2 Comparison of monocyte-derived and iPSC-derived macrophages

Collaboration note

The work described in this chapter has been published as “Transcriptional profiling of macrophages derived from monocytes and iPS cells identifies a conserved response to LPS and novel alternative transcription” (Alasoo et al., 2015). I performed the iPSC-derived macrophage experiments and analysed the data. Fernando O. Martinez from the University of Oxford performed the monocyte-derived macrophage experiments. Subhankar Mukhopadhyay and Gordon Dougan were involved in designing and optimising the experiments and interpreting the results. RNA-seq library construction and sequencing was done by DNA Pipelines core facility at Sanger. I thank Kosuke Yusa and Mariya Chhatriwala for fruitful discussions on troubleshooting iPSC culture.

2.1 Introduction

Macrophages are key cells associated with innate immunity, pathogen containment and modulation of the immune response (Murray and Wynn, 2011; Wynn et al., 2013). Commonly used model systems for studying macrophage biology have included macrophage-like leukemic cell lines, primary macrophages derived from model organisms and primary human macrophages differentiated from blood monocytes. Although these cells have provided important insights into macrophage-associated biology, they have some limitations. Immortalised cell lines often have accumulated multiple genetic aberrations and can exhibit functional defects compared to primary cells such as impaired cytokine production upon inflammatory stimulation (Adati et al., 2009; Schildberger et al., 2013), while multiple functional differences exist between macrophages from different species (Schroder et al., 2012). Additionally, human monocyte derived macrophages (MDMs) can be difficult to obtain in sufficient numbers for repeated experimental assays and it is currently challenging to introduce targeted mutations into their genomes, limiting their utility in genetic studies. For example, introduction of foreign nucleic acid into the cytosol induces a robust antiviral response that may make it difficult to interpret experimental data (Muruve et al., 2008).

Recently, methods have been developed to differentiate macrophage-like cells from human induced pluripotent stem cells (IPSCs) that have the potential to complement current approaches and overcome some of their limitations (Karlsson et al., 2008; van Wilgenburg et al., 2013). This approach is scalable and large numbers of highly pure iPSC-derived macrophages (IPSDMs) can be routinely obtained from any human donor following establishment of an iPSC line. IPSDMs also share striking phenotypic and functional similarities with primary human macrophages (Karlsson et al., 2008; van Wilgenburg et al., 2013). Since human iPSCs are amenable to genetic manipulation, this approach can provide large numbers of genetically modified human macrophages (van Wilgenburg et al., 2013). Previous studies have successfully used IPSDMs to model rare monogenic defects that severely impact macrophage function (Jiang et al., 2012). However, it remains unclear how closely IPSDMs resemble primary human monocyte-derived macrophages (MDMs) at the transcriptome level and to what extent they can be used as an alternative model for functional assays.

Here, we provide an in-depth comparison of the global transcriptional profiles of naïve and lipopolysaccharide (LPS) stimulated IPSDMs with MDMs using RNA-seq. We found that their transcriptional profiles were broadly similar in both naïve and LPS-stimulated conditions. However, certain chemokine genes as well as genes involved in antigen presentation and tissue remodelling were differentially regulated between MDMs and IPSDMs. Additionally, we identified novel changes in alternative transcript usage following LPS stimulation suggesting that alternative transcription may represent an important component of the macrophage immune response.

2.2 Methods

2.2.1 Samples

Human blood for monocyte-derived macrophages was obtained from NHS Blood and Transplant, UK and all experiments were performed according to guidelines of the University of Oxford ethics review committee. All IPSDMs were differentiated from four iPSC lines: CRL1, S7RE, FSPS10C and FSPS11B. CRL1 iPSC line was originally derived from a commercially available human fibroblast cell line and has been described before (Vallier et al., 2009). S7RE iPSC line was derived as part of an earlier study from our lab (Rouhani et al., 2014). FSPS10C

and FSPS11B iPSC lines were derived as part of the Human Induced Pluripotent Stem Cell Initiative (Kilpinen et al., 2016). All iPSC work was carried out in accordance to UK research ethics committee approvals (REC No. 09/H306/73 & REC No. 09/H0304/77).

2.2.2 Cell culture and reagents

IPSCs were grown on Mitomycin C-inactivated mouse embryonic fibroblast (MEF) feeder cells in Advanced DMEM F12 (Gibco) supplemented with 20% KnockOut Serum Replacement (Gibco, cat no 10828-028), 2mM L-glutamine, 50 IU/ml penicillin, 50 IU/ml streptomycin and 50 μM 2-mercaptoethanol (Sigma M6250) on 10 cm tissue-culture treated dishes (Corning). The medium was supplemented with 4 ng/ml rhFGF basic (R&D) and changed daily (10 ml per dish). Prior to passage, the cells were detached from the dish with 1:1 solution of 1 mg/ml collagenase and 1mg/ml dispase (both Gibco). Human macrophage colony stimulating factor (M-CSF) producing cell line CRL-10154 was obtained from ATCC. The cells were grown in T150 tissue culture flasks containing 40 ml of medium (90% alpha minimum essential medium (Sigma), 10% FBS, 2mM L-glutamine, 50 IU/ml penicillin, 50 IU/ml streptomycin). On day 9 the supernatant was sterile-filtered and stored at -80°C.

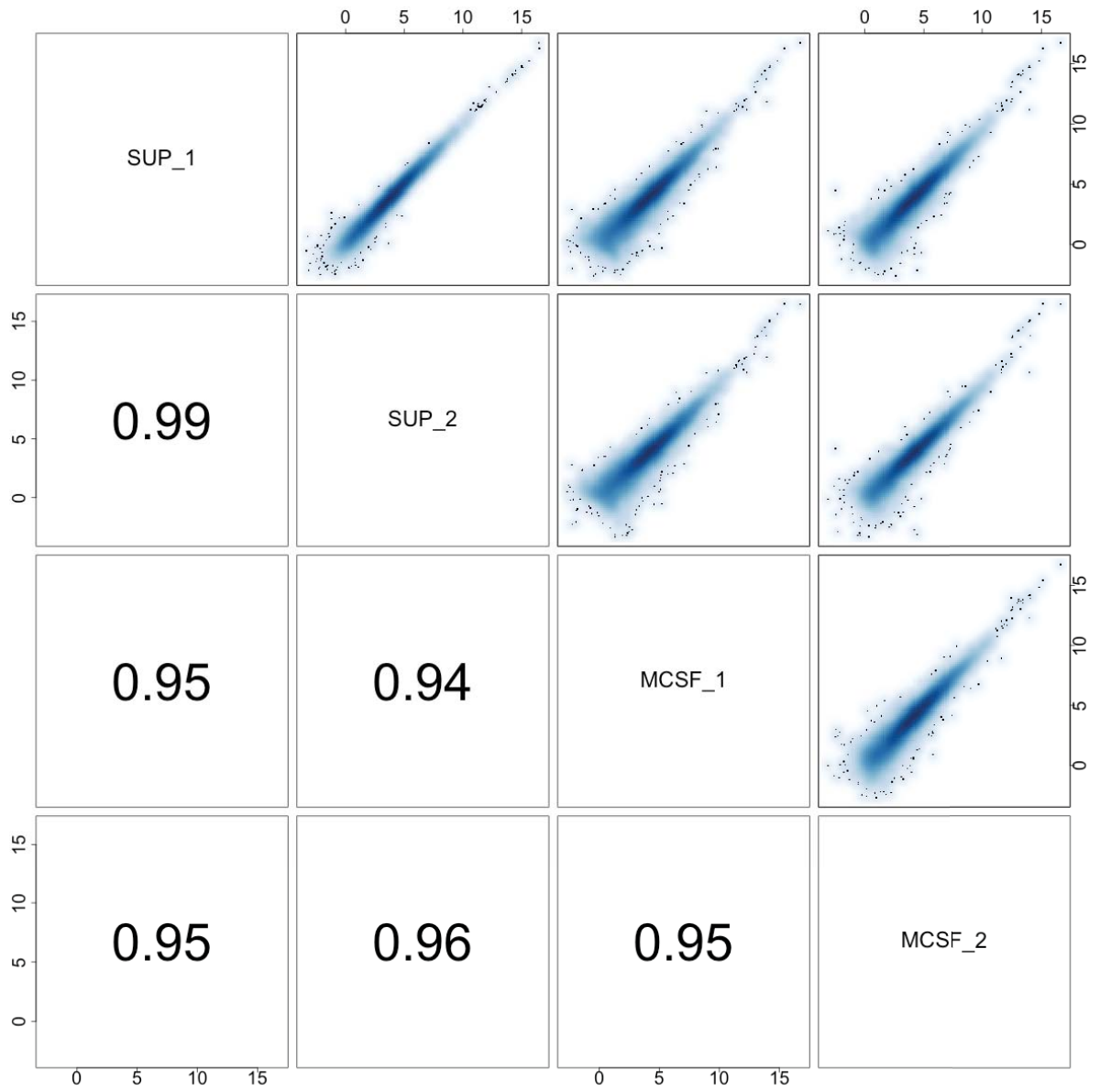

Figure 2.1. Biological reproducibility of IPSDM differentiation. Two biological replicates of FSPS10C-derived IPSDMs differentiated with either supernatant (SUP_1 and SUP_2) or recombinant M-CSF (MCSF_1 and MCSF_2). Above diagonal: pairwise scatterplots of expressed genes (transcripts per million (TPM) > 1) between all four samples. Below diagonal: pairwise Spearman's correlation of gene expression between all four samples.

