

Transcriptional Blood Signatures for Active and Amphotericin B Treated Visceral Leishmaniasis in India

Michaela Fakiola,^{1,2,†} Om-Prakash Singh,^{3,†} Genevieve Syn,⁴ Toolika Singh,³ Bhawana Singh,³ Jaya Chakravarty,³ Shyam Sundar^{3,*} and Jenefer M. Blackwell^{1,4,*,§}

¹Department of Pathology, University of Cambridge, Cambridge, UK

² INGM-National Institute of Molecular Genetics "Romeo ed Enrica Invernizzi" Milan, Milan, Italy

³Department of Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India

⁴Telethon Kids Institute, The University of Western Australia, Nedlands, Western Australia, Australia

§Corresponding Author: Professor Jenefer Blackwell, Telethon Kids Institute, PO Box 855, West Perth, Western Australia 6872, Australia. E-mail: jenefer.blackwell@telethonkids.org.au

† Equal first authors

* Equal senior authors

Data Availability Statement: Probe-level expression data have been deposited in NCBI's Gene Expression Omnibus (GEO) repository under accession number GSE125993.

Funding: This work was supported by the NIH as part of Tropical Medicine Research Centre award 1P50AI074321. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Short title: Transcriptional profiling in human visceral leishmaniasis

Abstract

Amphotericin B provides improved therapy for visceral leishmaniasis (VL) caused by *Leishmania donovani*, with single dose liposomal-encapsulated Ambisome providing the best cure rates. The VL elimination program aims to reduce the incidence rate in the Indian subcontinent to <1/10,000 population/year. Ability to predict which asymptomatic individuals (e.g. anti-leishmanial IgG and/or Leishmania-specific modified Quantiferon positive) will progress to clinical VL would help in monitoring disease outbreaks. Here we examined whole blood transcriptional profiles associated with asymptomatic infection, active disease, and in treated cases. Two independent microarray experiments were performed, with analysis focussed primarily on differentially expressed genes (DEGs) concordant across both experiments. No DEGs were identified for IgG or Quantiferon positive asymptomatic groups compared to negative healthy endemic controls. We therefore concentrated on comparing concordant DEGs from active cases with all healthy controls, and in examining differences in the transcriptome following different regimens of drug treatment. In these comparisons 6 major themes emerged: (i) expression of genes and enrichment of gene sets associated with erythrocyte function in active cases; (ii) strong evidence for enrichment of gene sets involved in cell cycle in comparing active cases with healthy controls; (iii) identification of *IFNG* encoding interferon- γ as the major hub gene in concordant gene expression patterns across experiments comparing active cases with healthy controls or with treated cases; (iv) enrichment for interleukin signalling (IL-1/3/4/6/7/8) and a prominent role for CXCL10/9/11 and chemokine signalling pathways in comparing active cases with treated cases; (v) the novel identification of Aryl Hydrocarbon Receptor signalling as a significant canonical pathway when comparing active cases with healthy controls or with treated cases; and (vi) global expression profiling support for more effective cure at day 30 post-treatment with a

single dose of liposomal encapsulated amphotericin B compared to multi-dose non-liposomal amphotericin B treatment over 30 days. (296 words; 300 words allowed)

Author Summary

Visceral leishmaniasis (VL), also known as kala-azar, is a potentially fatal disease caused by intracellular parasites of the *Leishmania donovani* species complex. VL is a serious public health problem in rural India, causing high morbidity and mortality, as well as major costs to local and national health budgets. Amphotericin B provides improved therapy for VL with single dose liposomal-encapsulated Ambisome, now affordable through WHO-negotiated price reductions, providing the best cure rates. The VL elimination program aims to reduce the incidence rate in the Indian subcontinent to <1/10,000 population/year. By assessing immune responses to parasites in people infected with *L. donovani*, but with different clinical status, we can determine the requirements for immune cell development and predict which asymptomatic individuals, for example healthy individuals with high anti-leishmanial antibody levels, will progress to clinical VL. This will help in monitoring disease outbreaks. In this study we looked at global gene expression patterns in whole blood to try to understand more about asymptomatic infection, active VL, and the progress to cure in cases treated with single or multi-dose amphotericin B. The signatures of gene expression identified aid in our understanding of disease pathogenesis and provide novel targets for therapeutic intervention in the future.

(200 words; 150-200 words allowed)

Introduction

Visceral leishmaniasis (VL), also known as kala-azar, is a potentially fatal disease caused by obligate intracellular parasites of the *Leishmania donovani* species complex. VL is a serious public health problem in indigenous and rural populations in India, causing high morbidity and mortality, as well as major costs to both local and national health budgets. The estimated annual global incidence of VL is 200,000 to 400,000, with up to 50,000 deaths annually occurring principally in India, Bangladesh, Sudan, South Sudan, Ethiopia and Brazil [1]. In India, improvements in drug therapy have been afforded through the introduction of amphotericin B treatment, with single dose liposomal encapsulated Ambisome providing the best cure rates and now being used as the preferred treatment regime in the VL elimination program [2]. However, with the potential development of drug resistance to each new therapeutic approach [3], there remains a continuing need for improved and more accurate methods of early diagnosis, as well as ability to monitor responses to treatment and to predict disease outcome. These objectives are also important in relation to the World Health Organization-supported VL elimination initiative in the Indian subcontinent, which aims at reducing the incidence rate of VL in the region to below 1 per 10,000 population per year by 2020 [4]. Monitoring disease outbreaks in the context of the elimination program will be an important goal, including the ability to determine which individuals displaying asymptomatic disease, as monitored by anti-leishmanial IgG [5, 6] and/or *Leishmania*-specific modified Quantiferon responses [7], will progress to clinical VL disease [6].

In recent years, the use of whole blood transcriptional profiling has provided a better understanding of the host response to infectious disease, leading to the identification of blood signatures and potential biomarkers for use in diagnosis, prognosis and treatment monitoring (reviewed [8]). This approach was successful in identifying a neutrophil-driven interferon

(IFN)-inducible blood transcriptional signature for active tuberculosis that involved both IFN- γ and type I IFN- α/β signalling [9] and was subsequently confirmed in multiple countries world-wide (reviewed [8]). This neutrophil-driven interferon signature was present in active disease but absent in both latent infection and in healthy controls [9]. While an IFN-inducible signature was also identified in patients with the autoimmune disease systemic lupus erythematosus, there were differences in the signatures that also distinguished the two profiles from each other [9]. Viral infections [10] and bacterial infections like melioidosis [11] are also broadly characterised by IFN-inducible gene expression, but whole blood signatures have been identified that are able to discriminate between bacterial and viral infections [10, 12], as well as between different viral infections [10]. In HIV, blood transcriptional signatures have been identified that distinguish between rapid compared to slow progression to disease [13]. Blood signatures have also been identified which distinguish between children who acquire dengue virus fever compared to those who develop dengue haemorrhagic fever [14, 15]. There are also signatures that distinguish between pulmonary and extra-pulmonary tuberculosis [16], as well as between pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers [17]. A transcriptional signature that can be used to monitor treatment response is also a valuable goal in infectious disease. Again, studies from two cohorts followed longitudinally in South Africa show that the transcriptional signature of active tuberculosis disease rapidly diminishes with successful treatment [18, 19].

In this study we set out to determine whole blood transcriptional profiles that might distinguish asymptomatic infection with *L. donovani* from active disease, as well as to monitor the changes in transcriptional profiles that accompanied drug treatment. Whilst we were unable to detect a signature that distinguished asymptomatic (IgG antibody positive [5,

6], or modified Quantiferon positive [7]) individuals from healthy endemic controls who were negative by these two assays, we were able to determine the transcriptional profile of active VL cases, and to demonstrate interesting differences in return to baseline between patients treated with non-liposomal compared to liposomal-encapsulated (Ambisome™) amphotericin B.

Methods

Study Subjects

In this study two independent microarray experiments were performed. For experiment 1, samples were collected between February and April 2011. For experiment 2, samples were collected between April and July 2012. Samples were collected at the Kala-azar Medical Research Center (KAMRC), Muzaffarpur, Bihar, India, or in nearby field sites for some asymptomatic individuals and endemic controls. Active VL cases were diagnosed by experienced clinicians based on clinical signs, including fever (>2 weeks), splenomegaly, positive serology for recombinant antigen (r)-K39 and/or by microscopic demonstration of *Leishmania* amastigotes in splenic aspirate smears. VL patients were treated according to routine clinical care with either (a) experiment 1: 0.75 mg/kg non-liposomal amphotericin B daily for 15 days (N=3), or on alternate days over 30 days (N=7), by infusion (i.e. 15 doses in all; total dose 11.25 mg/kg over 30 days); or (b) experiment 2: 10 mg/kg of Ambisome (liposome-encapsulated amphotericin B) as a single dose by infusion. Blood samples were collected pre- (N=10 experiment 1; N=11 experiment 2) and post- (day 30; N=10 experiment 1; N=11 experiment 2) treatment. There were 9 paired pre-/post-treatment samples for experiment 1; 10 for experiment 2. Healthy control subjects included (i) asymptomatic individuals (N=2 experiment 1; N=6 experiment 2) who had high anti-leishmanial antibody

levels by direct agglutination test (DAT titer $\geq 1:25,600$) [6]; (ii) asymptomatic individuals who were positive by *Leishmania*-specific modified quantiferon assays [7] (N=8 experiment 1; N=9 experiment 2); and (iii) Serology (DAT titer $\leq 1:1600$) and quantiferon negative healthy endemic controls (N=6 experiment 1; N=10 experiment 2). Sample sizes are for post-QC samples used in expression profiling studies (cf. below). Further clinical and demographic details on participants are provided in **S1 Table**.

Ethics

The enrolment of human subjects complies with the principles laid down in the Helsinki declaration. Institutional ethical approval (reference numbers: Dean/2012-2013/89) was obtained from the ethical review board of Banaras Hindu University (BHU), Varanasi, India. Informed written consent was obtained from each participant at the time of enrolment, or from their legal guardian if they were under 18 years old. Only patients who had not previously received treatment and who agreed to participate in the study were enrolled. All clinical treatment and follow-up records were maintained using standardised case report forms on an electronic server. All patient data were analysed anonymously.

RNA extraction and microarray analysis

Whole blood (5 mL) collected by venepuncture was immediately placed into Paxgene tubes (QIAGEN GmbH, Germany) and stored at -80°C for later processing for RNA. RNA was extracted using PAXgene Blood RNA kits (QIAGEN GmbH, Germany) according to manufacturer's instructions. RNA integrity and purity were checked using Tape Station 4200 (Agilent Technologies, USA). Only RNAs with RNA integrity (RIN) values >5.5 were taken forward for beadchip analysis. Globin mRNA was depleted using GLOBINclear™-Human kits (ThermoFischer Scientific, USA). RNA was reverse transcribed and biotin-labelled

using the Illumina TotalPrep™ RNA Amplification kit (ThermoFischer Scientific, USA). The resulting biotinylated cRNA was hybridised to Illumina HT12v4 Expression BeadChips, specifically HumanHT-12_V4_0_R2_15002873_B, containing 47,323 genome wide gene probes, and 887 control probes. Samples from different control or clinical groups were distributed evenly across 3 (experiment 1) or 4 (experiment 2) beadchips. All RNA preparation and processing of samples over beadchips was carried out at Sandor Lifesciences Pvt. Ltd. (Hyderabad, India).

