS-4, TACGAGCAGCTGAACAGGAAA; FBXL7-1, ATGGGCGCGAACAATGGCAAA; FBXL7-2, AAGGTCTCACTAGGAAATTTA; FBXL7-3, UCAGGAUGCUCCAAAGUGA; FBXL7-4, GTCTCAGGCTGTTACAATA. Four siRNAs directed against TLR7 were pooled: TLR7-1, ATGGTATGCCTCCAAATCTAA; TLR7-2, CTGGAGGTATTCCCACGAACA; TLR7-3, TAACCTCTCGCCATTACATAA; TLR7-4, CAGACCTTGGATCTAAGTAAA. siRNA sequences for additional TLR7/9 co-factors can be found in Table S7. Negative control siRNAs were previously described (Konig et al., 2008). HEK293T and THP-1 cells were transfected via Lipofectamine RNAiMAX (Invitrogen) and Hiperfect (Qiagen), respectively, according to the manufacturers' protocols. 72 hours post-transfection, the cells were used for further experiments.

Real-Time PCR

Total RNA from cells was extracted by using RNeasy Mini Kit according to the manufacturer's instructions (Qiagen). RNA samples were reverse transcribed using the QuantiTect Reverse transcription Kit (Qiagen). PCR products were detected using the Power SYBR® Green PCR Master Mix (Applied Biosystems). Formation of a unique DNA product was confirmed by verifying that products had a single melting temperature. Fluorescence monitored PCR values were normalized to TATBP values to account for any differences in cDNA recovery between samples. RT-PCR was performed with the following primers amplifying human genes: TATBP, 5'-CCACTCACAGACTCTCACAAC-3' and 5'-CTGCGGTACAATCCCAGAACT-3'; ICAM-1, 5'-TGGCCCTCCATAGACATGTGT-3' and 5'-TGGCATCCGTCAGGAAGTG-3'; HRS, 5'- TTCGAGCGTCTCCTAGACAAG-3' and 5'-GCAAAATGGACTCCCAATCTGT-3'; IL-8, 5'-TTTTGCCAAGGAGTGCTAAAGA-3' and 5'- AACCCTCTGCACCCAGTTTTC-3'; RIG-I, 5′-AAAGCCTTGGCATGTTACAC-3′ and 5′-GGCTTGGGATGTGGTCTACT-3′; MyD88, 5′-CGACACTGCTGGGGCAGCTT-3′ and 5′-TGCTTCCCCACCTGCCCTGT-3′; IFN-α 5′-CTGAATGACTTGGAAGCCTG-3′ and 5′- ATTTCTGCTCTGACAACCTC-3′; TNF-α, 5′-AGTGAAGTGCTGGCAACCAC-3′ and 5′-GAGGAAGGCCTAAGGTCCAC-3′; rps11 5′-GCCGAGACTATCTGCACTAC-3′ and 5′-ATGTCCAGCCTCAGAACTTC-3′; UNC93B1, 5'-TGCTCACCTACGGCGTCTA-3' and 5'-GATGGGAGTCACGTTGATGC-3'; ZNF700, 5'-CAGGAAGAGTGGACATTGCTG-3' and 5'-AATGGATGGGCGATACCTGAA-3'; UQCRC1, 5'-GGGGCACAAGTGCTATTGC-3' and 5'-GTTGTCCAGCAGGCTAACC-3'; UNC13D, 5'-TCTACGAGGACGCACTCTACA-3' and 5'-GGCCTCCGTCACATGGTTG-3'; HIST1H2BO, 5'-GACCCGGCTAAATCTGCTCC-3' and 5'-GGCCTTGGTTACGGCTTTC-3'; CTSL1, 5'-CTTTTGCCTGGGAATTGCCTC-3' and 5'-CATCGCCTTCCACTTGGTC-3'; PDK2, 5'-CGCCCAAGTACATAGAGCACT-3' and 5'-GGAAGCAGGTTGATCTCTTTCA-3'; FBXL7, 5'-GATCACACGCCCACTAAAGC-3' and 5'-CATGCTCAGGTCGGAGTCTTC-3'; TLR7, 5'-AGAACCATGTGATCGTGGACT-3' and 5'-TGTTCGTGGGAATACCTCCAG-3'.

High Content Imaging of p65 Nuclear Translocation

The cells labeled for p65 nuclear translocation were imaged using the Opera (PerkinElmer), an automated confocal microscope, using 488nm and 635nm laser lines to excite Alexa488 immunolabeled p65 and Draq5, respectively. A total of 18 fields of view were collected from each well (covering on average ~7,200 cells/well) at a resolution of 0.66µm/pixel using an LCPlanFL 20x/0.40na Objective (Olympus). Images were analyzed using a custom Acapella (PerkinElmer) script available on Google.Code at:

http://code.google.com/p/operahci/source/browse/#svn%2FAcapella%2FAcapellaScripts%2Ftrunk%2FGNF.Scripts%2FNuclearTranslocation.
In brief, the script was parameterized to detect nuclei and cytoplasm using the Draq5 stain. A donut limited by the cell cytoplasm was then drawn. For each cell donut a Pearson correlation coefficient was calculated between labeled p65 and labeled Nuclei (Draq5). Cells with a Pearson correlation >0.2 were classified as translocated and well signal defined as the median value of their Pearson correlation coefficient was used to establish whether the siRNA affected p65 nuclear translocation or not. p65 nuclear translocation was considered impaired if the well signal is 4 standard deviations lower than the average of the negative controls. For some genes where the state of p65 nuclear translocation was ambiguous, images from replicate wells were evaluated visually.

Intracellular Staining of TNF-α

2 x 10^5 immortalized macrophages were plated on 24-well plates. The following day, cells were stimulated; 30 minutes after stimulation cells were treated with brefeldin A. Six hours after stimulation cells were fixed using the BD Cytofix/cytoperm kit according to the manufacturer's instructions. Cells were stained with αTNF-PE clone MP6-XT22 (eBiosciences).

Phagosome Isolation

Phagosome isolation was carried out as described previously (Ewald et al., 2008).

TLR9 Immunofluorescence

RAW264.7 macrophages were transduced using the retroviral vector MigR2. Transduced cell lines were grown overnight on glass coverslips, then fixed with 4% paraformadehyde, permeablized with 0.5% saponin, and treated with 0.1% sodium borohydride/0.1% saponin. HA and LAMP1 were detected using Y11 (Santa Cruz Biotechnology) and 1D4B (BD Biosciences). Donkey anti-Rabbit AlexaFluor 488 and Donkey anti-Rat Cy3 conjugated secondary antibodies were purchased from Jackson ImmunoResearch. Stained slides were analyzed by epifluorescence microscopy. ImageJ was used for colocalization analysis.

Macrophage intracellular killing assay

TLR9$^{-/-}$ immortalized macrophages stably expressing either WT-TLR9-HA or KallR-TLR9-HA were seeded at 5 x 10^5 cells per well in 24 well plates the day prior to the assay. GAS serotype M49 strain NZ131 (Simon and Ferretti, 1991) were added at multiplicity of infection (MOI) of 1 bacteria per macrophage in 350 ml of assay media (RPMI 2% FBS), and plates were centrifuged at 2000 rpm to ensure bacterial contact with the macrophages. Plates were incubated for 30 min at 37ºC in 5% CO_2, then 100 µg/ml penicillin G and 100 µg/ml gentamicin were

added to kill extracellular bacteria, and the plates incubated for an additional 30 min. Each well was then rinsed with PBS to remove extracellular killed bacteria, and cells in a subset of wells were lysed with 0.025% Triton-X100 in water and serial dilutions plated on agar for enumeration of surviving bacterial colony forming units (cfu) at time 0 h. For the other wells, 500µl of fresh media containing 20 µg/ml gentamicin was added for an additional 3 hours, at which time macrophages were lysed, dilutions plated, and intracellular cfu quantified as above. The % intracellular survival was calculated as GAS cfu at 3 h / GAS cfu at 0 h x 100%.

Newcastle Disease Virus Infection of Human Monocyte Derived Dendritic Cells

Peripheral blood mononuclear cells (PBMC) were isolated by Ficoll density gradient centrifugation (Histopaque; Sigma Aldrich) from buffy coats of healthy human donors (Mount Sinai Blood Donor Center and New York Blood Center). $CD14^+$ cells were purified using anti-human CD14 antibody-labeled magnetic beads and iron-based MiniMACS liquid separation columns (Miltenyi Biotech). After elution, $CD14^+$ cells were incubated at 37°C for 5 days at a concentration of 10^6 cells/ml in RPMI medium containing 10% fetal bovine serum (FBS) (HyClone; Thermo Scientific), 2mM L-glutamine, 1mM sodium pyruvate, and 100U/ml penicillin–100µg/ml streptomycin (Gibco, Invitrogen) (complete DC medium) and supplemented with 500U/ml human granulocyte-macrophage colony-stimulating factor (hGM-CSF) and 1,000U/ml human interleukin 4 (hIL-4) (Peprotech), generating a pool of monocyte-derived dendritic cells (MDDCs).
Subsequently, 25,000 MDDCs were seeded per well in 96 well plates and transfected with siRNAs using the StemFect RNA transfection kit (Stemgent), according to the manufacturer instructions. Forty-eight hours after transfection, cells were infected with New Castle Disease Virus (NDV, B1 strain) (MOI of 2) as previously described (Fernandez-Sesma et al., 2006), with some modifications. NDV was grown in 9-day-old embryonated chicken eggs (SPAFAS; Charles River Laboratories) and titrated by immunostaining using the anti-HN antibody 7B1 (Mount Sinai Hybridoma Shared Research Facility) on Vero cells by following standard procedures. Briefly, for infection, cells were centrifuged (400 × *g,* 10 min), the media was removed and 25ul of RPMI containing the appropriate amount of virus was added and the plates were incubated for 45 min at 37°C. Then, 75ul of RPMI with 10% FBS was added and cells were incubated at 37°C for 8h. Cells were recovered by centrifugation for 10 min at 400 × *g*, and the cell pellets were lysed for RNA isolation. Finally, four replicates of each condition of transfected and infected MDDCs were pooled for RNA isolation and RT-PCR analysis.

Supplemental References

Bader, G.D., Betel, D., and Hogue, C.W. (2003). BIND: the Biomolecular Interaction Network Database. Nucleic acids research *31*, 248-250.

Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S.*, et al.* (2004). A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol *6*, 97-105

Ceol, A., Chatr Aryamontri, A., Licata, L., Peluso, D., Briganti, L., Perfetto, L., Castagnoli, L., and Cesareni, G. (2010). MINT, the molecular interaction database: 2009 update. Nucleic acids research 38, D532-539.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011). Reactome: a database of reactions, pathways andological processes. Nucleic acids research 39, D691-697.

Fernandez-Sesma, A., Marukian, S., Ebersole, B.J., Kaminski, D., Park, M.S., Yuen, T., Sealfon, S.C., Garcia-Sastre, A., and Moran, T.M. (2006). Influenza virus evades innate and adaptive immunity via the NS1 protein. J Virol *80*, 6295-6304.

Goldberg, D.E. (1989). Genetic algorithms in search, optimization, and machine learning (Reading, Mass., Addison-Wesley Pub. Co.)