IPSCs were differentiated into macrophages following a previously published protocol consisting of three steps: i) embryoid body (EB) formation, ii) production of myeloid progenitors from the EBs and iii) terminal differentiation of myeloid progenitors into mature macrophages (van

Wilgenburg et al., 2013). For EB formation, intact iPSC colonies were separated from MEFs using collagenase-dispase solution, transferred to 10 cm low-adherence bacteriological dishes (Sterilin) and cultured in 25 ml iPSC medium without rhFGF for 3 days. Mature EBs were resuspended in myeloid progenitor differentiation medium (90% X-VIVO 15 (Lonza), 10% FBS, 2mM L-glutamine, 50 IU/ml penicillin, 50 IU/ml streptomycin and 50 μM 2-mercaptoethanol (Sigma M6250), 50 ng/ml hM-CSF (R&D), 25 ng/ml hIL-3 (R&D)) and plated on 10 cm gelatinised tissue-culture treated dishes. Medium was changed every 4-7 days. After 3-4 weeks, floating progenitor cells were isolated from the adherent EBs, filtered using a 40 μm cell strainer (Falcon) and resuspended in macrophage differentiation medium (90 % RPMI 1640, 10% FBS, 50 IU/ml penicillin and 50 IU/ml streptomycin) supplemented with 20% supernatant from CRL-10154 cell line. Approximately 7×10^5 cells in 15 ml of media were plated on a 10 cm tissue-culture treated dish and cultured for 7 days until final differentiation. We observed that using supernatant instead of 100 ng/ml M-CSF as specified in the original protocol (van Wilgenburg et al., 2013) did not alter macrophage gene expression profile. The variation between cells differentiated with supernatant or M-CSF was comparable to the variation between two biological replicates of macrophages differentiated with M-CSF (Figure 2.1).

Human monocytes (90-95% purity) were obtained from healthy donor leukocyte cones (corresponding to 450 ml of total blood) by 2-step gradient centrifugation (Martinez, 2012; Martinez et al., 2006). The monocyte fraction in this type of preparation is on average 98% $CD14^+$, 13% $CD16^+$ by single staining. The isolated monocytes were cultured for 7 days in the same macrophage differentiation medium as IPSDMs. The same seeding density and tissue-culture treated plastic was used as for IPSDMs. Non-adherent contaminating cells were removed by vigorous washing before cell lysis at day 7.

On day 7 of macrophage differentiation, medium was replaced with either 10 ml of fresh macrophage medium (without M-CSF) or medium supplemented with 2.5 ng/ml LPS (E. coli). After 6 hours, cells were lifted from the plate using lidocaine solution (6 mg/ml lidocaine, PBS, 0.0002% EDTA), counted with haemocytometer (C-Chip) and lysed in 600 μl RLT buffer (Qiagen). All cells from a dish were used for lysis and subsequent RNA extraction.

2.2.3 Flow cytometry

Flow cytometry was used to characterise the IPSDM cell populations used in the experiments. Approximately 1×10^6 cells were resuspended in flow cytometry buffer (D-PBS, 2% BSA, 0.001%

EDTA) supplemented with Human TruStain FcX (Biolegend) and incubated for 45 minutes on ice to block the Fc receptors. Next, cells were washed once and resuspended in buffer containing one of the antibodies or isotype control. After 1 hour, cells were washed three times with flow cytometry buffer and immediately measured on BD LSRFortessa cell analyser. The following antibodies (BD) were used (cat no): CD14-Pacific Blue (558121), CD32-FITC (552883), CD163-PE (556018), CD4-PE (561844), CD206-APC (550889) and PE isotype control (555749). The data were analysed using FlowJo. The raw data are available on figshare (doi: 10.6084/m9.figshare.1119735).

2.2.4 RNA extraction and sequencing

RNA was extracted with RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. After extraction, the sample was incubated with Turbo DNase at 37°C for 30 minutes and subsequently re-purified using RNeasy clean-up protocol. Standard Illumina unstranded poly-A enriched libraries were prepared and then sequenced 5-plex on Illumina HiSeq 2500 generating 20-50 million 75bp paired-end reads per sample. RNA-seq data from six iPSC samples was taken from a previous study (Rouhani et al., 2014). Sample information together with the total number of aligned fragments are detailed in Table 2.1.

Table 2.1: General information about the RNA-seq samples. Library size column contains the total number of aligned fragments per sample.

Sample	Donor	Cell type	Treatment	Library size
S7_RE15	S7RE	IPSC	control	83280070
S7_RE11	S7RE	IPSC	control	72411619
S4_SF5	S4SF	IPSC	control	72167859
S4_SF3	S4SF	IPSC	control	72427265
S5_SF1	S5SF	IPSC	control	90998616
S5_SF3	S5SF	IPSC	control	83746320
CRL1_ctrl	CRL1	IPSDM	control	47052432
S7RE_ctrl	S7RE	IPSDM	control	25322078
FSPS10C_ctrl	FSPS10C	IPSDM	control	23443481
FSPS11B_ctrl	FSPS11B	IPSDM	control	19933949
CRL1_LPS	CRL1	IPSDM	LPS	33985920

S7RE_LPS	S7RE	IPSDM	LPS	24349911
FSPS10C_LPS	FSPS10C	IPSDM	LPS	24570506
FSPS11B_LPS	FSPS11B	IPSDM	LPS	24394255
B1_ctrl	B1	MDM	control	23381545
B4_ctrl	B4	MDM	control	47790764
B5_ctrl	B5	MDM	control	26056124
B2_ctrl	B2	MDM	control	20901894
B3_ctrl	B3	MDM	control	26059134
B1_LPS	B1	MDM	LPS	20748290
B4_LPS	B4	MDM	LPS	25538994
B5_LPS	B5	MDM	LPS	56227352
B2_LPS	B2	MDM	LPS	24456569
B3_LPS	B3	MDM	LPS	24075743

2.2.5 RNA-seq data analysis

Differential expression

Sequencing reads were aligned to GRCh37 reference genome with Ensembl 74 annotations using TopHat v2.0.8b (Kim et al., 2013). Reads overlapping gene annotations were counted using featureCounts (Liao et al., 2014) and DESeq2 (Love et al., 2014) was used to identify differentially expressed genes. Genes with FDR < 0.01 and fold-change > 2 were identified as differentially expressed. We used g:Profiler to perform Gene Ontology and pathway enrichment analysis (Reimand et al., 2011). For conditional enrichment analysis of the genes differentially regulated in LPS response we used all LPS-responsive genes as the background set. All analysis was performed on genes classified as expressed in at least one condition (TPM > 2) except where noted otherwise. The bedtools (Quinlan and Hall, 2010) suite was used to construct BigWig files with genome-wide read coverage. All downstream analysis was carried out in R and ggplot2 was used for figures.

Effect of genetic differences on differential expression analysis

To estimate the contribution that genetic differences between IPSDMs and MDMs might have on the differential expression analysis, I obtained gene level RNA-seq read counts from

lymphoblastoid cell lines (LCLs) from 84 British individuals from a previously published study (Lappalainen et al., 2013). To mimic our experimental design, I repeatedly (100 times) sampled 9 individuals from the pool of 84, assigned them randomly into two groups (four and five individuals) and used DESeq2 to estimate the number of differentially expressed genes between the groups that satisfied the same thresholds that I used in the main analysis (FDR < 0.01, fold change > 2).

Alternative transcript usage

To quantify alternative transcript usage, reads were aligned to Ensembl 74 transcriptome using bowtie v1.0.0 (Langmead et al., 2009). Next, I used mmseq and mmdiff to quantify transcript expression and identify transcripts whose proportions had significantly changed (Turro et al., 2011, 2014). For each transcript I estimated the posterior probability of five models (i) no difference in isoform proportion (null model), (ii) difference between LPS treatment and control (LPS effect), (iii) difference between IPSDMs and MDMs (macrophage type effect), (iv) independent treatment and cell type effects (both effects), (v) LPS response different between MDMs and IPSDMs (interaction effect). I specified the prior probabilities as (0.6, 0.1, 0.1, 0.1, 0.1) reflecting the prior belief that most transcripts were not likely to be differentially expressed. Transcripts with posterior probability of the null model < 0.05 were considered significantly changed.

