



Article

Single-Cell RNA Sequencing for the Detection of Clonotypic V(D)J Rearrangements in Multiple Myeloma

Antonio Matera ^{1,†} , Alessio Marella ^{1,†}, Akihiro Maeda ¹ , Matteo C. Da Vià ², Francesca Lazzaroni ², Sonia Fabris ², Stefania Pioggia ², Laura Porretti ² , Federico Colombo ², Federica Torricelli ³, Antonino Neri ⁴, Elisa Taiana ², Giuseppina Fabbiano ², Valentina Traini ¹, Elisa Genuardi ⁵ , Daniela Drandi ⁵ , Niccolò Bolli ^{1,2,*} and Marta Lionetti ¹

¹ Department of Oncology and Hemato-Oncology, University of Milan, 20122 Milan, Italy

² Hematology Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, 20122 Milan, Italy

³ Laboratory of Translational Research, Azienda USL-IRCCS di Reggio Emilia, 42123 Reggio Emilia, Italy

⁴ Scientific Directorate, Azienda USL-IRCCS di Reggio Emilia, 42123 Reggio Emilia, Italy

⁵ Department of Molecular Biotechnologies and Health Sciences, University of Torino, 10124 Torino, Italy

* Correspondence: niccolo.bolli@unimi.it

† These authors contributed equally to this work.

Abstract: Multiple myeloma (MM) has a highly heterogeneous genetic background, which complicates its molecular tracking over time. Nevertheless, each MM patient's malignant plasma cells (PCs) share unique V(D)J rearranged sequences at immunoglobulin loci, which represent ideal disease biomarkers. Because the tumor-specific V(D)J sequence is highly expressed in bulk RNA in MM patients, we wondered whether it can be identified by single-cell RNA sequencing (scRNA-seq). To this end we analyzed CD138⁺ cells purified from bone marrow aspirates of 19 samples with PC dyscrasias by both a standard method based on bulk DNA and by an implementation of the standard 10x Genomics protocol to detect expressed V(D)J sequences. A dominant clonotype was easily identified in each sample, accounting on average for 83.65% of V(D)J-rearranged cells. Compared with standard methods, scRNA-seq analysis proved highly concordant and even more effective in identifying clonal productive rearrangements, by-passing limitations related to the misannealing of consensus primers in hypermutated regions. We next validated its accuracy to track 5 clonal cells with absolute sensitivity in a virtual sample containing 3180 polyclonal cells. This shows that single-cell V(D)J analysis may be used to find rare clonal cells, laying the foundations for functional single-cell dissection of minimal residual disease.

Keywords: multiple myeloma; single-cell RNA sequencing; clonal biomarker; V(D)J rearrangement



Citation: Matera, A.; Marella, A.; Maeda, A.; Da Vià, M.C.; Lazzaroni, F.; Fabris, S.; Pioggia, S.; Porretti, L.; Colombo, F.; Torricelli, F.; et al. Single-Cell RNA Sequencing for the Detection of Clonotypic V(D)J Rearrangements in Multiple Myeloma. *Int. J. Mol. Sci.* **2022**, *23*, 15691. <https://doi.org/10.3390/ijms232415691>

Academic Editor: Beate Heissig

Received: 15 November 2022

Accepted: 6 December 2022

Published: 10 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multiple myeloma (MM) is a malignant proliferation of antibody-secreting bone marrow (BM) plasma cells (PCs) that accounts for slightly more than 17% of all hematological malignancies in the United States [1]. It is characterized by a highly heterogeneous genetic background and clinical course, and remains an incurable disease [2,3]. Mutations and copy number changes are not stable during the disease course from asymptomatic stages [4,5], newly diagnosed myeloma [6], and relapsed-refractory myeloma [7,8]. The clonal B cell origin of MM ensures that all malignant PCs share the same immunoglobulin (Ig) heavy and light chain variable regions. Due to the great diversity introduced during formation of the mature Ig gene through V(D)J recombination, junctional insertions/deletions and somatic hypermutation, it is nearly impossible that independent B cell clones share identical variable regions. The uniqueness of these tumor-specific sequences, along with their stability over time despite differential clonal evolution, make them ideal biomarkers for minimal residual disease (MRD) monitoring [9].