Data analysis

All data analysis was carried out in R Version 3.4.3 (Smooth Sidewalk - <https://www.r-project.org/>) and RStudio (version 1.1.383). The Bioconductor package *Lumi* [20] was used to read in raw expression values and perform quality control. Background correction and quantile normalisation of the data was carried out using the Bioconductor package *Limma* [21]. Pre-processing of the microarray data and removal of non-expressed (detection *P*-value > 0.05 in all arrays) and poor quality probes previously shown to have unreliable annotation [22] provided 21,959 and 23,466 probe sets which passed QC in experiments 1 and 2, respectively. Principal components analysis (PCA) and unsupervised cluster analysis (Pearson's correlation coefficient; hclust=complete) of normalised data was performed in R. Data was visualised using the R packages *ggplot2* (3.1.2) [23] and *pheatmap* (1.0.12) [24]. Differential expression analysis using linear modelling and empirical Bayes methods was carried out in the Bioconductor package *Limma* [21] for comparisons between control and clinical groups, as indicated. The threshold for differential expression was a log₂-fold-change ≥0 (i.e. ≥2-fold) and/or Benjamini-Hochberg [25] adjusted *p*-value (*P*_{adj}) ≤0.05, as indicated. Genes achieving these thresholds were taken forward in analyses using the gene set enrichment tool Enrichr [26], and using Ingenuity Pathway Analysis (IPA) (Ingenuity®

Systems, www.ingenuity.com) to identify canonical pathways, upstream regulators, and gene networks. Enrichr [26, 27] accesses a wide range of open access databases to identify pathways for which the gene set is enriched. It uses four scores to report enrichment: a p-value (reported here as P_{nominal}) calculated using Fisher's exact test; a q-value (reported here as P_{adj}) which is the Benjamin-Hochberg adjusted p-value; a rank or z-score computed using a modification to Fisher's exact test to compute deviation from an expected rank; and a combined score which is a combination of the p-value and z-score calculated by multiplying the two scores using the formula $c=\ln(p)*z$. IPA uses the Ingenuity Knowledge Base, an extensive database comprising biological pathways and functional annotations derived from the interactions between genes, proteins, complexes, drugs, tissues and disease, to carry out all its analyses. Benjamini-Hochberg correction was applied where applicable and $P_{\text{adj}} \leq 0.05$ was used to filter all results. Upstream Regulator Analysis within IPA was employed to predict if there were any endogenous genes/cytokines/transcription factors which may be responsible for the observed gene expression patterns. If an upstream regulator is identified, an activation Z-score is calculated based on the fold change values of its target genes within the dataset. A Z-score ≥ 2 suggests that an upstream regulator is activated whereas a Z-score ≤ -2 suggests it is inhibited, with active VL cases being the experimental group of baseline comparator. IPA also generates a "Top Tox List" pathway which provides an indication of toxic or pathogenic pathways that could be amenable to therapeutic intervention. Networks were constructed in IPA using the "Connect" option under the "Build" functionality. Genes with no previously documented interactions were removed from the diagram and the functions of each network were inferred from the remaining connected genes in each time-point.

Results

Comparative transcriptomics across clinical groups

Two independent microarray experiments were carried out to compare transcriptional profiles across clinical groups that included active VL cases pre-treatment (N=10 experiment 1; N=11 experiment 2), drug treated VL cases (N=10 experiment 1; N=11 experiment 2), modified quantiferon [7] positive asymptomatic individuals (N=8 experiment 1; N=9 experiment 2), high Leishmania-specific antibody positive (by DAT) asymptomatic individuals (N=2 experiment 1; N=6 experiment 2), and endemic healthy controls (N=6 experiment 1; N=10 experiment 2) who were both modified quantiferon negative and antibody negative by DAT. PCA of the top 500 most variable probes (Fig 1) across all pairwise comparisons of samples showed that principal component 1 (PC1) accounted for 45% (experiment 1; Fig 1A and B) and 31% (experiment 2; Fig 1D and E) of the variation and resolved active cases compared to endemic healthy control and asymptomatic groups. The latter were not well resolved from each other in either experiment. Treated patients sat intermediate between, and overlapping with, both active cases and control/asymptomatic groups in experiment 1 but showed greater overlap with control/asymptomatic groups in experiment 2 (cf. below). This is particularly apparent when comparing plots of PC1 by PC3 (Fig 1B and 1E). Unsupervised hierarchical cluster analysis also (Fig 1C and 1F) provided discrete clusters of active cases compared to control and asymptomatic individuals, with treated cases interspersed with both active cases and control groups and not falling into a single discrete cluster in either experiment.

Consistent with the PCA plots (Fig 1), there were no differentially expressed probes representing genes (i.e. Benjamini-Hochberg [25] $P_{adj} \leq 0.05$) when comparing either modified quantiferon positive asymptomatic individuals with endemic healthy controls, or when comparing high antibody titer individuals with endemic healthy controls, in either experiment 1 or experiment 2 (data not shown). These groups were therefore analysed as one

group referred to as “controls” or “healthy controls” in all further differential expression analyses.

Differential expression analysis focused on the comparison of (i) active VL cases *versus* controls, (ii) treated VL cases *versus* controls, and (iii) active *versus* treated VL cases. Log₂-fold-change in experiment 1 was highly correlated with log₂-fold-change in experiment 2 across all probes (data not shown). **Table 1** shows the number of differentially expressed probes representing genes in experiment 1 and experiment 2 for each comparison as well as the number of differentially expressed probes that replicated and were concordant for direction of effect between the two cohorts. At $P_{\text{adj}} \leq 0.05$, there are 2,584 concordant differentially expressed probes in common when comparing active cases with controls, 37 concordant probes when comparing treated cases with controls, and 221 concordant probes when comparing active and treated cases. At the more stringent threshold of ≥ 2 -fold change there were 439, 8, and 42 concordant probes for these comparisons, respectively.

Of note, we found a greater number of transcriptional differences between treated cases and controls in experiment 1 compared to experiment 2 (**Table 1**; differentially expressed probes are 1132 and 126, respectively, at $P_{\text{adj}} \leq 0.05$). One explanation for this could be the different treatment regimen employed in the two cohorts. VL patients of the first experiment were treated with 15 doses of a non-liposomal form of amphotericin B over 30 days. In experiment 2 patients received a single dose of liposomal amphotericin B, which has shown better efficacy for the treatment of VL [28, 29]. The effect of treatment regimen on whole blood transcriptional profiles is further indicated by the comparison of active and treated cases. In this case, fewer differences in transcriptional regulation are observed between active and treated cases in experiment 1 as opposed to experiment 2 (**Table 1**; differentially expressed probes are 654 and 1317, respectively, at $P_{\text{adj}} \leq 0.05$), in which patients have received a more efficacious therapy. These findings agree with the PCA results (**Fig 1**), in which treated cases

of the experiment 1 cohort form a more discrete group between active cases and controls (Fig 1A and 1B) whereas treated cases of the experiment 2 cohort are grouped more closely to controls (Fig 1D and 1E).

Due to the small number of concordant differentially expressed genes identified for the treated cases *versus* controls (Table 1; 37 at $P_{adj} \leq 0.05$, 8 at $P_{adj} \leq 0.05$ and ≥ 2 -fold change), only the concordant differentially expressed gene sets for active cases *versus* controls and active cases *versus* treated cases were used in subsequent pathway and gene set enrichment analyses.

Network, pathway and gene set enrichment analyses comparing active cases and healthy controls

Heatmaps were generated for individual expression levels for probes representing the top 10 concordant genes expressed at a higher (“induced”) level (Fig 2A), and the top 10 concordant genes expressed at a lower (“repressed”) level (Fig 2B), in active cases compared to controls in experiment 1. Heatmaps for the same “induced” and “repressed” probes/genes in experiment 2 are presented in Fig 2C and 2D. Of note 8/10 “repressed” genes were also in the top 10 most highly differentially expressed “repressed” genes in experiment 2; all 10 genes achieved ≥ 2 -fold change in both experiments. Amongst these 10 most “repressed” genes were: peptidase inhibitor 3 (*PI3*), a known antimicrobial peptide for bacteria and fungi that is upregulated by lipopolysaccharide and cytokines; the C-C chemokine ligand 23 (*CCL23*; represented by 2 probes) which acts as a chemoattractant for resting (but not active) T cells, monocytes, and to a lesser extent neutrophils; G-protein-coupled C-C motif chemokine receptor 3 (*CCR3*) which binds CCL10 (eotaxin), CCL26 (eotaxin-3), CCL7 (MCP3), CCL13 (MCP4) and CCL5 (RANTES) that likewise act as chemoattractants for eosinophils, monocytes and neutrophils; *ALOX15* which is a lipoxygenase known to regulate

inflammation and immunity; and the G-protein-coupled prostaglandin D2 receptor 2 (*PTGDR2* alias *GPR44*) that is preferentially expressed in CD4 effector T helper 2 (Th2) cells and mediates pro-inflammatory chemotaxis of eosinophils, basophils and Th2 cells. For the “induced” genes (Fig 2A and 2C), only 2/10 (the top 2 in both experiments) were also in the top 10 “induced” genes in experiment 2, but all achieved ≥ 2 -fold change in both experiments. In addition to type I interferon inducible 27 (*IFI27*) and complement C1q B chain (*CIQB*) genes, there was a bias amongst the most strongly “induced” genes towards genes involved in erythrocyte function, including: glycophorin B (*GYPB*), a major sialoglycoprotein of the human erythrocyte membrane; Rh D blood group antigens (*RHD*); hemoglobin subunit delta (*HBD*); 5'-aminolevulinic synthase 2 (*ALAS2*) an erythroid-specific enzyme located in the mitochondrion and involved in heme biosynthesis; carbonic anhydrase 1 (*CA1*) which is found at its highest level in erythrocytes; atypical chemokine receptor 1 (Duffy blood group) (*ACHR1* alias *DARC*) known for its role as the erythrocyte receptor for *Plasmodium vivax* and *P. knowlesi*; and 2,3-diphosphoglycerate (2,3-DPG) (*BPGM*) found at high concentrations in red blood cells where it binds to and decreases the oxygen affinity of haemoglobin.

To gain a more global picture of the impact of differential gene expression, the 391 genes represented by 439 probes that were concordant for differential gene expression ($P_{\text{adj}} \leq 0.05$; ≥ 2 -fold change) between active cases and controls in experiments 1 and 2 were taken forward in Ingenuity Pathway (IPA) and gene-set enrichment (Enrichr) analyses. IPA network analysis indicated that 254 of these genes are joined in a single network (Fig 3), with *IFNG* as the major hub gene (i.e. with most connections to other genes in the network), and other major hub genes including *CCNA2*, *CXCL10*, *SPI1*, *SNCA*, *CHEK1*, *MCM2*, *AURKB*, *RARA*, *CDK1*, *CDC20*, and *FOXMI*. The top Ingenuity Canonical Pathways for the 391 genes that achieved $P_{\text{adj}} < 0.05$ and ≥ 2 -fold change (Table 2) were Estrogen-mediated S-phase Entry

($P=6.46 \times 10^{-5}$; $P_{\text{adj}}=0.019$), Mitotic Roles of Polo-Like Kinases ($P=1.20 \times 10^{-4}$; $P_{\text{adj}}=0.019$), Aryl Hydrocarbon Receptor (AHR) Signalling ($P=1.51 \times 10^{-4}$; $P_{\text{adj}}=0.019$), and Heme Biosynthesis II ($P=3.72 \times 10^{-4}$; $P_{\text{adj}}=0.035$). Although not achieving $P_{\text{adj}} \leq 0.05$, identification of the Th2 pathway ($P=1.22 \times 10^{-3}$; $P_{\text{adj}}=0.074$) and Activation of Th1 and Th2 Pathway ($P=1.32 \times 10^{-3}$; $P_{\text{adj}}=0.074$) as nominally significant canonical pathways is consistent with prior knowledge of immune responses to leishmaniasis.

Identification of AHR signalling as the top Ingenuity Canonical Pathway is reflective of increasing recognition of the role of AHR signalling in immunity, including the ability of AHR ligands to significantly induce cell secretion of IL-10 and inhibit IL-1 β and IL-6 production in dendritic cells, and to promote IL-10 production and suppress IL-17 expression in CD4(+) T cells [30-32], and with the identification of *RARA*, *CCNA2* and *CHEK1* genes from the AHR pathway (Table 2) as major hub genes (Fig 3). AHR signalling was also identified as top in the Ingenuity “Top Tox List” pathway ($P=4.16 \times 10^{-4}$) indicative of its role as a toxic pathology endpoint that could be amenable to therapeutic intervention. Schematic representation of the core AHR canonical pathway overlaid with concordant gene expression data ($P_{\text{adj}} < 0.05$) for experiment 1 (Fig 4) for active cases relative to healthy controls shows differential gene expression that includes core players AHR and the AHR nuclear translocator (ARNT) in the AHR pathway, as well as for key phase I metabolising enzymes (CYPB1, ALDH5A1, ALD3B1 and ALD3A2). The full pathway, including cross-talk between AHR and other signalling pathways that lead to noncanonical mechanisms of action of AHR and its ligands, overlaid with expression data from experiments 1 (S1 Fig) and 2 (S2 Fig), highlight a total of 28 concordant genes that all achieve differential gene expression at $P_{\text{adj}} < 0.05$. These demonstrate the interplay between the top IPA-identified canonical pathways, with AHR function influencing cell proliferation and estrogen receptor signalling pathways, while heme derivatives biliverdin and bilirubin are known to act as endogenous ligands for AHR [33, 34].