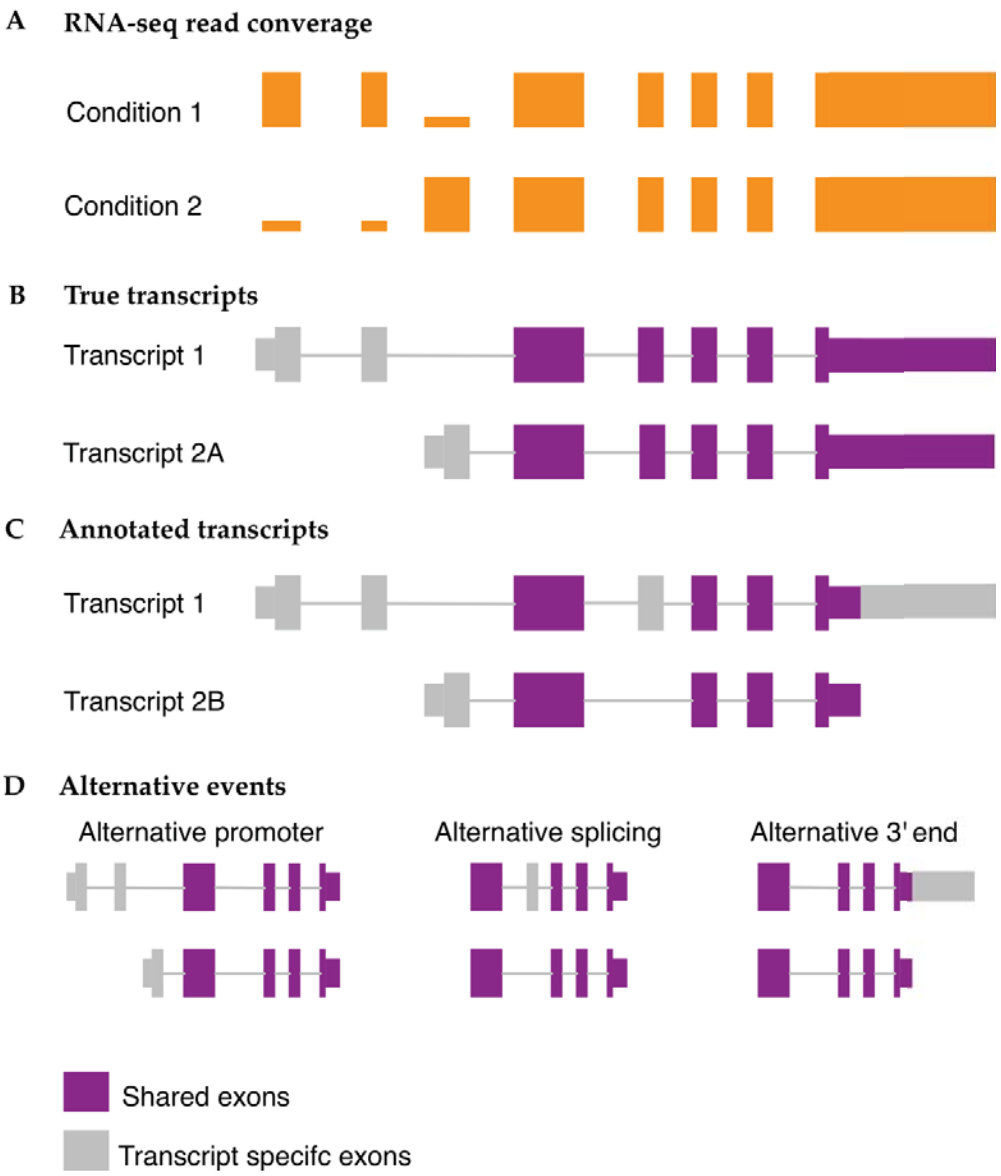

Figure 2.2. Constructing alternative transcription events from annotated transcripts. (A) Hypothetical RNA-seq read coverage over a gene indicating that there is switch from proximal to distal promoter between conditions 1 and 2. **(B)** True transcript annotations generating the read coverage observed on panel A. **(C)** Hypothetical reference transcripts detected to be differentially expressed between conditions 1 and 2. Note that the true transcript 2A from which the reads were generated was not present in the annotated transcripts. Consequently, different transcript 2B was detected to be differentially expressed that also had a skipped exon 4 and shorter 3′ UTR. Comparing transcript 1 to transcript 2B gives the wrong impression that exon 4 and the 3′ UTR are also differentially expressed although their read coverage has not changed between the conditions. **(D)** Three alternative transcription events constructed from transcripts 1

and 2B using the reviseAnnotations package. Estimating the differential expression of these alternative events separately correctly identifies that only the promoter usage changes between conditions.

Next, I used a two-step process to identify the exact alternative transcription events (alternative promoter usage, alternative splicing or alternative 3′ end usage) that were responsible for the observed changes in transcript proportions. First, to identify all potential alternative transcription events in each gene, I compared the transcript whose proportion changed the most between the two conditions to the most highly expressed transcript of the gene (Figure 2.2). This analysis revealed that for 93% of the genes the two selected transcripts differed from each other in more than one location, for example both the promoters and alternative 3′ ends were different between the two transcripts. However, visual inspection of the read coverage plots suggested that in majority of these cases there was only one change between the two transcripts and the other changes were false positives caused by missing or incomplete transcript annotations. To identify which one of the changes was responsible for the alternative transcription signal, I developed the reviseAnnotations R package (https://github.com/kauralasoo/reviseAnnotations) to split the two identified transcripts into individual alternative transcription events (Figure 2.2). Next, I reanalysed the RNA-seq data using exactly the same strategy as described above (bowtie + mmseq + mmdiff) but substituted Ensembl 74 annotations with the identified transcription events. Finally, I required events to change at least 10% in proportion between the two conditions to be considered for downstream analysis. This analysis revealed that instead of the 93% suggested by the transcript level analysis, only 4% of the genes had more than one event whose proportion changed at least 10%, indicating that transcript level analysis leads to a large number of false positives. Our event-based approach is similar to the one used by the Mixture of Isoforms (MISO) model (Katz et al., 2010).

Visualising alternative transcript usage

I developed the wiggleplotr R package (https://github.com/kauralasoo/wiggleplotr) to aid the visualisation of RNA-seq read coverage across alternative transcription events. A key feature of the software is that it allows introns to be shortened to constant width thus making it easier to see differences in read coverage between neighbouring exons in genes with long introns.

2.3 Gene expression variation between iPSCs, IPSDMs and MDMs

2.3.1 Global patterns of gene expression

RNA-seq was used to profile the transcriptomes of MDMs derived from five and IPSDMs derived from four different individuals (Methods). Identical preparation, sequencing and analytical methodologies were used for all samples. Initially, I used Principal Component Analysis (PCA) to generate a genome-wide overview of the similarities and differences between naïve and LPS-stimulated IPSDMs and MDMs as well as undifferentiated iPSCs. The first principal component (PC1) explained 50% of the variance and clearly separated iPSCs from all macrophage samples (Figure 2.3A) illustrating that IPSDMs are transcriptionally much more similar to MDMs compared to undifferentiated iPSCs. This was further confirmed by high expression of macrophage specific markers and low expression of pluripotency factors in IPSDMs (Figure 2.3B). The second PC separated naïve cells from LPS-stimulated cells and explained 16% of the variance, while the third PC, explaining 8% of the variance, separated IPSDMs from MDMs. The principal component that separated IPSDMs from MDMs (PC3) was different from that separating macrophages from iPSCs (PC1). Since principal components are orthogonal to one another, this suggests that the differences between MDMs and IPSDMs are beyond the simple explanation of incomplete gene activation or silencing compared to iPSCs.

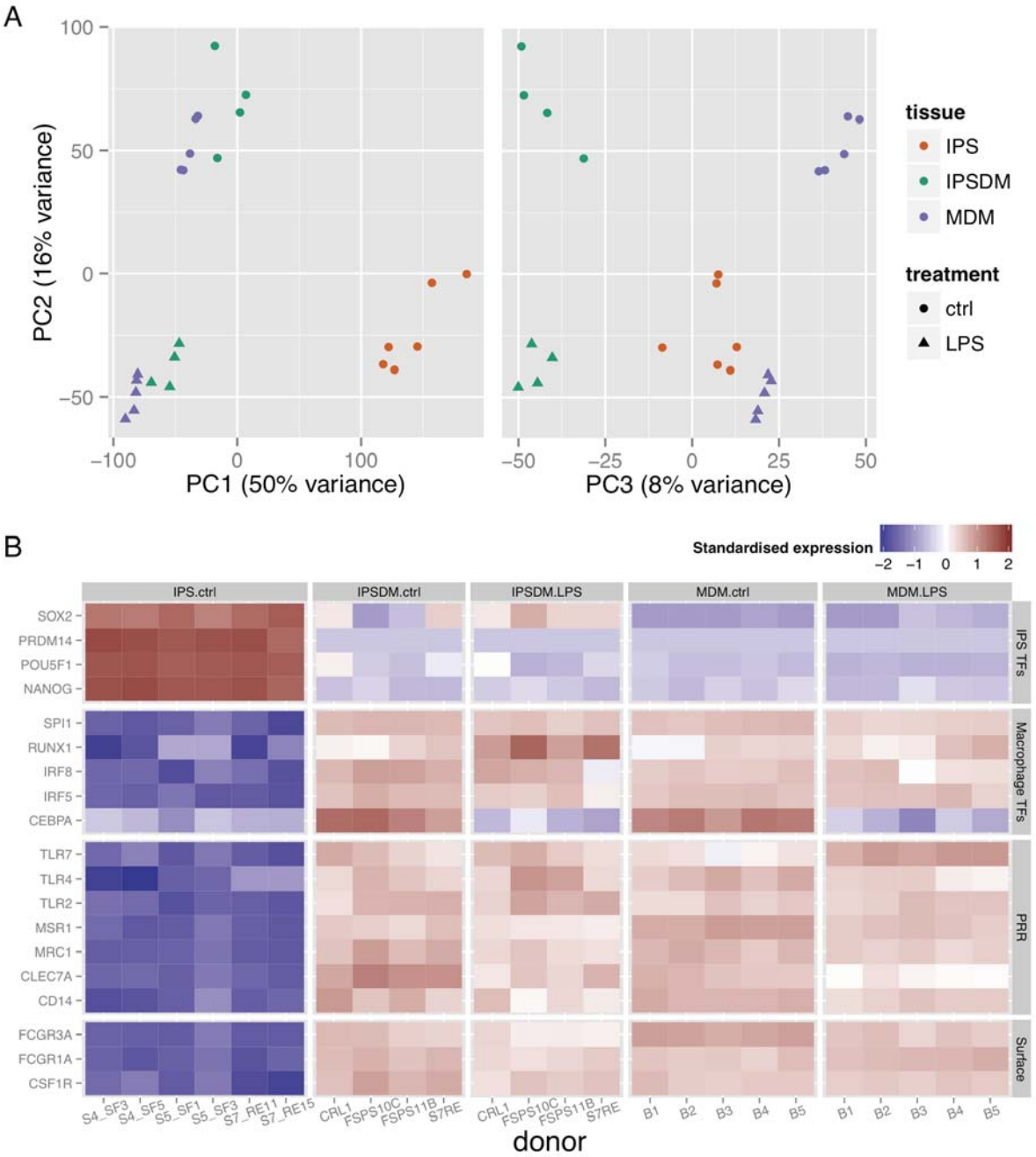

Figure 2.3. Gene expression variation between iPSCs, IPSDMs and MDMs. (A) Principal Component Analysis of expressed genes (TPM > 2) in iPSCs, IPSDMs and MDMs. **(B)** Heatmap showing the gene expression of selected iPSC-specific transcription factors (TFs), macrophage specific TFs, pattern recognition receptors (PRRs) and canonical macrophage cell surface markers. Rectangles correspond to measurements from independent biological replicates.