Konig, R., Zhou, Y., Elleder, D., Diamond, T.L., Bonamy, G.M., Irelan, J.T., Chiang, C.Y., Tu, B.P., De Jesus, P.D., Lilley, C.E.*, et al.* (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. Cell *135*, 49-60.

Peri, S., Navarro, J.D., Amanchy, R., Kristiansen, T.Z., Jonnalagadda, C.K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T.K., Gronborg, M.*, et al.* (2003). Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome research *13*, 2363-2371.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes--2009. Nucleic acids research *38*, D497-501.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

Tross, D., Petrenko, L., Klaschik, S., Zhu, Q., and Klinman, D.M. (2009). Global changes in gene expression and synergistic interactions induced by TLR9 and TLR3. Mol Immunol *46*, 2557-2564.

A

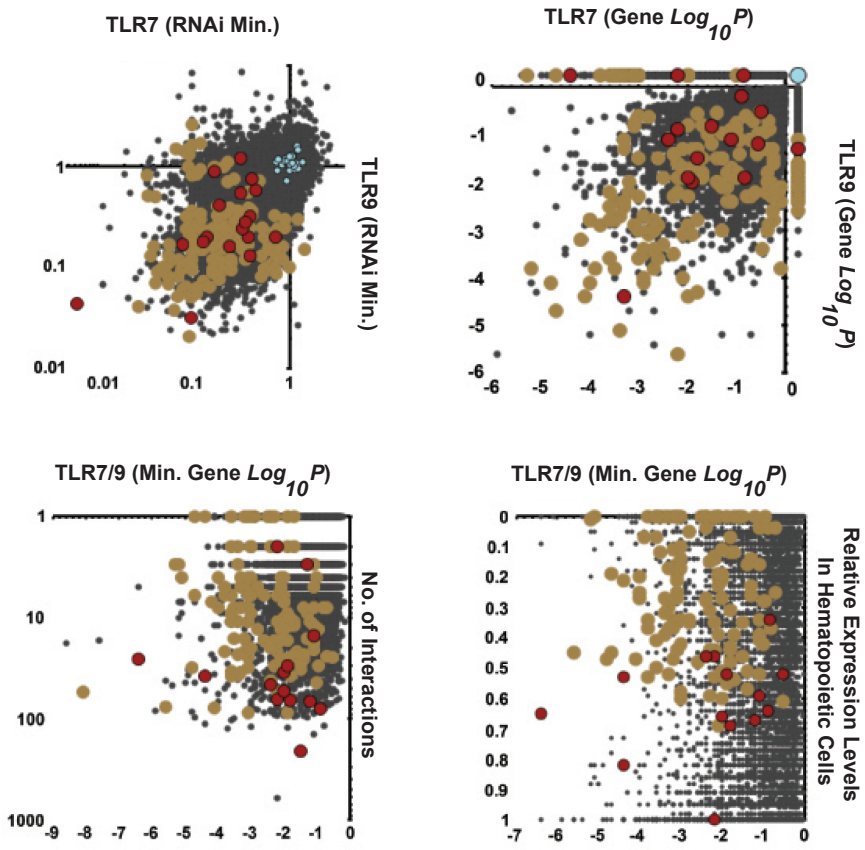

TLR7 (RNAi Min.)
TLR9 (RNAi Min.)
1
0.1
0.01
0.01
0.1
1
TLR7 (Gene Log10P)
TLR9 (Gene Log10P)
0
-1
-2
-3
-4
-5
-6
-6
-5
-4
-3
-2
-1
0
TLR7/9 (Min. Gene Log10P)
No. of Interactions
1
10
100
1000
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
TLR7/9 (Min. Gene Log10P)
Relative Expression Levels In Hematopoietic Cells
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
-7
-6
-5
-4
-3
-2
-1
0

Figure S1A

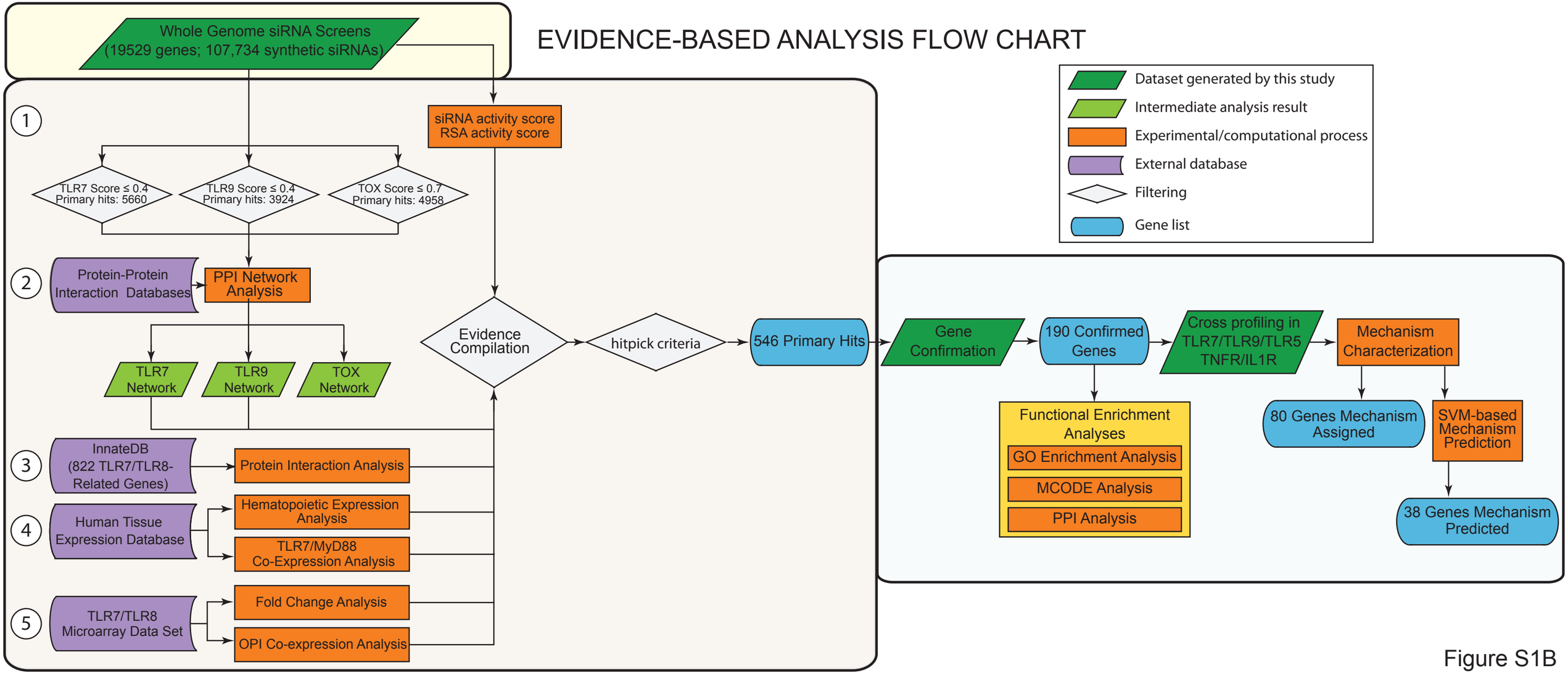

EVIDENCE-BASED ANALYSIS FLOW CHART
Whole Genome siRNA Screens
(19529 genes; 107,734 synthetic siRNAs)
siRNA activity score
RSA activity score
1
TLR7 Score ≤ 0.4
Primary hits: 5660
TLR9 Score ≤ 0.4
Primary hits: 3924
TOX Score ≤ 0.7
Primary hits: 4958
2
Protein-Protein
Interaction Databases
PPI Network
Analysis
TLR7
Network
TLR9
Network
TOX
Network
Evidence
Compilation
hitpick criteria
546 Primary Hits
3
InnateDB
(822 TLR7/TLR8-
Related Genes)
Protein Interaction Analysis
4
Human Tissue
Expression Database
Hematopoietic Expression
Analysis
TLR7/MyD88
Co-Expression Analysis
5
TLR7/TLR8
Microarray Data Set
Fold Change Analysis
OPI Co-expression Analysis
Dataset generated by this study
Intermediate analysis result
Experimental/computational process
External database
Filtering
Gene list
Gene
Confirmation
190 Confirmed
Genes
Cross profiling in
TLR7/TLR9/TLR5
TNFR/IL1R
Mechanism
Characterization
Functional Enrichment
Analyses
GO Enrichment Analysis
MCODE Analysis
PPI Analysis
80 Genes Mechanism
Assigned
SVM-based
Mechanism
Prediction
38 Genes Mechanism
Predicted

Figure S1B

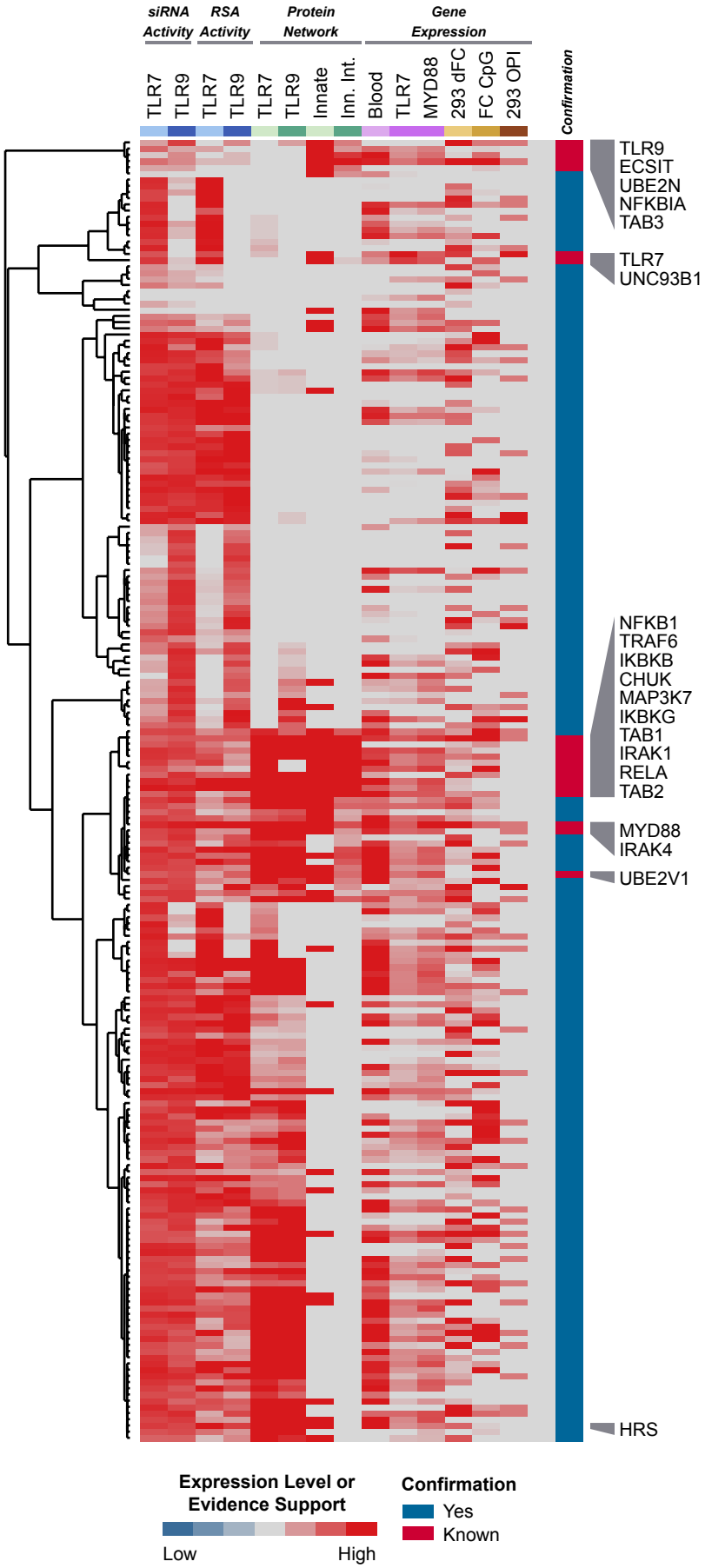

siRNA Activity
RSA Activity
Protein Network
Gene Expression
TLR7
TLR9
TLR7
TLR9
TLR7
TLR9
Innate
Inn. Int.
Blood
TLR7
MYD88
293 dFC
FC CpG
293 OPI
Confirmation
TLR9
ECSIT
UBE2N
NFKBIA
TAB3
TLR7
UNC93B1
NFKB1
TRAF6
IKBKB
CHUK
MAP3K7
IKBKG
TAB1
IRAK1
RELA
TAB2
MYD88
IRAK4
UBE2V1
HRS
Expression Level or Evidence Support
Low
High
Confirmation
Yes
Known