Identification of Mitotic Roles of Polo-Like Kinases as a top canonical pathway is indicative of cell proliferative activity that is consistent with *CDC20* and *CDK1* (Table 2) as major hub genes in the network (Fig 3), and with identification of the cyclin-dependent kinase inhibitor *CDKN1A* as the top inhibited upstream regulator (Activation Z-score = -2.764; $P=5.4 \times 10^{-26}$) in IPA.

Using Enrichr (S2 Table), signalling pathways involved in cell cycle predominated amongst the top pathways using the Reactome 2016 (“Cell cycle_ *Homo sapiens*”, “Cell Cycle, Mitotic_ *Homo sapiens*”, and multiple other pathways involved in cell cycle), WikiPathways 2016 (“Cell Cycle *Homo sapiens*”), KEGG 2016 (“Cell cycle_ *Homo sapiens*”), and NCI-Nature 2016 (“Aurora B signalling”, “Aurora A signalling” and the “FOXO1 transcription factor network for *Homo sapiens*”, all of which play key roles in cell cycle progression) databases. *CDK1* was also identified as the top PPI Hub Protein using Enrichr (S2 Table). Consistent with our top 10 “induced” gene list, other database comparisons using Enrichr (S2 Table) identified gene sets associated with erythrocyte function including “erythroid cell” (Jensen Tissues Table), “abnormal erythrocyte morphology” and multiple other erythrocyte-related phenotypes (MGI Mammalian Phenotype 2017), “CD71+Early Erythroid” (Human Gene Atlas), “congenital haemolytic anaemia” (Jensen Diseases), and “Haemoglobin’s Chaperone pathway” (BIOCARTA_2016).

Network, pathway and gene set enrichment analyses comparing active cases and treated cases

Heatmaps were generated for individual expression levels for the top 10 concordant genes expressed at a higher level (Fig 5A), and the top 10 concordant genes expressed at a lower level (Fig 5B), in active cases compared to treated cases in experiment 1. Heatmaps were also generated for the same “induced” and “repressed” probes/genes in experiment 2 (Fig 5C)

and 5D). In this case, 6/10 and 7/10 top genes from experiment 1 were also in the top 10 most highly differentially expressed genes for “induced” and “repressed” gene sets in experiment 2, respectively, and all achieved fold-change >2 in both experiments. Amongst the 10 most “repressed” genes in experiments 1 and 2 were 3 genes also observed in the comparison of active cases with healthy controls: peptidase inhibitor 3 (*PI3*), as noted above known as an antimicrobial peptide for bacteria and fungi; *ALPL* which encodes an alkaline phosphatase that plays a role in bone mineralization; and *CACNA2D3* which encodes the alpha2delta3 subunit of the voltage-dependent calcium channel complex. Of additional interest in this comparison were “repressed” genes: *CHI3L1* which encodes a chitinase-like protein that lacks chitinase activity but is secreted by activated macrophages and neutrophils; *EMR3* (*ADGRE3*) encoding an adhesion G protein-coupled receptor expressed predominantly in cells of the immune system and playing a role in myeloid-myeloid interactions during inflammation; and *MMP25* that encodes matrix metalloproteinase 25 which inactivates alpha-1 proteinase inhibitor produced by activated neutrophils during inflammation thereby facilitating transendothelial migration of neutrophils to inflammatory sites. Of interest amongst the top 10 “induced” genes in both experiments were: *CXCL10* encoding a chemokine of the CXC subfamily that is a ligand for CXCR3, binding to which results in stimulation of monocyte, natural killer and T-cell migration; *IFNG* encoding interferon- γ , well known for its role in macrophage activation for anti-leishmanial activity; and *GBP1* that encodes a guanylate binding protein induced by interferon.

As there were only 42 concordant genes that achieved ≥ 2 -fold change in expression, a more global picture of the impact of differential gene expression was obtained by performing IPA and Enrichr analyses using the full set of 210 genes represented by 221 probes that were concordant for differential gene expression at $P_{\text{adj}} \leq 0.05$. IPA network analysis indicated that 85 of these genes are joined in a single network (Fig 6), with *IFNG* as the major hub gene

(i.e. with most connections to other genes in the network), and other major hub genes including *STAT1*, *SPI1*, *RARA*, *NOTCH1* and *MAPK3*. The top canonical pathways included pathogenesis of multiple sclerosis (Table 2), consistent with interconnections between *CXCL10/CXCL9/CXCL11* and major hub genes *IFNG* and *STAT1* (Fig 6), and the Notch signalling pathway. Aryl hydrocarbon receptor signalling was also identified as a canonical pathway in this analysis at a nominal $P=0.003$ (Table 2). Enrichment for chemokine signalling and Notch signalling pathways were also supported by analyses undertaken using Enrichr (Reactome 2016; WikiPathways 2016, and KEGG 2016 pathways; S3 Table). Consistent with this were top LINCS_L1000_Ligand_Perturbations_Up (S3 Table) for which perturbations of TNFA, IFNG, IL1, IFNA, and HGF were all significant at $P_{adj}<0.01$. These ligand perturbations were all associated with differential expression at *CXCL10*, and commonly also at *CXCL11*, *CXCL9*, and *STAT1*. The major cell types associated with the treatment response were CD14⁺ monocytes and CD33⁺ myeloid cell populations (Human Gene Atlas; S3 Table).

Analysis of discordant genes for active cases compared to treated cases

As noted above, we found more differentially expressed probes between treated cases and controls, along with fewer differentially expressed probes between active and treated cases, in experiment 1 compared to experiment 2 (Table 1). We hypothesize that this is due to more effective treatment using liposome encapsulated amphotericin B in experiment 2 compared to the non-liposomal form of the drug employed during experiment 1. We therefore examined the genes that were discordant between active cases and treated cases across the two experiments to understand differences in the cure response. In support of the more efficient cure rate in experiment 2, 7/10 of the top “repressed” genes (namely: *OLIG1*, *OLIG2*, *PTGDR2* alias *GPR44*, *CCR3*, *CCL23*, *ALOX15*, *SLC29A1*) identified as differentially

expressed between active cases and treated cases in experiment 2 but not experiment 1 were the same genes that were most repressed in the concordant genes comparing active cases with healthy controls. In comparison, 0/10 of the top “repressed” genes identified as differentially expressed between active cases and treated cases in experiment 1 (but not experiment 2) matched the comparison of concordant genes for active cases and healthy controls. That is, treated cases in experiment 2 were behaving more like healthy controls than were treated cases in experiment 1.

To gain a more global picture of differential gene expression that might inform mechanistic differences in cure rates between the two therapeutic regimes, the 417 genes (from 443 probes) that were differentially expressed between active cases and treated cases in experiment 1 but not experiment 2, and the 988 genes (from 1096 probes) that were differentially expressed between active cases and treated cases in experiment 2 but not experiment 1, were analysed in Enrichr for gene-set enrichment. S4 Table and S5 Table present details of the pathways and gene sets that contrast molecular events that characterise the two different treatment groups. These are summarised in Table 3. For the 988 genes that were differentially expressed between active cases and treated cases in experiment 2 but not in experiment 1 (S4 Table) signalling pathways involved in cell cycle predominated amongst the top pathways using the Reactome 2016 (“Cell cycle_Homo sapiens”), Wiki Pathways 2016 (“Cell Cycle_Homo sapiens”), KEGG 2016 (“Cell cycle_Homo sapiens”), and NCI-Nature 2016 (“Aurora B signalling”) databases. In every case there were multiple other pathways involved in cell cycle that achieved Z-scores <-1 and $P_{adj} \leq 0.01$. This pattern recapitulates the results obtained in the earlier comparison of concordant genes for active cases and healthy controls (S2 Table), with *CDKI* again identified as the top PPI Hub Protein for this gene set (S5 Table). Consistent with an enhanced rate of cure, multiple immune response signalling pathways (Table 3 and S4 Table) were also identified in this gene set,

including IL-1, IL-3, IL-4, IL-6, IL-7 and IL-8 signalling pathways (Reactome 2016, Wiki Pathways 2016, and NCI Nature 2016 databases), Delta-Notch signalling (Wiki Pathways database), chemokine signalling (Wiki Pathways 2016 and KEGG 2016 databases) including specifically IL-8/CXCR2-mediated and CXCR4-mediated signalling (NCI Nature 2016 database), and Fc gamma R-mediated phagocytosis *Homo sapiens* (KEGG 2016 database). Of note, IL-4 was identified as the most significantly down-regulated perturbed ligand pathway (LINCS_L1000_Perturbed_Down; Z-score -1.8, P_{adj} 6.36×10^{-11}) in this set of genes differentially expressed in active versus treated cases in experiment 2 but not experiment 1. None of these databases showed significant gene set enrichment when interrogated with the 417 genes identified as differentially expressed between active cases and treated controls in experiment 1 but not in experiment 2, i.e. they are not present in Table 3 or S5 Table which compare other enriched gene sets showing differences of interest between experiments 1 and 2. For example, all PPI Hub Proteins identified as significant for the 988 genes that were differentially expressed between active cases and treated cases in experiment 2 but not experiment 1 were related to cell cycle (S5 Table). In contrast, the 5 significant matches to gene sets for PPI Hub Proteins for the 417 genes differentially expressed between active cases and treated cases in experiment 1 but not experiment 2 included the inhibitor of NFκB NFKBIA and the SMAD-signalling pathway gene SMAD9 which transduces signals from members of the TGFβ family. Mutations in NFKBIA are associated with T-cell immunodeficiency [35]. SMAD9 (aliases SMAD8, SMAD8A, SMAD8B, SMAD8/9) transduces signals following ligation of TGFβ family members known as bone morphogenesis proteins (BMPs) to specific BMP (TGFβ family) receptors. Enrichment identified enrichment for a gene set matching genes differentially expressed in BMP4-treated cells (SILAC-Phosphoproteomic Database; $P=5.4 \times 10^{-5}$, $P_{adj}=0.003$, Z-score = -1.74) from the 417 but not the 988 genes (S5 Table). Another difference was enrichment of the

“CD71+Early Erythroid” (Human Gene Atlas) gene set in the 417 genes, while the 988 genes were enriched for gene sets (S5 Table Human Gene Atlas database) associated with B lymphoblasts, CD105⁺ endothelial, CD33⁺ myeloid, and CD14⁺ monocytes but not erythroid cells. Overall these analyses of discordant gene sets between experiments 1 and 2 are consistent with our hypothesis that patients treated with a single dose of liposomal amphotericin B (experiment 2) were at a more advanced stage of cure at day 30 post treatment than patients treated with multi-dose non-liposomal amphotericin B (experiment 1).