In recent years, the evaluation of the transcriptome of individual cells has become possible thanks to the development of single-cell RNA sequencing (scRNA-seq), a powerful technology exploited by several commercial platforms, all of which forecast single-cell isolation before capturing of RNA molecules, reverse transcription, cDNA amplification, library preparation and next-generation sequencing (NGS). ScRNA-seq is able to dissect cell-to-cell variation in tumors and microenvironments, and therefore its application has the potential to provide the greatest insights in cell populations characterized by high heterogeneity, such as the malignant PC clone in MM. It is therefore natural that great efforts of scRNA-seq in MM take place immediately, making it possible to obtain a detailed molecular characterization of tumor and immune cells in symptomatic and asymptomatic patients [10–13]. Because the tumor-specific V(D)J sequence is highly expressed in bulk RNA in MM patients [14], the question is whether it can also be identified in single-cells and with what accuracy as compared to clinical grade V(D)J diagnostics based on consensus primers and a DNA template. This would allow a whole new field of study based on single-cell analysis of MRD positive (MRD+) cells.

Here, we evaluated the efficacy of scRNA-seq in identifying the MM-specific V(D)J rearrangement by means of an implementation of the standard 10x Genomics protocol. The output thus generated from single BM PCs from 19 patients' samples was then compared with the sequences of the rearrangements determined using a standard method based on the EuroClonality-NGS Working Group standard operating procedure applied on bulk BM PC DNA [15] and submitting data to the Vidjil web platform for analysis of high-throughput immune repertoire sequencing [16].

2. Results

We purified CD138⁺ BM PCs with magnetic beads separation in 19 samples from 18 patients affected by MGUS (6), SMM (11), and MM (1). We analyzed a total of 61,022 barcodes estimated to be associated with cells that express targeted V(D)J transcripts, with an average of 3212 per sample (range: 606–7379) (Figure 1a). A dominant clonotype, i.e., one that was over-represented in each sample PC population, was easily identified in each sample. On average, 83.65% of V(D)J rearranged cells (range: 49.79–99.44% in each sample) belonged to the dominant clonotype (Figure 1b); in absolute numbers, the 19 clonotypes accounted for 53,935 of these 61,022 barcodes ($53,935/61,022 = 88.39\%$). As expected, the most abundant IGHV family was IGHV3, with IGHV3-43 as the most represented gene, and IGHJ4 the prevalent IGHJ family (Figure 1c, d, respectively). In CellRanger output, we also found 8806 barcodes not associated with cells expressing targeted V(D)J transcripts (Figure 1e). According to automated transcriptome-based cell type assignment, these cells with unrearranged Ig heavy and light chain loci are components of the BM immune microenvironment (i.e., T cells, monocytes, and NK cells, detailed in Figure 1f), and the presence of their transcriptome in the 5' gene expression libraries reflects unwanted contamination of the positive selection of CD138⁺ cells.

We then compared scRNA-seq-based IGH analysis results with those obtained by applying the EuroClonality-NGS Working Group standard operating procedure for two-step Ig NGS-based marker identification, starting from genomic DNA of CD138⁺ cells. The V(D)J rearranged sequences of dominant clonotypes identified by the two methods were identical in 17/19 samples (Figure 2a, Table 1). Relative clonal fractions were heterogeneous, very moderately correlated inter-assay (Pearson's product-moment correlation: $R^2 = 0.19$; p value = 0.08), and constantly higher in scRNA-seq analysis (Figure 2b).

Dominant IGH rearrangement sequences determined by scRNA-seq and amplicon-based NGS from bulk DNA shared an exact match, supporting the robustness of the single cell-based approach. However, in 6/17 cases the V gene recognized as the most similar to the one involved in the patient-specific rearrangement by Vidjil and CellRanger was not the same (Table 1). This can be explained by slight differences in alignment among the two algorithms, made more likely by the low degree of homology of these genes with the germline repertoire due to their somatic hypermutation.