Figure S1C

D

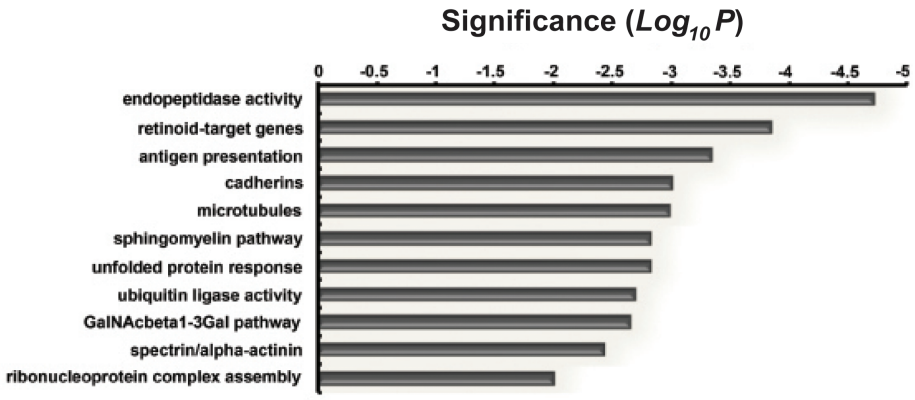

Significance ($Log_{10}P$)
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
-5
endopeptidase activity
retinoid-target genes
antigen presentation
cadherins
microtubules
sphingomyelin pathway
unfolded protein response
ubiquitin ligase activity
GalNAcbeta1-3Gal pathway
spectrin/alpha-actinin
ribonucleoprotein complex assembly

Figure S1D

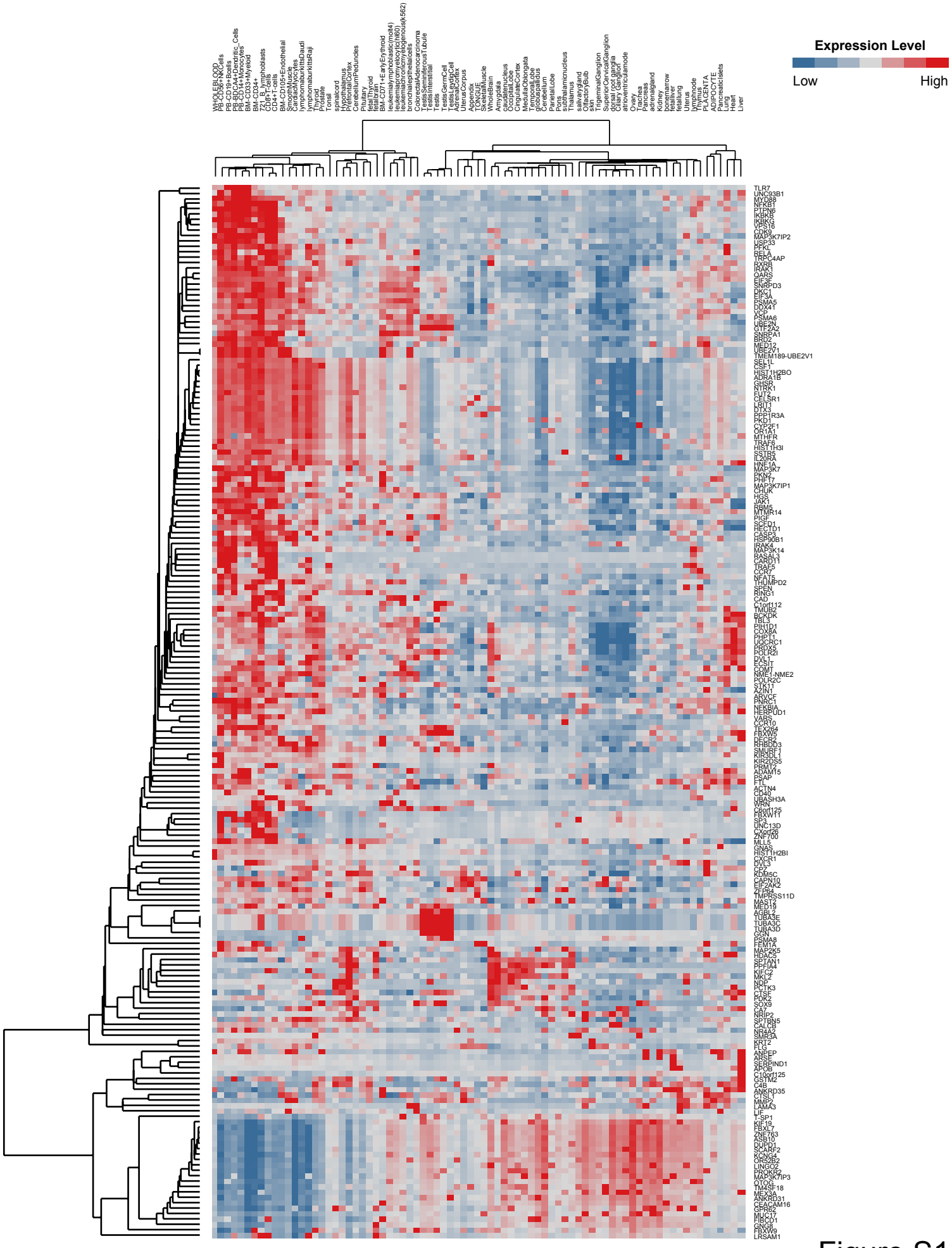
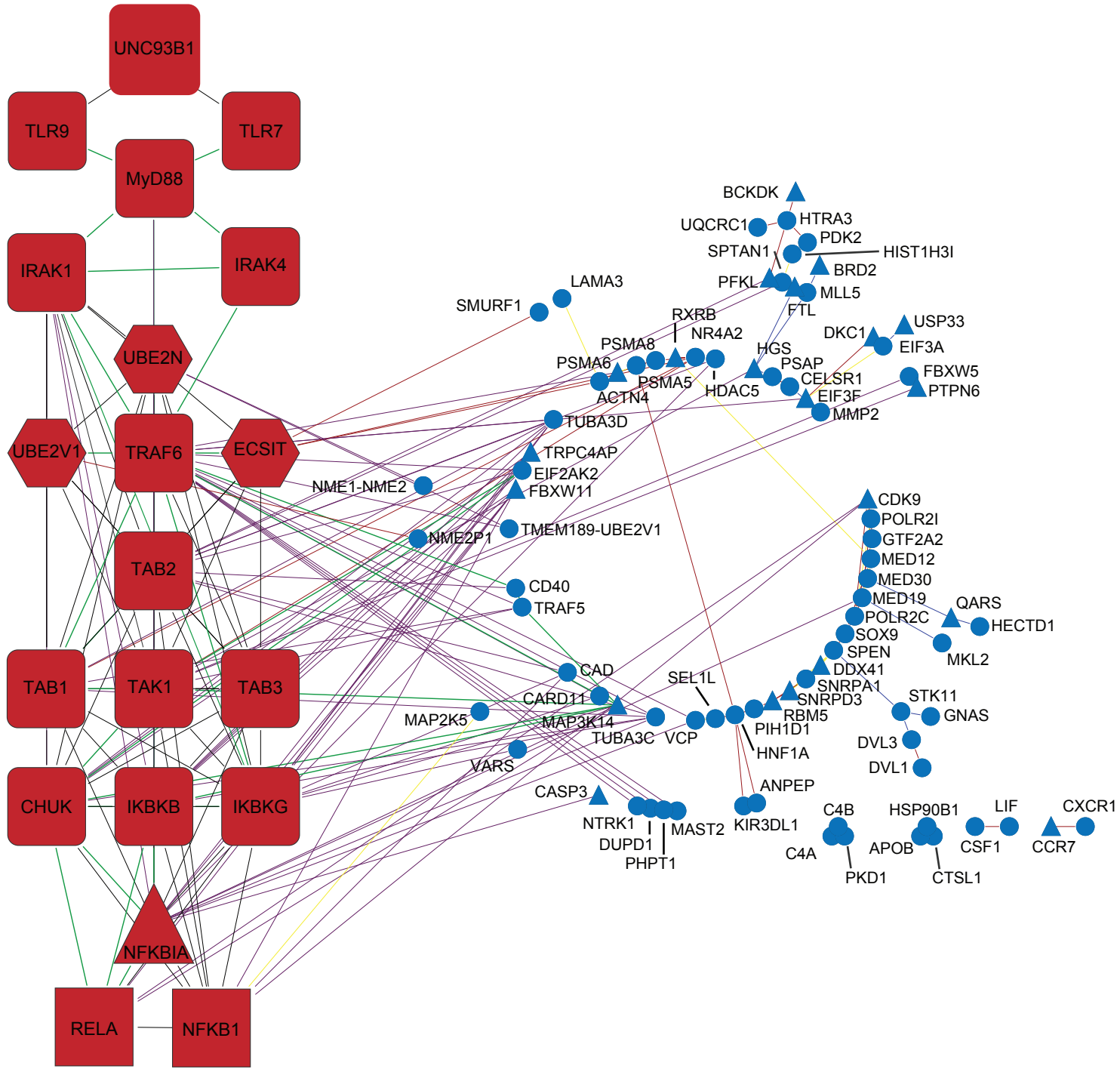


Figure S1E



p-value<0.001

Interaction Type (By edge color)