2.3.2 Differential expression analysis of IPSDMs vs MDMs

Table 2.2. Selection of enriched Gene Ontology terms and KEGG pathways for different groups of differentially expressed genes.

Upregulated in LPS response			
Term ID	**Domain**	**Term name**	**p-value**
GO:0045087	BP	innate immune response	7.31E-45
GO:0009617	BP	response to bacterium	2.42E-28
GO:0032496	BP	response to lipopolysaccharide	4.38E-28
KEGG:04668	ke	TNF signaling pathway	1.71E-20
KEGG:04064	ke	NF-kappa B signaling pathway	3.56E-14
Downregulated in LPS response			
Term ID	**Domain**	**Term name**	**p-value**
GO:0005096	MF	GTPase activator activity	1.01E-09
GO:0007264	BP	small GTPase mediated signal transduction	3.14E-09
More highly expressed in MDMs compared to IPSDMs			
Term ID	**Domain**	**Term name**	**p-value**
GO:0050778	BP	positive regulation of immune response	1.97E-21
GO:0003823	MF	antigen binding	2.55E-18
GO:0005764	CC	lysosome	1.42E-17
GO:0034341	BP	response to interferon-gamma	2.17E-16
GO:0042611	CC	MHC protein complex	3.67E-16
KEGG:04612	ke	Antigen processing and presentation	3.47E-13
KEGG:04145	ke	Phagosome	2.46E-11
More highly expressed in IPSDMs compared to MDMs			
Term ID	**Domain**	**Term name**	**p-value**
GO:0030198	BP	extracellular matrix organization	3.05E-45
GO:0016477	BP	cell migration	1.50E-40
GO:0001568	BP	blood vessel development	4.89E-36
GO:0016337	BP	cell-cell adhesion	6.27E-25
GO:0001525	BP	angiogenesis	1.34E-24

Although PCA provides a clear picture of global patterns and sources of transcriptional variation across all genes in the genome, important signals at individual genes might be missed. To better understand transcriptional changes at the gene level I used a two factor linear model implemented in the DESeq2 package (Love et al., 2014). The model included an LPS effect, capturing differences between unstimulated and stimulated macrophages and a macrophage

type effect capturing differences between MDMs and IPSDMs. Our model also included an interaction term that identified genes whose response to LPS differed between MDMs and IPSDMs. I defined significantly differentially expressed genes as having a fold-change of >2 between two conditions using a p-value threshold set to control our false discovery rate (FDR) to 0.01.

Using these thresholds, I identified 2977 genes that were differentially expressed between unstimulated IPSDMs and MDMs. Among these genes, 2080 were more highly expressed in IPSDMs and 897 were more highly expressed in MDMs (Figure 2.4A). Genes that were more highly expressed in MDMs such as HLA-B, LYZ, MARCO and HLA-DRB1 (Figure 2.4C), were significantly enriched for antigen binding, phagosome and lysosome pathways (Table 2.2). This result is consistent with a previous report that MDMs have higher cell surface expression of MHC-II compared to IPSDMs (Karlsson et al., 2008; van Wilgenburg et al., 2013). Genes that were more highly expressed in IPSDMs, such as MMP2, VEGFC and TGFB2 (Figure 2.4C) were significantly enriched for cell adhesion, extracellular matrix, angiogenesis, and multiple developmental processes (Table 2).

In the LPS response I identified 2638 genes that were differentially expressed in both MDMs and IPSDMs, of which 1525 genes were upregulated while 1113 were downregulated. As might be expected, Gene Ontology and KEGG pathway analysis revealed large enrichment for terms associated with innate immune and LPS response, NF-κB and TNF signalling (Table 2.2). I also identified 569 genes whose response to LPS was significantly different between IPSDMs and MDMs. The majority of these genes (365) responded in the same direction in both IPSDMs and MDMs, but the magnitude of change was significantly different. The remaining 229 genes showed a change in the opposite direction (8.7% of the LPS-responsive genes) (Figure 2.4B). This set of 229 were much weaker responders to LPS overall (2.3-fold compared to 4.7-fold). Additionally, I could not find convincing pathway or Gene Ontology enrichment signals in either gene set (229 and 569 genes) compared to all LPS-responsive genes. Overall, I found that the fold change of the genes that responded to LPS was highly correlated between MDMs and IPSDMs ($r = 0.82$, Figure 2.4B) indicating that the LPS response in these two macrophage types was broadly conserved. Interestingly, I also found that mean fold change was marginally (10%) higher in MDMs (4.95) compared to IPSDMs (4.43). The behaviour of some canonical LPS response genes is illustrated in Figure 2.4D.

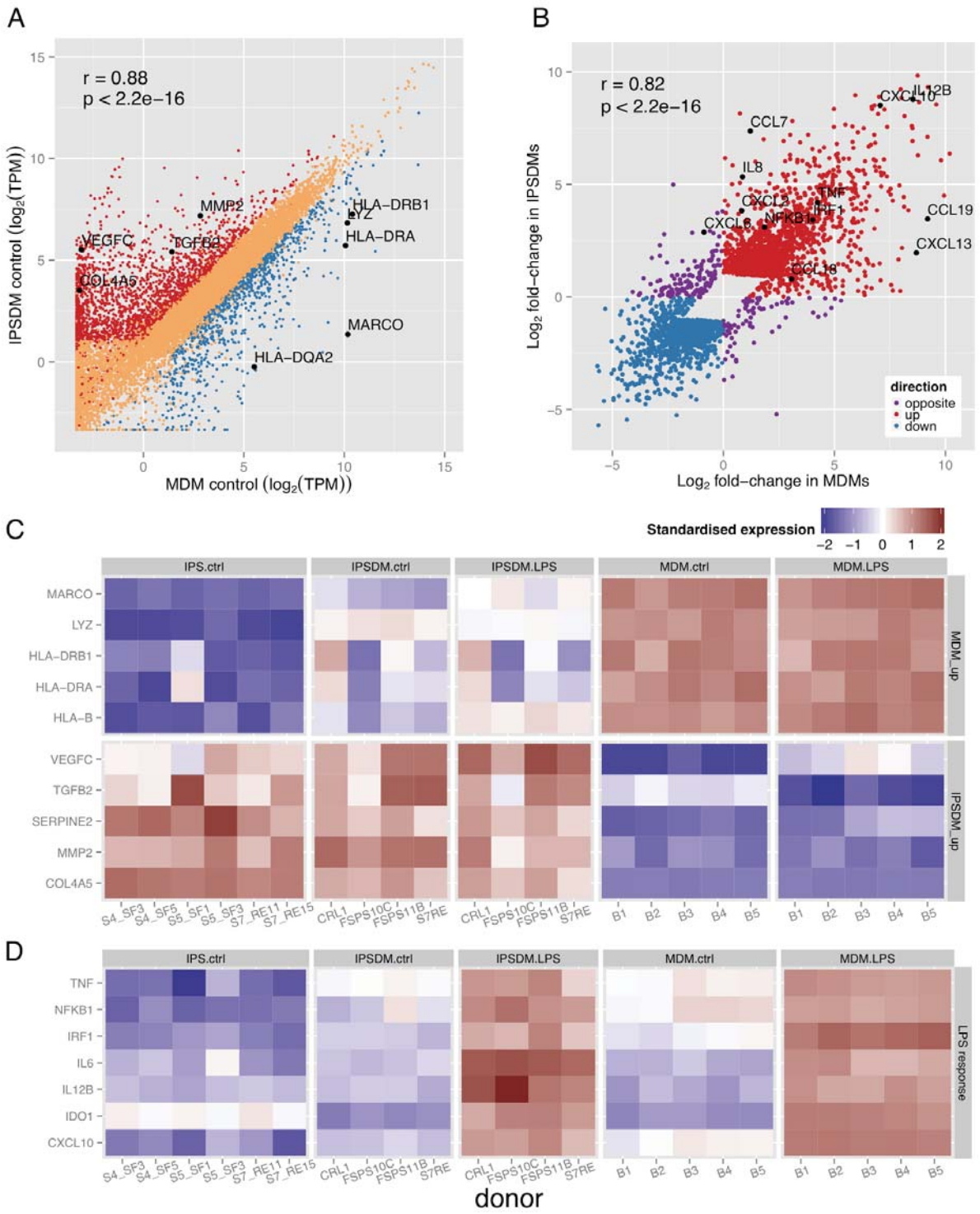

Figure 2.4. Differential expression analysis of IPSDMs vs MDMs. (A) Scatter plot of gene expression levels between MDMs and IPSDMs. Genes that are significantly more highly expressed in IPSDMs are shown in red and genes that are significantly more highly expressed in MDMs are shown in blue. **(B)** Scatter plot of fold change in response to LPS between MDMs

(x-axis) and IPSDMs (y-axis). Only genes with significant LPS or interaction term in the linear model are shown. Genes with LPS response fold change in the opposite direction between MDMs and IPSDMs are highlighted in purple. **(C)** Heatmap of genes differentially expressed between MDMs and IPSDMs. Representative genes from significantly overrepresented Gene Ontology terms (Table 1) include antigen presentation (HLA genes), lysosome formation (LYZ), angiogenesis (VEGFC, TGFB2), and extracellular matrix (SERPINE2, MMP2 COL4A5). The same genes are also marked in panel A. **(D)** Heatmap of example genes upregulated in LPS response.

