Using a multi-omics approach to investigate the reversal of proteasome inhibitor drug resistance in multiple myeloma with epigenetic inhibitors

Anna James-Bott

St Hilda's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

Acknowledgements

I would like to thank my supervisors Dr Adam Cribbs, Professor Udo Oppermann and Dr Sarah Gooding etc etc...

GSK... DTC... Family... Friends...

Abstract

Background: Multiple myeloma (MM) is an incurable cancer of plasma cells. Novel therapeutics, including proteasome inhibitors (PI) and immunomodulatory imide drugs, have almost doubled median survival time of MM patients. However, most patients relapse and become resistant to drugs they previously have been treated with. Acquired anti-cancer drug resistance remains one of the biggest barriers in the treatment of myeloma.

Aims: PI resistance mechanisms in MM will be investigated with the aim of reversing the resistance phenotype, making MM cells sensitive to proteasome inhibition. Standardised robust wet-lab and computational single-cell workflows will be established to characterise drug-resistant MM cells and their surrounding microenvironment at different points in disease progression.

Results: Cured cancer mate

Contents

List of Figures

List of Tables

List of Abbreviations

MM . . . . . Multiple Myeloma

BM . . . . . Bone marrow

MGUS . . . . Monoclonal gammopathy of unknown significance

SMM . . . . . Smoldering multiple myeloma

PI . . . . . Proteasome inhibitor

IMiDs . . . . . Immunomodulatory imide drugs

ER . . . . . Endoplasmic reticulum

UPS . . . . . Ubiquitin proteasome system

UPR . . . . . Unfolded protein response

RNA-Seq . . . Ribonucleic acid sequencing

scRNA-Seq . . Single cell RNA-Seq

dscRNA-Seq . Droplet-based scRNA-Seq

CB . . . . . Cellular barcode

UMI . . . . . Unique molecular identifier

WGS . . . . . Whole genome sequencing

1 Introduction

1.1 Overview

Multiple myeloma accounts for 1-2% of all cancers and has the second highest incidence of hematological malignancies, after non-Hodgkin's lymphoma [1].

1.2 The adaptive immune system

Humans are exposed to millions of potential pathogens every day and therefore require defences to be able to protect themselves against infection. These defences can be innate or adaptive. An example of an innate defence is the skin acting as a physical barrier between the outside world and the body. Another example of an innate defence is non-specific engulfing (phagocytosis) of foreign pathogens by macrophages (a type of white blood cell). Innate responses are relied upon as the first line of defence, however sometimes a more sophisticated, specialised response is required- called the adaptive immune response. (REF-mol biology of the cell).

Adaptive immune responses are specific to the pathogen that induced the response and are dependent on B cells and T cells, two major classes of lymphocytes (a class of white blood cell). Two classes of adaptive immune responses exist: antibody responses, co-ordinated by B cells, and cell mediated immune responses, co-ordinated by T cells. T-cell-mediated immune responses recognise foreign antigens (antibody generators; substances capable of eliciting an immune response by stimulating B or T cell activation) on the surface of cells and can either kill

the pathogen-infected cells or stimulate B cells or phagocytes to help eliminate the pathogen. In antibody responses, B cells and plasma cells secrete antibodies, also known as immunoglobulins. Immunoglobulins are large Y-shaped proteins, which recognise and bind to the specific foreign antigen on the pathogen which stimulated their production. Binding of immunoglobulins to antigens renders the virus or microbial toxin inactive as it blocks their ability to bind to host cells. Additionally, antibody binding makes it easier for phagocytic cells to ingest the pathogen.

1.2.1 Plasma cells

Plasma cell development

Stem cells are precursor cells which can give rise to at least one type of differentiated (mature) cell, with the capability of indefinite self-renewal. Hematopoietic stem cells (HSC) are stem cells that give rise to all the cells of the hematopoietic system. Two predominant cell populations are produced by HSCs: the common myeloid progenitor (CMP) and the common lymphocyte progenitor (CLP). CMP differentiation produces erythrocytes (red blood cells), mast cells, monocytes, macrophages, neutrophils, eosinophils, basophils and myeloid dendritic cells. CLP differentiation results in B cells, T cells, natural killer (NK) cells and lymphoid dendritic cells.

Most B cells die in the bone marrow soon after developing, however some will develop in the bone marrow, where initial stages of maturation occur and then migrate to secondary lymphoid organs, such as the spleen. Within secondary lymphoid organs, numerous critical decisions on B cell fate are made, involving complex transcriptional networks, cell interactions, gene rearrangements, and mutations (roth2014tracking; jourdan2011characterization). Terminally differentiated plasma cells are the final effectors of the B cell lineage, each dedicated to secreting large amounts of a single type of antibody. Plasma cells have an extensive rough endoplasmic reticulum (ER), and have numerous genes involved in immunoglobulin secretion upregulated, including XBP-1 and CHOP (shapiro2004plasma), to enable the production of the copious amounts of antibody required.