Discussion

In this study we have analysed whole blood transcriptomic data to further understand the pathogenesis of VL. One original goal of the study was to identify transcriptomic signatures that might differentiate asymptomatic infections from uninfected controls. In our attempt to achieve this we compared both modified quantiferon positive asymptomatic individuals and high antibody positive asymptomatic individuals with healthy endemic controls who were negative for these assays. In the event, we did not find signatures that would be diagnostic for either of these asymptomatic groups compared to negative controls. This was despite longitudinal epidemiological evidence from our study area showing that high antibody individuals are the group at most risk of progressing to clinical VL [6]. Our attention therefore focussed on understanding disease pathogenesis by comparing whole blood transcriptomes from active cases with all healthy controls, and in examining differences in the transcriptome following different regimens of drug treatment. In these comparisons 6 major themes emerged: (i) expression of genes and enrichment of gene sets associated with erythrocyte function in active cases; (ii) strong evidence for enrichment of gene sets involved in cell cycle in comparing active cases with healthy controls (or with more effective cure in

experiment 2); (iii) identification of *IFNG* encoding interferon- γ as the major hub gene in concordant gene expression patterns across experiments comparing active cases with healthy controls or with treated cases; (iv) enrichment for interleukin signalling (IL-1/3/4/6/7/8) and a prominent role for CXCL10/9/11 and chemokine signalling pathways in the comparison of active cases with treated cases; (v) the novel identification of AHR signalling as a significant IPA canonical pathway identified from concordant gene expression patterns across experiments comparing active cases with healthy controls or with treated cases; and (vi) global expression profiling support for more effective cure at day 30 post-treatment with a single dose of liposomal encapsulated amphotericin B compared to multi-dose treatment over 30 days.

Interesting in our analysis of top differentially expressed genes and enriched gene sets/pathways between active cases and healthy controls was the predominance of gene sets associated with erythroid cells and function. A recent systematic review [36] found that anaemia has an overall prevalence higher than 90% in VL. Pathogenesis of anaemia based on clinical observations included the presence of anti-erythrocyte antibodies, dysfunction in erythropoiesis, and hemophagocytosis in spleen or bone marrow. Of these, the authors of this review conclude that hemophagocytosis is the most likely cause [36]. The results of our study indicate differential regulation of gene sets associated with abnormal erythrocyte morphology, erythropoiesis, erythrocyte physiology, erythrocyte osmotic lysis, along with decreased haematocrit, spherocytosis and reticulocytosis. The gene sets defining these erythrocyte phenotypes therefore suggest mechanisms other than just hemophagocytosis and could provide important signatures to monitor clinical cure. This is especially relevant given our observation that erythroid related genes were present amongst the discordant genes that were differentially expressed between active cases and cases treated with multi-dose

amphotericin B (experiment 1) in which the degree of clinical cure was not as progressed for the same period of treatment with a single dose of liposomal amphotericin B (experiment 2).

A common feature of both the comparison of active cases with healthy controls, and of active cases with treated cases, was the identification of *IFNG* encoding interferon- γ as the major hub gene. This was not itself surprising since interferon- γ plays a key role in activating macrophages to kill *L. donovani* parasites [37]. Studies across the leishmaniasis have generally supported the notion that type 1 immune responses and the production of interferon- γ are vital for macrophage activation and parasite elimination [38-40]. It was interesting in our study that transcript levels for *IFNG* were higher in active cases than treated cases, where enhanced interferon- γ responses might have been expected to accompany drug cure. Nonetheless, it concurs with our observations that CD4⁺ T cells in whole blood from active VL patients and treated patients secrete high levels of interferon- γ following stimulation with crude *Leishmania* antigen [41, 42], the difference being that only active VL cases secreted IL-10 concurrently with interferon- γ [42]. The higher transcript abundance for *IFNG* in active compared to treated cases in our study suggests return to baseline with treatment in the latter.

Accompanying the central role of *IFNG* as a hub gene when comparing active cases with treated cases was evidence for perturbation of multiple cytokines, including IFNG, IFNA, IL-1, IL-6, and TNF, all of which were supported by differentially expressed gene signatures that generally included CXCL10/11/9 and STAT1. This CXCL10/11/9 chemokine gene expression signature also accounted for the identification of “pathogenesis of multiple sclerosis” [43] as the top disease-related canonical pathway identified using IPA, consistent with a proinflammatory response contributing to disease pathology in active VL. “Pathogenesis of multiple sclerosis” was also identified as a top canonical pathway in spleen tissue and splenic macrophages from *L. donovani* infected hamsters [44], a study in which the

authors also noted high interferon- γ expression that was ineffective in directing macrophage activation and parasite killing. STAT1 is a transcription factor activated by ligation of interferon- γ receptors. CXCL10/11/9 are all induced by interferon- γ , all bind to CXCR3, and between them have multiple roles as chemoattractants for monocytes and macrophages, T cells, NK cells, and dendritic cells, and in promoting T cell adhesion. CXCL10 and CXCL9 were also identified as the most highly “induced” genes in comparing lesion transcript profiles with normal skin of patients with American cutaneous leishmaniasis, consistent with their roles in inflammatory cell recruitment [45]. Cxcl9, Gbp1 (encoding the interferon- γ -induced guanylate binding protein *GBP1* identified here as one of the top 10 induced genes when comparing active *versus* treated cases), and Ifng were also identified as part of a common signature of 26 genes upregulated in blood, spleen and liver throughout the course of experimental infection with *L. donovani* in susceptible BALB/c mice, with Cxcl9 and Gbp1 reported as hub genes from a STRING analysis [46].

Given the many studies that have identified the importance of regulatory IL-10 in VL pathogenesis [42, 47-50], it was of some interest in our study that *IL10* was not identified as a top differentially expressed gene or as a significantly enriched signalling pathway in either comparison of active cases with healthy controls, or of active cases with treated cases. Nor did we observed perturbation of IL10R as has been reported in experimental transcriptional profiling studies of VL [46]. Indeed, downregulated expression of the type 2 cytokine gene *IL4* was the strongest response associated with effective cure in liposome-encapsulated amphotericin B treated cases, in line with previous studies showing that IL-4 levels were two-fold higher in VL patients who had failed treatment compared to previously untreated patients, whereas IL-10 levels were comparable in both [49].

One novel observation of our study was identification of AHR signalling as the top canonical pathway when comparing transcriptomes between active cases and healthy controls

or treated cases. Through crosstalk between signalling pathways, AHR ligands have been shown to significantly induce IL-10 secretion and inhibit IL-1 β and IL-6 production in dendritic cells, and to promote IL-10 production and suppress IL-17 expression in CD4⁺ T cells [30-32]. IL-17 is a potent activator of neutrophils, both through lineage expansion and through their recruitment by regulating chemokine expression. While IL-17 perturbation was not identified in our whole blood transcriptional profiles associated with human VL, evidence from murine models [51] demonstrate a strong role for IL-17 and neutrophils in parasite clearance from liver and spleen. Duthie and coworkers [50] have shown that both IL-10 and IL-17 cytokines are elevated in the serum of active VL patients, reverting to baseline levels with standard antimonial treatments. AHR activation has also been shown to inhibit inflammation through upregulation of IL-22 [52], another cytokine that has been shown to be significantly higher in *Leishmania* antigen stimulated peripheral blood mononuclear cells from active VL cases compared to treated cases [53]. AHR activation during VL may underpin the complex regulation of pro- and anti-inflammatory responses during disease pathogenesis and during response to therapy.

Of potential translational importance in our study was the additional identification of AHR signalling pathway at the top of the Ingenuity “Top Tox List” indicative of its role as a toxic pathology endpoint that could be amenable to therapeutic intervention. AHR locates to the cytoplasm in a stable complex that includes HSP90 observed as a differentially regulated gene in our comparison of active cases with healthy controls. Ligand binding occurs in the cytoplasm and triggers AHR translocation to the nucleus where it binds with ARNT to act as a transcription factor. Both AHR and ARNT were differentially expressed between active VL cases and controls in our study. The AHR response was first associated with xenobiotic induction of metabolizing enzymes, such as the induction of cytochrome P450, family 1, subfamily A, polypeptide 1 (Cyp1a1) following exposure to the polychlorinated dibenzo-p-

dioxin 2,3,7,8-Tetrachlorodibenzo-p-dioxin [54]. Multiple AHR ligands are known to induce a “gene battery” of metabolizing enzymes involved in oxidative stress response, cell cycle and apoptosis [55], amongst which are CYP1B1, ALD3B1, ALD3B and ALDH5A1 that were differentially expressed between active VL cases and healthy controls. Transcriptomic profiling of *M. tuberculosis* infected macrophages uncovered evidence for the generation of endogenous AHR ligands through induction of enzymes controlling tryptophan catabolism [56]. The generation of endogenous AHR ligands may likewise explain the role of AHR signalling in VL. For example, heme derivatives biliverdin and bilirubin have both been shown to act as endogenous ligands for AHR, as have arachidonic acid metabolites such as prostaglandins and leukotrienes [33, 34]. The former would be consistent with the strong perturbation of erythrocyte function between active VL cases and controls observed in our study. Importantly, addition of exogenous AHR ligands enhanced *M. tuberculosis* infection associated AHR transactivation to stimulate expression of AHR target genes, including IL-1 β and IL-23 which stimulate T cell subsets to produce IL-22. This suggests that administration of exogenous ligands could be used as a therapeutic intervention, especially in the knowledge that different exogenous AHR ligands can modulate either regulatory T cell or inflammatory T helper 17 cell differentiation in a ligand-specific fashion to suppress or exacerbate autoimmune disease [57].

Overall, our study has made some novel observations in relation to gene signatures that accompany both active VL disease and clinical cure in treated cases that could provide translatable targets for the development of novel or drug repurposed therapeutic interventions. Furthermore, by studying in more detail the discordant gene patterns that accompanied treatment with single dose liposome encapsulated amphotericin B versus multi-dose non-liposomal amphotericin B we were able to define gene signatures that could be used to monitor progress towards clinical cure.

Supporting Information

S1 Table. Demographic and clinical information on study participants.

S2 Table. Gene set enrichment analysis for concordant DEGs for active cases versus healthy controls. The table shows Enrichr results for analysis of 391 genes concordant for differential expression across two experiments comparing active VL cases with healthy controls.

S3 Table. Gene set enrichment analysis for concordant DEGs for active cases versus treated cases. The table shows Enrichr results for analysis of 210 genes concordant for differential expression across two experiments comparing active VL cases with treated VL cases.

S4 Table. Gene set enrichment analysis for discordant genes. The table shows Enrichr results for analysis of 988 genes differentially expressed between active VL cases and treated VL cases in experiment 2 but not experiment 1.

S5 Table. Comparison of gene sets enriched for discordant genes. The table shows Enrichr results for 417 genes differentially expressed between active VL cases and treated VL cases in experiment 1 but not experiment 2, with additional gene sets (see also S4 Table) enriched between active VL cases and treated VL cases in experiment 2 but not experiment 1.

S1 Fig. Schematic representation of the Aryl Hydrocarbon Receptor (AHR) Signalling

in Experiment 1. The diagram includes all AHR-interacting pathways generated in IPA using data for concordant differentially expressed ($P_{\text{adj}} < 0.05$) genes across the two experiments. Molecules outlined in purple achieved fold-change > 2 . Genes in green have decreased expression in active cases compared to healthy controls, genes in red have increased expression. The more intense the colour the larger the fold change values. Expression values are based in experiment 1, representative of similar results obtained for concordant genes across the two experiments (see S2 Fig).

S2 Fig. Schematic representation of the Aryl Hydrocarbon Receptor (AHR) Signalling

in Experiment 2. The diagram includes all AHR-interacting pathways generated in IPA using data for concordant differentially expressed ($P_{\text{adj}} < 0.05$) genes across the two experiments. Molecules outlined in purple achieved fold-change > 2 . Genes in green have decreased expression in active cases compared to healthy controls, genes in red have increased expression. The more intense the colour the larger the fold change values. Expression values are based in experiment 2, representative of similar results obtained for concordant genes across the two experiments (see S1 Fig).

Acknowledgements

We would like to thank the hospital staffs at Kala-azar Medical Research Centre, Muzaffarpur for their assistance in the collection of samples and all research scholars of Infectious Disease Research Laboratory, Banaras Hindu University for their kind help during the study.

Author Contributions

Conceptualization: Michaela Fakiola, Om-Prakash Singh, Jenefer M. Blackwell

Data curation: Michaela Fakiola, Om-Prakash Singh,

Formal analysis: Michaela Fakiola, Om-Prakash Singh, Genevieve Syn, Jenefer M. Blackwell

Funding acquisition: Shyam Sundar, Jenefer M. Blackwell

Investigation: Michaela Fakiola, Om-Prakash Singh, Genevieve Syn, Toolika Singh, Bhawana Singh, Jaya Chakravarty

Project administration: Michaela Fakiola, Om-Prakash Singh, Jaya Chakravarty

Supervision: Shyam Sundar, Jenefer M. Blackwell

Writing – original draft: Jenefer M. Blackwell

Writing – reviewing and editing: Jenefer M. Blackwell, Michaela Fakiola, Om-Prakash Singh.