Although genes with significantly different response to LPS between MDMs and IPSDMs were not enriched for particular Gene Ontology terms or pathways, IL8 and CCL7 mRNAs were more strongly upregulated in IPSDMs compared to MDMs (Figure 2.4B). Consequently, I looked at the response of all canonical chemokines in an unbiased manner. I observed relatively higher induction of further CXC subfamily monocyte and neutrophil attracting chemokines in IPSDMs (Figure 2.3). Moreover, five out of seven CXCR2 ligands (Zlotnik and Yoshie, 2012) were more strongly induced in IPSDMs (FDR < 0.1, fold-change difference between MDMs and IPSDMs > 2) which is significantly more than is expected by chance (Fisher's exact test $p = 4.5\times10^{-6}$) (Figure 2.5). These genes were also expressed at substantial levels (TPM > 100), with IL8 being one of the most highly expressed gene in IPSDMs after LPS stimulation. On the other hand, MDMs displayed relatively higher induction of three chemokines involved in attracting B-cells, T-cells and dendritic cells (CCL18, CCL19, CXCL13) (Figure 2.5).

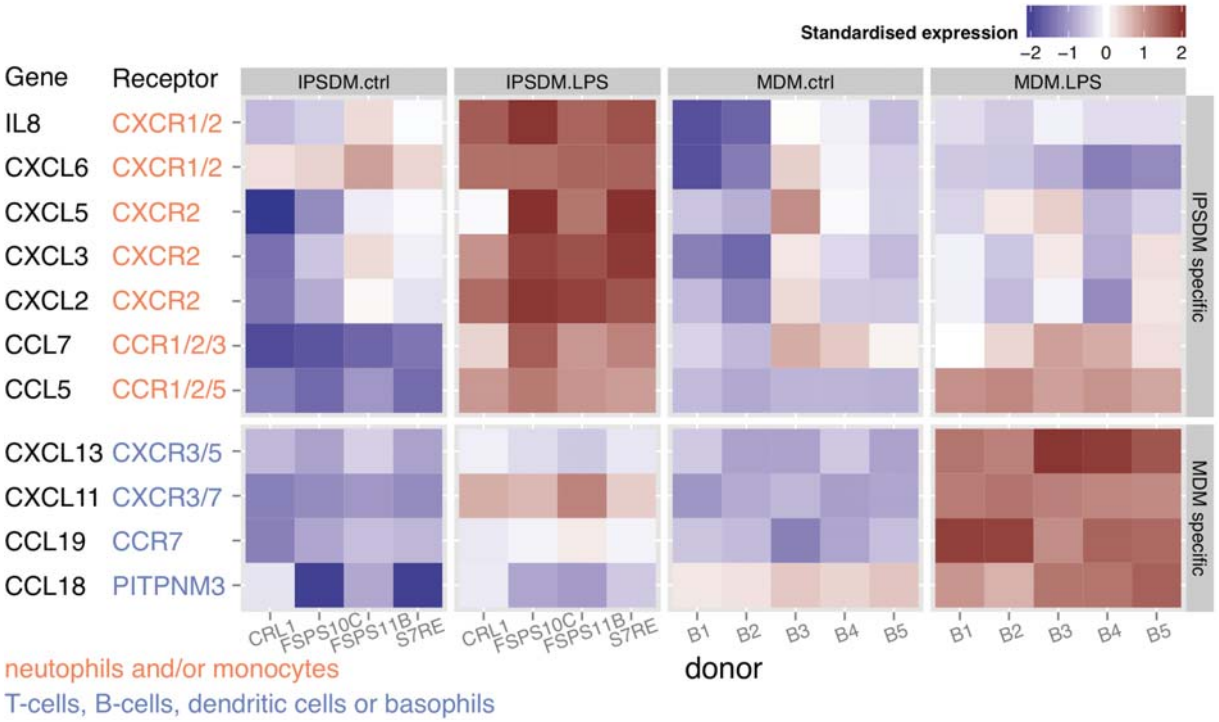

Figure 2.5. Chemokine genes that were particularly upregulated in either IPSDMs or MDMs in LPS response. Their annotated receptors and target cell types were taken from the literature (Soehnlein and Lindbom, 2010; Zlotnik and Yoshie, 2012).

2.3.3 Mechanisms underlying differences between MDMs and IPSDMs

To understand the mechanisms that might underlie the gene expression differences between MDMs and IPSDMs, I focussed on three hypotheses: (1) a minority contaminating cell population in IPSDM samples that is absent in MDMs, (2) genetic differences between donors from which the IPSDMs and MDMs were derived, and (3) incomplete differentiation from iPSCs resulting in developmentally immature macrophages that might exhibit some properties of the iPSCs. The high purity of our IPSDM samples (92-98%) (Table 2.3) and MDM samples (routinely 90-95% pure) suggested that there was no obvious contaminating cell type present that did not express the canonical macrophage markers. Furthermore, even the 99% pure IPSDM samples retained most of the differential expression with MDMs (Figure 2.6A) suggesting contamination is not a major source of IPSDM-MDM differences.

Table 2.3. Purity of iPSC-derived macrophages. We used flow cytometry to estimate the percentage of cells expressing five cell surface markers in IPSDMs differentiated from three IPSC lines.

Marker / Cell line	FSPS10C	FSPS11B	S7RE
CD14	98.6	90.4	91.2
CD206	99.5	85.1	
CD4	99.5	92.8	92.9
CD32	94.8		87.6
CD163	74.1	92	85.6

Alternatively, IPSDMs could be incompletely differentiated from iPSCs. Under this model, genes that are expressed in iPSCs but repressed in mature macrophages would be more highly expressed in IPSDMs compared to MDMs. Consistent with this hypothesis, genes that were more highly expressed in IPSDMs were often also expressed in iPSCs (Figure 2.4C, Figure 2.6A). Furthermore, while the majority of the genes that were more highly expressed in MDMs had mean expression > 2 TPM in both cell types, a large proportion of the genes that were more highly expressed in IPSDMs had mean expression < 1 TPM across both cell types (Figure 2.6B), suggesting that their expression level in IPSDMs might be too low to be functional. Moreover, the promoters of the upregulated genes were highly enriched for repressive H3K27me3 histone marks in CD14+ monocytes (The ENCODE Project Consortium, 2012) (Figure 2.6C), suggesting that these genes normally become silenced prior to monocyte-macrophage differentiation *in vivo* and may not have been completely silenced in IPSDMs.

Finally, it is possible that some of the differences between IPSDMs and MDMs could be confounded with genetic differences between the donors. For example, by chance, the different individuals from which the IPSDMs and MDMs were derived could be fixed for alternate alleles of a cis-regulatory variant that changes the expression of a given gene, which would appear to be differentially expressed between the two cell types. However, since all our IPSDM and MDM donors were randomly sampled from the same population, strong clustering of IPSDM and MDM samples in the PCA analysis (Figure 2.3A) suggests that genetics is not a major source of differences between these cell types. To address this quantitatively, I reanalysed an independent RNA-seq data from 84 British individuals (Lappalainen et al., 2013). I found only a median of three differentially expressed genes between any two random samples of 4 and 5

individuals (Figure 2.6D). This suggests that only a small fraction of the differences between MDMs and IPSDMs are likely to be due to genetics.