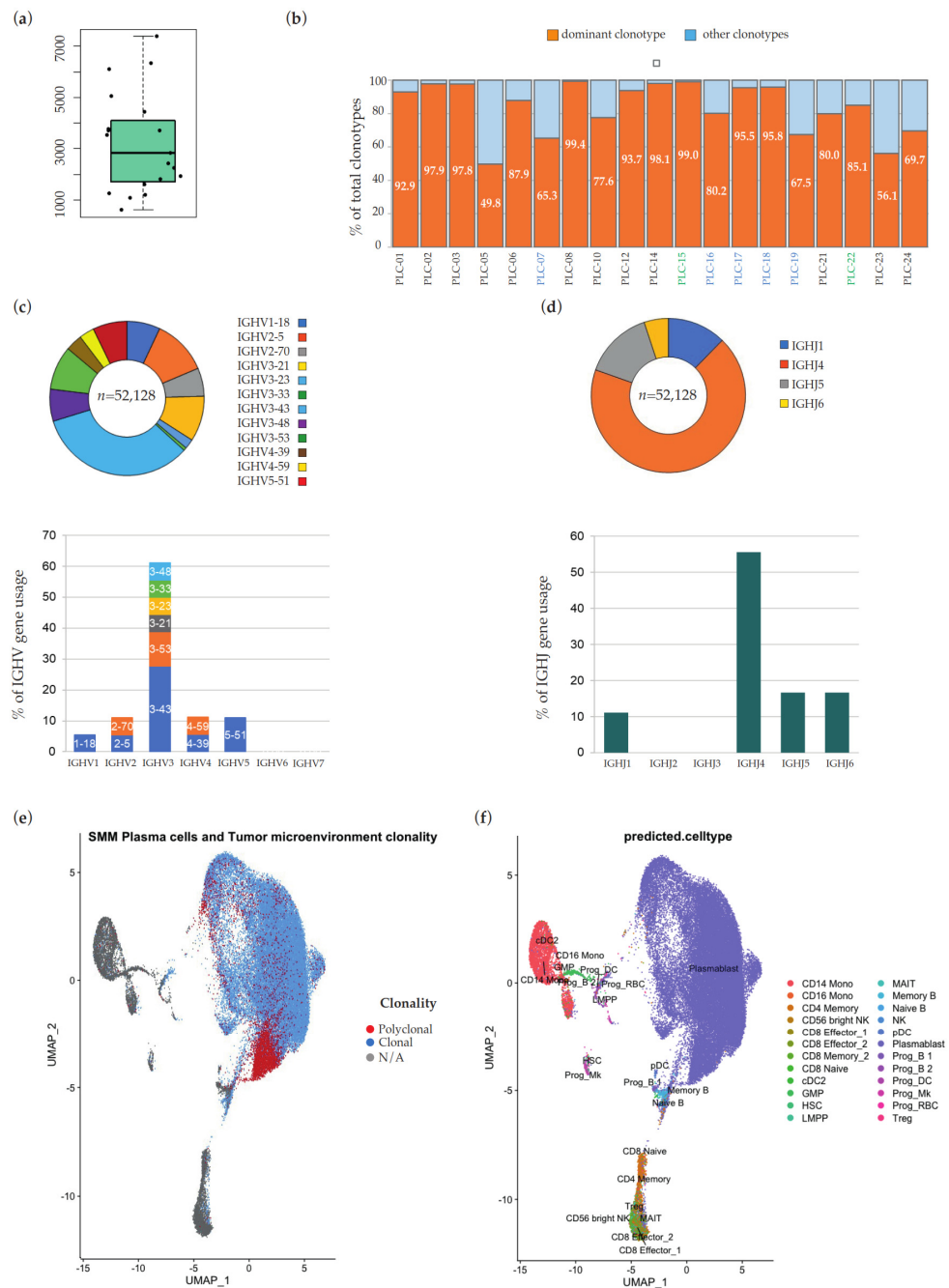


Figure 1. (a) Distribution of the number of barcodes estimated to be associated with cells that express targeted V(D)J transcripts in the 19 patients' samples. (b) Percentage of barcodes associated with the dominant clonotype and with all other clonotypes, respectively, in the 19 patients' samples. On the x-axis, IDs of MGUS samples are in blue, IDs of SMM samples are in black and IDs of MM samples are in green. VH (c) and JH (d) gene usage in 18/19 MM samples with clonal heavy chain rearrangements. The histograms show the sample-level rate of use of each IgH variable region gene family on total of clonal rearrangements. Within each V-family, discrete bands represent each of the individual genes, as indicated. Above each histogram, the representativeness of relative IGH genes/families is plotted at single-cell-level, as number of associated cellular barcodes. (e,f) UMAP dimension reduction of all the barcodes sequenced from samples obtained after CD138-based magnetic beads positive selection. In (e), light blue dots represent barcodes associated with the dominant clonotype of each sample, red dots barcodes associated with other rearrangements, and grey dots barcodes with unrearranged Ig heavy and light chain loci. In (f), dots are color-coded according to transcriptome-based cell type assignment, as indicated in the legend.