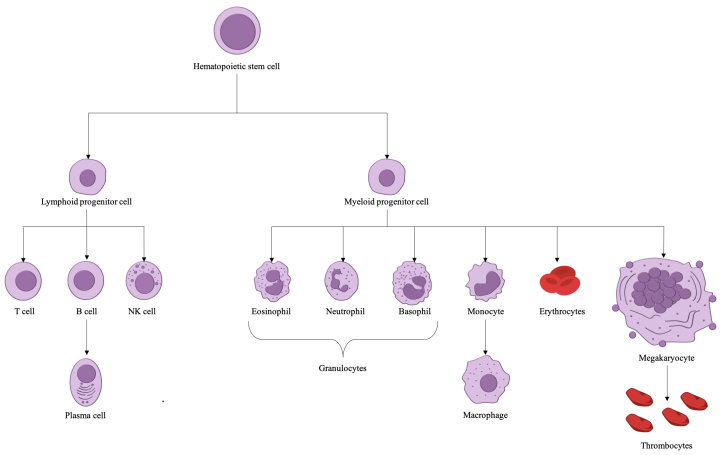

Figure 1.1: Hematopoietic stem cell (HSC) cell differentiation. HSCs divide into myeloid or lymphoid progenitor cells. Dendritic cells and a number of precursor states have been ommitted.

1.3 Clinical MM

1.3.1 Epidemiology

Multiple myeloma accounts for 1-2% of all cancers and has the second highest incidence of hematological malignancies, after non-Hodgkin's lymphoma[1]. MM is rare in individuals under the age of 40, with the average age at time of diagnosis centering around 70[2, 3]. MM is more prevalent in males than females and is around twice as common in black populations than in Caucasian or Asian populations[4]. The average incidence rate is approximately 1-6 cases per 100000 individuals[2, 3, 5], with the highest age-standardised incidence rates in the regions of Australasia, North America, and Western Europe[6]. Five-year survival rate of MM patients is approximately 49%, whilst approximately a third of MM patients survive ten years or greater[7, 8].

1.3.2 Presentation

Precursor states

All cases of MM are preceded by asymptomatic precursor states, monoclonal gammopathy of unknown significance (MGUS) and smoldering multiple myeloma (SMM). However, only some patients with SMM or MGUS progress to active MM.

MGUS is a pre-malignant condition where patients have the presence of monoclonal immunoglobulins in their blood or urine, <10% clonal plasma cells in their bone marrow, but lack any myeloma-related end-organ damage[9]. Patients with SMM have between 10 and 60% clonal plasma cells in their bone marrow, serum monoclonal immunoglobulin of ≥3 g/dL, and like MGUS, have no signs of end-organ damage[10]. Progression risk of MGUS into symptomatic MM is about 1% per year, whilst progression risk of SMM to MM is higher, at around 10% per year for the first 5 years, after which it decreases[11, 12].

Active MM

There are multiple classifications of active MM. The International Myeloma Working Group's definition[13] is as follows: Greater than 10% clonal plasma cells located in the bone marrow and one or more myeloma-defining event or biomarker of malignancy. Myeloma defining events consist of evidence of end-organ damage that can be attributed to the surplus of M protein and clonal plasma cells, namely the CRAB features:

- Hypercalcemia
  - Serum calcium > 1 mg/dL higher than the upper limit of normal, or
  - Serum calcium > 11 mg/dL
- Renal insufficiency
  - Creatinine clearance < 40 mL per min, or
  - Serum creatine > 2 mg/dL
- Anemia
  - Hemoglobin value of > 20 g/L below the lower limit of normal, or
  - Hemoglobin value < 100 g/L
- Bone lesions

– One or more osteolytic lesions on skeletal radiography, CT or PET-CT

Biomarkers of malignancy include greater than or equalt to 60% clonal plasma cells in the bone marrow, an involved:uninvolved serum free light chain ratio greater than or equal to 100, and more than one focal lesion on an MRI study[13].

It is currently unclear what causes the malignant transformation between precursor states and active MM. However certain factors have been identified as risk factors, including point mutations, a large array of up-regulated transcription factors, and numerous immune events.

1.3.3 Treatment of multiple myeloma

Multiple myeloma may be an incurable disease, however it is treatable. In fact, in the last decade median survival time for newly diagnosed MM patients has almost doubled[14]. Novel therapeutic advances have contributed to this improvement (Table1.1).