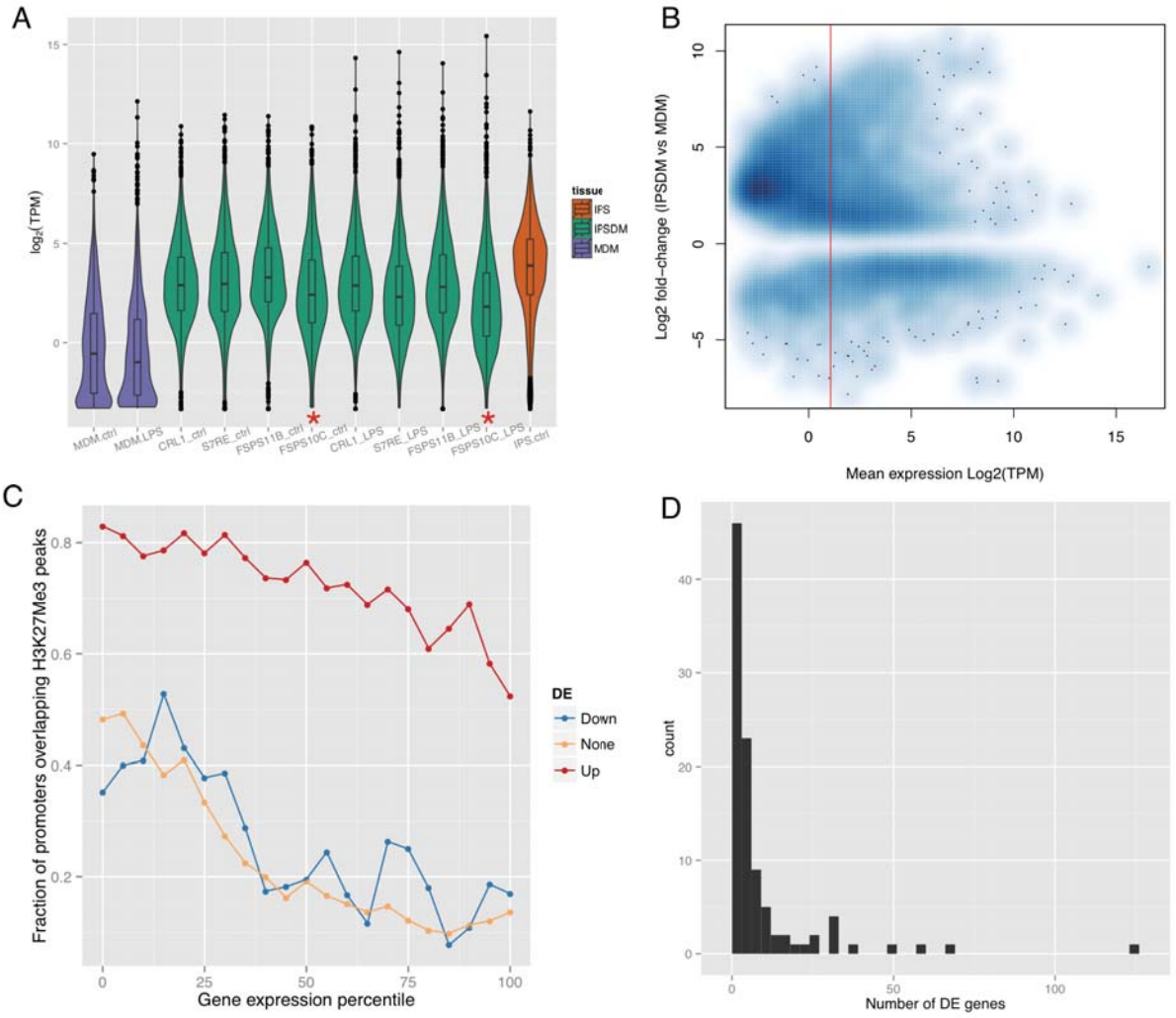

Figure 2.6: Mechanisms underlying differential expression between MDMs and IPSDMs. **(A)** Expression levels of genes that were more highly expressed in IPSDMs compared to MDMs (TPM > 2). Purple violin plots show the mean expression of these genes in MDMs and orange in IPS cells. Red asterisks mark IPSDM samples (FSPS10C) that stained > 99% positive for CD14, CD206 and CD4 while S7RE and FSPS11B samples were ~91% positive. **(B)** MA-plot of differentially expressed genes between MDMs and IPSDMs (without TPM cut-off). On the y-axis is the DESeq2 estimate of fold-change between MDMs and IPSDMs. Red line denotes the 2 TPM cut-off used in most analyses. **(C)** Fraction of gene promoters overlapping H3K27Me3 peaks in ENCODE CD14+ monocyte samples stratified by the percentile of gene expression

level. Up - genes upregulated in IPSDMs; Down - downregulated in IPSDMs; None - not differentially expressed between MDMs and IPSDMs. **(D)** Histogram of the number of differentially expressed genes between two groups of randomly selected individuals.

2.4 Global variation in alternative transcript usage

Many human genes express multiple transcripts that can differ from each other in terms of function, stability or subcellular localisation of the protein product (Carpenter et al., 2014; Wang et al., 2008). Considering expression only at a whole gene level can hide some of these important differences. Therefore, we sought to quantify how similar were naïve and stimulated IPSDMs and MDMs at the individual transcript expression level. Here, we first used mmseq (Turro et al., 2011) to estimate the most likely expression level of each annotated transcript that would best fit the observed pattern of RNA-seq reads across the gene. Next, we calculated the proportion of total expression accounted for by each transcript by dividing transcript expression by the overall expression level of the gene, only including genes that were expressed over two transcripts per million (TPM) (Wagner et al., 2012) in all experimental conditions (8284 genes). Since the proportions of all transcripts of a gene sum to one and most genes express one dominant transcript (Gonzàlez-Porta et al., 2013), I used the proportion of the most highly expressed transcript as a proxy to capture variation in transcript proportions within a gene. In this context and similarly to gene level analysis, the first PC explained 31% of the variance and clearly separated IPSCs from macrophages (Figure 2.7A). However, the second PC (11% of variance) not only separated unstimulated cells from stimulated cells but also IPSDMs from MDMs. One interpretation of this result is that the changes in transcript proportions between IPSDMs and MDMs, to some extent, also resemble those induced in the LPS response. Further analysis (below) highlighted that much of this variation can be explained by changes in 3′ untranslated region (UTR) usage.

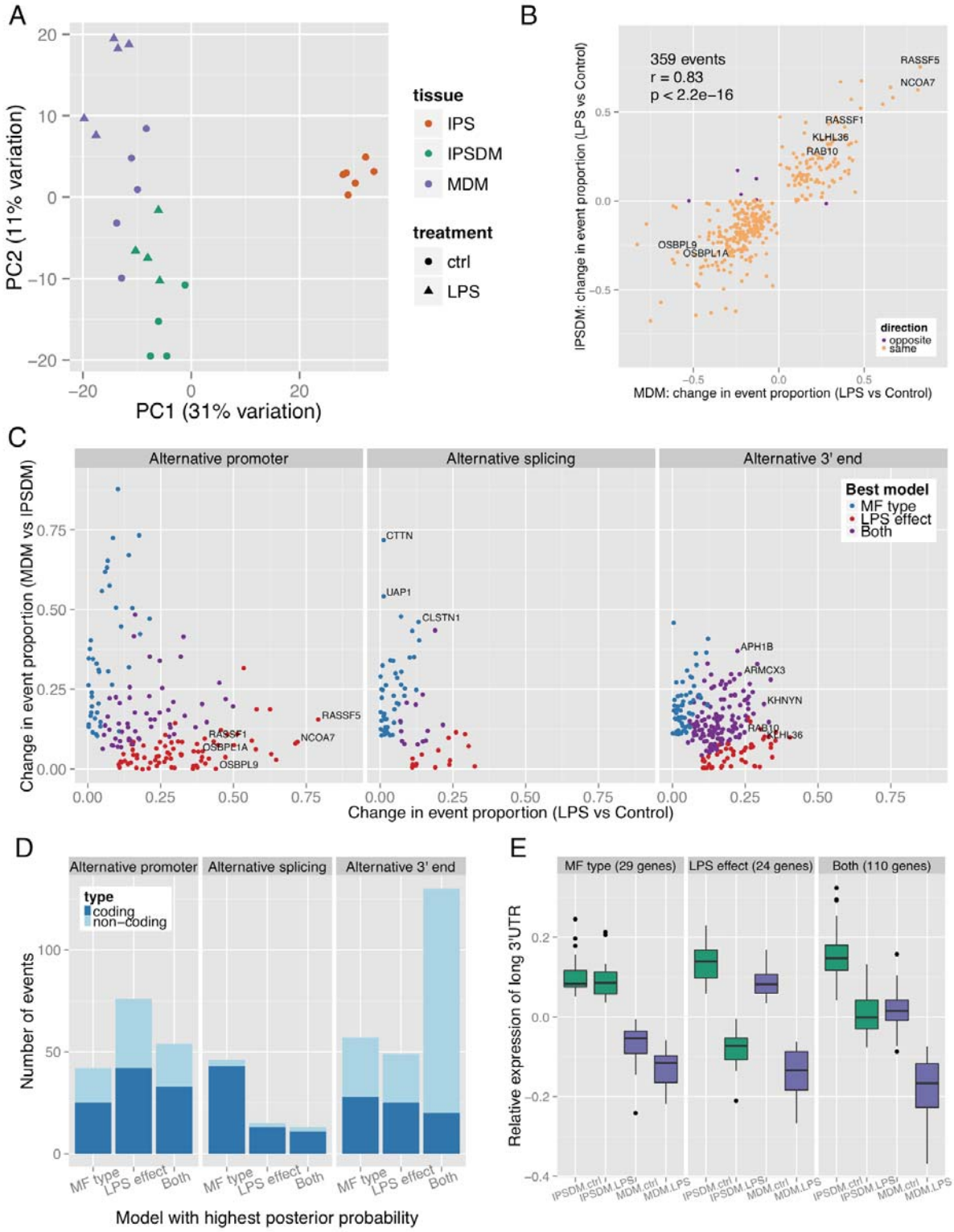

Figure 2.7. Alternative transcription in IPSDMs and MDMs. (A) PCA of relative transcript proportions in iPSCs, IPSDMs and MDMs. Only genes with mean TPM > 2 in all conditions were

included. **(B)** Alternative transcription events detected in LPS response. Each point corresponds to an alternative transcription event and shows the absolute change in the proportion of the most highly expressed transcript (across all samples) in LPS response in MDMs (x-axis) and IPSDMs (y-axis). **(C)** All detected alternative transcription events were divided into three groups based on whether they affected alternative promoter, alternative splicing or alternative 3′ end of the transcript. For each event, we plotted its change in proportion in LPS response (x-axis) against its change between macrophage types (y-axis). The events are coloured by the most parsimonious model of change selected by mmseq: LPS effect (difference between naïve and LPS-stimulated cells only); macrophage (MF) type (difference between IPSDMs and MDMs only); both (data support both MF type and LPS effects). **(D)** Number of alternative transcription events form panel C grouped by position in the gene (alternative promoter, alternative splicing, alternative 3′ end) and most parsimonious model selected by mmseq. (e) Relative expression of long alternative 3′ UTRs in genes showing a change between IPSDM and MDMs (MF type), between naïve and LPS-stimulated cells (LPS effect) and for genes showing both types of change.

