3 Large-scale differentiation of macrophages from human iPSCs

Collaboration note

The macrophage differentiation work in this chapter was performed in collaboration with Julia Rodrigues who was a research assistant in Daniel Gaffney's lab at the time. I designed the experiments, performed *Salmonella* infection and IFNɣ stimulation assays, took care of sample logistics and performed all of the data analysis. Julia was mainly responsible for tissue culture required for macrophage differentiation and preparing cells for stimulation experiments. Julia also prepared and stained the cells for flow cytometry experiments. Subhankar Mukhopadhyay and Gordon Dougan provided valuable feedback in designing and optimising *Salmonella* infection and IFNɣ stimulation conditions. RNA-seq library construction and sequencing was done by DNA Pipelines core facility at Sanger.

3.1 Introduction

Human induced pluripotent cells (iPSCs) can be derived from almost any individual with many differentiation protocols available for different cell lineages, including macrophages (van Wilgenburg et al., 2013), neurons (Rigamonti et al., 2016) and cardiomyocytes (Kempf et al., 2015). However, typical published differentiation protocols have been developed and used on a few iPSC lines. Hence, the expected range of normal variability between iPSC lines regarding many aspects of these protocols including success rate, duration of differentiation, yield, and purity of the differentiated cells is generally not well understood. If iPSCs are to be used for studying the functions of common genetic variation in differentiated cell types, differentiation protocols need to be robust enough to facilitate large-scale studies in tens or hundreds of lines. However, for most differentiation protocols systematic studies of critical iPSC differentiation parameters are not available.

The factors that influences iPSC differentiation success and yield are not well understood. In one of the largest studies to date, (Koyanagi-Aoi et al., 2013) performed neural differentiation from 10 human embryonic stem cell lines (ESCs) and 40 human iPSC lines. They observed that

7/40 iPSC lines showed aberrant gene expression profiles that correlated with defects in neural differentiation. A smaller study looking at five human ESC lines and 12 iPSC lines observed that iPSCs showed higher variability in their potency to differentiate into neurons compared to ESCs, but was unable uncover a specific cause (Hu et al., 2010). A study of 28 iPSC lines found that variations in hepatic differentiation could largely be attributed to differences between donors (Kajiwara et al., 2012) and other work has found that the method used to form embryoid bodies can have a large effect on differentiation propensity (Paull et al., 2015). Finally, a study in mouse iPSCs showed that the cell type of origin might influence differentiation propensity in early passage iPSCs, but these effects disappeared after 10-16 passages (Polo et al., 2010). Thus, there are many factors influencing differentiation success and their relative importance is likely to vary between protocols.

Additionally, when we differentiate iPSCs into a cell type of interest, we typically have a specific phenotype of interest, such as difference in gene expression level between two conditions, that we want to measure. Ideal experimental design should control for all other sources of variability in differentiation to maximise the chance of detecting the signal of interest. However, controlling all potential sources of variability is often impractical or even unfeasible. Hence, there is great interest in knowing which sources of variability have a strong effect on the phenotype (and should be controlled for) and which are so weak that they can be ignored. Variance component analysis is an effective approach to understand the relative contribution of both technical and biological factors on a phenotype of interest such as gene expression levels (’t Hoen et al., 2013; Rouhani et al., 2014). For example, two recent studies have used this approach to highlight the importance of genetic differences between donors as a major factor underlying gene expression variation in human iPSCs (Kilpinen et al., 2016; Rouhani et al., 2014).

We performed 138 macrophage differentiation attempts from 123 IPSC lines selected randomly from the HipSci project (Kilpinen et al., 2016), making it one of the largest directed differentiation study from human iPSCs. However, some of the differentiated lines did not produce enough macrophages to perform all of the experimental assays or the cells were not pure enough to be used in stimulation experiments. In total, we sequenced the RNA from 84 of these lines in four experimental conditions. We focussed on three questions: (i) how reliable and reproducible was the macrophage differentiation protocol (ii) which sources of variation had a strong effect on macrophage gene expression levels (iii) because flow cytometry is often used as a quality

control step in cellular differentiation assays, what factors are responsible for variability in the expression of cell surface markers in iPSC-derived macrophages.

We were able to successfully differentiate macrophages from 101/123 iPSC cell lines, with an overall success rate of 82%. Combining gene expression data with extensive sample metadata, we were able to estimate the relative proportion of gene expression variance explained by different experimental factors. Our results highlight the importance of maintaining high purity and constant cell density of the differentiated cells. We also showed that using live bacteria can lead to larger stimulation-specific batch effects than using well-defined molecular stimuli such as IFNɣ. Finally, we have shown that expression of CD14 and CD16 cell surface markers can be highly variable between genetically distinct cell lines and in the case of CD14, most of this variation can be attributed to a genetic variant upstream of the CD14 gene. This highlights the importance of accounting for genetic differences when comparing primary and iPSC-derived cells from different individuals.

3.2 Methods

3.2.1 Cell culture and reagents

Donors and cell lines

Human induced pluripotent stem cells (iPSCs) from 123 healthy donors (72 females and 51 males) were obtained from the HipSci project (Kilpinen et al., 2016). Of these lines, 57 were initially grown in feeder-dependent medium and 66 were grown in feeder-free E8 medium.

Feeder-free iPSC culture

Feeder-free iPSCs were grown on tissue culture treated plates coated with vitronectin (VTN-N) (Gibco, cat. no. A14700) in Essential 8 (E8) medium (Gibco). The cells were dissociated from the plates using Gentle Cell Dissociation Buffer (Stemcell Technologies, cat. no. 07174) and passaged every 3-5 days. Prior to macrophage differentiation, the feeder-free iPSCs were first transferred to feeder-dependent media and propagated for at least two passages. This step was necessary because multiple attempts to differentiate macrophage directly from feeder-free iPSCs with our protocol failed.

Feeder-dependent iPSC culture

Feeder-dependent iPSCs were grown on irradiated CF-1 mouse embryonic fibroblast (MEF) feeder cells (AMS Biotechnology) in Advanced DMEM-F12 (Gibco) supplemented with 20% KnockOut Serum Replacement (KSR) (Gibco), 2mM L-glutamine (Sigma), 50 IU/ml penicillin (Sigma), 50 IU/ml Streptomycin (Sigma) and 50μM β-Mercaptoethanol (Sigma M6250). The media was supplemented with 4 ng/ml recombinant human fibroblast growth factor (rhFGF) basic (R&D, 233-FB-025) to maintain pluripotency and was changed daily. MEFs were seeded on 0.1% gelatine-coated tissue-culture treated plates (Corning 6-well or 10 cm plates) 24 hours prior to passaging iPSCs at a cell density of 2 million cells per 6-well or 10-cm plate in Advanced DMEM-F12 supplemented with 10% FBS (Labtech), 2mM L-glutamine (Sigma), 50IU/ml Penicillin & 50IU/ml Streptomycin (Sigma). Prior to passaging or embryoid body formation, iPSCs were dissociated from the plates using 1:1 mixture of collagenase (1 mg/ml) and dispase (1 mg/ml) (both Gibco).

Macrophage differentiation

iPSCs were differentiated into macrophages using a previously published protocol (van Wilgenburg et al., 2013) involving 3 stages: i) embryoid body (EB) formation, ii) generation of monocyte-like myeloid progenitors from the EBs and iii) terminal differentiation of the progenitors into macrophages. For EB formation, iPSC colonies were treated with 1:1 mixture of collagenase (1 mg/ml) and dispase (1 mg/ml) and intact colonies were transferred to low-adherence plates (Sterilin). The colonies were cultured in feeder-dependent iPSC medium without rhFGF for 3 days. On day 3, the EBs were harvested and transferred to gelatinised tissue-culture treated 10 cm plates in serum-free haematopoietic medium (Lonza X-VIVO 15), supplemented with 2 mM L-glutamine (Sigma), 50 IU/ml penicillin, 50 IU/ml streptomycin (Sigma), 50 μM β-Mercaptoethanol (Sigma M6250), 50 ng/ml macrophage colony stimulating factor (M-CSF) (R&D) and 25 ng/ml interleukin-3 (IL-3) (R&D). EBs were maintained in these plates with media changes every 3-5 days for 4-6 weeks until the progenitor cells appeared in the supernatant. Progenitor cells were harvested from the supernatant, filtered through a 40μm cell strainer (BD 352340), centrifuged at 1200 rpm for 5 minutes, counted, and plated in RPMI 1640 (Gibco) supplemented with 10% FBS (Labtech), 2mM L-glutamine (Sigma) and 100 ng/ml hM-CSF (R&D) at a cell density of 150,000 cells per 6-well plate or 1,000,000 cells per 10 cm plate and differentiated for another 7 days.

3.2.2 Macrophage stimulation assays

After harvesting, macrophage progenitors were seeded on 6-well plates at 150,000 cells/well. Two wells were used per condition to ensure sufficient amount of RNA. On day 6 of macrophage differentiation, medium was changed for all wells with half of the wells receiving macrophage differentiation media (with M-CSF) and half of the cells receiving macrophage differentiation media supplemented with 20 ng/ml IFNɣ (R&D) and M-CSF. After 18 hours, cells from two wells of the naive and IFNɣ conditions were harvested for RNA extraction. The remaining two wells from each condition were additionally infected with *Salmonella* Typhimurium SL1344 (hereafter 'SL1344') for 5 hours. For RNA extraction, cells were washed once with PBS and lysed in 300 µl of RLT buffer (Qiagen) per one well of a 6-well plate. Lysates from two wells were immediately pooled and stored at -80°C. RNA was extracted using RNA Mini Kit (Qiagen) following manufacturer's instructions and eluted in 35 µl nuclease-free water. RNA concentration was measured using NanoDrop and RNA integrity was measured on Agilent 2100 Bioanalyzer using RNA 6000 Nano total RNA kit.