Approval of manuscript: All authors read and approved the manuscript.

References

1. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis worldwide and global estimates of its incidence. PLoS ONE. 2012;7(5):e35671. Epub 2012/06/14. doi: 10.1371/journal.pone.0035671 PONE-D-11-24894 [pii]. PMID: 22693548; PMCID: PMC3365071.
2. Sundar S, Singh A. Recent developments and future prospects in the treatment of visceral leishmaniasis. Ther Adv Infect Dis. 2016;3(3-4):98-109. doi: 10.1177/2049936116646063. PMID: 27536354; PMCID: PMC4971590.
3. Croft SL, Sundar S, Fairlamb AH. Drug resistance in leishmaniasis. Clin Microbiol Rev. 2006;19(1):111-26. Epub 2006/01/19. doi: 10.1128/CMR.19.1.111-126.2006. PMID: 16418526.
4. Singh OP, Hasker E, Boelaert M, Sundar S. Elimination of visceral leishmaniasis on the Indian subcontinent. Lancet Infect Dis. 2016;16(12):e304-e9. doi: 10.1016/S1473-3099(16)30140-2. PMID: 27692643; PMCID: PMC5177523.
5. Hasker E, Kansal S, Malaviya P, Gidwani K, Picado A, Singh RP, et al. Latent infection with Leishmania donovani in highly endemic villages in Bihar, India. PLoS Negl Trop Dis. 2013;7(2):e2053. doi: 10.1371/journal.pntd.0002053. PMID: 23459501; PMCID: PMC3573094.
6. Hasker E, Malaviya P, Gidwani K, Picado A, Ostyn B, Kansal S, et al. Strong association between serological status and probability of progression to clinical visceral leishmaniasis in prospective cohort studies in India and Nepal. PLoS Negl Trop Dis. 2014;8(1):e2657. doi: 10.1371/journal.pntd.0002657. PMID: 24466361; PMCID: PMC3900391.
7. Gidwani K, Jones S, Kumar R, Boelaert M, Sundar S. Interferon-gamma release assay (modified QuantiFERON) as a potential marker of infection for Leishmania donovani, a proof of concept study. PLoS Negl Trop Dis. 2011;5(4):e1042. Epub 2011/04/29. doi: 10.1371/journal.pntd.0001042. PMID: 21526219; PMCID: PMC3079582.
8. Blankley S, Berry MP, Graham CM, Bloom CI, Lipman M, O'Garra A. The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. Philos Trans R Soc Lond B Biol Sci. 2014;369(1645):20130427. doi: 10.1098/rstb.2013.0427. PMID: 24821914; PMCID: PMC4024221.
9. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010;466(7309):973-7. Epub 2010/08/21. doi: nature09247 [pii] 10.1038/nature09247. PMID: 20725040.
10. Hu X, Yu J, Crosby SD, Storch GA. Gene expression profiles in febrile children with defined viral and bacterial infection. Proc Natl Acad Sci U S A. 2013;110(31):12792-7. doi: 10.1073/pnas.1302968110. PMID: 23858444; PMCID: PMC3732941.
11. Koh GC, Schreiber MF, Bautista R, Maude RR, Dunachie S, Limmathurotsakul D, et al. Host responses to melioidosis and tuberculosis are both dominated by interferon-mediated signaling. PLoS ONE. 2013;8(1):e54961. doi: 10.1371/journal.pone.0054961. PMID: 23383015; PMCID: PMC3558488.

12. Ramilo O, Allman W, Chung W, Mejias A, Ardura M, Glaser C, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood*. 2007;109(5):2066-77. doi: 10.1182/blood-2006-02-002477. PMID: 17105821; PMCID: PMC1801073.
13. Zhang ZN, Xu JJ, Fu YJ, Liu J, Jiang YJ, Cui HL, et al. Transcriptomic analysis of peripheral blood mononuclear cells in rapid progressors in early HIV infection identifies a signature closely correlated with disease progression. *Clin Chem*. 2013;59(8):1175-86. doi: 10.1373/clinchem.2012.197335. PMID: 23592504.
14. Ubol S, Masrinoul P, Chaijaruwanich J, Kalayanarooj S, Charoensirisuthikul T, Kasisith J. Differences in global gene expression in peripheral blood mononuclear cells indicate a significant role of the innate responses in progression of dengue fever but not dengue hemorrhagic fever. *J Infect Dis*. 2008;197(10):1459-67. doi: 10.1086/587699. PMID: 18444802.
15. Nascimento EJ, Braga-Neto U, Calzavara-Silva CE, Gomes AL, Abath FG, Brito CA, et al. Gene expression profiling during early acute febrile stage of dengue infection can predict the disease outcome. *PLoS ONE*. 2009;4(11):e7892. doi: 10.1371/journal.pone.0007892. PMID: 19936257; PMCID: PMC2775946.
16. Blankley S, Graham CM, Turner J, Berry MP, Bloom CI, Xu Z, et al. The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. *PLoS ONE*. 2016;11(10):e0162220. doi: 10.1371/journal.pone.0162220. PMID: 27706152; PMCID: PMC5051928.
17. Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS, Wang Y, et al. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS ONE*. 2013;8(8):e70630. doi: 10.1371/journal.pone.0070630. PMID: 23940611; PMCID: PMC3734176.
18. Cliff JM, Lee JS, Constantinou N, Cho JE, Clark TG, Ronacher K, et al. Distinct phases of blood gene expression pattern through tuberculosis treatment reflect modulation of the humoral immune response. *J Infect Dis*. 2013;207(1):18-29. doi: 10.1093/infdis/jis499. PMID: 22872737.
19. Bloom CI, Graham CM, Berry MP, Wilkinson KA, Oni T, Rozakeas F, et al. Detectable changes in the blood transcriptome are present after two weeks of antituberculosis therapy. *PLoS ONE*. 2012;7(10):e46191. doi: 10.1371/journal.pone.0046191. PMID: 23056259; PMCID: PMC3462772.
20. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* (Oxford, England). 2008;24(13):1547-8. Epub 2008/05/10. doi: 10.1093/bioinformatics/btn224. PMID: 18467348.
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. Epub 2015/01/22. doi: 10.1093/nar/gkv007 gkv007 [pii]. PMID: 25605792; PMCID: PMC4402510.
22. Barbosa-Morais NL, Dunning MJ, Samarajiwa SA, Darot JF, Ritchie ME, Lynch AG, et al. A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res*. 2010;38(3):e17. Epub 2009/11/20. doi: 10.1093/nar/gkp942. PMID: 19923232; PMCID: PMC2817484.

23. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
24. Kolde R. pheatmap: Pretty Heat Maps. R package version 1.0.12. 2019. Available from: <https://CRAN.R-project.org/package=pheatmap>.
25. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. J Roy Stat Soc Series B (Methodological). 1995;57(1):289-300.
26. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7. Epub 2016/05/05. doi: 10.1093/nar/gkw377 gkw377 [pii]. PMID: 27141961; PMCID: PMC4987924.
27. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128. Epub 2013/04/17. doi: 10.1186/1471-2105-14-128 1471-2105-14-128 [pii]. PMID: 23586463; PMCID: PMC3637064.
28. Sundar S, Singh A, Rai M, Chakravarty J. Single-dose indigenous liposomal amphotericin B in the treatment of Indian visceral leishmaniasis: a phase 2 study. Am J Trop Med Hyg. 2015;92(3):513-7. Epub 2014/12/17. doi: 10.4269/ajtmh.14-0259. PMID: 25510715; PMCID: PMC4350540.
29. Mondal D, Alvar J, Hasnain MG, Hossain MS, Ghosh D, Huda MM, et al. Efficacy and safety of single-dose liposomal amphotericin B for visceral leishmaniasis in a rural public hospital in Bangladesh: a feasibility study. The Lancet Global health. 2014;2(1):e51-7. doi: 10.1016/S2214-109X(13)70118-9. PMID: 25104636.
30. Wei P, Hu GH, Kang HY, Yao HB, Kou W, Liu H, et al. Increased aryl hydrocarbon receptor expression in patients with allergic rhinitis. QJM : monthly journal of the Association of Physicians. 2014;107(2):107-13. Epub 2013/09/21. doi: 10.1093/qjmed/hct188. PMID: 24049053.
31. Wei P, Hu GH, Kang HY, Yao HB, Kou W, Liu H, et al. An aryl hydrocarbon receptor ligand acts on dendritic cells and T cells to suppress the Th17 response in allergic rhinitis patients. Lab Invest. 2014;94(5):528-35. Epub 2014/02/12. doi: 10.1038/labinvest.2014.8. PMID: 24514067.
32. Wei P, Hu GH, Kang HY, Yao HB, Kou W, Zhang C, et al. Role of the aryl hydrocarbon receptor in the pathogenesis of chronic rhinosinusitis with nasal polyps. Inflammation. 2014;37(2):387-95. Epub 2013/10/05. doi: 10.1007/s10753-013-9751-7. PMID: 24092408.
33. Nguyen LP, Bradfield CA. The search for endogenous activators of the aryl hydrocarbon receptor. Chem Res Toxicol. 2008;21(1):102-16. Epub 2007/12/14. doi: 10.1021/tx7001965. PMID: 18076143; PMCID: PMC2572005.
34. Esser C, Rannug A. The aryl hydrocarbon receptor in barrier organ physiology, immunology, and toxicology. Pharmacol Rev. 2015;67(2):259-79. Epub 2015/02/07. doi: 10.1124/pr.114.009001. PMID: 25657351.
35. Lopez-Granados E, Keenan JE, Kinney MC, Leo H, Jain N, Ma CA, et al. A novel mutation in NFKBIA/IKBA results in a degradation-resistant N-truncated protein and is associated with ectodermal dysplasia with immunodeficiency. Hum Mutat.

- 2008;29(6):861-8. Epub 2008/04/17. doi: 10.1002/humu.20740. PMID: 18412279; PMCID: PMC3179847.
36. Goto Y, Cheng J, Omachi S, Morimoto A. Prevalence, severity, and pathogenesis of anemia in visceral leishmaniasis. *Parasitology research*. 2017;116(2):457-64. Epub 2016/11/09. doi: 10.1007/s00436-016-5313-x. PMID: 27822583.
37. Murray HW, Rubin BY, Rothermel CD. Killing of intracellular *Leishmania donovani* by lymphokine-stimulated human mononuclear phagocytes. Evidence that interferon-gamma is the activating lymphokine. *J Clin Invest*. 1983;72(4):1506-10. PMID: 6415111.
38. Scott P, Pearce E, Cheever AW, Coffman RL, Sher A. Role of cytokines and CD4+ T-cell subsets in the regulation of parasite immunity and disease. *Immunol Rev*. 1989;112:162-82.
39. Reed SG, Scott P. T-cell and cytokine responses in leishmaniasis. *Curr Opin Immunol*. 1993;5:524-31.
40. Locksley RM, Scott P. Helper T-cell subsets in mouse leishmaniasis induction, expansion and effector function. *Immunol Today*. 1991;12:A58-A61.
41. Kumar R, Singh N, Gautam S, Singh OP, Gidwani K, Rai M, et al. Leishmania specific CD4 T cells release IFN-gamma that limits parasite replication in patients with visceral leishmaniasis. *PLoS Negl Trop Dis*. 2014;8(10):e3198. doi: 10.1371/journal.pntd.0003198. PMID: 25275531; PMCID: PMC4183461.
42. Singh OP, Gidwani K, Kumar R, Nylen S, Jones SL, Boelaert M, et al. Reassessment of immune correlates in human visceral leishmaniasis as defined by cytokine release in whole blood. *Clin Vaccine Immunol*. 2012;19(6):961-6. Epub 2012/04/28. doi: CVI.00143-12 [pii] 10.1128/CVI.00143-12. PMID: 22539471; PMCID: PMC3370446.
43. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011;476(7359):214-9. Epub 2011/08/13. doi: nature10251 [pii] 10.1038/nature10251. PMID: 21833088; PMCID: PMC3182531.
44. Kong F, Saldarriaga OA, Spratt H, Osorio EY, Travi BL, Luxon BA, et al. Transcriptional Profiling in Experimental Visceral Leishmaniasis Reveals a Broad Splenic Inflammatory Environment that Conditions Macrophages toward a Disease-Promoting Phenotype. *PLoS Pathog*. 2017;13(1):e1006165. Epub 2017/02/01. doi: 10.1371/journal.ppat.1006165. PMID: 28141856; PMCID: PMC5283737.
45. Novais FO, Carvalho LP, Passos S, Roos DS, Carvalho EM, Scott P, et al. Genomic profiling of human *Leishmania braziliensis* lesions identifies transcriptional modules associated with cutaneous immunopathology. *J Invest Dermatol*. 2015;135(1):94-101. doi: 10.1038/jid.2014.305. PMID: 25036052; PMCID: PMC4268311.
46. Ashwin H, Seifert K, Forrester S, Brown N, MacDonald S, James S, et al. Tissue and host species-specific transcriptional changes in models of experimental visceral leishmaniasis. *Wellcome Open Res*. 2018;3:135. Epub 2018/12/14. doi: 10.12688/wellcomeopenres.14867.1. PMID: 30542664; PMCID: PMC6248268.
47. Nylen S, Maurya R, Eidsmo L, Manandhar KD, Sundar S, Sacks D. Splenic accumulation of IL-10 mRNA in T cells distinct from CD4+CD25+ (Foxp3) regulatory T cells in human visceral leishmaniasis. *J Exp Med*. 2007;204(4):805-17. Epub