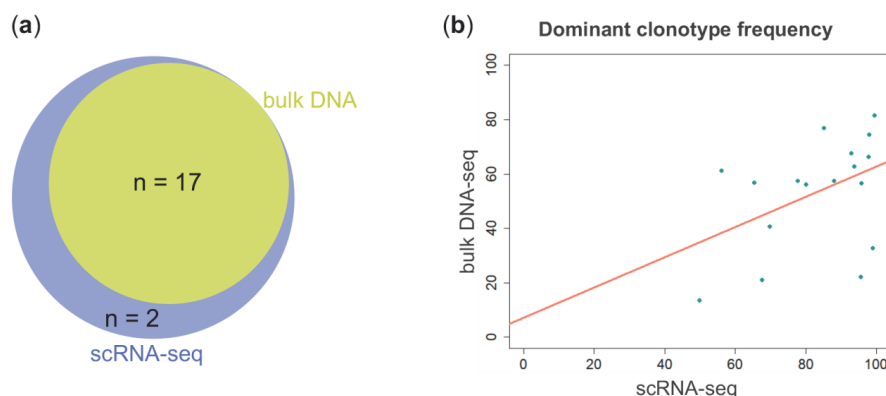


Figure 2. (a) Venn diagram showing the number of unique and shared clonotypes between scRNA-seq (blue circle) and bulk DNA (yellow circle) analyses. (b) Scatterplot of the percentage frequencies of the 17 dominant clonotypes in (a) as determined by scRNA-seq (x axis) and bulk DNA-seq (y axis) analyses.

Table 1. Comparison of V(D)J clonal rearrangements at IGH locus identified by scRNA-seq and amplicon-based NGS from bulk DNA in the 17/19 samples in which the two methods gave matching results.

Patient ID	scRNA-Seq			Amplicon-Based NGS from Bulk DNA			Common CDR3
	IGHV	IGHJ	Proportion	IGHV	IGHJ	Proportion	
PLC-01	IGHV3-48	IGHJ4	92.90%	IGHV3-30*09	IGHJ4*02	67.55%	CARDSYEDYVYW
PLC-02	IGHV3-21	IGHJ4	97.93%	IGHV3-21*01	IGHJ4*02	74.39%	CARYQLDAVAGKWGHYFDYW
PLC-03	IGHV3-43	IGHJ5	97.77%	IGHV3-9*01	IGHJ4*02	66.23%	CAKARLPLVGGGLDSW
PLC-05	IGHV3-43	IGHJ6	49.79%	IGHV3-9*01	IGHJ6*02	13.62%	CTRIVIGSGASCYDCYYHGMDVW
PLC-06	IGHV3-53	IGHJ5	87.91%	IGHV3-53*01	IGHJ4*02	57.58%	CARGLTAPGFPLDSW
PLC-07	IGHV3-23	IGHJ6	65.26%	IGHV3-23*01	IGHJ6*01	56.92%	CAKGRADCTDGVCYRRYGMVDVW
PLC-08	IGHV3-43	IGHJ4	99.44%	IGHV3-43*01	IGHJ4*02	81.44%	CVKGQGGYTYGGFDCW
PLC-10	IGHV5-51	IGHJ4	77.63%	IGHV5-51*01	IGHJ4*02	57.39%	CARTNWPYYFDHW
PLC-12	IGHV3-43	IGHJ4	93.73%	IGHV3-9*01	IGHJ4*02	62.86%	CARDRYQLIYYFDRW
PLC-15	IGHV3-43	IGHJ4	98.97%	IGHV3-9*01	IGHJ4*02	32.77%	CAKDVRVRYGYGSTQSGAFDYW
PLC-17	IGHV4-39	IGHJ4	95.53%	IGHV4-39*07	IGHJ4*02	22.10%	CARDKTTMTFSSPIFDYW
PLC-18	IGHV2-5	IGHJ1	95.78%	IGHV2-5*02	IGHJ1*01	56.58%	CAHSGSMWSGYAGTEYFQHW
PLC-19	IGHV4-59	IGHJ4	67.50%	IGHV4-59*01	IGHJ4*02	21.08%	CARAGDYDLLLLDYW
PLC-21	IGHV5-51	IGHJ6	80.02%	IGHV5-51*03	IGHJ6*03	56.06%	CARLPQGGYYMDVW
PLC-22	IGHV3-53	IGHJ5	85.10%	IGHV3-53*01	IGHJ4*02	77.01%	CARGLTAPGFPLDSW
PLC-23	IGHV3-33	IGHJ1	56.11%	IGHV3-30-3*02	IGHJ1*01	61.13%	CAFAIGADGEYFQHW
PLC-24	IGHV2-70	IGHJ4	69.68%	IGHV2-70*01	IGHJ4*02	40.67%	CARGASETQVAMSTAELYFFDSW