Two days before infection, *Salmonella* Typhimurium SL1344 culture was inoculated in 10 ml low salt LB broth and incubated overnight in a shaking incubator (200 rpm) at 37°C. Next morning, the culture was diluted 1:100 into 10 ml of fresh LB broth and incubated again in a shaking incubator. In the afternoon the culture was diluted once more 1:100 into 45 ml of LB broth and kept overnight in a static incubator. In the morning before infection, the culture was centrifuged at 4000 rpm for 10 minutes, washed once with 4°C PBS and re-suspended in 30 ml of PBS. Subsequently, optical density at 600 nm was measured and *Salmonella* was diluted in macrophage differentiation media (without M-CSF) at multiplicity of infection (MOI) 10 assuming 300,000 cells per well. To infect the cells, old media was removed and replaced with 1 ml of media containing *Salmonella* for 45 minutes. Subsequently, the cells were washed twice with PBS and replaced in fresh medium with 50 ng/ml gentamicin (Sigma) to kill extracellular bacteria. After 45 minutes, the medium was changed once again to fresh medium containing 10 ng/ml gentamicin.

3.2.3 RNA sequencing

All of the RNA-seq libraries were constructed using poly-A selection. The first 120 RNA-seq libraries from 30 donors were constructed manually using the Illumina TruSeq stranded library preparation kit. The TruSeq libraries were quantified using Bioanalyzer and manually pooled for

sequencing. For the remaining 216 samples, we used an automated library construction protocol that was based on the KAPA stranded mRNA-seq kit. The KAPA libraries were quantified using Quant-iT plate reader and pooled automatically using the Beckman Coulter NX-8. The first 16 samples were sequenced on Illumina HiSeq 2500 using V3 chemistry and multiplexed at 4 samples/lane. All of the other samples were sequenced on Illumina HiSeq 2000 using V4 chemistry and multiplexed at 6 samples/lane.

RNA-seq pre-processing and quality control

I aligned RNA-seq data to the GRCh38 reference genome and Ensembl 79 transcript annotations using STAR v2.4.0j (Dobin et al., 2013). I then used VerifyBamID v1.1.2 (Jun et al., 2012) to detect and correct any potential sample swaps and cross-contamination between donors. I did not detect any cross-contamination, but I did identify one sample swap between two donors. I used featureCounts v1.5.0 (Liao et al., 2014) to count the number of uniquely mapping fragments overlapping GENCODE (Harrow et al., 2012) basic annotation from Ensembl 79. I excluded short RNAs and pseudogenes from the analysis leaving 35,033 unique genes of which 19,796 were protein coding. I only used 15,797 genes with mean expression in at least one of the conditions greater than 0.5 transcripts per million (TPM) (Wagner et al., 2012) in all downstream analyses. I also quantile-normalised the data and corrected for sample-specific GC content bias using the conditional quantile normalisation (cqn) (Hansen et al., 2012) R package as recommended previously (Ellis et al., 2013). To detect hidden confounders in gene expression, I applied PEER (Stegle et al., 2012) on each condition separately allowing for at most 10 hidden factors. I found that the first 3-5 factors explained the most variation in the data and the others remained close to zero.

Variance component analysis

I used a linear mixed model implemented in the lme4 (Bates et al., 2015) package to estimate the proportion of variance explained by various biological and technical factors in the expression levels of 15,797 genes across 336 samples. The 14 factors that I included in the model are listed below. Continuous variables were binned into a small number of categories as described.

1. **Salmonella** - Salmonella infection status (yes or no) (binary)
2. **IFNɣ** - IFNɣ stimulation status (yes or no) (binary)
3. **IFNɣ:Salmonella** - interaction term between Salmonella and IFNɣ stimulations (binary)
4. **Line** - the iPSC cell line from which the macrophages were derived. All lines used in the analysis were from 84 unique donors. This component should capture genetic

differences between donors, but can also capture line and differentiation specific effects. (84 categories)

5. **Cell density** - I used mean RNA concentration across the four conditions as proxy of the total number of cells on a plate, because counting the cells prior to lysis and RNA extraction was not feasible. (categorical: 0-100 ng/ul, 100-200 ng/ul, 200-300 ng/ul, 300-500 ng/ul)
6. **Library type** - type of the RNA library construction method used (manual or automatic) (binary)
7. **Sex** - sex of the donor (binary)
8. **Purity** - purity of the differentiated macrophages as quantified by flow cytometry. This is a noisy measurement, because RNA-seq and flow cytometry were not performed from the same plate of cells and they were often performed on different days (up to 2 weeks apart) due to logistical reasons (categorical: 90-95%, 95-97.5%, 97.5-100%).
9. **Chemistry** - chemistry of the Illumina RNA-seq protocol (V3 or V4).
10. **Stimulation date** - date of the stimulation assays and cell lysis (categorical: 32 levels)
11. **Library pool** - RNA-seq library construction batch (categorical: 10 levels)
12. **RNA extraction** - RNA extraction batch (categorical: 31 levels)
13. **Differentiation duration** - Number of days from the start of the differentiation until cell lysis (5 categories: 20-30 days, 31-40 days, 41-50 days, 51-60 days, 61+ days).
14. **Passage** - passage of the IPSC line at the start of the differentiation (4 categories: 0-25, 26-35, 36-45, 46-60)

First, I analysed all 15,797 expressed genes from all of the 336 samples across the four conditions together using a single linear mixed model with all of the 14 factors included as random effects. The following model was fit to each gene independently, using `lme4`:

```
expression ~ (1|Salmonella) + (1|IFNɣ) + (1|IFNɣ:Salmonella) + (1| Line) +
      (1| Cell_density) + (1| Library_type) + (1| Stimulation_date) +
      (1| Sex) + (1| Chemistry) + (1| Purity) + (1| Passage) +
      (1| Diff_duration) + (1| Library_pool) + (1| RNA_extraction)
```

To better understand the relative contribution of weaker technical factors and how their effects might vary between conditions, I also performed variance component analysis in each condition separately by only including the ten technical factors in the model as random effects:

```
expression ~ (1| Cell_density) + (1| Library_type) + (1| Stimulation_date) +
      (1| Sex) + (1| Chemistry) + (1| Purity) + (1| Passage) +
      (1| Diff_duration) + (1| Library_pool) + (1| RNA_extraction)
```

Next, I used the `VarCorr` function from the `lme4` package to calculate the amount of variance attributed to each of the factors. I then estimated the proportion of variance explained by each factor by dividing the variance attributed to each factor by the total variance of the gene. As a result, for each factor I obtained a distribution of the proportion of variance explained estimates across 15,797 genes.

3.2.4 Flow cytometry

Measuring macrophage cell surface marker expression using flow cytometry

We used flow cytometry to measure the cell surface expression of three canonical macrophage markers: CD14, CD16 (FCGR3A/FCGR3B) and CD206 (MRC1). Macrophages were cultured in 10 cm tissue-culture treated plates and detached from the plates by incubation in 6 mg/ml lidocaine-PBS solution (Sigma L5647) for 30 minutes followed by gentle scraping. From each cell line we harvested between 300,000-500,000 cells. Detached cells were washed in media, centrifuged at 1200 rpm for 5 minutes and resuspended in flow cytometry buffer (2% BSA, 0.001% EDTA in D-PBS) and split into two wells of a 96-well plate. Nonspecific antibody binding sites were blocked by incubating cells with Human TruStain FcX (Biolegend) for 45 minutes and washing with flow cytometry buffer. Half of the cells were stained for 1 hour with the PE-isotype control (BD 555749) antibody. The other half of the cells were co-stained for 1 hour with following three antibodies: CD14-Pacific Blue (BD 558121), CD16-PE (BD 555407), CD206-APC (BD 550889). After staining, the cells were washed three times. Resuspended cells were filtered through cell strainer cap tubes (BD 352235) and measured on the BD LSRFortessa Cell Analyzer.

Flow cytometry data analysis

I used the flow cytometry data for two purposes: to estimate the proportion of cells expressing macrophage surface markers CD14, CD16 and CD206 and to quantify the relative intensity of these markers compared to unstained cells. I imported the raw FCS files into R using the OpenCyto (Finak et al., 2014) package. First, I logicle-transformed (Herzenberg et al., 2006) the

intensity values for all three channels in both stained and isotype control samples using the `estimateLogicle` function. I then performed two automated gating steps to exclude debris and identify the main cell population using the `mindensity` (max = 150,000) and `flowCust` (K=2, target=c(1e5,5e4), level=0.9) functions. For pure macrophage samples the distribution of intensity values for all three cell surface markers looked bimodal with stained and unstained cells in two separate peaks (Figure 3.1). Samples with moderate contamination had an additional low intensity peak both in stained and unstained cells (Figure 3.1) corresponding to the contaminating cells. Since all of the peaks were approximately normally distributed, I decided to model the data for each mark as a mixture of Gaussian distributions and used the mclust (Fraley and Raftery, 1999) R package to estimate the optimal number of components (2 or 3) as well as the mean and standard deviation of each component. I used the Bayesian Information Criterion to choose between two or three components. I then compared the mean of the highest intensity peak ($\mu_{stained}$) to the mean of the second highest intensity peak ($\mu_{unstained}$) to estimate the relative fluorescent intensity of each cell surface marker (Figure 3.1). I also measured sample purity by estimating the proportion of cells whose intensity was greater than the threshold $t = \mu_{stained} - 3\times\sigma_{stained}$ (Figure 3.1), where $\sigma_{stained}$ is the standard deviation of the stained population.