1.3.4 Proteasome inhibitors

The ubiquitin-proteasome system

1.4 Drug resistance in multiple myeloma

1.5 Transcriptomics, Epigenomics

1.6

Year	Treatment	Usage	Ref
1958	Melphalan	The alkylating agent melphalan was first used in plasma cell myeloma in 1958.	[15]
1960s	Corticosteroids	Placebo-controlled double-blind trial of prednisone in multiple myeloma. Combinations of prednisone and melphalan showed an increased survival over melphalan alone. Dexamethasone and prednisone have become a cornerstone in the treatment of multiple myeloma.	[16, 17]
1980s	Stem-cell transplantations	Numerous successful allogenic and autologous bone marrow transplantations in patients with multiple myeloma	[18–21]
2003	Proteasome inhibitors	Bortezomib, a first-in-class proteasome inhibitor, was first approved by the FDA for use in relapsed and refractory multiple myeloma. In 2008 it was approved for patients with no prior treatment. Carfilzomib was approved in 2012 for advanced MM and later in 2015 for treatment of relapsed MM. The oral proteasome inhibitor, ixazomib, was approved as a combination treatment with lenalidomide and dexamethasone in 2016 for people who have received at least one previous treatment.	[22–24]
2006	IMiDs	The antitumour activity of thalidomide was demonstrated in 1999, this led to the development of lenalidomide, the first approved immunomodulatory imide drug (IMiD) for use in multiple myeloma. Currently, thalidomide, lenalidomide and pomalidomide are approved for use in multiple myeloma	[25–27]
2015	Monoclonal antibodies	In 2015, daratumumab, an anti-CD38 monoclonal antibody and elotuzumab, an anti-SLAMF7 monoclonal antibody, were approved for MM treatment.	[28, 29]

Table 1.1: Timeline of treatment options for multiple myeloma. Listed by first usage or FDA approval for MM.

2
Background

2.1 Drug resistance in MM

2.1.1 Genomic changes in drug resistant MM

2.1.2 Epigenetic changes in drug resistant MM

3
Methods

3.1 Cell culture

3.1.1 AMO-1 cells

AMO-1 cells, plasma cells from a 64-year old female myeloma patient, were used as a model cell-line for multiple myeloma. Bortezomib and carfilzomib resistant AMO-1 cells were generated and kindly gifted by the Driessen lab[30]. Bortezomib, carfilzomib, pomolidimide and bortezomib plus pomolidimide resistant AMO-1 cells were also generated by Dr James Dunford by continual and escalating drug exposure of drug-sensitive (WT) cells. AMO-1 cells were cultivated in RPMI-1640 medium (WHERE we get it from), supplemented with 10% fetal bovine serum (FBS), 100µg ml^{-1} streptomycin and 100 U/ml penicillin (P/S) and 2mM L-glutamine (Invitrogen, UK). Cells were passaged when they reached approximately 1.5-2 million cells per ml (IS THIS THE RIGHT MEASURE??). Media was replaced twice a week.

3.2 Bulk RNA-seq

3.2.1 RNA extraction

RNA was extracted and purified using the Direct-Zol RNA MiniPrep kit (Zymo), following the manufacturer's protocol. In brief, for each sample, approximately 100,000 cells were lysed in 300µl of TRIzol and the lysate was transferred to a microcentrifuge tube. 300µl of ethanol was added to the lysed samples and vortexed.

The mixture was transferred to miniPrep columns and centrifuged at 10,000-16,000g for 30 seconds. The column was washed twice with 400µl of Direct-Zol pre-wash and once with 700µl of RNA wash buffer The column was transferred to an RNase-free tube and eluted with 50µl of nuclease-free water and centrifuged.

The RNA concentration was quantified using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, USA), and samples were stored at -80°C. samples were normalised to 100ng with nuclease-free water.

3.2.2 RNA library preparation

NEBNext® Ultra II directional RNA library prep kit for Illumina® with TruSeq indexes was used to prepare RNA libraries, following the manufacturer's protocol. RNA concentration was normalised to 100ng with nuclease-free water, made up to 50µl. The NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, USA) was used to enrich poly-adenylated RNA. READ booklet in lab

The molarities of the libraries were determined by electrophoresis on a TapeStation (Agilent, USA).

3.3 Single-cell RNA-seq

3.3.1 Cell encapsulation

The Drop-Seq protocol[31] was followed for single-cell RNA-Seq sample preparation. Cells were loaded into a microfluidics cartridge. Nadia, an automated microfluidics device (Dolomite Bio, UK), performed cell capture, cell lysis and reverse transcription. Reverse transcription reactions were performed using ChemGene beads or (ATDBio beads 2020 onwards!!!!! might need to change if reperform).

3.3.2 Library preparation

Beads were collected from the device and cDNA amplification was performed. The beads were treated with Exo-I prior to PCR. The amplified, purified cDNA then underwent tagmentation reactions. A TapeStation (Agilent, USA) was used

to assess library quality. The samples were pooled together and split across multiple sequencing runs.