Two patients showed a discordant rearrangement between DNA-based NGS and scRNA-seq. For patient sample PLC-14, V(D)J analysis from bulk DNA identified only one clonotype corresponding to an unproductive rearrangement due to stop codons (IGHV3-23*01/IGHD3-9*01/IGHJ5*02) in 78% of reads. Conversely, scRNA-seq detected a productive rearrangement involving IGHV1-18 and IGHJ4 genes in 98.08% of barcodes. Each of these V(D)J sequences is exclusive to one type of approach, and there is no trace of it in the other. The failure of scRNA-seq to detect the unproductive rearrangement is expected and attributable to the degradation of the transcript containing premature translation-termination codons by the nonsense-mediated decay (NMD) mechanism. A mutation in the J gene involved in the productive rearrangement, on the other hand, is in all likelihood at the origin of its refractoriness to amplification by the EuroClonality-NGS protocol. The mutated nucleotide, in fact, causes a misannealing of input DNA with the 3' end of both reverse primers used in the first-step PCR, and mismatches in this position are known to be detrimental to PCR priming (Figure 3).

	IGHV1-18	IGHJ4	IGHJ4	IGHJ4	IGHJ4	IGHG1	IGHG1
Universal Reference	CGAGAGA	ACTACTTTGACTACTGGGGCCAGGGAACCTGGTACACGTCTCTCAG				CCTCCACCAAGGGCCCAT	
Donor Reference	CGAGAGA						
Consensus	C	CC	CGA	ACCTTGACTACTGGGGCCAGGGAACCT	GGTACACGTCTCTCAG	CCTCCACCAAGGGCCCAT	

3'CCAGTGGCAGAGGAGTCCAYTC-5'

Figure 3. Alignment between the DNA sequence of EuroClonality-NGS protocol’s reverse primers IGH-J-A-1 and IGH-J-A-2 (written in 3’-to-5’ direction) and the rearranged sequence of the dominant clonotype (“Consensus”) as displayed by Loupe V(D)J browser for patient sample PLC-14. Nucleotides highlighted in orange in the Consensus sequence represent mutated positions compared with the germline IGHJ4 gene (“Universal Reference”). The gray background indicates where the sequences align to the reference. The purple background corresponds to a nucleotide deletion. The red oval indicates the mutated nucleotide causing a misannealing of input DNA with the 3’ end of both reverse primers used in the first-step PCR, and thus prevents PCR priming. Nucleotides indicated in grey in the primers’ sequence belong to the intron separating the J segment of the rearranged V-region and the C-region sequence. This intron is removed after transcription by RNA splicing joining the V-region exon to the C-region sequence, and for this reason, this portion of the reverse primers does not match with the consensus sequence of mature RNA transcript.

Furthermore, our analysis from bulk DNA was unable to identify any productive clonotype in patient sample PLC-16, where only one dominant unproductive rearrangement was detected (IGHV3-11*05/IGHD3-3*01/IGHJ5*02, 29.97% of reads). ScRNA-seq of this sample found one light chain-only dominant clonotype (IGKV1-33/IGKJ3; 79% of reads). This was consistent with serum immunofixation electrophoresis results for this patient, showing a k light-chain myeloma.

Our results therefore suggest that scRNA-seq can accurately identify dominant clonotypes in PC dyscrasias at diagnosis. We next asked if the same methodology can be applied to remission samples after treatment, i.e., if scRNA-seq can identify few residual cells of a known clonotype and therefore be a suitable methodology for transcriptomic characterization of MRD+ cells. To this end, we informatically generated a virtual patient sample where 5 known clonal PCs were admixed with 3180 cells with a polyclonal V(D)J sequence. Remarkably, the tool successfully identified all 5 cells belonging to the MRD clonotype within the polyclonal background (Figure 4).