- BHMRRS
- CORUM
- Canonical
- Hynet
- InnateDB

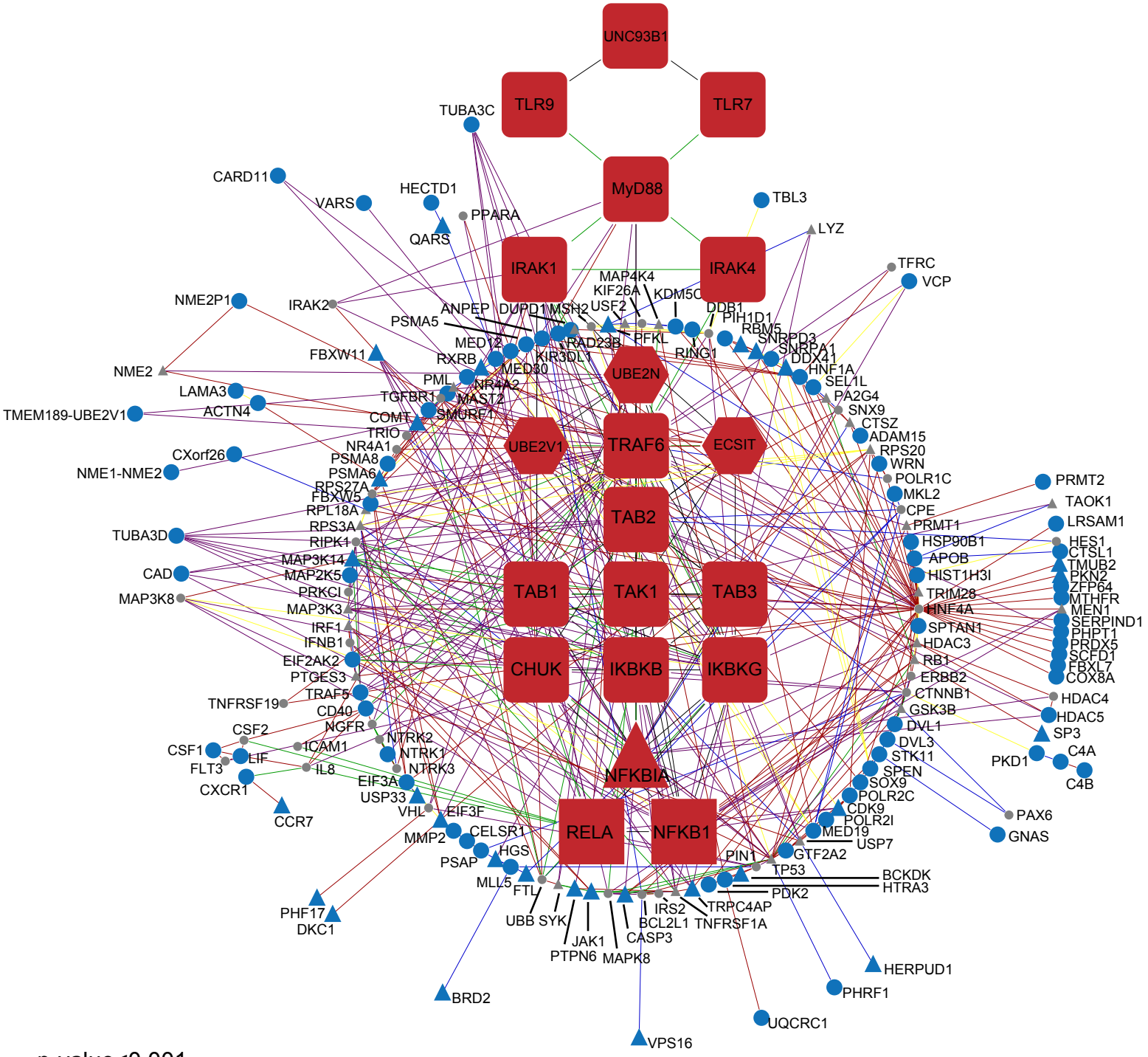
Protein Type (By node color)

- Known pathway member
- Confirmed hit

Expression pattern (By node shape)

- ◯ Poor correlation with lymphoid tissues and cell types
- △ Good correlation with lymphoid tissues and cell types

Figure S1F



p-value<0.001

Interaction Type (By edge color)

- BHMRRS
- CORUM
- Canonical
- Hynet
- InnateDB

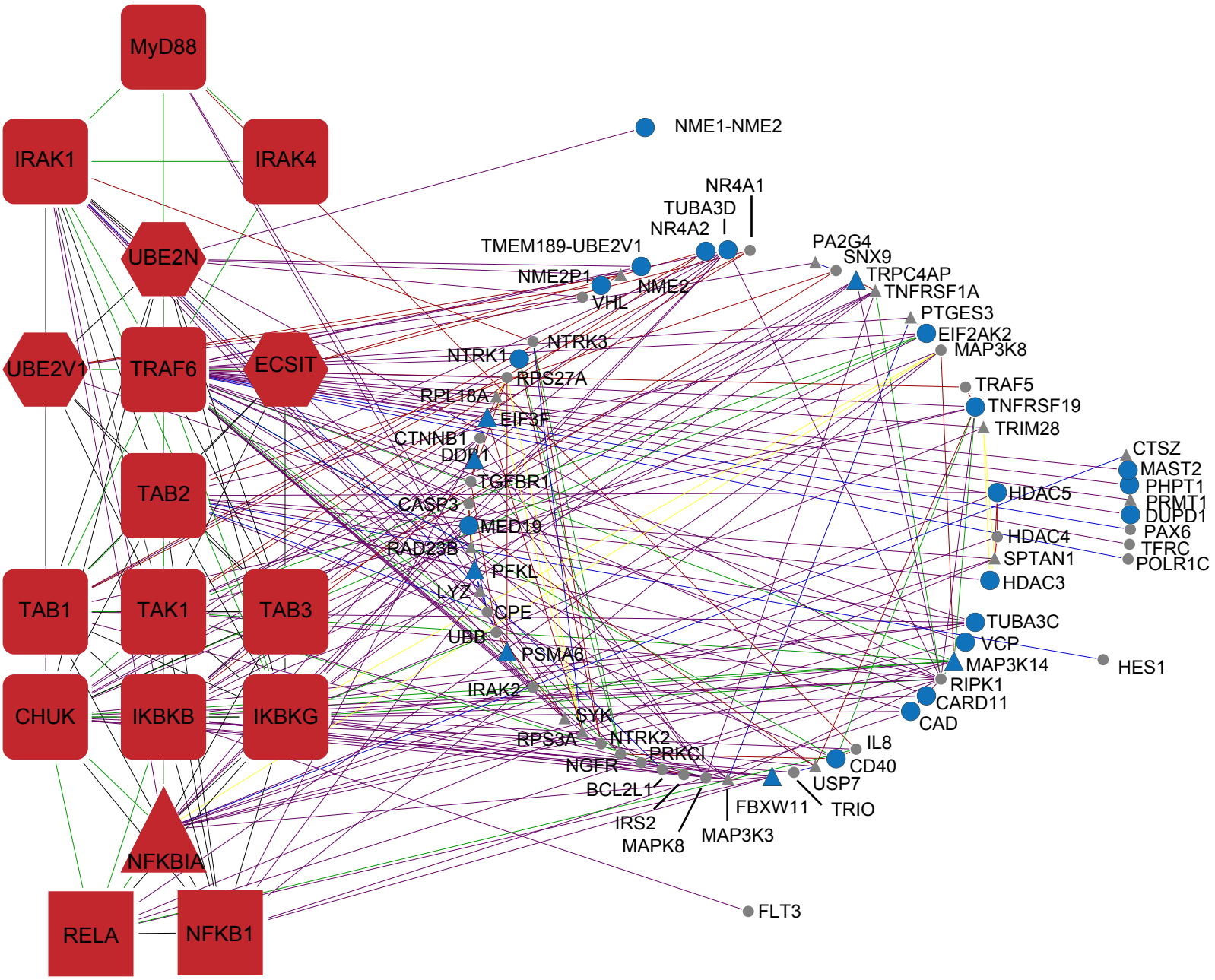
Protein Type (By node color)

- Known pathway member
- Primary hit
- Confirmed hit

Expression Pattern (By node shape)

- ○ Poor correlation with lymphoid tissues and cell types
- △ Good correlation with lymphoid tissues and cell types

Figure S1G



p-value<0.001

Interaction Type (By edge color)

- BHMRRS
- CORUM
- Canonical
- Hynet
- InnateDB

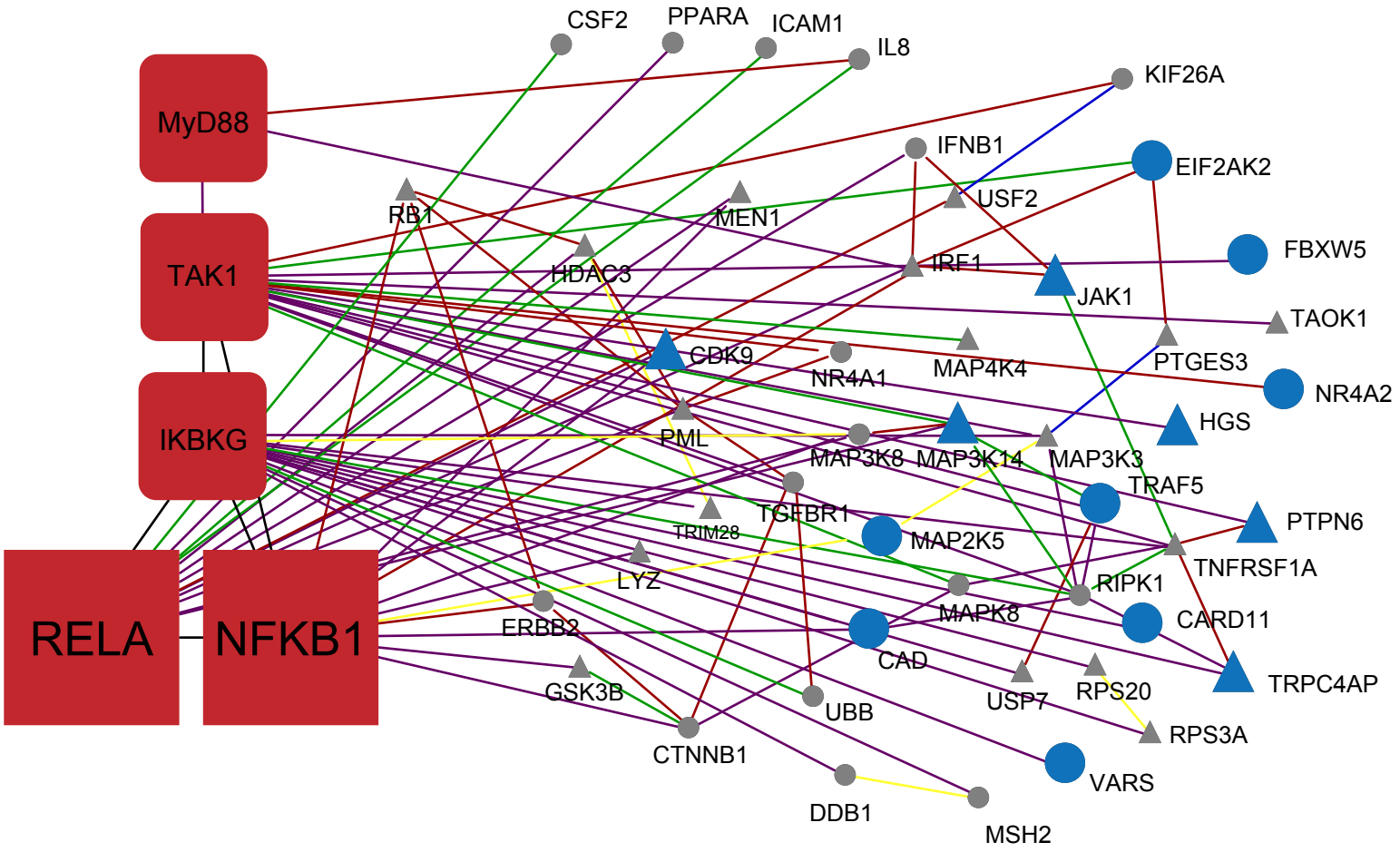
Protein Type (By node color)