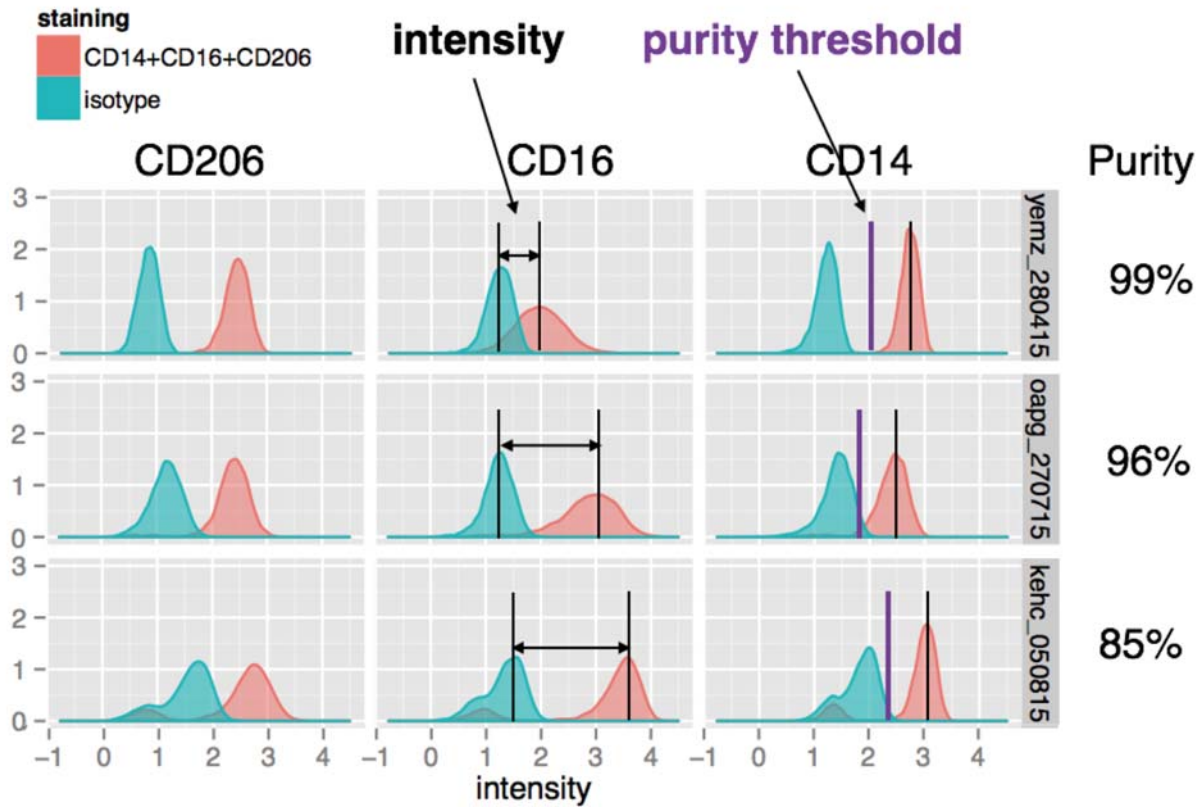

Figure 3.1: Quantifying cell purity and relative fluorescent intensity of macrophage CD206, CD16 and CD14 markers from the flow cytometry data. The rows correspond to three different iPSC lines and the columns represent three macrophage markers. X-axis shows the logicle-transformed absolute intensity values from the flow cytometer and values on the y-axis correspond the density of the cells with that intensity value. Red designates cells stained with antibodies against the three markers, blue indicates cells stained with isotype control (unstained). Marker relative fluorescent intensity is defined as the difference in mean intensity between the stained and unstained cell populations (middle panel). Purity is measure by estimating the proportion of stained cells (red) whose intensity is greater than the purity threshold (purple) $t = \mu_{stained} - 3 \times \sigma_{stained}$.

Variance component analysis and QTL mapping

I used linear mixed model implemented in `lme4` (Bates et al., 2015) package in R to characterise the observed variation in the relative fluorescent intensity measurements of the three macrophage markers. For each marker, I estimated the proportion of variance explained by differences between the iPSC lines (hereafter 'line effect') as well as the batch effect represented by the date when the cells were harvested, stained and measured on the flow

cytometer ('date effect'). I used the following `lme4` model specification: `intensity ~ (1|date) + (1|line)`.

I used FastQTL (Ongen et al., 2016) to test for association between relative fluorescent intensity and common genetic variants (minor allele frequency > 0.05, IMP2 > 0.7) in the +/- 200kb region around the corresponding genes: CD14, FCGR3A and FCGR3B for CD16, and MRC1 for CD206. I used measurements from 95 unique lines (donors) for QTL mapping. If a particular line had multiple measurements, then I picked one randomly. After permutation testing (n=10,000), I identified significant cis QTLs for CD14 and CD16 markers. Subsequently, I redid the variance component analysis for each marker and included the lead QTL variants into the model: 'intensity ~ (1|date) + (1|line) + (1|rs2569177) + (1|rs4657019)'.

3.3 Large-scale differentiation of macrophages for genomics assays

We aimed to develop a robust and standardised differentiation pipeline that would allow us to produce at least 3 million macrophages from each donor for four different experimental assays: (1) Flow cytometry (this chapter), *Salmonella* RNA-seq (Chapter 4), *Salmonella* ATAC-seq (Chapter 5) and acLDL RNA-seq (not described here). We relied on a previously published macrophage differentiation protocol (van Wilgenburg et al., 2013) that I compared to monocyte-derived macrophages in Chapter 2. The timeline of the differentiation protocol is illustrated in Figure 3.2 and the full details of the protocol are given in the Methods. Briefly, the main steps of the differentiation are (1) expansion of iPSCs in feeder-dependent medium (median 19 days), (2) embryoid body (EB) formation (3 days), (3) differentiating EBs into macrophage progenitors (median 27 days) and (4) harvesting and final differentiation of progenitors into macrophage (7 days). One attractive feature of this system is that differentiated EBs can be kept in culture for prolonged period of time and progenitors can be harvested in every 4-5 days making it possible to perform additional assays on the cells without increasing the amount of tissue culture needed for the initial steps of the differentiation (van Wilgenburg et al., 2013).

Although other protocols exist that can be used to differentiate macrophages in a shorter period of time (Zhang et al., 2015), a major advantage of our protocol is that the bulk of the differentiation and maintenance is performed in single medium containing only two cytokines (interleukin-3 (IL-3) and macrophage colony stimulating factor (M-CSF)) and the exact timing

between medium changes can be varied without significantly influencing differentiation success. This property made the protocol scalable to differentiating many iPSC lines in parallel without a large increase in complexity, because all of the dishes receive the same media and medium changes could be conveniently scheduled.

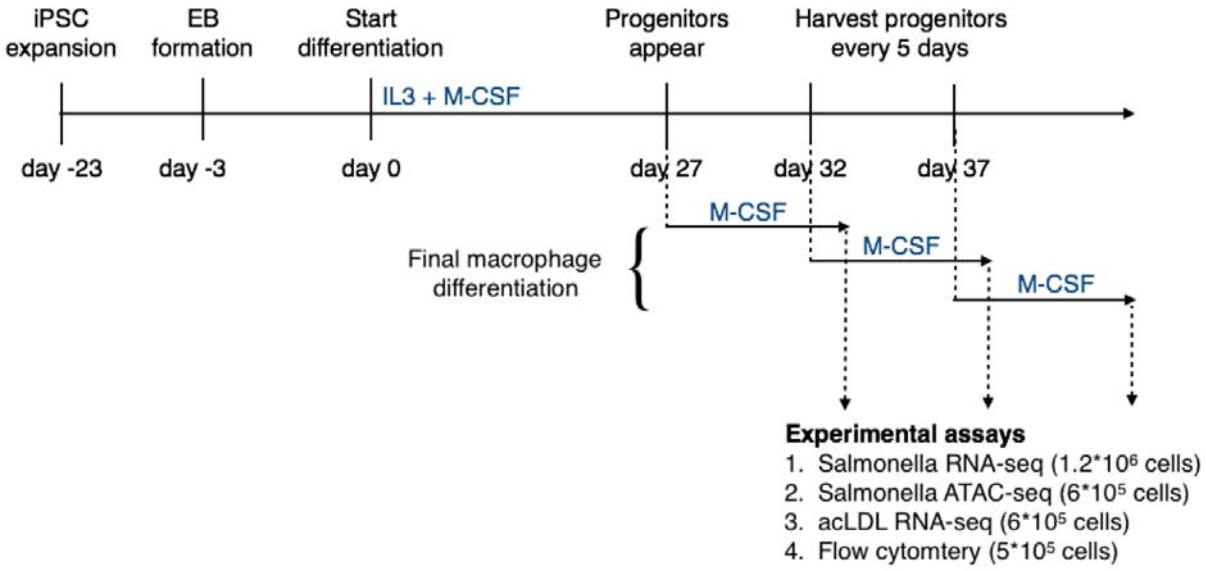

Figure 3.2: Timeline of macrophage differentiation from iPSCs. The protocol starts with the expansion of IPSCs followed by embryoid body formation. The bulk of the differentiation is performed in X-VIVO 15 media supplemented with IL-3 and M-CSF cytokines. The differentiation takes usually 4-5 weeks (median 27 days) until macrophages progenitors appear. During this time the medium has to be changed in every 4-5 days. Once the macrophage progenitors appear, they are harvested at every medium change and differentiated in the presence of M-CSF for another 7 days until the cells are ready for experimental assays.

We differentiated macrophages from batches of multiple iPSC lines in parallel. In addition to logistical convenience, this approach enabled us to estimate and control for batch-to-batch variation in gene expression and differentiation success measurements.

3.3.1 Variability in success rate

We performed 138 macrophage differentiation attempts from 123 different HipSci iPSC lines. We were able to successfully differentiate macrophages from 101/123 (82%) of the iPSC lines. Here successful differentiation is defined as obtaining at least some proportion of cells that exhibited characteristic spindle-like macrophage morphology. For 97/101 lines, we further

confirmed the expression of CD14, CD16 and CD206 macrophage cell surface markers with flow cytometry.

To understand what was responsible for the failed differentiation, we tried to re-differentiate 8 iPSC lines that had failed on the first attempt. Surprisingly, 7/8 failed also at the second attempt. This was over 6 times higher than the global 18% failure rate observed across all lines (Fisher's exact test $p = 0.002$), suggesting that there might be a line specific bias against macrophage differentiation. However, six of these lines (all of which failed) were all re-differentiated in the same month (January 2015), meaning that this observation might have also arisen from a shared batch effect. We note though, that 3/4 lines cultured concurrently with the 6 failed lines differentiated successfully into macrophages. Hence, this suggests that there might be a line-specific (or donor-specific) bias against macrophage differentiation but further experiments on more iPSC lines are needed to confirm this.