- 2007/03/29. doi: jem.20061141 [pii] 10.1084/jem.20061141. PMID: 17389235; PMCID: PMC2118563.
48. Nylen S, Sacks D. Interleukin-10 and the pathogenesis of human visceral leishmaniasis. *Trends Immunol.* 2007;28(9):378-84. Epub 2007/08/11. doi: S1471-4906(07)00183-4 [pii] 10.1016/j.it.2007.07.004. PMID: 17689290.
49. Sundar S, Reed SG, Sharma S, Mehrotra A, Murray HW. Circulating T helper 1 (Th1) cell- and Th2 cell-associated cytokines in Indian patients with visceral leishmaniasis. *Am J Trop Med Hyg.* 1997;56(5):522-5.
50. Duthie MS, Guderian J, Vallur A, Bhatia A, Lima dos Santos P, Vieira de Melo E, et al. Alteration of the serum biomarker profiles of visceral leishmaniasis during treatment. *Eur J Clin Microbiol Infect Dis.* 2014;33(4):639-49. Epub 2013/11/01. doi: 10.1007/s10096-013-1999-1. PMID: 24173820; PMCID: PMC3955411.
51. Goncalves-de-Albuquerque SDC, Pessoa ESR, Trajano-Silva LAM, de Goes TC, de Moraes RCS, da COCN, et al. The Equivocal Role of Th17 Cells and Neutrophils on Immunopathogenesis of Leishmaniasis. *Front Immunol.* 2017;8:1437. Epub 2017/11/23. doi: 10.3389/fimmu.2017.01437. PMID: 29163510; PMCID: PMC5670345.
52. Monteleone I, Rizzo A, Sarra M, Sica G, Sileri P, Biancone L, et al. Aryl hydrocarbon receptor-induced signals up-regulate IL-22 production and inhibit inflammation in the gastrointestinal tract. *Gastroenterology.* 2011;141(1):237-48, 48 e1. Epub 2011/05/24. doi: 10.1053/j.gastro.2011.04.007. PMID: 21600206.
53. Nateghi Rostami M, Seyyedani Jasbi E, Khamesipour A, Mohammadi AM. Tumour Necrosis Factor-alpha (TNF-alpha) and its soluble receptor type 1 (sTNFR I) in human active and healed leishmaniasis. *Parasite Immunol.* 2016;38(4):255-60. Epub 2016/01/28. doi: 10.1111/pim.12305. PMID: 26813918.
54. Israel DI, Whitlock JP, Jr. Regulation of cytochrome P1-450 gene transcription by 2,3,7,8-tetrachlorodibenzo-p-dioxin in wild type and variant mouse hepatoma cells. *J Biol Chem.* 1984;259(9):5400-2. Epub 1984/05/10. PMID: 6715350.
55. Nebert DW, Roe AL, Dieter MZ, Solis WA, Yang Y, Dalton TP. Role of the aromatic hydrocarbon receptor and [Ah] gene battery in the oxidative stress response, cell cycle control, and apoptosis. *Biochem Pharmacol.* 2000;59(1):65-85. Epub 1999/12/22. PMID: 10605936.
56. Memari B, Bouttier M, Dimitrov V, Ouellette M, Behr MA, Fritz JH, et al. Engagement of the Aryl Hydrocarbon Receptor in Mycobacterium tuberculosis-Infected Macrophages Has Pleiotropic Effects on Innate Immune Signaling. *J Immunol.* 2015;195(9):4479-91. Epub 2015/09/30. doi: 10.4049/jimmunol.1501141. PMID: 26416282.
57. Quintana FJ, Basso AS, Iglesias AH, Korn T, Farez MF, Bettelli E, et al. Control of T(reg) and T(H)17 cell differentiation by the aryl hydrocarbon receptor. *Nature.* 2008;453(7191):65-71. Epub 2008/03/26. doi: 10.1038/nature06880. PMID: 18362915.

Table 1. Summary of numbers of between group DEGs. DEGs at adjusted P -value ≤ 0.05 (top panel) and fold-change of expression ≥ 2 (bottom panel) for the comparison of the three main phenotype groups.

adj.$P \leq 0.05$	Experiment 1	Experiment 2	Concordant
Case <i>vs</i> control	4596	4651	2584
Treated <i>vs</i> control	1132	126	37
Case <i>vs</i> treated	654	1317	221

adj.$P \leq 0.05$ and fold-change ≥ 2	Experiment 1	Experiment 2	Concordant
Case <i>vs</i> control	683	783	439
Treated <i>vs</i> control	120	27	8
Case <i>vs</i> treated	94	337	42

Table 2. List of pathways identified by IPA Canonical Pathway Analysis. The table shows results for 439 probes representing 391 genes concordant for differential expression (adjusted P-value <0.05; >2-fold change) when comparing active VL cases with healthy controls across experiments 1 and 2, and for 221 probes representing 210 genes concordant for differential expression (adjusted P-value <0.05) when comparing active VL cases with treated VL cases across the two experiments.

Ingenuity Canonical Pathways	P_{nominal}	P_{adj}	Genes
Active cases <i>versus</i> controls (391 top concordant genes; $P_{\text{adj}} < 0.05$; fold-change > 2)			
Estrogen-mediated S-phase Entry	6.46×10^{-5}	0.019	↑CCNA2, ↑CCNE2, ↑TFDP1, ↑CDK1, ↑CDC25A
Mitotic Roles of Polo-Like Kinase	1.20×10^{-4}	0.019	↑PLK4, ↑CDC20, ↑PTTG1, ↑CCNB2, ↑CDK1, ↑KIF11, ↑CDC25A
Aryl Hydrocarbon Receptor Signaling	1.51×10^{-4}	0.019	↑TGM2, ↑CCNA2, ↑CCNE2, ↑NFIX, ↑TFDP1, ↓RARA, ↑ALDH5A1, ↓CYP1B1, ↑HSPB1, ↑CHEK1
Heme Biosynthesis II	3.72×10^{-4}	0.035	↑FECH, ↑ALAS2, ↑HMBS
Th2 Pathway	1.12×10^{-3}	0.074*	↓CCR1, ↑IFNG, ↓CCR3, ↓PTGDR2, ↑CXCR6, ↓PIK3R6, ↑HLA-DQA1, ↓SPI1, ↑TIMD4
Th1 and Th2 Activation Pathway	1.32×10^{-3}	0.074*	↓CCR1, ↑IFNG, ↑SOCS1, ↓CCR3, ↓PTGDR2, ↑CXCR6, ↓PIK3R6, ↑HLA-DQA1, ↓SPI1, ↑TIMD4
Active cases <i>versus</i> treated cases (210 Top concordant genes; $P_{\text{adj}} < 0.05$)			
Pathogenesis of Multiple Sclerosis	9.33×10^{-5}	0.032	↑CXCL10, ↑CXCL11, ↑CXCL9
Acute Myeloid Leukemia Signaling	3.55×10^{-4}	0.049	↓CSF3R, ↑NRAS, ↓MAPK3, ↓RARA, ↓SPI1, ↓CSF1R
Oncostatin M Signaling	4.37×10^{-4}	0.049	↑NRAS, ↓MAPK3, ↓CHI3L1, ↑STAT1
Notch Signaling	6.03×10^{-4}	0.05	↓MAML3, ↓LFNG, ↓RBPJ, ↓NOTCH1
Thyroid Cancer Signaling	7.41×10^{-4}	0.05	↑CXCL10, ↑NRAS, ↓MAPK3, ↓RXRA
Aryl Hydrocarbon Receptor Signaling	3.09×10^{-3}	0.105**	↑CCNE2, ↑NCOA7, ↓MAPK3, ↓RARA, ↓ALDH3B1, ↓RXRA

*Did not achieve adjusted P-value <0.05 but supported *a priori* by previous knowledge of immune response to leishmaniasis. **Did not achieve adjusted P-value <0.05 but supports importance of the pathway identified by the comparison of active cases with healthy controls. Arrows indicated expression level in active cases relative to the relevant comparator.

Table 3. Comparison of Enrichr results for discordant gene sets. The table compares genes sets that were enriched in 417 genes differentially expressed ($P_{adj} \leq 0.05$) between active cases and treated cases in experiment 1 but not experiment 2 with those enriched in 988 genes differentially expressed ($P_{adj} \leq 0.05$) in experiment 2 but not in experiment 1. Only results where the Z score is < -1 or > 1 , and the enrichment $P_{adj} \leq 0.05$, are included. See methods for an explanation of the combined score. Full data provided in S4 and S5 Tables.

Database	Term	<i>P</i>	<i>P</i> _{adj}	Z-score	Combined Score	Overlap *
Gene sets identified experiment 1 but not experiment 2						
Human Gene Atlas	CD71+_EarlyErythroid	3.0E-04	0.016	-1.96	16.17	31/553
SILAC Phosphoproteomics	BMP4 vs control	5.4E-05	0.003	-1.74	17.13	27/561
PPI Hub Proteins	NFKBIA	1.17E-04	0.009	-1.81	16.35	11/134
	SMAD9	1.55E-03	0.047	-1.63	10.54	9/129
Gene sets identified experiment 2 but not experiment 1						
Reactome 2016	Cell Cycle, Mitotic**	3.5E-34	3.95E-31	-2.48	190.81	96/462
	Interleukin-3, 5; GM-CSF signalling	0.003	0.032	-1.71	10.15	24/261
	Interleukin-2 signalling	0.003	0.041	-1.66	9.36	23/252
Wiki Pathways 2016	Cell Cycle**	2.3E-11	3.89E-09	-1.87	45.76	25/103
	IL-1 Signalling Pathway	3.03E-04	0.007	-1.53	12.43	8/36
	IL-3 Signalling Pathway	5.83E-04	0.013	-1.78	13.25	9/49
	IL-7 Signalling Pathway	0.001	0.018	-1.36	9.30	8/43
	IL-4 Signalling Pathway	0.001	0.019	-1.50	10.10	9/54
	Apoptosis	0.002	0.028	-1.43	8.84	12/93
	Delta-Notch Signalling	0.002	0.028	-1.32	8.14	11/81
	IL-3 Signalling	0.003	0.036	-1.41	8.20	12/97
	IL-6 signalling	0.003	0.036	-1.37	7.99	12/97
	Chemokine signalling	0.003	0.038	-1.37	7.81	17/165
KEGG 2016	Cell cycle**	1.6E-09	4.25E-07	-1.73	35.13	25/124
	Chemokine signalling	0.002	0.050	-1.67	10.11	19/187
	Fc gamma R-mediated phagocytosis	0.002	0.050	-1.63	10.10	12/93
NCI_Nature 2016	Aurora B signalling**	1.4E-11	2.26E-09	-1.39	34.76	16/39
	IL8- and CXCR2-signaling	1.98E-04	0.005	-1.23	10.50	9/34
	IL4-mediated signalling	6.57E-04	0.012	-1.28	9.34	10/60
	CXCR4-signalling	0.004	0.047	-1.10	6.13	12/100
LINCS_L1000_ligand_perturbations_down	IL4	6.6E-13	6.36E-11	-1.81	50.84	34/163
	MSP	4.51E-04	0.022	-1.78	13.73	19/163
	TGFA	0.001	0.033	-1.48	10.16	17/148
Human Gene Atlas	721_B_lymphoblasts	6E-09	2.38E-07	-2.61	49.32	172/1543
	CD105+_Endothelial	1.8E-07	4.67E-06	-1.96	30.52	65/370
	CD33+_Myeloid	3.6E-05	5.78E-04	-2.14	21.90	76/679
	CD14+_Monocytes	3.7E-05	5.78E-04	-1.76	17.95	53/385
PPI Hub Proteins	CDK1**	3.2E-15	1.12E-12	-2.15	71.82	83/659

* Overlap indicates the number of genes in our dataset relative to the number of genes in the public domain dataset with which our data are being compared. **Only the top cell cycle-related pathways are shown; many more significant cell cycle-related pathways not shown.