- Known pathway member
- Primary hit
- Confirmed hit

Expression pattern (By node shape)

- ○ Poor correlation with lymphoid tissues and cell types
- △ Good correlation with lymphoid tissues and cell types

Figure S1H



p-value<0.001

Interaction Type (By edge color)

- BHMRRS
- CORUM
- Canonical
- Hynet
- InnateDB

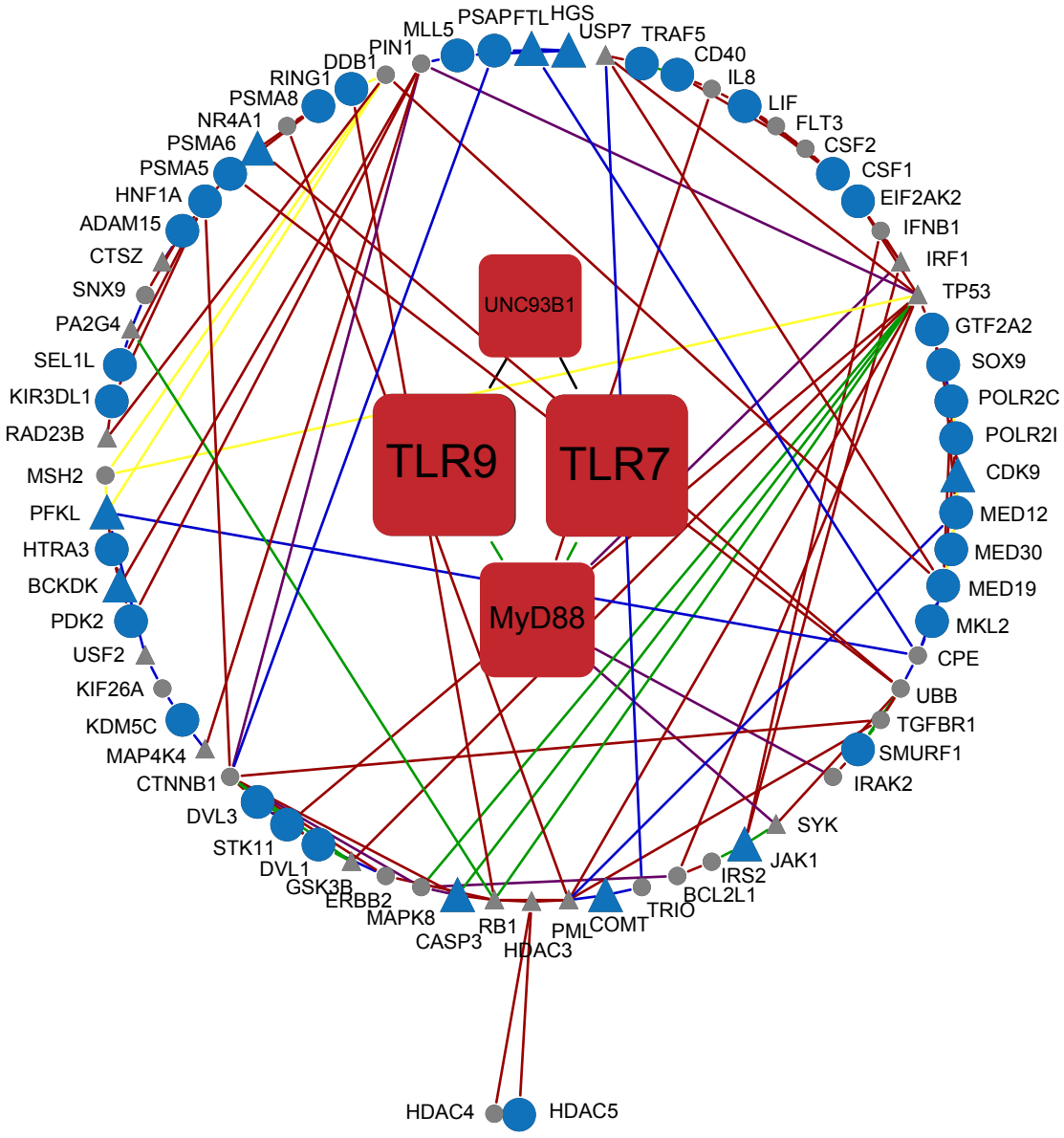
Protein Type (By node color)

- Known pathway member
- Primary hit
- Confirmed hit

Expression pattern (By node shape)

- ○ Poor correlation with lymphoid tissues and cell types
- △ Good correlation with lymphoid tissues and cell types

Figure S1I



p-value<0.001

Interaction Type (By edge color)

- BHMRRS
- CORUM
- Canonical
- Hynet
- InnateDB

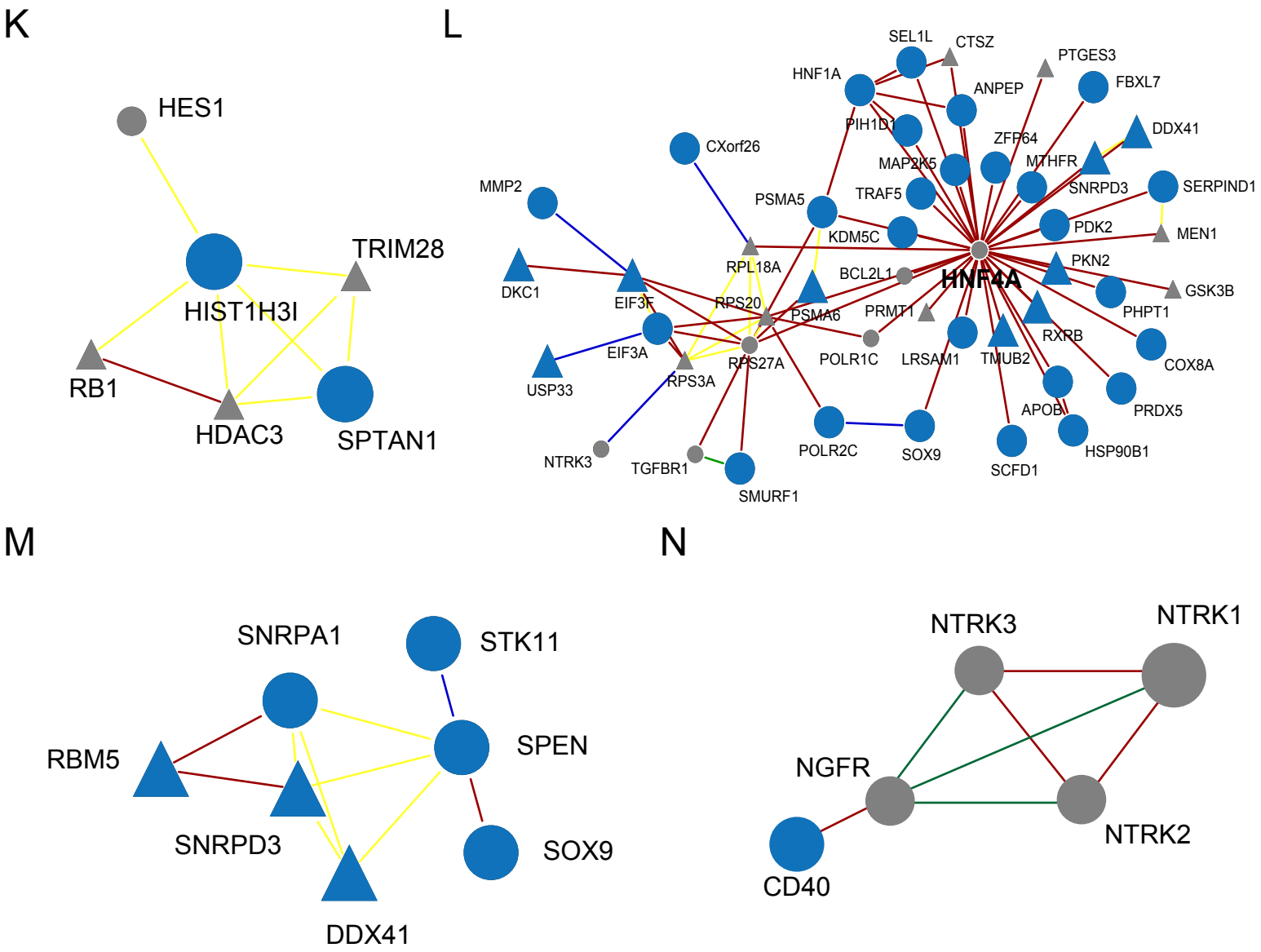
Protein Type (By node color)

- Known pathway member
- Primary hit
- Confirmed hit

Expression pattern (By node shape)

- ○ Poor correlation with lymphoid tissues and cell types
- △ Good correlation with lymphoid tissues and cell types

Figure S1J



p-value<0.001

Interaction Type (By edge color)

- BHMRRS
- CORUM
- Canonical
- Hynet

Protein Type (By node color)

- Primary hit
- Confirmed hit

Expression pattern (By node shape)

- ○ Poor correlation with lymphoid tissues and cell types
- △ Good correlation with lymphoid tissues and cell types