3.3.2 Variability in the duration of the differentiation

Throughout our experiments we observed considerable variation in the time from initial iPSC culture to the production of mature macrophages. This variation was influenced by a variety of experimental factors, most importantly whether the differentiation was started from live or frozen cells. Initially, we received live cells in feeder dependent media from Wellcome Trust Sanger Institute core facilities. These live cell cultures required only a single passage before EB formation could be initiated (Figure 3.3A). Subsequently, however, for operational reasons we switched to cryopreserved cells cultured either on feeder-dependent or feeder-free E8 medium. Since our attempts to differentiate macrophages directly from feeder-free iPSCs were not successful, we had to transfer feeder-free cells to feeder-dependent medium for at least two passages. This added approximately 7-10 days to the time required for initial iPSC culture and expansion. However, the total time needed for iPSC expansion was comparable for feeder-free and feeder-dependent cryopreserved cells, because thawing feeder-dependent iPSCs generally took much longer than thawing feeder-free iPSCs (Figure 3.3A). We did not observe any discrepancy in the differentiation success rate between IPSCs initially grown either on feeder-dependent or feeder-free media.

The median time from the start of the differentiation (3 days after EB formation) until the appearance of first macrophage progenitors was 27 days (Figure 3.3B), and 96% of the lines that successfully differentiated into macrophages did so within 40 days. Thus, for this protocol, a

40-day threshold provided a useful guideline for deciding when a differentiation attempt had failed and should be aborted. Final macrophage differentiation added another 7 days to the protocol and for logistical reasons we were not always able to perform the stimulation assays on the first batch of cells that we harvested. This increased the median time from differentiation start to cell lysis to 38 days (Figure 3.3C). We recorded this information for each cell line to assess retrospectively if the time spent in culture had an effect on the macrophage transcriptome.

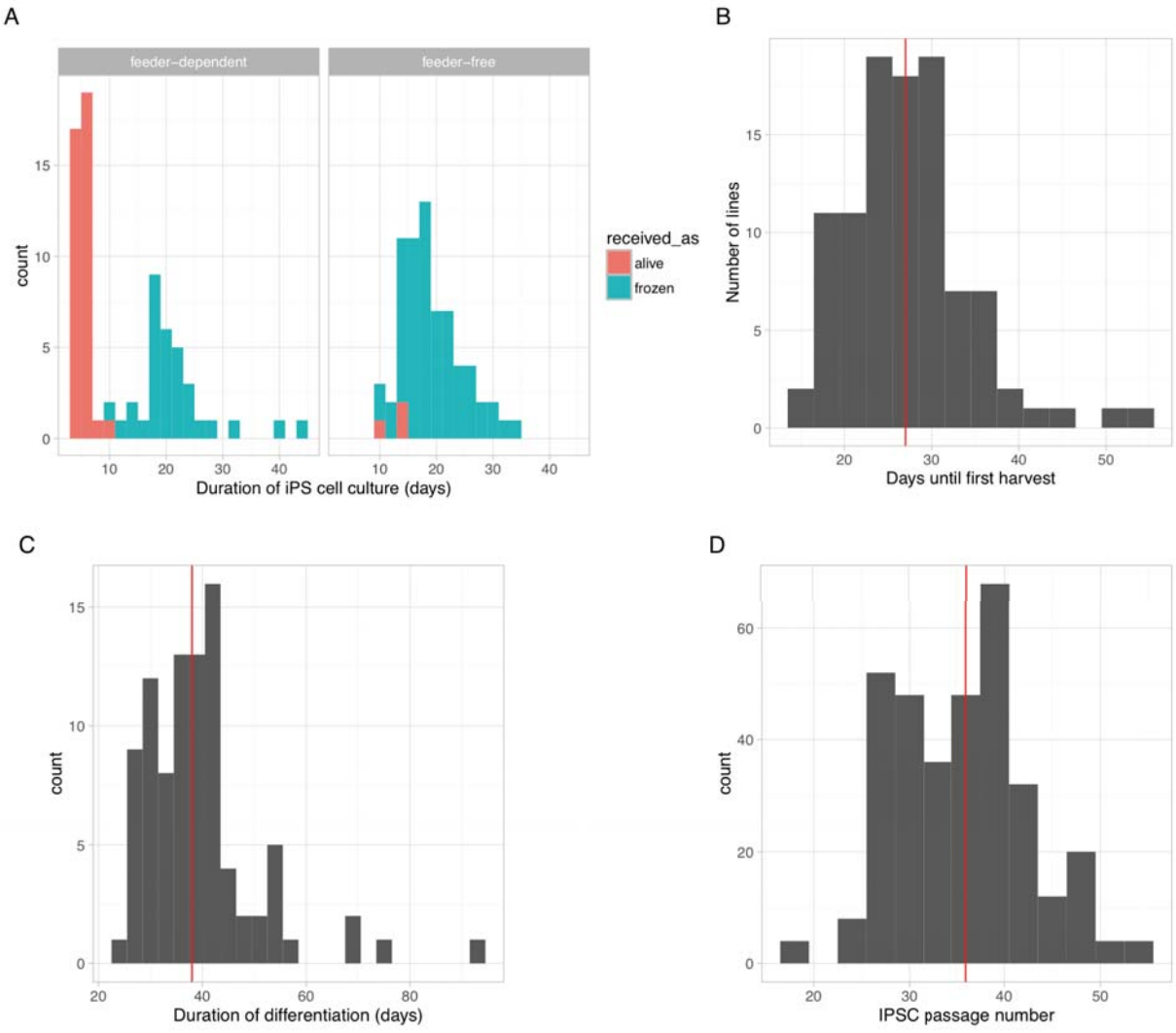

Figure 3.3: Variation in the duration of macrophage differentiation. (A) Duration of iPSC culture prior to the start of the differentiation. The two panels correspond to iPSC lines that were initially either on feeder-dependent medium or feeder-free medium. The colour represents whether the cell lines were received as live culture or cryopreserved stock. (**B)** Number of days

from the start of the differentiation until the harvest of first macrophages. Red line corresponds to the median of the distribution (27 days). **(C)** Histogram of the number of days from the start of the differentiation until the *Salmonella* infection experiment and cell lysis (median 38 days). **(D)** Histogram of the number of passages iPSCs had been propagated prior to the start of the differentiation.

3.3.3 Variability in cell numbers

Although we differentiated all lines in the same number of tissue culture plates, we observed an order of magnitude variation between lines in the mean number of macrophage progenitors produced per harvest (min 3×10^5, median 3×10^6, max 15×10^6) (Figure 3.4A). Most of the variation was likely caused by differences in the size and number of EBs per line, which was challenging to control during differentiations. Our approach to deal with this variation was to use more than minimally required cells for EB formation, thus ensuring that even differentiation with lower yield would produce enough cells for all of the planned experimental assays.

For the final macrophage differentiation, we always seeded 150,000 progenitors into a single well of a 6-well plate. However, due to variation in the fraction of adherent cells and their proliferation rate between iPSC lines, we observed substantial variation in the numbers of cells on the plate at the time of the stimulation assays. Since this variation was hard to control for experimentally (macrophages are strongly adherent cell type making them difficult to replate), we decided to measure the mean RNA concentration for each line as a proxy of the cell count (Figure 3.4B).

3.3.4 Variability in macrophage purity

Finally, we examined the purity of the differentiated macrophages. Despite not using cell sorting or other methods to experimentally enrich for macrophages, we found that 88% of the differentiations produced macrophages that were >90% pure based on the cell surface expression of CD14, CD16 and CD206 markers (Figure 3.4C). Although we did not use flow cytometry to directly select samples for RNA sequencing (flow cytometry was often performed after RNA had been collected), we found that only 4/84 of the selected samples had purity below 90% (Figure 3.4C).