3.4 ATAC maybe

3.4.1 ATAC stuff

3.5 ChIP maybe

3.5.1 ChIP stuff

3.6 Sequencing

Sequencing of the resultant libraries was performed on the NextSeq 500 (Illumina, USA) platform using a paired-end run, according to the manufacturer's instructions.

3.7 Phosphoproteomics

3.7.1 Cell lysis

Approximately 20 million cells for each condition in triplicate were pelleted and stored in 500µl of PBS at -80°C. 300µl of fresh lysis buffer (10ml RIPA buffer, 3µl benzonase, 1 tablet phos stop) was added to each pellet, vortexed and left for 10 minutes on ice and then sonicated. The supernatant was transferred to a fresh (WHAT TYPE) tube.

3.7.2 Protein quantification

Protein concentrations were determined by BCA protein assay (Thermofisher, UK). 400µg of protein was taken from each sample. Samples were made up to a volume of 200µl with MilliQ-H_2O.

3.7.3 Protein Digestion

Kessler lab protocols were followed (`https://www.tdi.ox.ac.uk/research/research/tdi-mass-spectrometry-laboratory/mass-spectrometry/protocols-and-tools`). The lysed samples were reduced with 5µl of 200 mM DTT in 0.1 M Tris buffer

and incubated for 40 minutes at room temperature. The reduced samples were alkylated with 20µl of 200mM iodoacetamide in 0.1M Tris buffer, vortexed and then incubated for 45 minutes in the dark at room temperature. The protein was precipitated using methanol/chloroform extraction. The alkylated samples were transferred to 2ml eppendorfs. 600µl of methanol was added to each sample, followed by 150µl of chloroform and then vortexed gently. 450µl of MilliQ-H_2O was then added and vortexed gently. The samples were centrifuged at maximum speed on a table top centrifuge for one minute. The upper aqueous phase was removed, without disturbing the precipitate at the interface. 450µl of methanol was added to each sample, without disturbing the disc and centrifuged for two minutes. Protein pellets were resuspended, one sample at a time: the supernatant was removed and 100µl of 6M urea in 0.1M Tris buffer was added. The samples were vortexed and then sonicated (???). Samples were diluted with 500µl MilliQ-H_2O, to ensure the final urea concentration was below 1M. Porcine trypsin (Sequencing Grade Modified Trypsin; Promega, USA) was added in a 1:50 ratio of enzyme:total protein content of sample, such that 40µl of trypsin solution containing 8µg trypsin in 0.1M Tris buffer was added to each sample. Samples were left to digest overnight at 37°C in an incubator shaker.

3.7.4 Peptide purification

The following day, the reaction was stopped, acidifying samples to 1% Trifluoroacetic acid (TFA). Samples were desalted and concentrated using 1ml C-18 Sep-Pak (Waters) cartridges. Two reagents were used: solution A (98% MilliQ-H_2O, 2% Acetonitrile (CH_3CN) and 0.1% TFA) for washing and solution B (65% Acetonitrile, 35% MilliQ-H_2O and 0.1% TFA) for activation and elution. The columns were flushed with 1ml of solution B and then washed with 1ml of solution A. The digested samples were added to the columns and vacuumed through slowly. Two 1ml washes with solution A were performed. Fresh, labelled eppendorfs were placed beneath the columns and peptides were eluted with 500µl of solution B. For phosphopeptide-enrichment, 90% of the peptides were removed for Immobilized

Metal Affinity Chromatography (IMAC) on a Bravo Automated Liquid Handling Platform (Agilent). 10% of the peptides were used for total proteome analysis. Eluted peptides were dried using a vacuum concentrator (Speedvac, Eppendorf) and stored at -20°Cuntil analysis by mass spectrometry (MS). Prior to MS analysis, dried peptides were resuspended in solution A.

3.8 Ubiquitinomics

3.8.1 Cell lysis

Approximately 100 million cells for each condition in triplicate were pelleted and stored in 500µl of PBS at -80°C. PMTScan Ubiquitin Remnant Motif Kit (K-ε-GG; Cell signalling) was used, following the manufacturer's protocol (REF). Pellets were solubilized and denatured in 4ml urea lysis buffer (20mM HEPES, pH 8.0, 9M urea, 1mM sodium orthovanadate, 2.5mM sodium pyrophosphate, 1mM β-glycerophosphate). The lysates were sonicated on ice, with two bursts of 15 seconds with a one minute break in-between.

3.8.2 Protein quantification

Protein concentrations were determined by BCA protein assay (Thermofisher, UK). All samples were found to contain between 10mg and 20mg of protein, so all of the available protein was used, with no normalisation.