2.4.1 Identification and characterisation of alternative transcription events

Alternative transcription can manifest in many forms, including alternative promoter usage, alternative splicing and alternative 3′ end choice, each likely to be regulated by independent biological pathways. Thus, I sought to characterise and quantify how these different classes of alternative transcription events were regulated in the LPS response, and between MDMs and IPSDMs. Using a linear model implemented in the mmdiff (Turro et al., 2014) package followed by a series of downstream filtering steps (Methods) we identified 504 alternative transcription events (ATEs) in 485 genes. Out of those, 145 events changed between unstimulated IPSDMs and MDMs (macrophage (MF) type effect) while 156 events changed between naive and LPS stimulated cells across macrophage types (LPS effect). Further 197 events had different baseline expression between macrophage types, but also changed in the same direction after LPS stimulation (Both effects). Finally, only 6 events change in the opposite direction after LPS stimulation between MDMs and IPSDMs (Figure 2.7B). Next, I focussed on the 359 events that changed in the LPS response in at least one macrophage type (156 + 197 events with LPS response in the same direction and 6 events with LPS response in the opposite direction). I found that the LPS-induced change in the proportion of the most highly expressed transcript was highly correlated between MDMs and IPSDMs (Pearson r = 0.83) (Figure 2.7B), further confirming that the LPS response in both macrophage types is conserved.

Perhaps surprisingly, although the transcriptional response to LPS at the whole gene level is relatively well understood, the effect of LPS on transcript usage has remained largely unexplored. Therefore, I decided to investigate the types of alternative transcription events identified in LPS response as well as between MDMs and IPSDMs (See Methods for details). Most protein coding changes in LPS response were generated by alternative promoter usage (Figure 2.7C-D). In total, I identified 180 alternative promoter events, 51 of which changed the coding sequence by more than 100 bp in LPS response. Strikingly, alternative promoter events displayed larger change in proportion than other events so that often the most highly expressed transcript of the gene changed between cell types and conditions (Figure 2.7C). Alternative promoter usage for three example genes is illustrated on Figure 2.8.

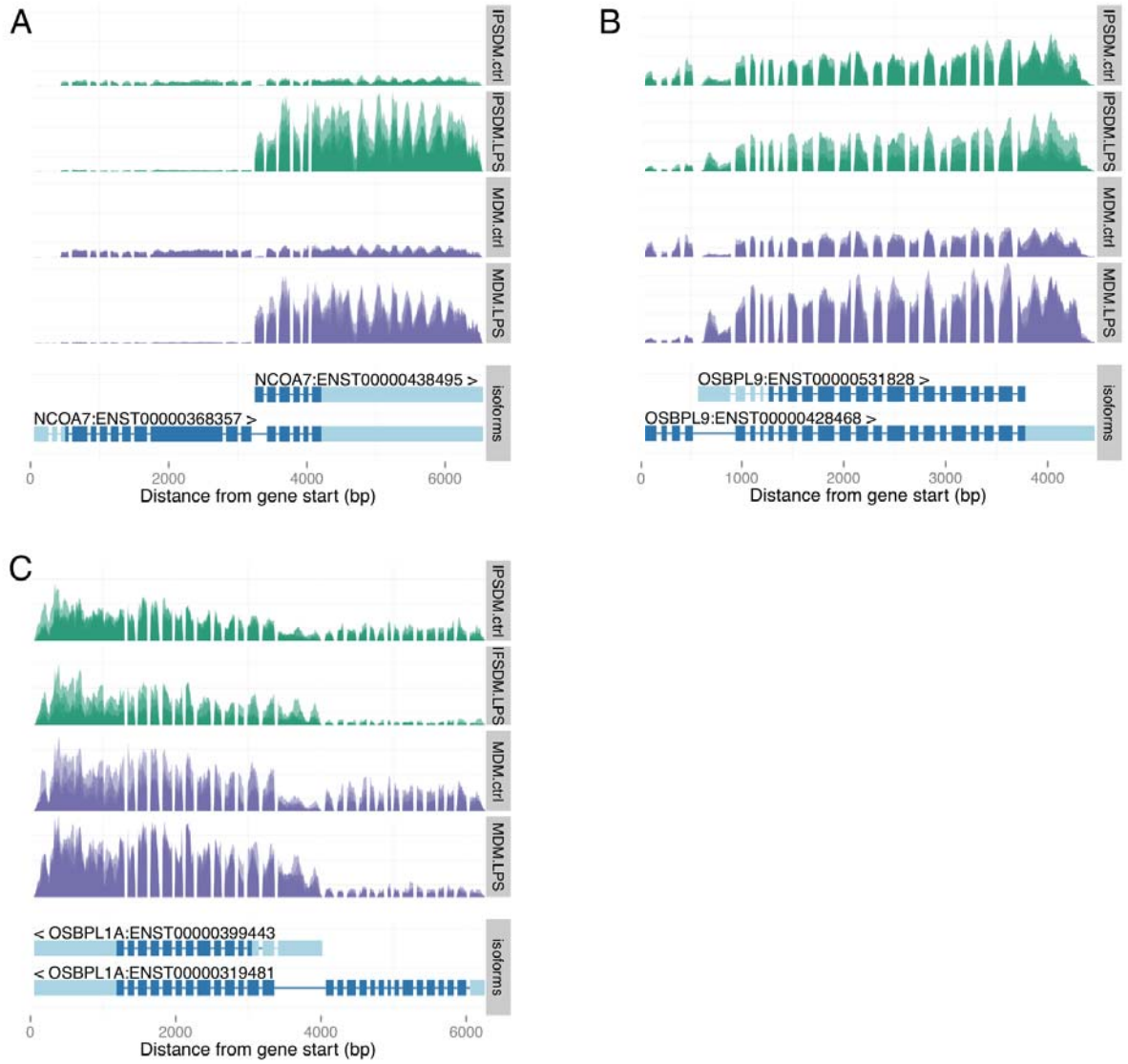

Figure 2.8. Examples of alternative promoter usage in LPS response. Each plot shows normalised read depth across the gene body in IPSDMs (green) and MDMs (purple) with gene structure in the panel beneath each plot. Introns have been compressed relative to exons to facilitate visualisation. (A-C) Alternative promoter usage in NCOA7, OSBPL9 and OSBPL1A genes.

I also observed widespread alternative 3′ end usage both in the LPS response as well as between MDMs and IPSDMs (Figure 2.7C-D). In contrast to alternative promoters, most of the 3′ end events only changed the length of the 3′ UTR and not the coding sequence (Figure 2.7D). Changes in 3′ UTR usage were strongly asymmetric, with longer 3’ UTRs being more highly

expressed in IPSDMs relative to MDMs, and in unstimulated cells relative to stimulated cells (Figure 2.7E, Figure 2.9A). Notably, I also observed that the decrease in 3′ UTR length correlated with the second principal component of relative transcript expression (Figure 2.7A). Consistent with this observation, I found that genes with 3′ UTR events were enriched for high absolute weights in PC2 ($p < 2.2\times10^{-16}$, chi-square goodness-of-fit test), (Figure 2.9B) indicating that part of the transcriptional variation captured by PC2 manifests as changes in 3′ UTR usage. I found no convincing pathway or Gene Ontology enrichment signal in genes with alternative 3′ UTR events.

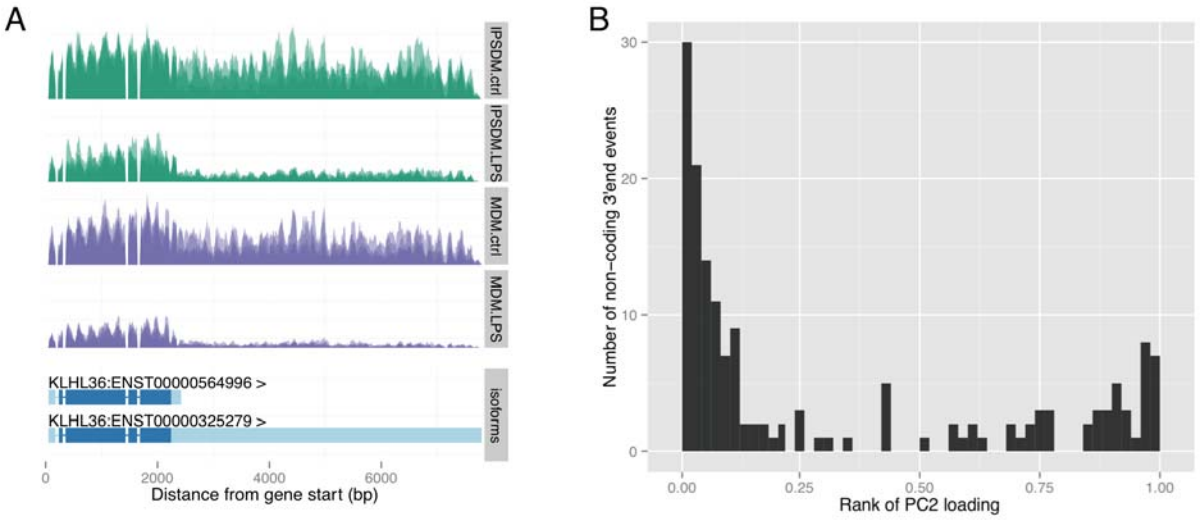

Figure 2.9. 3′ UTR shortening in LPS response. (A) Examples of 3′ UTR shortening in LPS response. The plot shows normalised read depth across the gene body in IPSDMs (green) and MDMs (purple) with gene structure in the panel beneath the plot. Introns have been compressed relative to exons to facilitate visualisation. **(B)** All genes were ranked based on their weights in PC2 (Figure 2.7A) and the relative ranks of the 162 genes with 3'UTR events are displayed on the histogram. The ranks of a randomly sampled set of genes should be uniformly distributed whereas genes that contribute strongly to the PC should be enriched for high and low relative ranks (corresponding to large positive and negative weights on the PC).