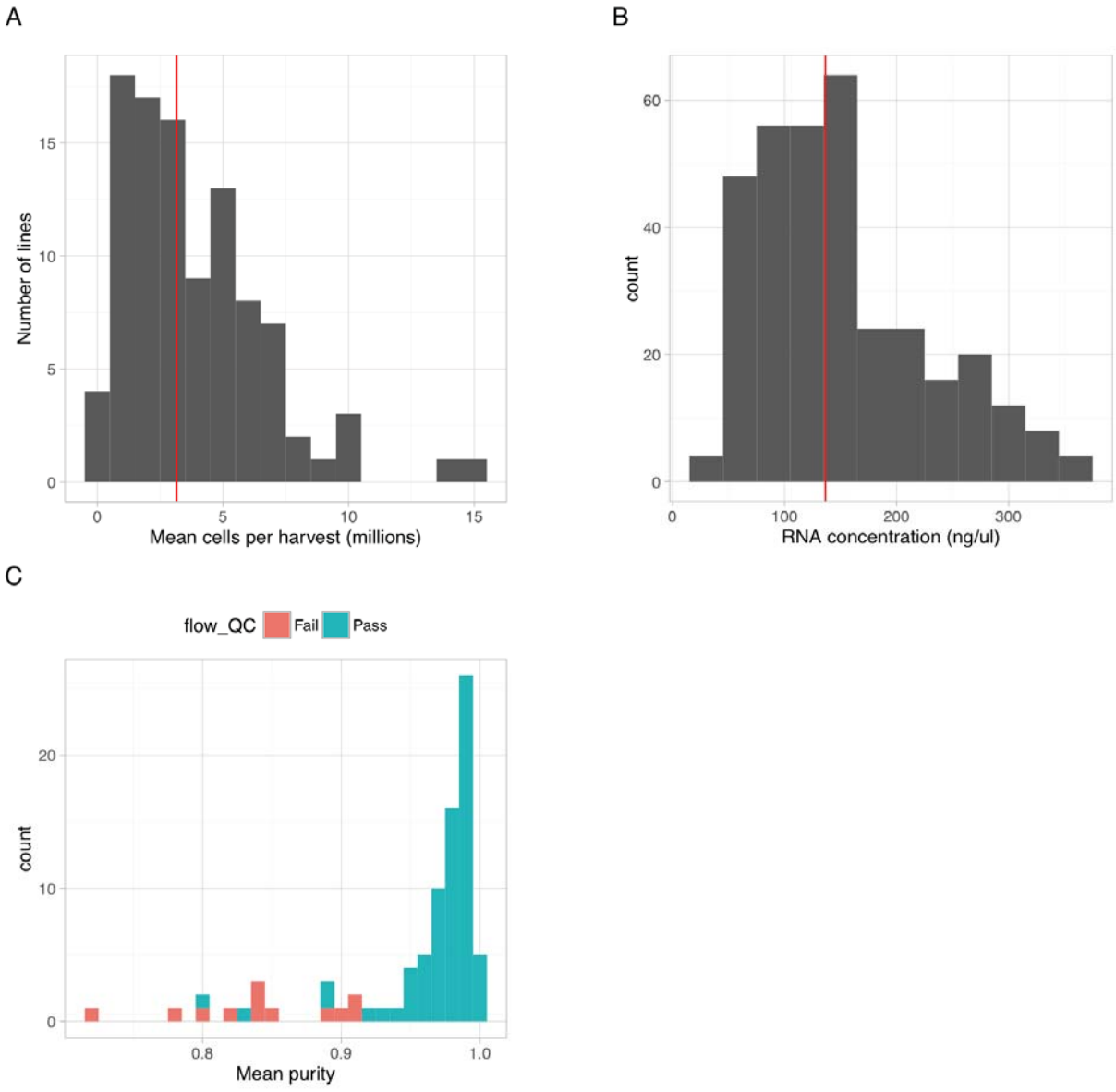

Figure 3.4: Distributions of some the key experimental variables the could influence the transcriptomes of differentiated macrophages. (A) Histogram of the mean number of cells obtained per harvest for all of the iPSC lines that successfully differentiated into macrophages. Red line corresponds to the median (3.16 million). **(B)** Histogram of the mean RNA concentration values for all of the cell lines across four experimental conditions. **(C)** Differentiated cells were stained with antibodies for CD14, CD16 and CD206 and the proportion

of cells staining positive for each of the three marks was estimated. The mean of the three marks was used as the purity score for each cell line. The figure shows the histogram of the purity scores across iPSC lines. Samples represented in green were used for RNA-seq experiments.

3.4 Variability in gene expression data

While many aspects of the differentiation might be variable between iPSCs lines, not all of them will have a significant effect on downstream macrophage gene expression levels. Thus, I decided to use variance component analysis to estimate the relative contribution of various biological and technical factors on macrophage gene expression levels.

3.4.1 Technical variability between RNA-seq samples

In addition to biological variability in the differentiation protocol described above, macrophage gene expression levels could also be influenced by technical variability in the way RNA samples were processed. The potential sources of variability that we identified were RNA extraction batch, RNA integrity, library construction batch, method of library construction used (manual or automatic) and sequencing chemistry used.

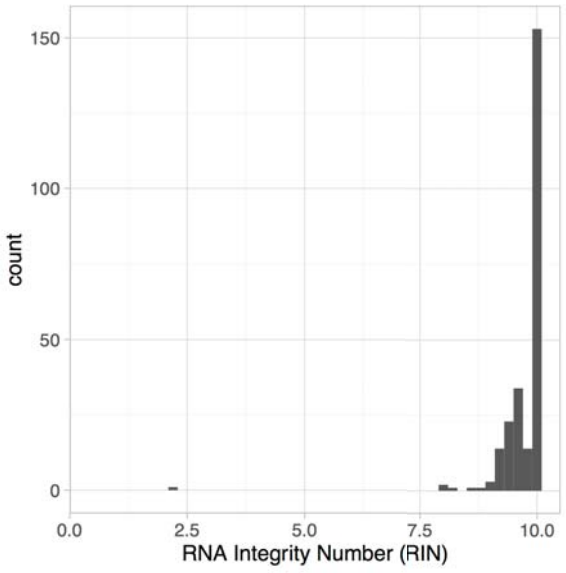

Figure 3.5: Distribution of RNA integrity (RIN) values for a subset of the RNA-seq samples.

Fortunately, I observed very little variation in RNA integrity between samples and for the vast majority of the samples the RNA integrity number (RIN) was greater than 9 out of 10 (Figure 3.5). However, I did observe some differences between automatic and manual library construction methods. First, the variability in total read coverage between samples was greatly reduced when the automatic protocol was used (Figure 3.6A). I also found that the automated protocol had lower GC content bias than the manual protocol which showed slight preference for low GC content fragments over high GC content fragments (Figure 3.6B).

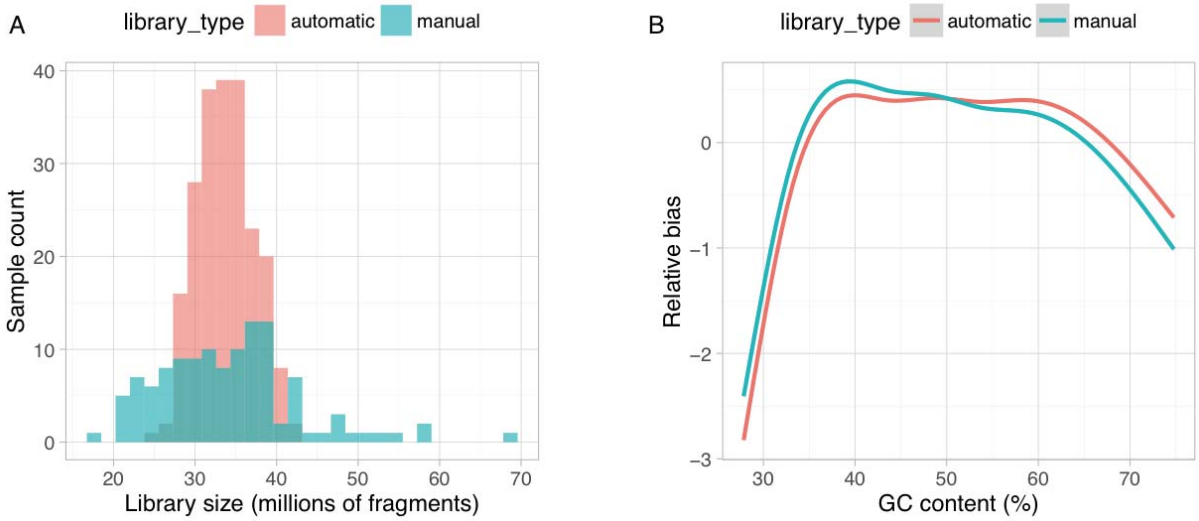

Figure 3.6: Comparison of manual and automated RNA-seq library construction protocols. (A) Histogram of total library size distribution for samples prepared either with manual or automatic protocol. **(B)** Mean GC content bias for the two library construction protocols. GC content bias was estimated from the raw read counts using the cqn (Hansen et al., 2012) package in R.

3.4.2 Variance component analysis of the RNA-seq data

Variance component analysis is a powerful approach to estimate the relative importance of various known experimental factors in an unbalanced experimental design (Rouhani et al. 2014; Kilpinen et al. 2016). When applied to our dataset, variance component analysis revealed that most of the variance in gene expression was explained by the three experimental stimuli: *Salmonella* infection (32.9%), IFNɣ stimulation (15.5%) and interaction between the two (11.4%) (Figure 3.7), highlighting the plasticity of the macrophage transcriptome in response to strong immunological stimuli. The second largest amount of variance explained (7.7%) was attributed

to differences between cell lines (hereafter ‘line effect’) while all of technical factors explained significantly less variance

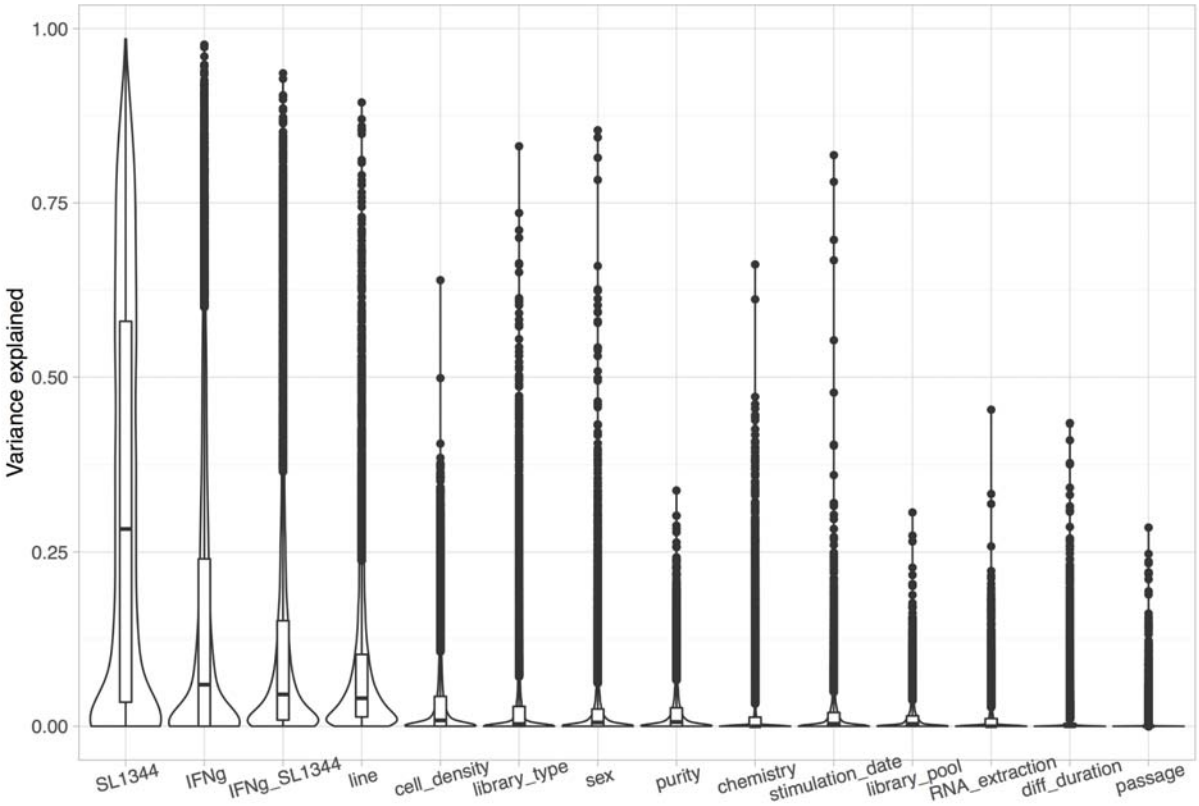

Figure 3.7: Variance component analysis of all four conditions in a joint model. We used a linear mixed model to estimate the proportion of variance explained by 14 different factors in the expression levels of 15,797 expressed genes (see Methods). For each factor on the x-axis, the violin plot shows the distribution of variance explained by that factor across all expressed genes. Factors are ordered by the mean variance explained across all genes.