3.8.3 Protein Digestion

Lysates were reduced using dithiothreitol (DTT) at a final concentration of 4.5 mM for 30 minutes at room temperature. The reduced samples were alkylated using iodoacetamide (100mM final) for 15 minutes in the dark at room temperature. The alkylated samples were diluted four-fold with 20mM HEPES (pH 8.0) and digested with 400µl trypsin solution, containing $1\,\mathrm{mg\,ml^{-1}}$ trypsin-TPCK (Worthington, LS003744) in 1mM HCl. Samples were left to digest overnight at room temperature on a rotator.

3.8.4 Peptide purification

The following day, the reaction was stopped, acidifying samples to 1% Trifluoroacetic acid (TFA). Samples were desalted and concentrated using 10ml C-18 Sep-Pak (Waters) cartridges. The columns were activated using 5ml of solution B, washed with 10ml of solution A. The samples were added to the columns and ran through slowly. The peptides were washed with 10ml of solution A. The cartridges were then removed from the vacuum and the peptides were eluted into fresh falcon tubes with 6ml of solution B, using the plunger of the syringes. 20µg of digested protein was removed from each sample for matching total proteome analysis. The eluate was kept at -80°C overnight. The frozen peptide solutions were lyophilized for two days and then stored at -80°C.

3.8.5 Immunoaffinity purification

10x immunoaffinity purification (IAP) buffer provided with PTMScan Kit was diluted to 1x concentration with MilliQ-H_2O. Purified peptides pellets were re-suspended in 1.4ml of IAP buffer by pipetting up and down and transferred to 1.7ml eppendorfs. The samples were centrifuged at 4°C for 5 minutes at 10000 g and kept on ice whilst preparing antibody beads. The anti-body bead slurry was centrifuged (30 seconds at 2000 g) and 1ml of PBS was added and then centrifuged. The supernatant was removed and the antibody beads were washed a further four times with PBS and resuspended in 40µl of PBS. The peptide solution was transferred to the antibody vial and the solution was incubated on a rotator for two hours at 4°C. The samples were centrifuged, put on ice and the supernatant was removed. The beads were washed twice with 1ml IAP, followed by three washes with 1ml chilled HLPC water. Immunoprecipitated material was eluted at room temperature in 55µl and 50µl 0.15% TFA in water, letting the sample stand for 10 minutes after each elution, with gentle mixing every two-three minutes. The eluates were centrifuged and the supernatant was transferred to new tubes. Peptide material was desalted and concentrated using 1ml C-18 Sep-Pak cartridges as above. Prior to mass spectrometry analysis, purified GlyGly-modified peptide

eluates and matching proteome material were dried by vacuum centrifugation, and re-suspended in solution A.

3.9 Liquid-chromatography-tandem mass spectrometry

Liquid-chromatography-tandem mass spectrometry (LC-MS/MS) analysis was performed using a Dionex Ultimate 3000 nano-ultra high pressure reverse-phase chromatography coupled on-line to an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) (REF: adan's 3-5 dropbox). In brief, samples were separated on an EASY-Spray PepMap RSLC C18 column (500mm × 75µm, 2µm particle size; Thermo Scientific) over a 60 min (120 min in the case of the matching proteome) gradient of 2–35% acetonitrile in 5% dimethyl sulfoxide (DMSO), 0.1% formic acid at 250nl min^{-1}. MS1 scans were acquired at a resolution of 60000 at m/z 200 and the top 12 most abundant precursor ions were selected for high collision dissociation (HCD) fragmentation.

3.10 Data Processing

3.10.1 Bulk RNA-Seq

Fasta files were processed using a CGAT-flow (REF) pipeline, workflow can be found at: `https://github.com/cgat-developers/cgat-flow/blob/master/cgatpipelines/tools/pipeline_rnaseqdiffexpression.py`. Pseudo-alignment tool, Kallisto (REF), was implemented to pseudo-align reads to the reference human genome sequence (GRCH38 (hg38) assembly) and to construct a counts matrix of samples against transcripts (/GENES??). DESeq2 (REFERENCE) was used for differential expression analysis of the generated counts matrix (using negative binomial generalized linear models) within the R statistical framework (v3.5.1).

3.10.2 Single-cell RNA-Seq

Pipeline etc...

3.10.3 LC-MS/MS

Mass-spectrometry raw data were searched against the UniProtKB human sequence data base and label-free quantitation (LFQ) was performed using MaxQuant Software (v1.5.5.1). Digestion was set to trypsin/P. Search parameters were set to include carbamidomethyl (C) as a fixed modification, oxidation (M), deamidation (NQ), and phosphorylation (STY) as variable modifications. A maximum of 2 missed cleavages were allowed for phosphoproteome analysis and 3 for the GlyGly peptidome analysis, with matching between runs. LFQ quantitation was performed using unique peptides only. Label-free interaction data analysis was performed using Perseus (v1.6.0.2). Results were exported to Microsoft Office Excel Sheets and imported into the R statistical framework (v3.5.1) for further analysis.