Figure S1K-N

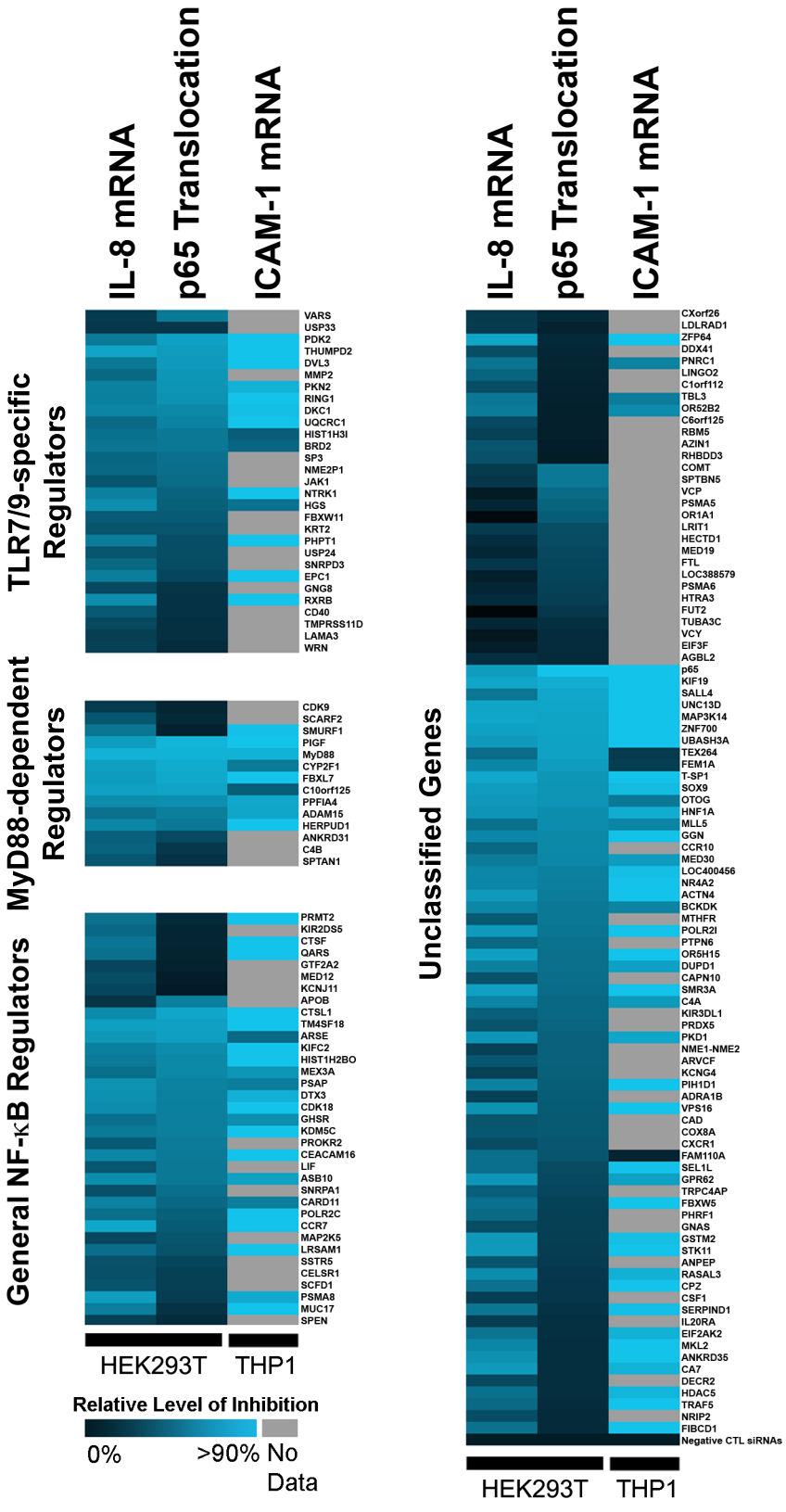

IL-8 mRNA
p65 Translocation
ICAM-1 mRNA
TLR7/9-specific Regulators
MyD88-dependent Regulators
General NF-κB Regulators
Unclassified Genes
VARS
USP33
PDK2
THUMPD2
DVL3
MMP2
PKN2
RING1
DKC1
UQCRC1
HIST1H3I
BRD2
SP3
NME2P1
JAK1
NTRK1
HGS
FBXW11
KRT2
PHPT1
USP24
SNRPD3
EPC1
GNG8
RXRB
CD40
TMPRSS11D
LAMA3
WRN
CDK9
SCARF2
SMURF1
PIGF
MyD88
CYP2F1
FBXL7
C10orf125
PPFIA4
ADAM15
HERPUD1
ANKRD31
C4B
SPTAN1
PRMT2
KIR2DS5
CTSF
QARS
GTF2A2
MED12
KCNJ11
APOB
CTSL1
TM4SF18
ARSE
KIFC2
HIST1H2BO
MEX3A
PSAP
DTX3
CDK18
GHSR
KDM5C
PROKR2
CEACAM16
LIF
ASB10
SNRPA1
CARD11
POLR2C
CCR7
MAP2K5
LRSAM1
SSTR5
CELSR1
SCFD1
PSMA8
MUC17
SPEN
CXorf26
LDLRAD1
ZFP64
DDX41
PNRC1
LINGO2
C1orf112
TBL3
OR52B2
C6orf125
RBM5
AZIN1
RHBDD3
COMT
SPTBN5
VCP
PSMA5
OR1A1
LRIT1
HECTD1
MED19
FTL
LOC388579
PSMA6
HTRA3
FUT2
TUBA3C
VCY
EIF3F
AGBL2
p65
KIF19
SALL4
UNC13D
MAP3K14
ZNF700
UBASH3A
TEX264
FEM1A
T-SP1
SOX9
OTOG
HNF1A
MLL5
GGN
CCR10
MED30
LOC400456
NR4A2
ACTN4
BCKDK
MTHFR
POLR2I
PTPN6
OR5H15
DUPD1
CAPN10
SMR3A
C4A
KIR3DL1
PRDX5
PKD1
NME1-NME2
ARVCF
KCNG4
PIH1D1
ADRA1B
VPS16
CAD
COX8A
CXCR1
FAM110A
SEL1L
GPR62
TRPC4AP
FBXW5
PHRF1
GNAS
GSTM2
STK11
ANPEP
RASAL3
CPZ
CSF1
SERPIND1
IL20RA
EIF2AK2
MKL2
ANKRD35
CA7
DECR2
HDAC5
TRAF5
NRIP2
FIBCD1
Negative CTL siRNAs
HEK293T
THP1
Relative Level of Inhibition
0%
>90%
No Data

Figure S2A

B

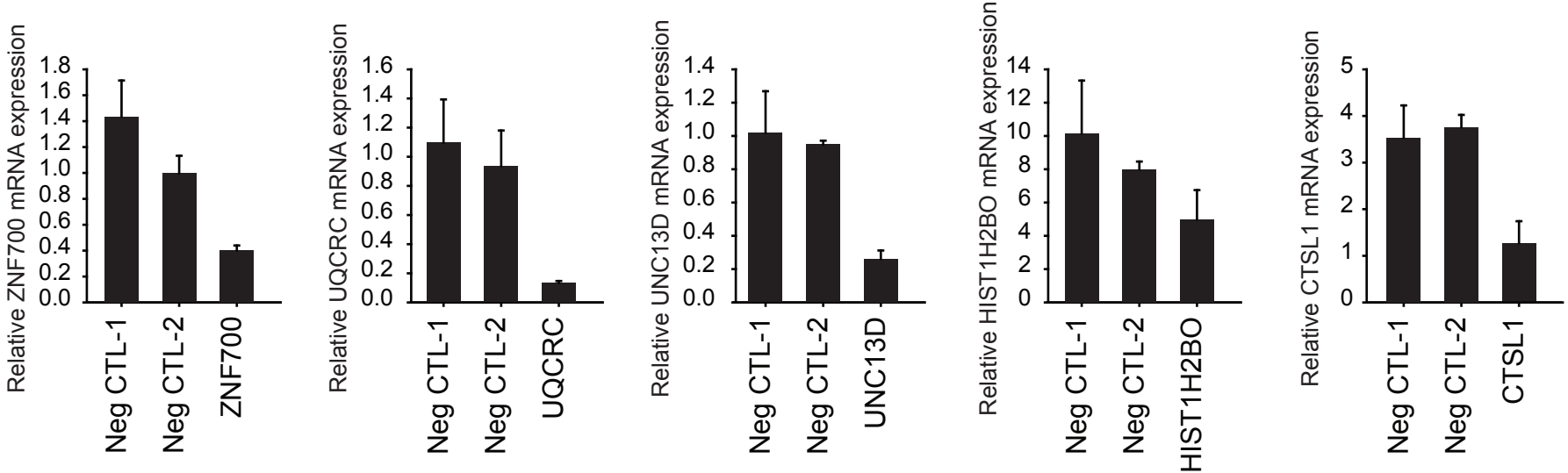

Relative ZNF700 mRNA expression
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Neg CTL-1
Neg CTL-2
ZNF700
Relative UQCRC mRNA expression
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Neg CTL-1
Neg CTL-2
UQCRC
Relative UNC13D mRNA expression
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Neg CTL-1
Neg CTL-2
UNC13D
Relative HIST1H2BO mRNA expression
14
12
10
8
6
4
2
0
Neg CTL-1
Neg CTL-2
HIST1H2BO
Relative CTSL1 mRNA expression
5
4
3
2
1
0
Neg CTL-1
Neg CTL-2
CTSL1

Figure S2B

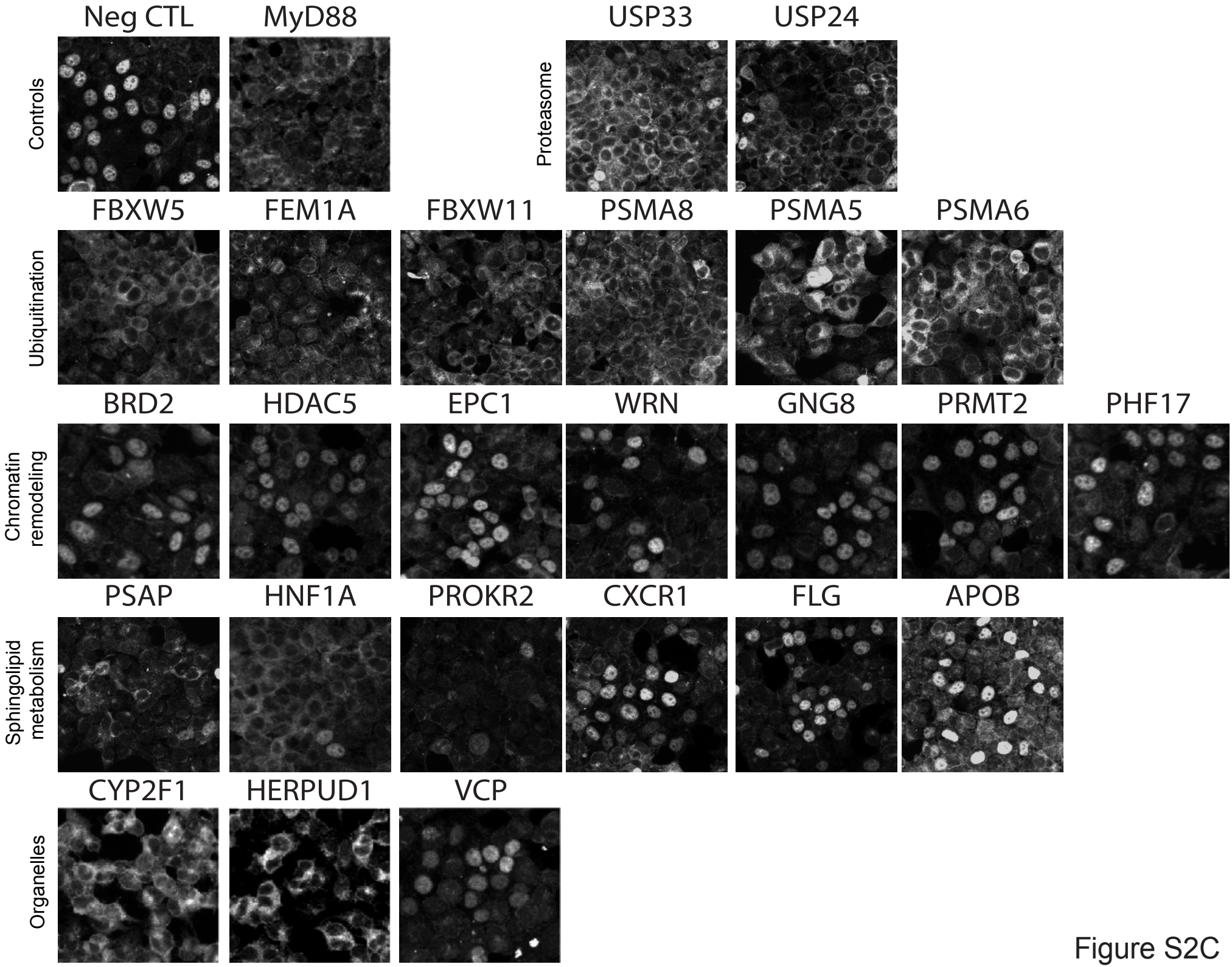

Neg CTL
MyD88
USP33
USP24
Controls
Proteasome
FBXW5
FEM1A
FBXW11
PSMA8
PSMA5
PSMA6
Ubiquitination
BRD2
HDAC5
EPC1
WRN
GNG8
PRMT2
PHF17
Chromatin remodeling
PSAP
HNF1A
PROKR2
CXCR1
FLG
APOB
Sphingolipid metabolism
CYP2F1
HERPUD1
VCP
Organelles