To see how the relative contribution of the weaker technical factors varied between conditions, I performed variance component analysis in each of the four conditions separately. Now that the differences between stimulations were controlled for, most of variance was explained by RNA-seq library type (automatic vs manual), cell density, sex and purity of the macrophage population (Figure 3.8A) and the estimates for most of the factors were similar in all four conditions. The large contribution of library type is likely to be at least partially explained by differences between GC bias reported above (Figure 3.6B). The date when macrophages were

stimulated with IFNɣ and infected with *Salmonella* (‘stimulation date’) explained almost double the variance in *Salmonella* and IFNɣ + *Salmonella* conditions than in naive and IFNɣ conditions (Figure 3.8B). This is probably because live *Salmonella* culture was prepared fresh for each day of infections whereas IFNɣ originated from single-use frozen aliquots. Indeed, both of the *Salmonella* conditions had an excess of highly variable genes compared to naive and IFNɣ conditions (Figure 3.8C), indicating that *Salmonella* batch introduced additional variability into the data. Finally, the passage number of the iPSCs prior to differentiation (Figure 3.3D) and the total duration of the differentiation (Figure 3.3C) (after accounting for differences in purity) explained less than 3% of the variance, suggesting that controlling for these factors during differentiation is less important.

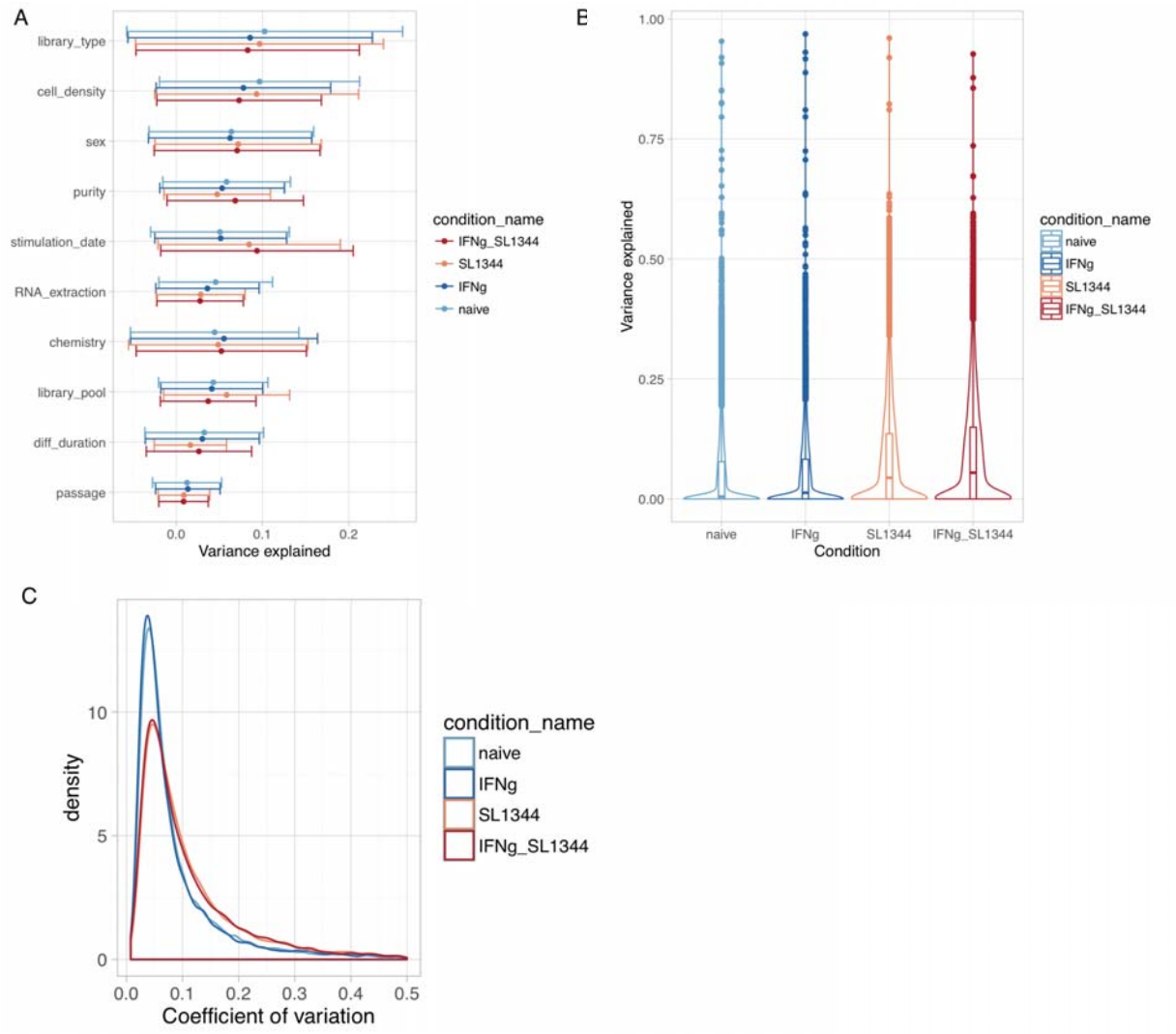

Figure 3.8: Variance component analysis of each of the four conditions separately. (A) Variance explained by ten technical factors in each of the four conditions. Points correspond to mean across all genes and horizontal lines represent standard deviation. Note that the lines are not true confidence intervals, because variance explained cannot be negative. **(B)** Proportion of variance explained by the date when the macrophages were stimulated with IFNɣ and infected with *Salmonella* ('stimulation date' effect). The violin plots show the distributions of the variance explained estimates across all genes in four experimental conditions. **(C)** Distribution of coefficient of variation (standard deviation / mean) for 10,000 most highly expressed genes in each condition.

3.4.3 Detecting hidden sources of variation

A complementary approach to dissect sources of variability in a large gene expression dataset are latent variable models (Leek and Storey, 2007; Stegle et al., 2010). Latent variable models are especially useful when the relevant covariates are not known beforehand or when they have not been measured accurately (Parts et al., 2011; Stegle et al., 2010). I applied PEER (Stegle et al., 2010) to the RNA-seq data from each condition to detect hidden sources of variation that affect many genes at the same time. I then calculated the proportion of variance explained by each hidden factor in each of the four experimental conditions (Figure 3.9). Note that PEER does not report residual variance and as a result these estimates are not directly comparable to the estimates from variance component analysis above. I found that in the naive cells 90% of the variance captured by PEER was explained by the first factor. Although macrophage purity and cell density (mean RNA concentration) were both correlated with the first factor (Figure 3.10A-B), in a joint linear model these two known covariates could explain only 42% of the variance captured by the first factor. This illustrates that PEER is able to capture additional variability beyond what can be explained by known covariates.

The second PEER factor explained an additional 3.1% of the variance and was correlated with the RNA-seq library type (Figure 3.10C) ($r = 0.79$, $p < 2.2 \times 10^{-16}$). However, as shown on Figure 3.6B, one of the differences between automatic and manual RNA-seq protocol was difference in GC bias. The quantile normalised gene expression values that we used as input to PEER were already corrected for sample-specific differences in GC bias. Therefore, the amount of variance explained by the second PEER factor might be higher in uncorrected samples.

I noticed that in stimulated conditions factors 2-5 explained much more variance than they did in the naive condition (Figure 3.9). This was especially prominent after *Salmonella* infection where ~40% of variance was explained by factors 2-5 (7.5% in naive). One possible interpretation is that stimulating cells introduces additional independent sources of variability ('batch effects') that are then captured by PEER as additional factors. This is consistent with the excess of highly variable genes observed after *Salmonella* infection (Figure 3.8C) and more variance explained by stimulation date (Figure 3.8B) reported above.

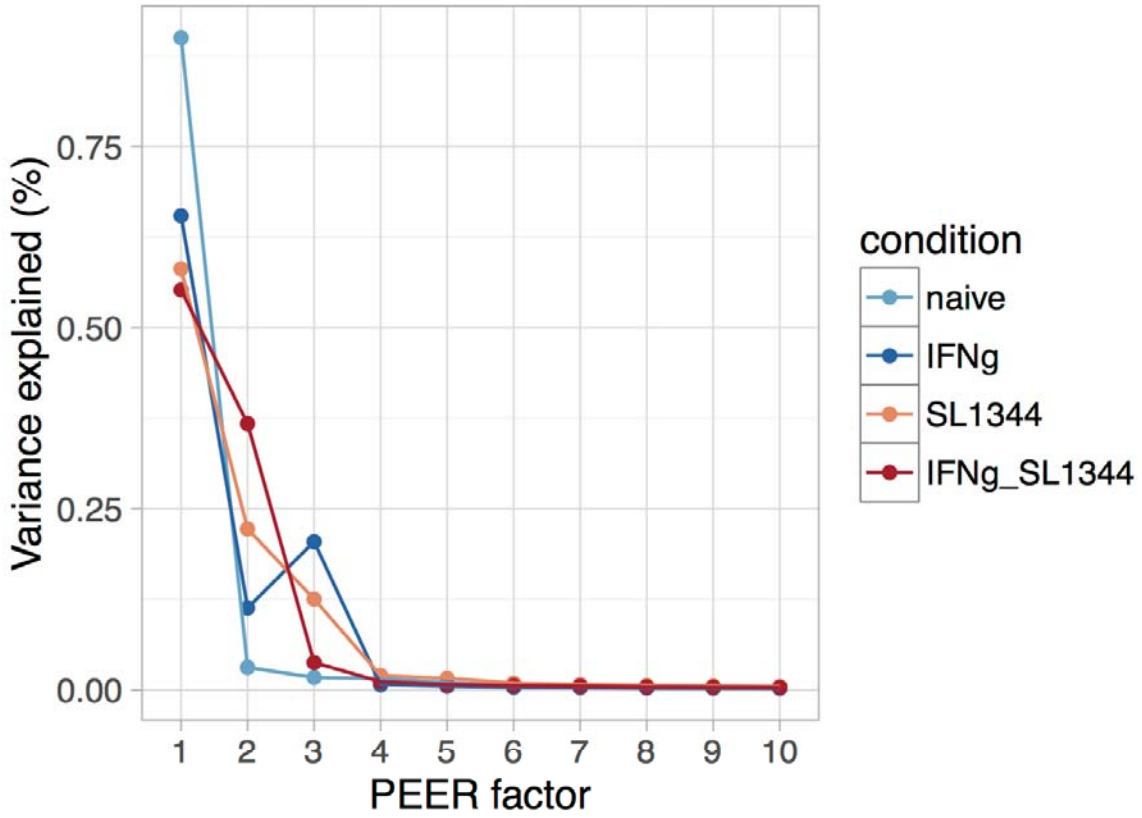

Figure 3.9: Proportion of variance explained by the first 10 PEER factors in each experimental condition. PEER was run on each condition separately, which mean that the factor names do not necessarily correspond to the same sources of variation in each condition.