3.11 CyTOF

Get data off ADAM

3.11.1 CyTOF stuff

4
Workflow Generation

4.1 Introduction

4.1.1 Reproducible workflows

In data analysis, particularly in bioinformatics, many users create simple bash or R scripts to execute the specific task at hand. However, if this is done often, the user can have an accumulation of these single-use scripts, which are often named uninformatively and never used again. Subsequently, the user may create scripts which perform the same function numerous times. Additionally, users may just use the command line alone to perform tasks. This means that exactly how they performed the analysis is difficult to find or not recorded. These are bad practices in terms of efficiency and reproducibility. It is much better practice to create well-documented, generalised workflows which can then be applied to multiple different experiments. This enables the user to reuse their code more easily and reproduce results, if need be. This also allows other researchers to reproduce results or apply the code to their own research.

In addition to creating generalised, reproducible workflows, it can be beneficial to create more extensive computational pipelines for jobs which require multiple tasks or actions to be performed sequentially.

4.1.2 Computational pipelines

A computational pipeline consists of a series of manipulations and transformations, where the output of one element is the input of the next. Often these elements are executed in parallel. Pipelining 'omics' data-processing means that tasks that are not interdependent can be executed simultaneously. Additionally, multiple samples can be processed in parallel, thereby reducing run time. There are many available pipelining frameworks, for example Snakemake[32], Luigi and Ruffus[33].

For this work, a series of computational pipelines and workflows were generated. Ruffus and CGAT-core[34] were used as the backbone for the pipelines developed.

4.2 scRNA-Seq pseudoalignment pipeline

Fewer pipelines exist for single-cell RNA-Seq compared to bulk RNA-Seq. For the Chromium 10X Genomics platform, most of the processing and analysis is automated by Cell Ranger; however for other technologies, the workflow is not as well defined. A single-cell analysis pipeline was constructed with the aim to produce an easy-to-use, robust and reproducible workflow that works for Drop-Seq as well as 10X technology, which utilises pseudoalignment rather than traditional mapping methods.

4.2.1 Psuedoaligment

Traditional mapping techniques such as Tophat[35] or STAR[36], rely on aligning each read to a reference genome. This is generally very time consuming and computationally expensive. Another challenge that arises with traditional mapping is the occurrence of multi-mapping, whereby a read cannot be uniquely aligned as it could map equally well to multiple sites in the genome[37]. More recently, a series of methods called pseudoaligners have been developed that overcome some of the issues associated with traditional mapping approaches. Pseudoalignment (sometimes referred to as quasi-mapping) methods provide a lightweight, alignment-free alternative to traditional mapping. It has been shown that information on where exactly inside transcripts sequencing reads may have originated is not required for

accurate quantification of transcript abundances[38]. Rather, only which transcript the read could have originated from is needed and transcript abundances are calculated by computing the compatibility of reads with different transcripts. This negates the need for alignment to a reference genome, alleviating the issue of multi-mapping and reducing the computational load. Pseudoaligners have been shown to complete data processing of RNA-seq datasets up to 250-times faster than traditional alignment and quantification approaches[39]. Kallisto[39] and Salmon[40] are tools which implement pseudoalignment. They have similar speed and accuracy for bulk RNA-seq data[1].

Pseudoalignment of scRNA-seq

Pseudoalignment tools have recently been developed for droplet-based scRNA-seq analysis (dscRNA-seq). Additional challenges come with dscRNA-seq data processing, having the extra complication of cellular barcodes (CBs) and unique molecular identifiers (UMIs). These tools must handle transcript abundance estimation, as with bulk RNA-seq analysis, but also perform CB detection, collapsing of UMIs (arising from PCR duplication of molecules) and barcode error correction. Kallisto BUS[41] has been developed as an analysis tool and file format specifically for single-cell analysis, alongside BUStools, for processing of the resultant BUS file[42]. Salmon Alevin[43] has also been developed for single-cell RNA-seq analysis.

Flow diagram of pipeline– placeholder

4.2.2 Benchmark

Benchmarking measures the performance of a method/software relative to other methods available. Run time and the accuracy of results are often the factors considered in a benchmark. To be able to calculate the accuracy of results, the 'true' results must be known. This is difficult in scRNA-seq analysis as no gold standard analysis protocol exists. Instead, methods are compared against simulated results which act as the underlying 'ground truth'.