Figure S2C

D
Relative PDK2 mRNA expression
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Neg CTL
MyD88
PDK2-1
PDK2-2
PDK2-3
PDK2-4
Relative IL-8 mRNA expression
0.8
0.6
0.4
0.2
0.0
Neg CTL
MyD88
PDK2-1
PDK2-2
PDK2-3
PDK2-4
E
Relative FBXL7 mRNA expression
0.4
0.3
0.2
0.1
0.0
Neg CTL
FBXL7-1
FBXL7-2
FBXL7-3
FBXL7-4
Relative IL-8 mRNA expression
0.4
0.3
0.2
0.1
0.0
Neg CTL
FBXL7-1
FBXL7-2
FBXL7-3
FBXL7-4

Figure S2D-E

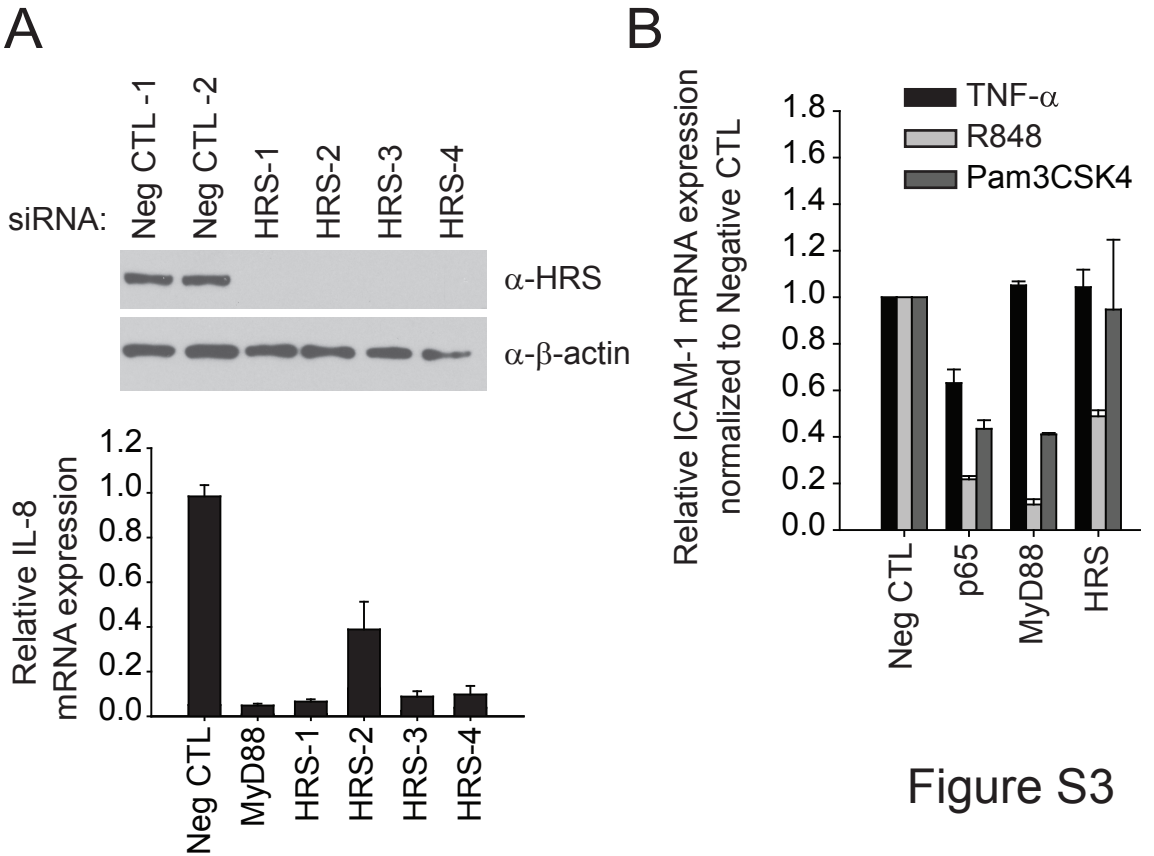

A
siRNA:
Neg CTL -1
Neg CTL -2
HRS-1
HRS-2
HRS-3
HRS-4
α-HRS
α-β-actin
Relative IL-8 mRNA expression
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Neg CTL
MyD88
HRS-1
HRS-2
HRS-3
HRS-4
B
Relative ICAM-1 mRNA expression normalized to Negative CTL
TNF-α
R848
Pam3CSK4
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Neg CTL
p65
MyD88
HRS

Figure S3

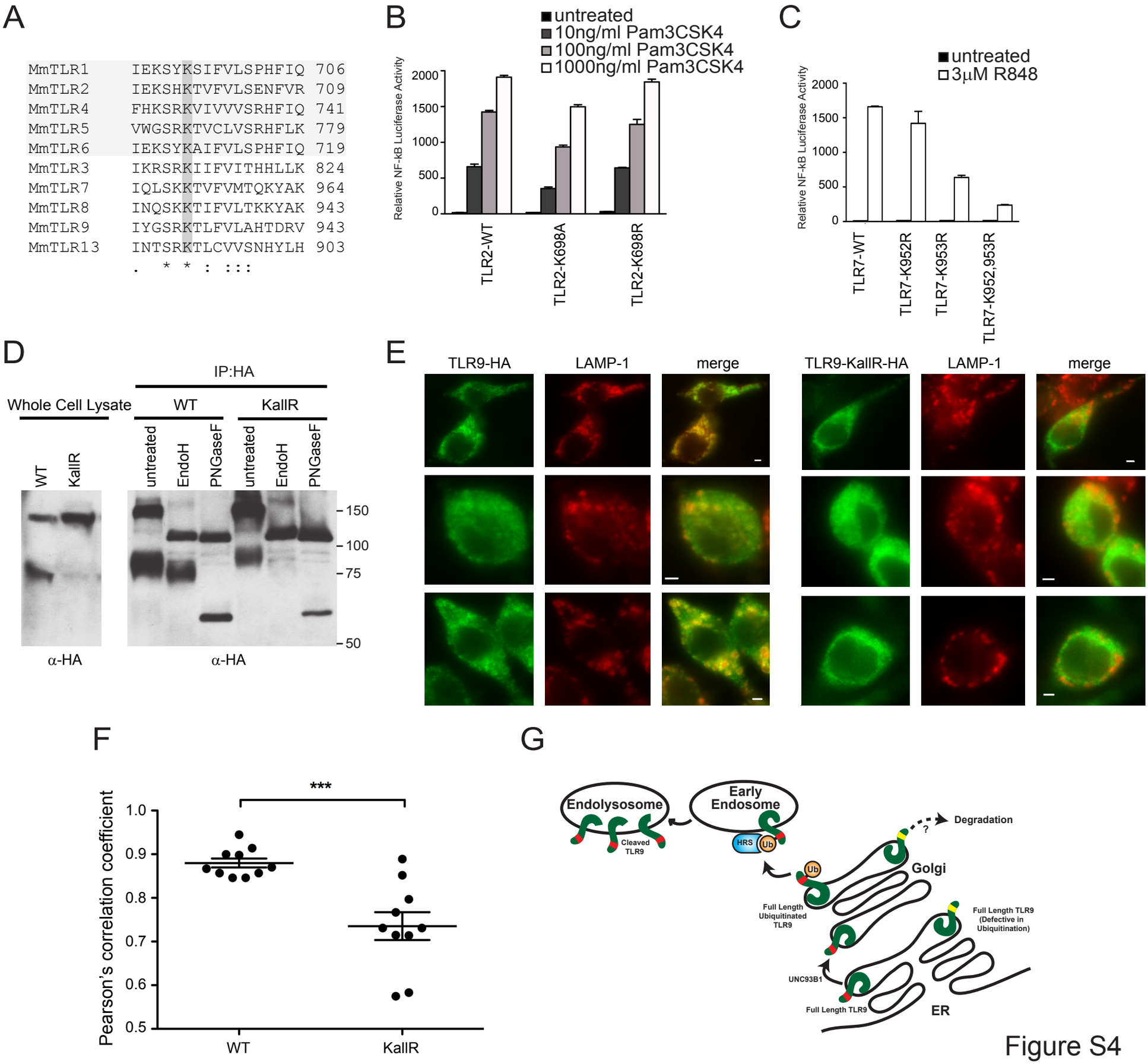

A
MmTLR1 IEKSYKSIFVLSPHFIQ 706
MmTLR2 IEKSHKTVFVLSENFVR 709
MmTLR4 FHKSRKVIVVVSRHFIQ 741
MmTLR5 VWGSRKTVCLVSRHFLK 779
MmTLR6 IEKSYKAIFVLSPHFIQ 719
MmTLR3 IKRSRKIIFVITHHLLK 824
MmTLR7 IQLSKKTVFVMTQKYAK 964
MmTLR8 INQSKKTIFVLTKKYAK 943
MmTLR9 IYGSRKTLFVLAHTDRV 943
MmTLR13 INTSRKTLCVVSNHYLH 903
B
untreated
10ng/ml Pam3CSK4
100ng/ml Pam3CSK4
1000ng/ml Pam3CSK4
Relative NF-kB Luciferase Activity
TLR2-WT
TLR2-K698A
TLR2-K698R
C
untreated
3μM R848
Relative NF-kB Luciferase Activity
TLR7-WT
TLR7-K952R
TLR7-K953R
TLR7-K952,953R
D
IP:HA
Whole Cell Lysate
WT
KallR
untreated
EndoH
PNGaseF
α-HA
E
TLR9-HA
LAMP-1
merge
TLR9-KallR-HA
F
Pearson's correlation coefficient
***
WT
KallR
G
Endolysosome
Early Endosome
Cleaved TLR9
HRS
Ub
Degradation
Golgi
Full Length Ubiquitinated TLR9
Full Length TLR9 (Defective in Ubiquitination)
UNC93B1
Full Length TLR9
ER