Legends to figures

Figure 1. Principal Components Analysis (PCA) and hierarchical clustering of top 500

most variable probes. (A) PC1 by PC2 and (B) PC1 by PC3 in experiment 1, and (D) PC1 by PC2 and (E) PC1 by PC3 in experiment 2. Z-score transformed expression levels of the 500 most variable probes across all samples are represented as a heatmap for (C) experiment 1 and (F) experiment 2. Hierarchical clustering results based on Pearson's correlation are shown as dendrograms on the top and left side of the matrix. Columns represent individual samples and rows individual probes. Experimental groups are color coded on the upper part of the heatmap. Active (=case) and treated (=treated) cases, as well as asymptomatics (=Quantiferon or HighAb positive individuals) and endemic healthy controls (=EHC), are colour coded as per the keys provided.

Figure 2. Heatmaps for top differentially expressed genes between active cases and

healthy controls. (A) top 10 "induced" and (B) top 10 "repressed" genes for differential expression between active cases (N=10) and healthy controls (N=16) in experiment 1. (C) and (D) show heatmaps for the same genes in active cases (N=10) and healthy controls (N=25) using data from experiment 2. Columns represent individuals and rows represent individual genes, coloured to indicate expression levels based on post-QC normalised and log₂-transformed data as indicated by the legend to the left of each figure. LogFC = log₂ fold-change.

Figure 3. Gene network for concordant genes comparing active cases and healthy

controls. The network was generated in IPA for 254 (of 391) genes concordant across experiments 1 and 2 for differential expression (adjusted p-value ≤0.05; ≥2-fold change)

when comparing active cases and healthy controls. Genes in red have increased expression and genes in green have decreased expression when comparing active cases with healthy controls. The more intense the colour the larger the fold change values. Expression values are based in experiment 1, representative of similar results obtained for concordant genes across the two experiments.

Figure 4. Schematic representation of the core Aryl Hydrocarbon Receptor (AHR)

Signalling pathway. The pathway was generated in IPA using data for concordant differentially expressed ($P_{\text{adj}} < 0.05$) genes across the two experiments. Molecules outlined in purple achieved fold-change > 2 . Genes in green have decreased expression in active cases compared to healthy controls, genes in red have increased expression. The more intense the colour the larger the fold change values. Expression values are based in experiment 1, representative of similar results obtained for concordant genes across the two experiments.

Figure 5. Heatmaps for top differentially expressed genes between active cases and treated cases. (A) top 10 “induced” and (B) top 10 “repressed” genes for differential expression between active cases (N=10) and treated cases (N=10) in experiment 1. (C) and (D) show heatmaps for the same genes in active cases (N=11) and treated cases (N=12) using data from experiment 2. Columns represent individuals and rows represent individual genes, coloured to indicate expression levels based on post-QC normalised and log₂-transformed data as indicated by the legend to the left of each figure. LogFC = log₂ fold-change.

Figure 6. Gene network for concordant genes comparing active cases and treated cases.

The network was generated in IPA for 85 (of 210) genes concordant across experiments 1 and 2 for differential expression (adjusted p-value ≤ 0.05) when comparing active cases and

991 treated cases. Genes in red have increased expression and genes in green have decreased
 992 expression when comparing active cases with treated cases. The more intense the colour the
 993 larger the fold change values. Expression values are based in experiment 1, representative of
 994 similar results obtained for concordant genes across the two experiments.
 995