Finally, I detected only a small number of alternative splicing events influencing middle exons, most of which occurred between MDMs and IPSDMs rather than in the LPS response (Figure 2.7C-D). Three of the events with largest changes in proportion affected cassette exons in UAP1, CTTN and CLSTN1 genes (Figure 2.10A-C). The inclusion of these exons has previously

been shown to be regulated by RNA-binding protein RBFOX2 that was also significantly more highly expressed in IPSDMs (Figure 2.10D) (Lambert et al., 2014; Venables et al., 2013).

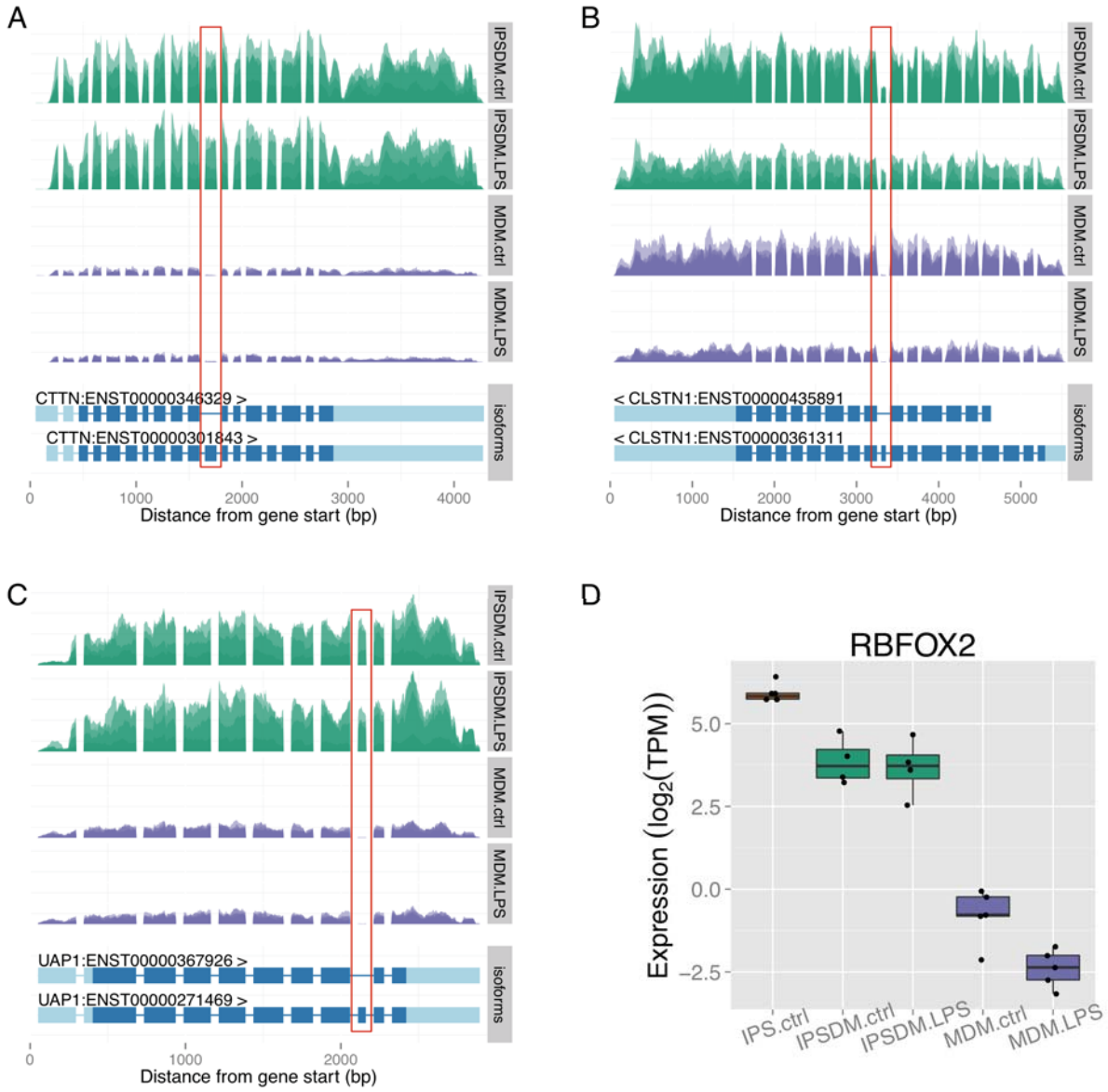

Figure 2.10. Alternative splicing between IPSDMs and MDMs. (A-C) Examples of alternative splicing between MDMs and IPSDMs. The alternatively spliced exon is marked with the red rectangle. **(D)** Expression of RBFOX2 gene in iPSCs, IPSDMs and MDMs.

2.5 Discussion

In this study, we used high-depth RNA-seq to investigate transcriptional similarities and differences between human monocyte and iPSC-derived macrophages. Our principal findings are that, relative to differences between MDMs and iPSCs, the transcriptomes of naïve and LPS stimulated MDMs and IPSDMs are broadly similar both at the whole gene and individual transcript levels. Concurrently with our study, another paper using a different macrophages differentiation protocol came to the same broad conclusion (Zhang et al., 2015). Although we have only examined steady-state mRNA levels, conservation of transcriptional response to LPS implies that the major components of regulatory network that coordinate LPS response on the protein level are likely to also be similarly conserved. We did, however, also observe intriguing differences in expression in specific sets of genes, including those involved in tissue remodelling, antigen presentation and neutrophil recruitment, suggesting that IPSDMs might possess some phenotypic differences from MDMs. Our analysis also revealed a rich diversity of alternative transcription changes suggesting widespread fine-tuning of regulation in macrophage LPS response.

We also looked at the mechanisms that might be underlying the observed differences between MDMs and IPSDMs. We were able to rule out genetic differences between MDMs and IPSDMs or contamination by some other cell type not expressing macrophage specific cell surface markers as a major source of these differences. However, we did find some evidence that IPSDMs might be developmentally less mature than MDMs. This was illustrated by the fact that IPSDMs expressed residual amounts of genes what were substantially more highly expressed in iPSCs and almost completely silenced in MDMs. Furthermore, we found that promoters of these genes were usually actively silenced by H3K27Me3 histone modifications in CD14+ monocytes suggesting that this silencing might be incomplete in IPSDMs.

Alternatively, IPSDMs might share some features with tissue resident macrophages that are developmentally and phenotypically distinct from MDMs (Gautier et al., 2012; Ginhoux et al., 2010; Gosselin et al., 2014; Lavin et al., 2014). In support of that, higher expression of tissue remodelling and neutrophil recruitment genes has previously been associated with tissue and tumour associated macrophages (Cailhier et al., 2005; Mantovani et al., 2013; Schmieder et al., 2012; Soehnlein and Lindbom, 2010). On the other hand, higher expression of antigen presentation genes in MDMs is consistent with the specialised role of monocyte-derived cells in

immune regulation and antigen presentation (Gundra et al., 2014; Jakubzick et al., 2013; Soehnlein and Lindbom, 2010). This is consistent with a previous study suggesting a shared developmental pathway between IPSDMs and foetal macrophages (Klimchenko et al., 2011). Nevertheless, it is likely that the exact characteristics of IPSDMs can be shaped by the addition of cytokines and other factors during differentiation and this could be an important area for further exploration.

In addition to showing that LPS response was broadly conserved between MDMs and IPSDMs both on gene and transcript level, we also identified hundreds of individual alternative transcription events, highlighting an important, but potentially overlooked, regulatory mechanism in innate immune response. A small number of the events have known functional consequences. For example, the LPS-induced short isoform of the NCOA7 (Figure 2.8A) gene is known to be regulated by Interferon β-1b and it is suggested to protect against inflammation-mediated oxidative stress (Yu et al., 2014) whereas the long isoform is a constitutively expressed coactivator of oestrogen receptor (Shao et al., 2002). Similarly, the two isoforms of the OSBPL1A gene (Figure 2.8C) have distinct intracellular localisation and function (Johansson et al., 2003) while the LPS-induced short transcript of the OSBPL9 gene (Figure 2.8B) codes for an inhibitory isoform of the protein (Ngo and Ridgway, 2009). Thus, alternative promoter usage has the potential to significantly alter gene function in LPS response and these changes can be missed in gene level analysis.

Widespread shortening of 3′ UTRs has previously been observed in proliferating cells and cancer as well as activated T-cells and monocytes (Mayr and Bartel, 2009; Sandberg et al., 2008). The functional consequences of 3′ UTR shortening are unclear, but extended 3′ UTRs are often enriched for binding sites for miRNAs or RNA-binding proteins that can regulate mRNA stability and translation efficiency (Gupta et al., 2014; Sandberg et al., 2008). The role of miRNAs in fine-tuning immune response is well established (O'Neill et al., 2011). Furthermore, interactions between alternative 3′ UTRs and miRNAs have recently been implicated in the brain (Miura et al., 2013; Wehrspaun et al., 2014). Therefore, it might be interesting to explore how 3′ UTR shortening affects miRNA-dependent regulation in LPS response.

In summary, we have performed an in depth comparison of an iPSC-derived immune cell with its primary counterpart. Our study suggests that iPSC-derived macrophages are potentially valuable alternative models for the study of innate immune stimuli in a genetically manipulable,

stable cell culture system. The ability to readily derive and store iPSCs potentially enables in-depth future studies of the innate immune response in both healthy and diseased individuals. A key advantage of this model will be the ability to study the impact of human genetic variation, both natural and engineered, in innate immunity.