Figure S4

Supplemental Figure Legends

Figure S1: Overall data analysis, integrative analyses, network analysis, and Molecular Complex Detection (MCODE) analysis of genes identified by genome-wide siRNA screening. (A) *Top left quadrant:* siRNA-Based Analysis - the scaled activities of each well (black circles; 2 siRNAs targeting 1 gene/well) in the primary TLR7 RNAi screen are shown on the x-axis, while the activities of each well in the primary TLR9 RNAi screen are shown on the y-axis. Red circles represent the activities of canonical TLR7/9 pathway members; gold circles indicate the activities of the 190 confirmed innate immune signaling co-factors; and blue circles reflect negative control measurements (see Table S1). *Top right quadrant:* Gene-based Analysis - the combined activities of all siRNAs targeting the same gene (RSA $\log_{10}$P values; see Supplemental Experimental Procedures) in the TLR7 screen are shown on the x-axis, while their activities in the TLR9 screen are shown on the y-axis. *Bottom left quadrant:* the minimal RSA $\log_{10}$P value for a gene in either the TLR7 or the TLR9 screen is shown (x-axis) in relation to the number of binary interactions attributed to each encoded protein within protein interaction databases (y-axis). *Bottom right quadrant:* the minimal RSA $\log_{10}$P value of a gene in either the TLR7 or TLR9 screen is shown on the x-axis, and is plotted against the relative expression level of that gene in hematopoietic cell types (y-axis; also see Supplemental Experimental Procedures). (B) Evidence-based analysis flow chart: 107,734 siRNAs were reverse transfected into the HEK293/TLR7/NF-κB or HEK293/TLR9/NF-κB luciferase reporter cell line. siRNA activity scores were generated by normalizing and transforming the scores for each well assayed; RSA activity scores were

generated by a redundant siRNA activity (RSA) algorithm that was applied to both assays, and p-values for each gene were assigned. *The genome-wide experimental dataset is indicated by the yellow shaded area.* To identify hits that were more likely to be physiologically relevant, a comprehensive evidence-based approach was employed. In this approach, both experimental and orthogonal datasets were complied. Experimental datasets are described above (genome-wide screens). Additional orthogonal data was extracted from protein-protein interaction network databases, human tissue expression datasets, TLR7/TLR8 microarray datasets, as well as others. (For a detailed description of this approach, see Supplemental Experimental Procedures.) *The orthogonal approaches and evidence-based analysis are indicated by the tan shaded area. Numbering along the left side of the flow chart corresponds to numbered evidences listed in Supplemental Experimental Procedures.* The evidences collected for each gene were compiled into a decision matrix and 546 genes were chosen for reconfirmation based on the hitpick criteria described in the Supplemental Experimental Procedures. By this approach, 190 genes were reconfirmed. Based on the activity of siRNAs directed against these genes in a ligand cross-profiling study, 80 genes were assigned proposed mechanisms associated with the TLR signaling pathway. Additionally, the mechanisms of 38 genes were predicted based on a Support Vector Machine (SVM) algorithm. *The confirmation studies are indicated by blue shaded area.* (C) A multi-scale approach was employed to select candidate genes for study by compiling the multiple lines of experimental and orthogonal evidences described in (B) (Also

see Table S2 for quantitative measures of each evidence, and Supplemental Experimental Procedures for detailed descriptions of how evidences were collected). Specifically, criteria included in the analyses are: 1) *siRNA activity* – The siRNA activity score of single wells containing 2 siRNAs targeting one gene in the genome-wide assay for TLR7 or TLR9; 2) *RSA activity* – The RSA activity score (combined activity of multiple wells targeting a single gene) in the genome-wide assay for TLR7 or TLR9; 3) *Protein Network-TLR7 & TLR9* – The number of interacting partners in the direct protein-protein networks constructed with primary hits in the TLR7 and TLR9 screens; *Protein Network-Innate* – Whether the candidate gene was previously annotated in InnateDB; *Protein Network-Inn. Int.* – The number of predicted interactions between the candidate protein and proteins in InnateDB; 4) *Gene Expression-Blood* – Overall percentage of blood tissue samples where the gene is expressed; *Gene Expression-TLR7 & MYD88* – The Pearson correlation coefficient of the tissue expression profiles between the candidate gene and that of TLR7 or MyD88; *Gene Expression-293dFC* – The largest change in the expression levels of a gene in HEK293/TLR7 cells or HEK293/TLR8 cells after stimulation for 2hr/4hr/8hr, whichever yielded the largest fold changes; *Gene Expression-FC CpG* – The gene expression fold change after CpG ODN treatment at 4hr or 12hr, whichever yielded larger changes; *Gene Expression-293 OPI* – Inclusion of a gene into an ontology-based pattern identification (OPI) gene cluster derived from a HEK293 expression panel of nine samples. All the above individual evidences were then transformed into a normalized score between zero and one, where zero stands for no support and

one stands for the strongest support (also see Table S2). (D) Statistically significant overrepresentation of selected functional classes and protein families based upon gene ontology (GO) analysis and Interpro (IPR) domain mapping of 190 innate signaling co-factors (also see Table S3). (E) mRNA transcriptional signatures were extracted from GNF Tissue Atlas. Tissues and cell types are depicted on the x-axis, while genes identified through our analyses are reflected on the y-axis. This hierarchical clustering revealed a subset of genes that are highly expressed in cells and tissues of myeloid and lymphoid origin, consistent with a role for these factors as important regulators of innate immune signaling (also see Figure 1B). A continuum of blue (low expression) to red (high expression) depicts evidence support or expression levels (Figure S1C and S1E). (F) Protein-protein interaction analysis of confirmed TLR7/9 co-factors in a highly condensed network with known TLR canonical pathway members. (G) Plot generated in Cytoscape representing the local connectivity of selected primary hits identified from the genome-wide siRNA screen, including 190 confirmed factors (see Figure 1C for a less detailed interaction map). For (F) and (G), interaction type, protein type, and expression patterns are distinguished by differential edge color, node color, and node shape, respectively, as indicated. Additionally, significance of these networks (p-value <0.001) was determined by comparing the complexity of iterative randomized (1000X) networks of the same size. (H-N) To identify unusually dense regions of protein-protein interconnectivity within Figure 1C and S1G, a graph theoretic clustering algorithm plugin for Cytoscape (MCODE) was applied to generate sub-networks. The

significance of each of the sub-networks was determined as described above (p-value < 0.001). Interaction type, protein type, and expression patterns are distinguished by differential edge color, node color, and node shape, respectively, as indicated. For protein type, 'Confirmed hit' depicts proteins confirmed by our analysis and 'Primary hit' depicts genes identified only in the primary screen.

Figure S2: Compilation of functional data for the confirmed genes. (A) Heatmap of the effect of siRNAs targeting confirmed factors on NF-κB-dependent gene expression and p65 nuclear translocation in response to R848 stimulation; *IL-8* mRNA expression levels in HEK293/TLR7/NF-κB luciferase cells, and *ICAM-1* mRNA expression levels in THP-1 cells were quantified by real-time PCR (RT-PCR) (also see Table S7). p65 nuclear translocation in HEK293/TLR7/NF-κB reporter cells was imaged and quantified by immunofluorescence (also see Table S7). The data from each individual assay was normalized and scaled to the negative controls, and then genes were clustered according to activity. The left column represents confirmed factors where proposed mechanisms were assigned based on ligand cross-profiling studies (Figure 1D). The right column represents identified factors where mechanisms could not be predicted. The gray blocks indicate no data is available. (B) Knockdown efficiency in THP-1 cells was measured by RT-PCR. Gene silencing was evaluated for those factors that most potently reduced R848-induced *ICAM-1* upregulation in THP-1 cells (see Figure 3A). THP-1 cells were

transfected with indicated siRNAs. Three days post transfection, mRNA expression levels for the indicated genes were measured by RT-PCR. mRNA expression values are shown relative to *TATA BP*. (C) Functional study of the role of siRNAs involved in ubiquitination and proteasomal degradation (called '*ubiquitination*' and '*proteasome*'); transcriptional control of the inflammatory response through chromatin remodeling (called '*chromatin remodeling*'); sphingolipid metabolism (called '*sphingolipid metabolism*'); and organelles and ancient stress responses (called '*organelles*') in TLR7/9 signaling. Individual siRNAs directed against confirmed co-factors are indicated above each image depicting p65 nuclear translocation. The functional data for the positive and negative control siRNAs (MyD88 and NegCTL) are shown in the top panel of images. All p65 nuclear translocation data was obtained as in Figure 3F. (D and E) R848-induced *IL-8* expression was correlated with knockdown of *PDK2* and *FBXL7* using 4 sequence-independent siRNAs. Briefly, siRNAs targeting indicated genes were transfected in HEK293/TLR7/NF-κB luciferase cells. Three days post-transfection, cells were stimulated with 3uM of R848 for 5 hours. *IL-8* mRNA levels and *PDK2* or *FBXL7* knockdown levels were quantified by RT-PCR. mRNA expression levels are shown relative to *TATA BP*. Bar graphs in (B), (D) and (E) are presented as mean ± SD from a representative experiment.

Figure S3: A role for HRS in endosomal TLR7/9 trafficking. (A) R848-induced *IL-8* expression was correlated with knockdown of *HRS* using 4 sequence-independent siRNAs. HRS knockdown was measured by western blot using the

indicated antibodies. (B) THP-1 cells were transfected with indicated siRNAs, and stimulated with either 5uM of R848, 10ng/ml of TNF-α or 10ng/ml of Pam3CSK4. Inhibition of each signaling pathway was measured by quantitative real-time PCR analysis of *ICAM-1* mRNA levels. Data shown is a representation of three independent replicates. The bar graph in (A) is presented as mean $\pm$ SD from a representative experiment. The bar graph in (B) is presented as a percentage relative to negative control siRNAs $\pm$ SD.

Figure S4: Cellular trafficking of WT TLR9 and KallR TLR9 mutant. (A) ClustalW alignment of a TIR-domain region of murine TLRs. TLRs capable of engaging ligands from the surface are grouped and shaded light gray; endosomal nucleic acid-sensing TLRs are not shaded. The conserved lysine is shaded in dark gray. (B-C) HEK293T were transfected with the indicated wild type or mutant TLRs together with an NF-κB luciferase reporter, and luciferase activity in cell lysates was measured after stimulation with the relevant TLR ligands (See Supplemental Experimental Procedures). (D) TLR9 was immunoprecipitated from *TLR9*$^{-/-}$ immortalized macrophages retrovirally transduced with wild type TLR9-HA or mutant TLR9KallR-HA and the resulting eluate was treated with endoglycosidases EndoH and PNGaseF. Whole cell lysates and immunoprecipitates were immunoblotted with antibodies against HA. (E) RAW264.7 macrophages expressing C-terminally HA tagged WT TLR9 or TLR9-KallR were costained for HA and LAMP-1. Representative images are depicted (n=70). Scale bars represent 2um. (F) Mean ± SEM of Pearson's

correlation coefficients are 0.88 ± 0.010 for WT TLR9 vs. LAMP-1 in B6 cells and 0.73 ± 0.032 for TLR9-KallR vs. LAMP-1. *; $P < 0.001$, unpaired t-test. (G) Model of TLR9 ubiquitination and endosomal trafficking. UNC93B1 facilitates the exit of TLR9 from the ER. After exit from the ER, TLR9 is ubiquitinated. HRS physically interacts with TLR9 in a ubiquitin-dependent manner and allows trafficking of TLR9 to early endosomes. Subsequently, TLR9 is proteolytically processed in the acidified endolysosomes and becomes competent to recognize and respond to PAMPs. The TLR9 KallR mutant is not ubiquitinated and is degraded via an undetermined mechanism instead of undergoing HRS-mediated sorting. The red band on TLR9 represents the wild-type context of cytoplasmic lysine residues. The yellow band on TLR9 indicates the arginine substitution of the cytoplasmic lysine residues (TLR9 KallR). Bar graphs in (B) and (C) are presented as mean ± SEM from a representative experiment.