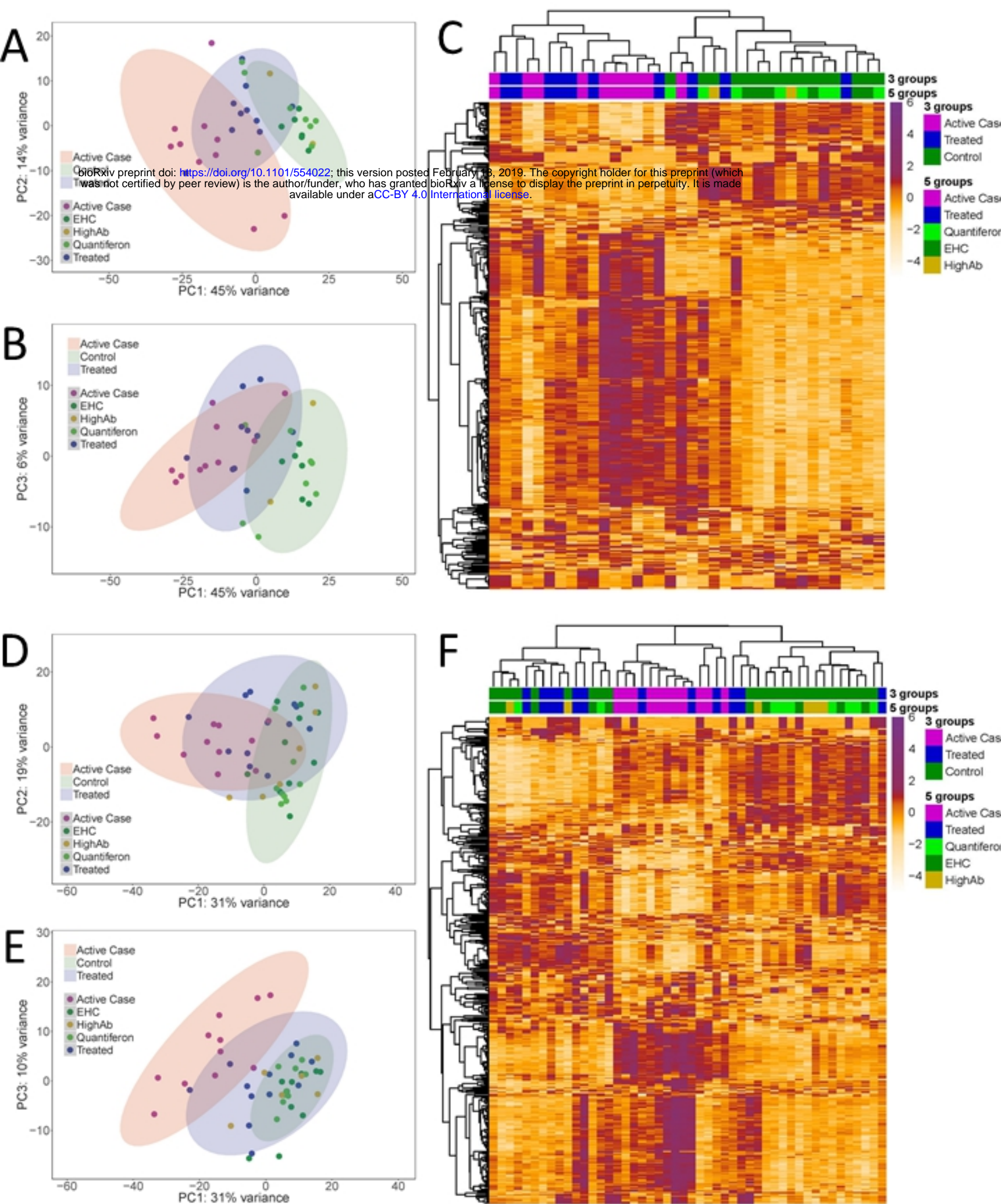


Figure 1

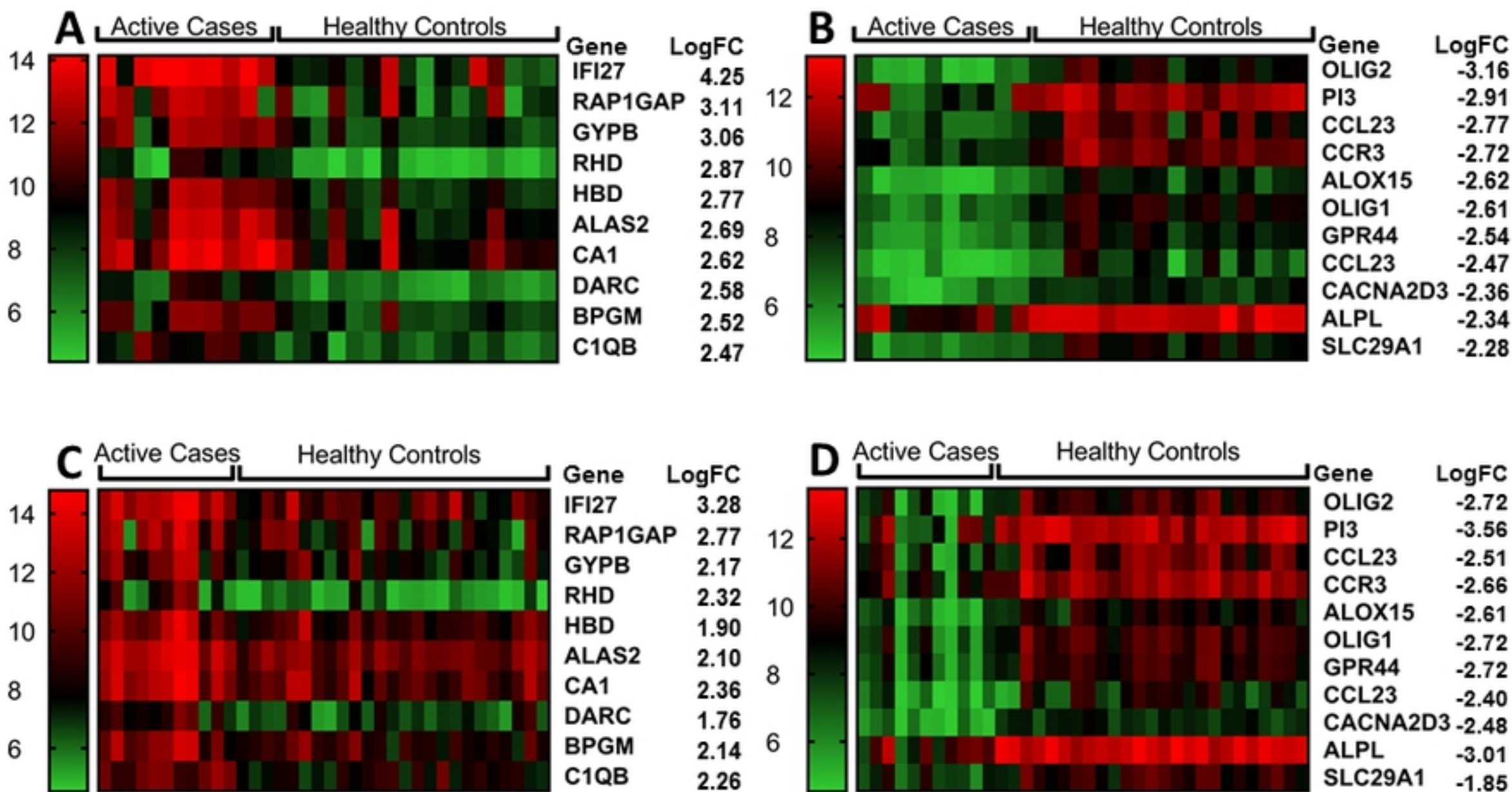


Figure 2

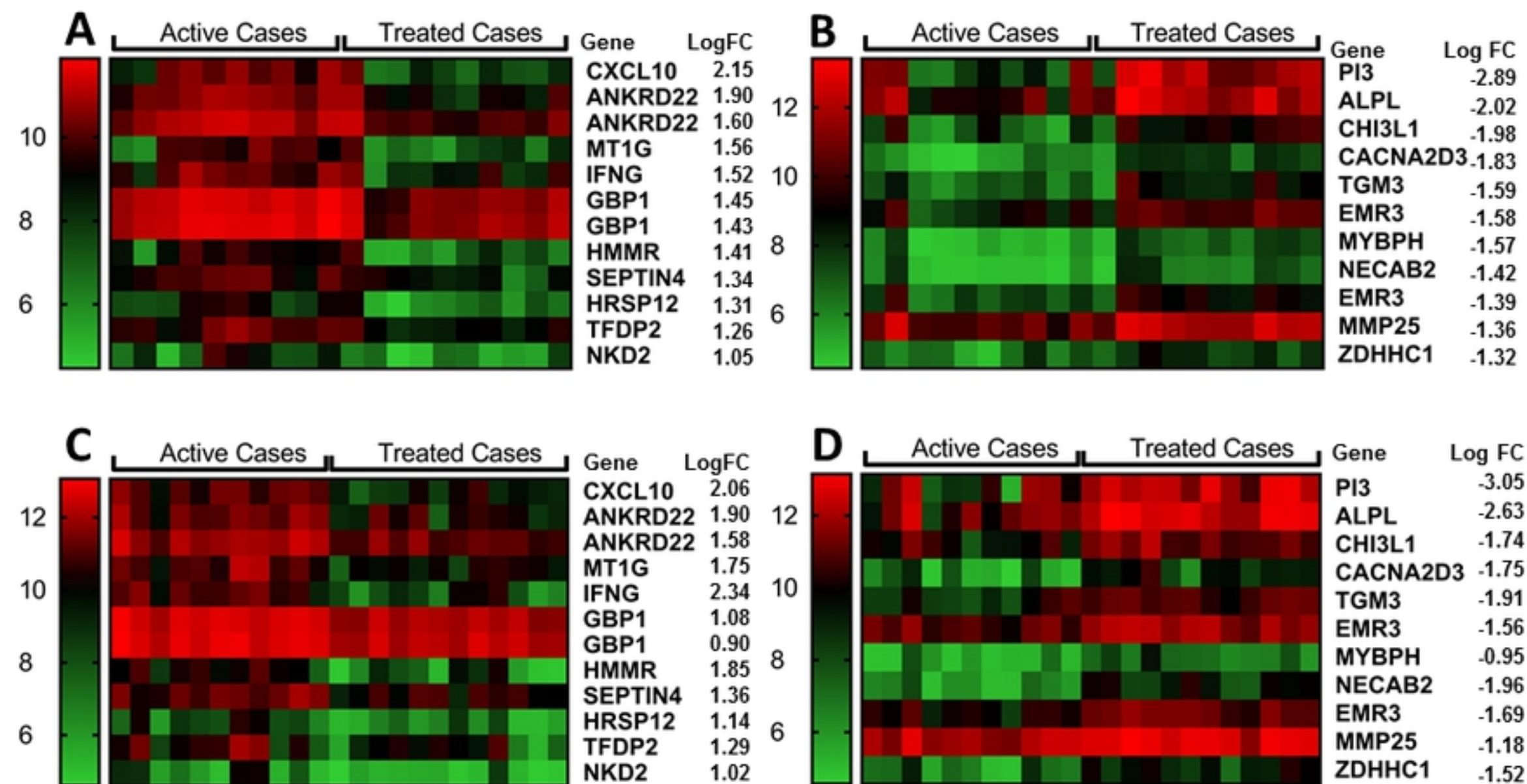


Figure 5

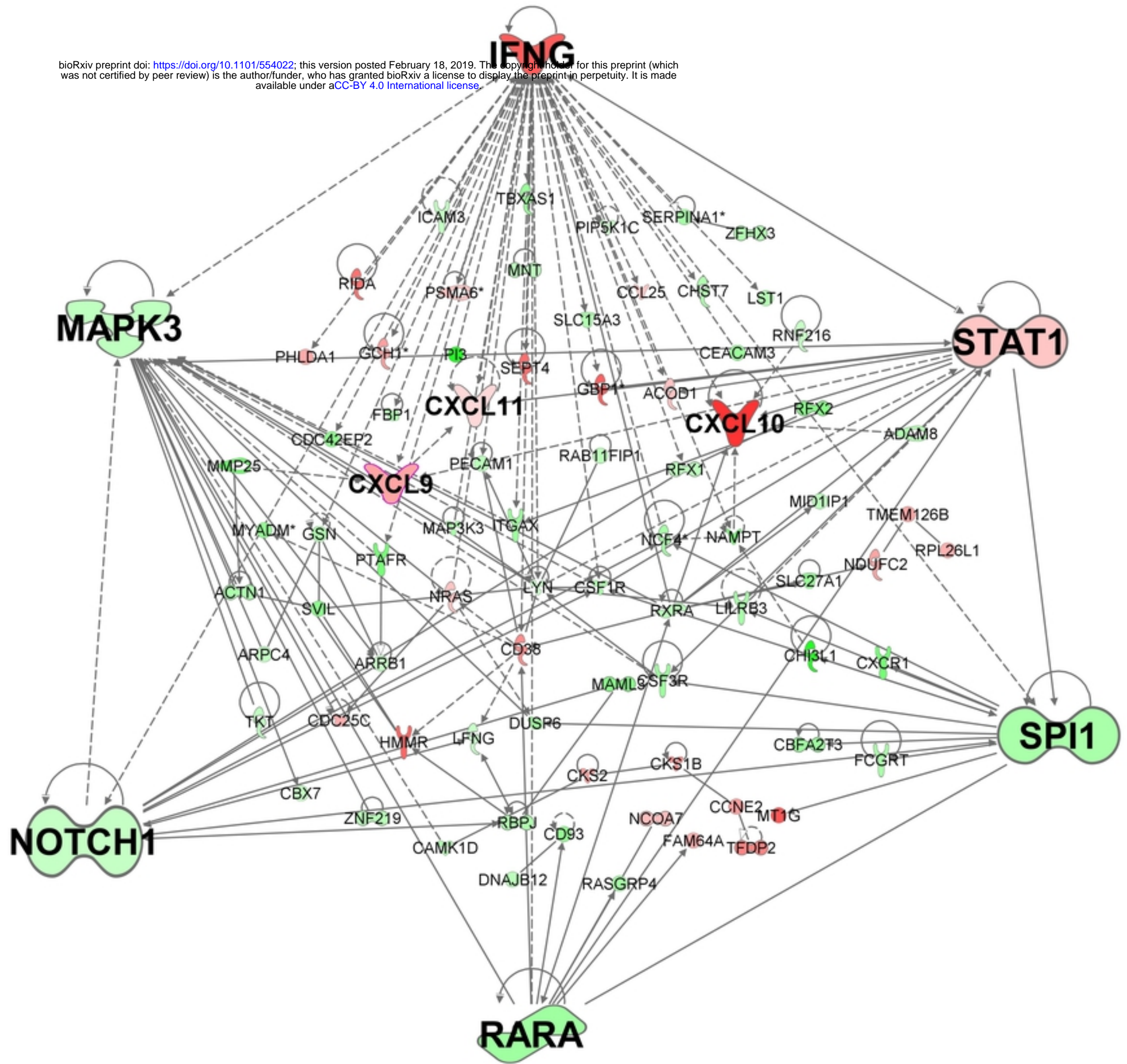


Figure 6

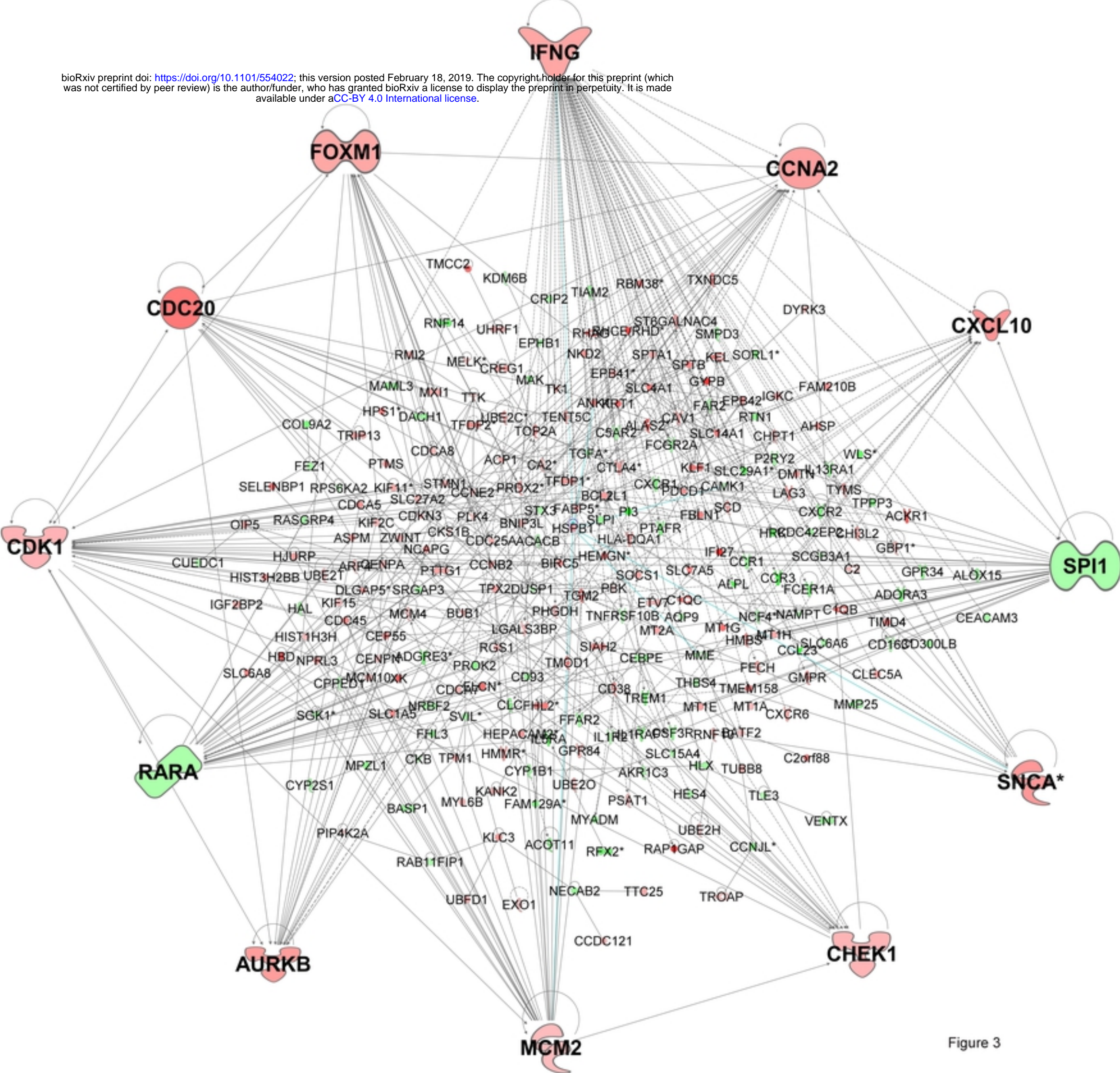


Figure 3

Figure 3

Nucleus