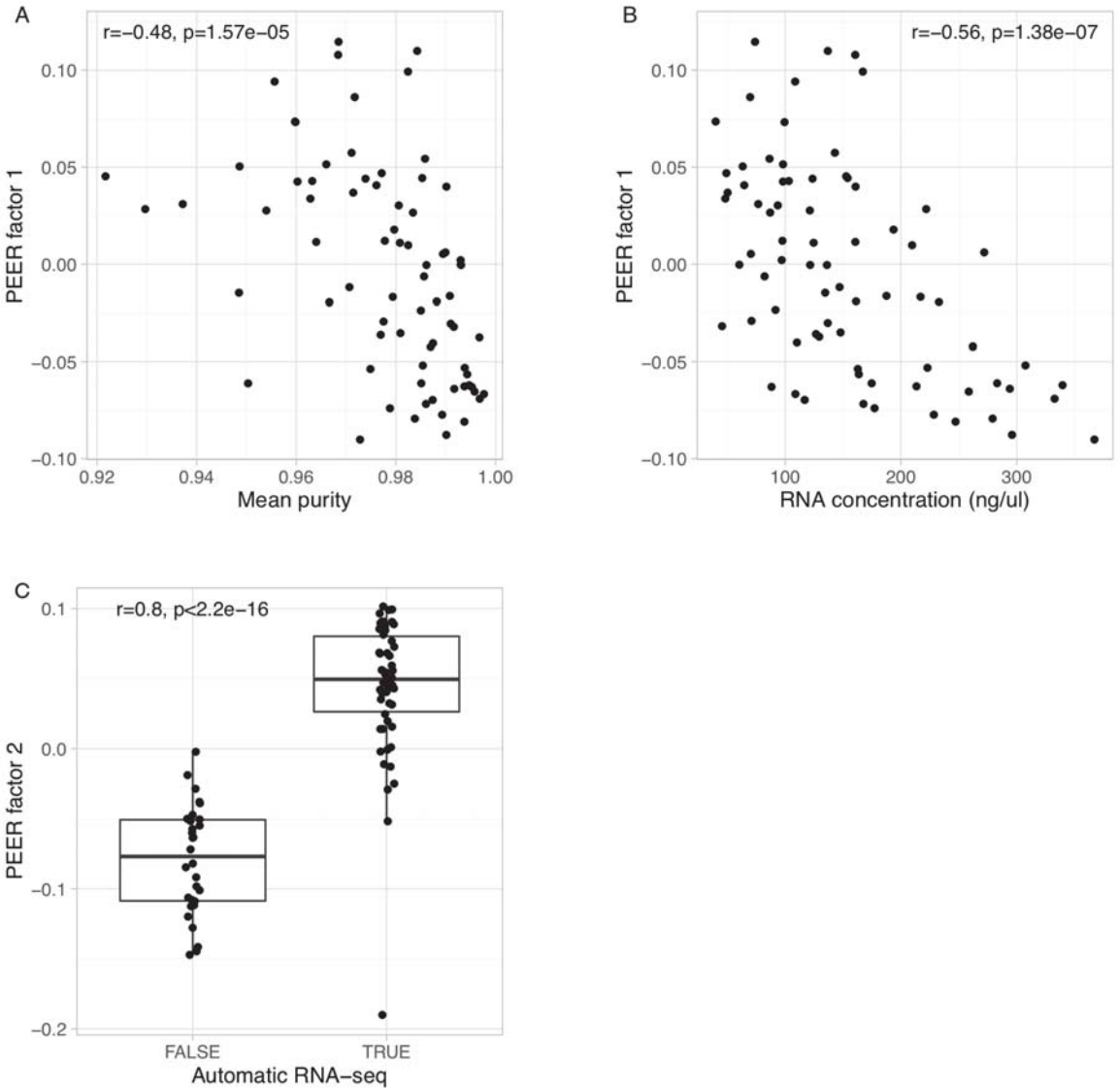

Figure 3.10: Correlation between first two PEER factors and experimental variables. (A) Correlation between mean purity and PEER factor 1 **(B)** Correlation between RNA concentration and PEER factor 1 **(C)** Correlation between RNA-seq library construction protocol (automatic vs manual) and PEER factor 2.

3.4.4 Reproducibility of differentiation

To assess how reproducible gene expression profiles were between differentiations, I analysed RNA-seq data from multiple independent differentiations from three different donors (three differentiations from donor ffdj and two from donors fpdl and ougl). We performed the same

stimulation experiments and RNA-seq on all of the samples. Although the differentiations were performed over the course of 10 months, I found that in the naive condition the samples clearly clustered together by donor (Figure 3.11), indicating that donor specific effects on gene expression are reproducible between differentiations. The clustering was not as clear in the stimulated conditions, possibly because of stronger batch effects induced by stimulation. For the ffdj donor we know that two of the differentiations were from the same iPSC line (samples fpdj_A and fpdj_A_2) whereas the third (nibo_A) was from a different line. For ougl and fpdl donors we unfortunately do not know if the two differentiations were from the same line or different lines, because we received these lines twice due to accidental sample swaps upstream and we only discovered the duplicate samples after matching genotypes in the RNA-seq data to the VCF file.

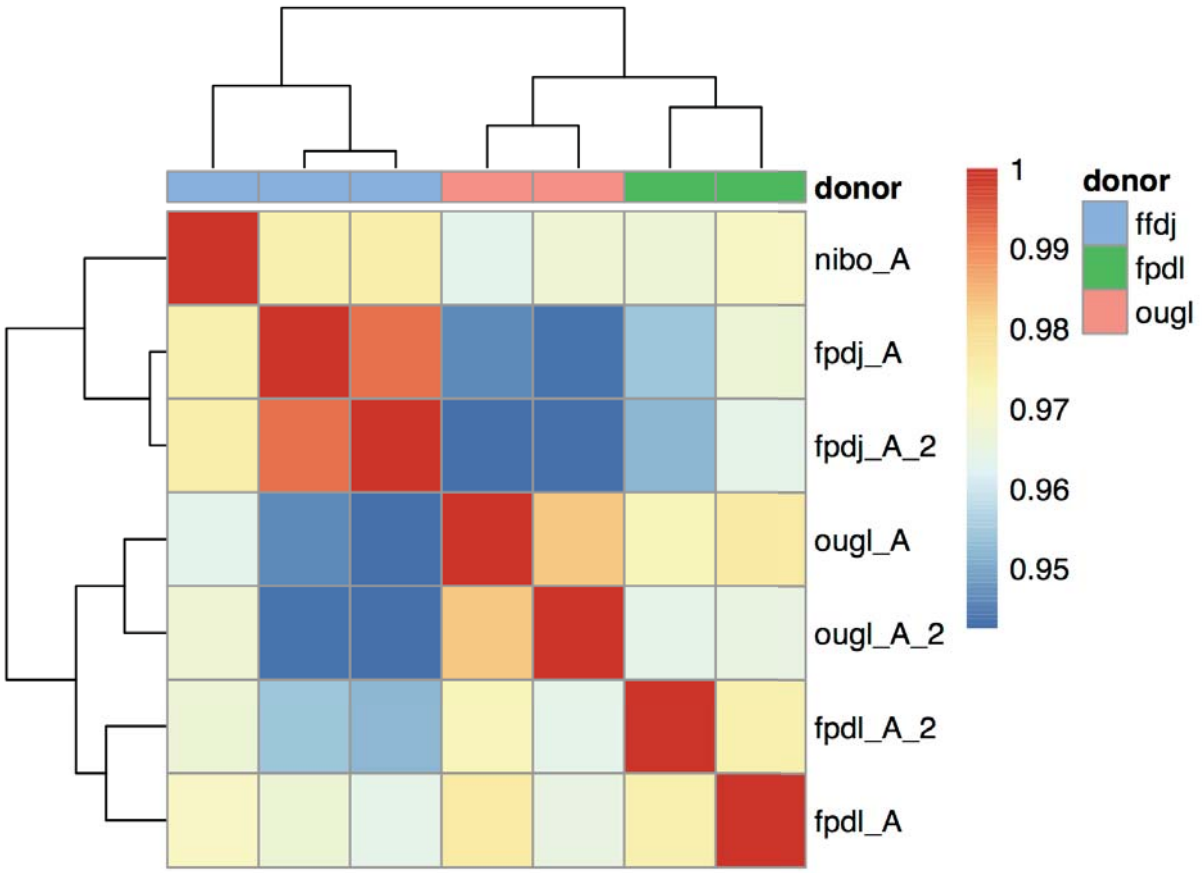

Figure 3.11: Gene expression reproducibility between independent differentiations. The heat map shows the Spearman correlation of gene expression profiles from 7 independent differentiations from 3 different donors.

3.5 Variability in cell surface marker expression

In addition to gene expression data, we also wanted to understand what is responsible for the variance in the cell surface expression of macrophage markers. Specifically, we wanted to know if, on average, the fluorescent intensity measurements of the same line on different days were more similar to each other than the measurements of different lines on the same day.