[1] `https://liorpachter.wordpress.com/2017/09/02/a-rebuttal/`

Simulated data

Splatter[44] was used to simulate single-cell counts matrices, using a real single-cell counts matrix to estimate simulation parameters from. The simulated counts matrix was randomly assigned cell barcodes and ensembl gene names for the column and rownames respectively, and then used as input for Minnow[45] as 'ground-truth' data. Minnow generates droplet-based scRNA-seq simulated reads, working backwards from a known counts matrix to generating raw sequencing files from which the counts matrix could have originated. Minnow accounts for core experimental dscRNA-seq characteristics, such as PCR amplification bias, barcode sequencing errors, the presence of doublets and ambiguously mapped reads, to try and emulate a realistic set of sequencing reads consistent with the provided counts matrix. The simulated reads were used as input for the scRNA-Seq pseudoalignment pipeline and the resulting count matrices outputted by Salmon Alevin and Kallisto BUS were compared to the 'ground-truth' data.

Results

4.2.3 Comparison to published data using other methods..

4.3 scRNA-Seq velocity analysis pipeline

4.3.1 RNA velocity

"RNA velocity is a high-dimensional vector that predicts the future state of individual cells on a timescale of hours"[46]. In combination with clustering analysis, the trajectory of a single-cell can be tracked.

Appendices

A
Epigenetic compound screen

A compound screen consisting of approximately 140 epigenetic inhibitors was performed for AMO-1 cells.

References

[1] International Myeloma Working Group. "Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group". In: *British journal of haematology* 121.5 (2003), pp. 749–757.

[2] Matthew Tsang et al. "Multiple myeloma epidemiology and patient geographic distribution in Canada: a population study". In: *Cancer* (2019).

[3] Antonio Palumbo and Kenneth Anderson. "Multiple Myeloma". In: *New England Journal of Medicine* 364.11 (2011), pp. 1046–1060.

[4] NHS UK. *Multiple Myeloma*. `https://www.nhs.uk/conditions/multiple-myeloma/`. Accessed: 06-2019.

[5] Lauren R Teras et al. "2016 US lymphoid malignancy statistics by World Health Organization subtypes". In: *CA: a cancer journal for clinicians* 66.6 (2016), pp. 443–459.

[6] Andrew J Cowan et al. "Global burden of multiple myeloma: a systematic analysis for the Global Burden of Disease Study 2016". In: *JAMA oncology* 4.9 (2018), pp. 1221–1227.

[7] Cancer Research UK. *Myeloma Survival Statistics*. `https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma/survival`. Accessed: 06-2019.

[8] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. "Cancer statistics, 2016". In: *CA: a cancer journal for clinicians* 66.1 (2016), pp. 7–30.

[9] Niels van Nieuwenhuijzen et al. "From MGUS to multiple myeloma, a paradigm for clonal evolution of premalignant cells". In: *Cancer research* 78.10 (2018), pp. 2449–2456.

[10] S Vincent Rajkumar, Ola Landgren, and Maria-Victoria Mateos. "Smoldering multiple myeloma". In: *Blood, The Journal of the American Society of Hematology* 125.20 (2015), pp. 3069–3075.

[11] Neha Korde, Sigurdur Y Kristinsson, and Ola Landgren. "Monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM): novel biological insights and development of early treatment strategies". In: *Blood* 117.21 (2011), pp. 5573–5581.

[12] Robert A Kyle et al. "Clinical course and prognosis of smoldering (asymptomatic) multiple myeloma". In: *New England Journal of Medicine* 356.25 (2007), pp. 2582–2590.

[13] S Vincent Rajkumar et al. "International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma". In: *The lancet oncology* 15.12 (2014), e538–e548.

[14] Dickran Kazandjian and Ola Landgren. "A look backward and forward in the regulatory and treatment history of multiple myeloma: approval of novel-novel agents, new drug development, and longer patient survival". In: *Seminars in oncology*. Vol. 43. 6. Elsevier. 2016, pp. 682–689.

[15] N Blokhin et al. "Clinical experiences with sarcolysin in neoplastic diseases". In: *Annals of the New York Academy of Sciences* 68.3 (1958), pp. 1128–1132.

[16] ROBERT E MASS. "A comparison of the effect of prednisone and a placebo in the treatment of multiple myeloma." In: *Cancer chemotherapy reports* 16 (1962), p. 257.

[17] Raymond Alexanian et al. "Treatment for multiple myeloma: combination chemotherapy with different melphalan dose regimens". In: *Jama* 208.9 (1969), pp. 1680–1685.

[18] TJ McElwain and RL Powles. "High-dose intravenous melphalan for plasma-cell leukaemia and myeloma". In: *The Lancet* 322.8354 (1983), pp. 822–824.

[19] Elliott F Osserman et al. "Identical twin marrow transplantation in multiple myeloma". In: *Acta haematologica* 68.3 (1982), pp. 215–223.

[20] Alexander Fefer, Martin A Cheever, and Philip D Greenberg. "Identical-twin (syngeneic) marrow transplantation for hematologic cancers". In: *Journal of the National Cancer Institute* 76.6 (1986), pp. 1269–1273.