We measured cell surface expression of CD14, CD16 and CD206 macrophage markers in 97 different cell lines. This included 19 cell lines where duplicate measurements were obtained on different days. We processed and measured a median of four cell lines in a single batch on the same day. This allowed me to use a linear mixed model to estimate the relative proportion of variance explained by cell line and date of the assay (batch) for each of the three markers. I found that 62% of the variance in CD14 surface expression and 52% of the variance in CD16 surface expression was explained by the line effect and almost no variance was attributed to the date of the assay (Figure 3.12A). On the other hand, 64% of the variation in CD206 measurement was explained by the date of the assay and there was almost no line effect, suggesting that this antibody might have been more susceptible to technical variation. Between 25-50% of the variance remained unexplained for all three marks.

Next, I tested whether there was a genetic basis for the observed variation in the surface expression of CD14, CD16 and CD206 by performing QTL mapping for each of the three markers in +/- 200kb cis window around the corresponding genes (CD14, FCGR3A and FCGR3B for CD16, and MRC1 for CD206). I detected a very strong association between CD14 surface expression and rs2569177 variant (MAF = 0.24) located 19 kb upstream of the CD14 gene (permutation FDR = 2.7×10^{-11}) (Figure 3.12B). I also detected a weak association between CD16 expression and rs4657019 variant (MAF = 0.28) located 120 kb upstream of the FCGR3A gene (permutation FDR = 0.047) (Figure 3.12C). There was no significant QTL for CD206 consistent with only a small fraction of the variance being attributed to differences between lines. I redid the variance component analysis with the two CD14 and CD16 QTL SNPs included in the model. I found that the CD14 QTL explained most of the variation in CD14 intensity that was previously attributed to line effect (3.12D). On the other hand, the CD16 QTL explained only ~⅓ of the CD16 line effect, suggesting that there might be additional cis or trans QTLs for this protein that we were unable to detect because of our small sample size (Figure 3.12D).

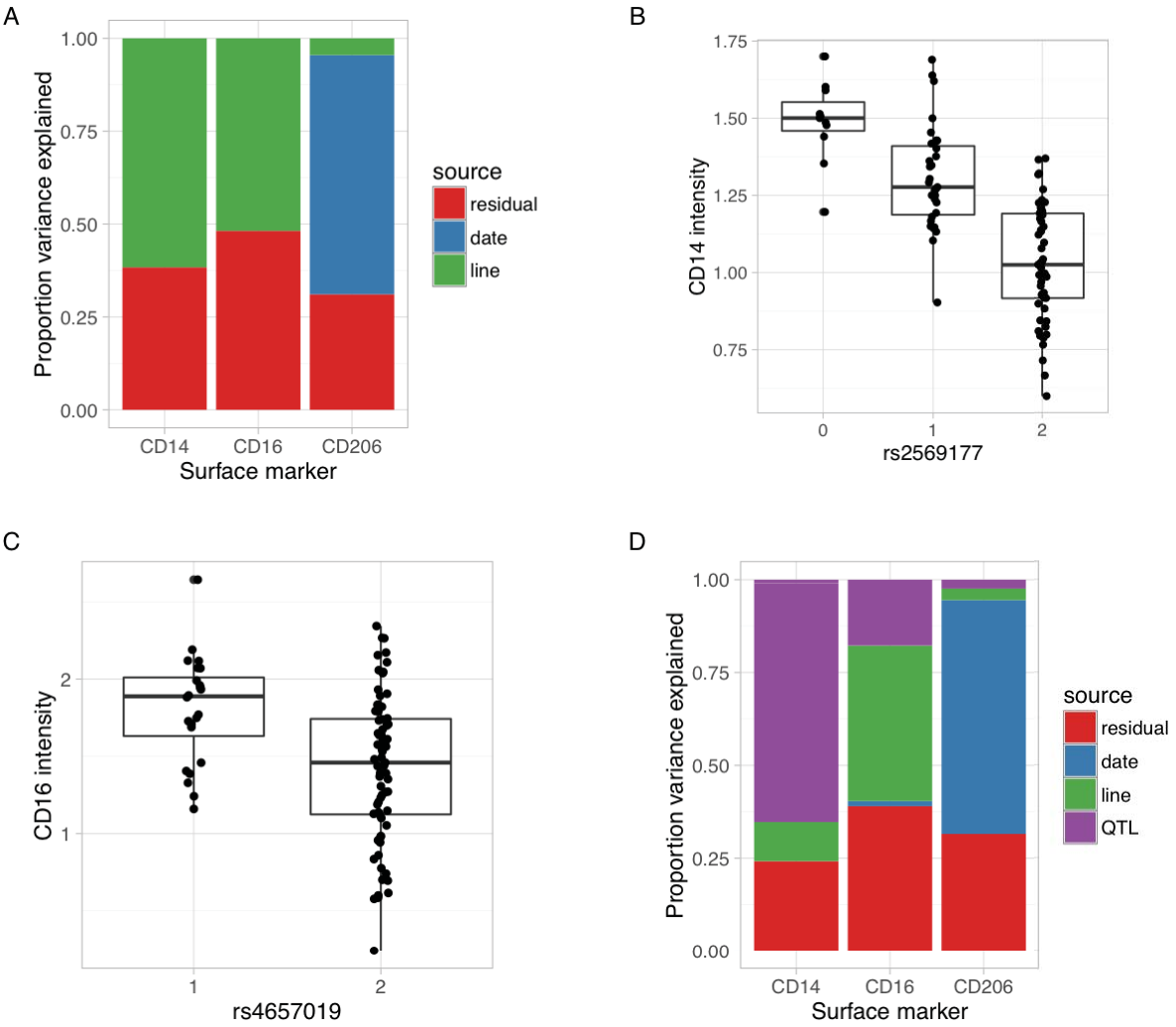

Figure 3.12: Variance of macrophage cell surface marker expression. (A) Variance of cell surface expression of CD14, CD16 and CD206 partitioned into three components: (1) iPSC line from which the macrophages were differentiated; (2) date of the flow cytometry assay; (3) residual variation. **(B)** Fluorescent intensity of CD14 cell surface expression stratified by the genotype of the lead QTL variant (FDR < 2.7×10^{-11}). **(C)** Fluorescent intensity of CD16 cell surface expression stratified by the genotype of the lead QTL variant (FDR < 0.048). **(D)** Variance partitioning after including CD14 and CD16 lead QTL variants into the model.

3.6 Discussion

In this chapter, we performed 138 macrophages differentiation attempts from 123 unique iPSC lines and we were able to successfully differentiate 101 (82%) of them. This makes our study

one of the largest long term directed differentiations of human iPSCs into another cell type. Extensive documentation of the differentiation attempts allowed us to characterise the extent of normal variation in multiple aspects of the differentiation protocol such as success rate, duration, yield and purity of the resulting macrophage population. We have shown that this differentiation protocol is highly reproducible at the level of gene expression, works on most iPSC lines and can be scaled to differentiate large numbers of cells in parallel.

An important open question is what underlies variability in iPSCs differentiation potential; are these genetic differences between donors, differences between clonal iPSC lines from the same donor or technical batch effects between independent differentiation attempts. Our experimental design of differentiating only one line per donor was optimised for detecting the maximal number of gene expression QTLs. As a result, we were not able to distinguish between donor and line effects. However, our observation that repeated differentiations are much more likely to fail for lines that failed the first differentiation than for lines that succeeded the first differentiation does suggest that there are some differences between iPSC lines (either genetic or epigenetic) that influence differentiation success.

We also collected RNA-seq data from most of the differentiated lines in four experimental conditions. Combining gene expression data with extensive metadata from the differentiations in a linear mixed model allowed us to identify important factors contributing to gene expression variation in iPSC-derived macrophages. In particular, we highlighted the importance of controlling for cell density and cell purity when performing genomics assays on iPSC-derived cells. The large effect of macrophage purity was unexpected, because the majority of the samples were already over 95% pure and we had discarded all samples that were less than 90% pure prior to RNA sequencing. On visual inspection the contaminating cells seemed larger than macrophages, and thus could have contributed relatively more RNA to the pool. We also observed that the date of stimulation explained double the variance in conditions where live *Salmonella* was used to infect cells compared to naive and IFNɣ conditions, highlighting an important trade-off between physiologically more accurate live infections and inherently less variable stimulations with well-defined molecular signals (such as IFNɣ and LPS).

Finally, we showed that variation in the intensity of expression of commonly used macrophage markers CD14 and CD16 on the cell surface is driven by common genetic variants. This was especially pronounced for CD14, where we identified a common genetic variant 19 kb upstream

of the gene that could explain almost all of the line-to-line variation in CD14 expression. Thus, it is important to take into account natural genetic variation when comparing the expression of cell type specific markers between primary cells, iPSC-derived cells and embryonic stem cell-derived cells. This is especially important because these different cell types can rarely be obtained from genetically matched donors. For example, CD14 has previously been highlighted as variably methylated gene in human ESCs and variably expressed in differentiated macrophages (Bock et al., 2011). The authors attributed this variability to defective methylation in some ESCs that interfered with macrophage differentiation. However, our results suggest that much of this variability is caused by segregation of a common genetic polymorphism. Flow cytometry on cell surface markers is also commonly used to quantify the relative abundance of different cell types in a tissue (such as blood). It is therefore important to take the natural variation in the expression of these markers into account when designing the experiments and setting threshold values so as not to mistake differences in cell surface expression of marker gene as differences of cell type proportion.

An important area for future work will be to optimise the differentiation protocol to work directly on feeder-free iPSCs without transferring them to feeder cells. This has the potential to greatly reduce the time and work needed for iPSC expansion prior to differentiation which currently takes ~20 days. With newer RNA-seq and chromatin assay requiring fewer cells, there is also potential to miniaturise the differentiation protocol making it feasible to differentiate hundreds of IPSCs in parallel. Here, alternative embryoid body formation protocols can be trialled (e.g. AggreWell plates (van Wilgenburg et al., 2013)) that have the potential to reduce variability in macrophage yield between differentiations.