[21] G Gahrton et al. "Bone marrow transplantation in multiple myeloma: report from the European Cooperative Group for Bone Marrow Transplantation". In: *Blood* 69.4 (1987), pp. 1262–1264.

[22] Robert C Kane et al. "Velcade®: US FDA approval for the treatment of multiple myeloma progressing on prior therapy". In: *The oncologist* 8.6 (2003), pp. 508–513.

[23] Paul G Richardson et al. "A phase 2 study of bortezomib in relapsed, refractory myeloma". In: *New England Journal of Medicine* 348.26 (2003), pp. 2609–2617.

[24] Alla Katsnelson. *Next-generation proteasome inhibitor approved in multiple myeloma*. 2012.

[25] Seema Singhal et al. "Antitumor activity of thalidomide in refractory multiple myeloma". In: *New England Journal of Medicine* 341.21 (1999), pp. 1565–1571.

[26] FDA Label. "Revlimid-lenalidomide capsule". In: *For Multiple Myeloma Myelodysplastic Syndrome and Mantle Cell Lymphoma* 47 ().

[27] Jesus San Miguel et al. "Pomalidomide plus low-dose dexamethasone versus high-dose dexamethasone alone for patients with relapsed and refractory multiple myeloma (MM-003): a randomised, open-label, phase 3 trial". In: *The lancet oncology* 14.11 (2013), pp. 1055–1066.

[28] Henk M Lokhorst et al. "Targeting CD38 with daratumumab monotherapy in multiple myeloma". In: *New England Journal of Medicine* 373.13 (2015), pp. 1207–1219.

[29] Sagar Lonial et al. "Elotuzumab therapy for relapsed or refractory multiple myeloma". In: *New England Journal of Medicine* 373.7 (2015), pp. 621–631.

[30] GP Soriano et al. “Proteasome inhibitor-adapted myeloma cells are largely independent from proteasome activity and show complex proteomic changes, in particular in redox and energy metabolism”. In: *Leukemia* 30.11 (2016), pp. 2198–2207.

[31] Evan Z Macosko et al. “Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets”. In: *Cell* 161.5 (2015), pp. 1202–1214.

[32] Johannes Köster and Sven Rahmann. “Snakemake—a scalable bioinformatics workflow engine”. In: *Bioinformatics* 28.19 (2012), pp. 2520–2522.

[33] Leo Goodstadt. “Ruffus: a lightweight Python library for computational pipelines”. In: *Bioinformatics* 26.21 (2010), pp. 2778–2779.

[34] Adam P Cribbs et al. “CGAT-core: a python framework for building scalable, reproducible computational biology workflows”. In: *F1000Research* 8 (2019).

[35] Cole Trapnell, Lior Pachter, and Steven L Salzberg. “TopHat: discovering splice junctions with RNA-Seq”. In: *Bioinformatics* 25.9 (2009), pp. 1105–1111.

[36] Alexander Dobin et al. “STAR: ultrafast universal RNA-seq aligner”. In: *Bioinformatics* 29.1 (2013), pp. 15–21.

[37] Ali Mortazavi et al. “Mapping and quantifying mammalian transcriptomes by RNA-Seq”. In: *Nature methods* 5.7 (2008), p. 621.

[38] Marius Nicolae et al. “Estimation of alternative splicing isoform frequencies from RNA-Seq data”. In: *International Workshop on Algorithms in Bioinformatics*. Springer. 2010, pp. 202–214.

[39] Nicolas L Bray et al. “Near-optimal probabilistic RNA-seq quantification”. In: *Nature biotechnology* 34.5 (2016), p. 525.

[40] Rob Patro et al. “Salmon provides fast and bias-aware quantification of transcript expression”. In: *Nature methods* 14.4 (2017), p. 417.

[41] Páll Melsted, Vasilis Ntranos, and Lior Pachter. “The barcode, UMI, set format and BUStools”. In: *bioRxiv* (2018), p. 472571.

[42] Páll Melsted et al. “Modular and efficient pre-processing of single-cell RNA-seq”. In: *BioRxiv* (2019), p. 673285.

[43] Avi Srivastava et al. “Alevin efficiently estimates accurate gene abundances from dscRNA-seq data”. In: *Genome biology* 20.1 (2019), p. 65.

[44] Luke Zappia, Belinda Phipson, and Alicia Oshlack. “Splatter: simulation of single-cell RNA sequencing data”. In: *Genome biology* 18.1 (2017), p. 174.

[45] Hirak Sarkar, Avi Srivastava, and Rob Patro. “Minnow: a principled framework for rapid simulation of dscRNA-seq data at the read level”. In: *Bioinformatics* 35.14 (2019), pp. i136–i144.

[46] Gioele La Manno et al. “RNA velocity of single cells”. In: *Nature* 560.7719 (2018), p. 494.