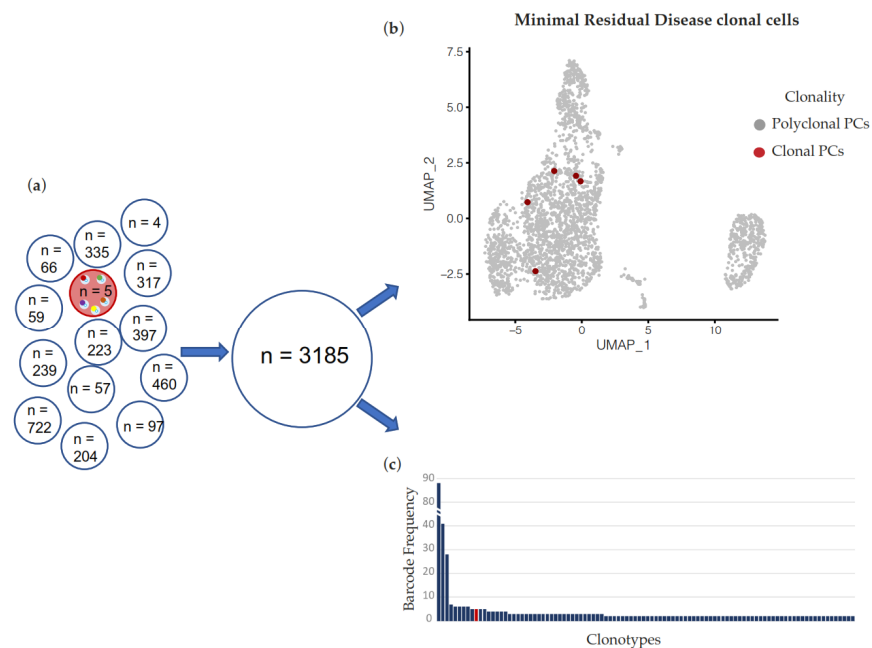


Figure 4. (a) In silico dilution of reads from 5 barcodes associated with the dominant clonotype of patient’s sample PLC-10 (red) into reads from 3180 barcodes of 13 different samples. (b) UMAP

dimension reduction of the simulated sample. Red dots represent barcodes associated with PLC-10-specific V(D)J rearrangement. (c) Bar plot of clonotypes' abundance in the virtual fastq file generated in (a) and analyzed by Loupe V(D)J Browser. Clonotypes after the 100th rank according to the barcode frequency are not plotted. The red bar represents the clonotype made up of the five barcodes sharing the myeloma-specific V(D)J rearrangement of PLC-10. The presence of the top-ranked clonotypes displayed at the left of the red bar is dependent on the procedure followed for the generation of the virtual sample, and is compatible with the composition of the V(D)J repertoire observable in a normal sample [17].

3. Discussion

The tumor cell-specific rearrangement in the immunoglobulin V(D)J gene region (particularly at the IGH locus) is already exploited as a target by NGS (both amplicon- and capture-based [18]) or allele specific oligonucleotide (ASO) PCR. Here, the identification of the clonotypic rearrangement is particularly worthwhile also in routine clinical practice, as it represents an extremely specific molecular marker that is useful for monitoring the disease burden along its various stages. Indeed, MRD evaluation is proving more and more useful due to the impact that the depth of the response has on the outcome, and it could also act as a rapidly assessable surrogate trial endpoint, thus bridging the increasing delay between drug development and approval.

Overall, our findings argue that scRNA-seq provides robust and accurate data to derive the sequence of such clonal V(D)J rearrangements in single-cells. Specifically, in our study cohort, this approach proved even more effective in identifying clonal productive Ig rearrangements than the V(D)J analysis from bulk DNA, which in two cases failed. The data emerging from scRNA-seq could in both cases be exploited to overcome the limitations demonstrated by the conventional method. In particular, scRNA-seq can overcome false negative results due to DNA mispriming. Furthermore, the comprehensive DNA-based genotyping of the B cell receptor requires a high amount of starting DNA and a high number of PCRs, which can be a limiting factor in some circumstances, while analysis of few thousand cells is feasible in most cases. On the other hand, it is worth underlining that in MM, where the V(D)J rearrangements are expected to be productive, the Ig marker screening by scRNA-seq does not present some limitations which, on the contrary, may be a limiting factor in other B neoplasms, such as acute lymphoblastic leukemia (ALL). Here, in fact, most of the rearrangements are unproductive, and therefore the Ig/TCR marker screening by RNA-seq, targeting productive transcripts, can be incomplete [19].

ScRNA-seq, in its standard version and even more in its various implementations (CITE-seq, scATAC-seq, etc.), is an approach capable of providing an unprecedented amount of information, but still extremely expensive and demanding at the level of data analysis. For these reasons, its application is far from widespread, but remains rather limited to research scenarios. To maximize the output of scRNA-seq analysis in MM, parallel to a 5' gene expression library, the generation of a V(D)J enriched library from amplified cDNA of the same cells constitutes a worthwhile solution that requires no additional input material, involves low wet-lab and sequencing costs, and an extremely simple data analysis. Indeed, the validation proposed here is of paramount importance as it opens new avenues of investigation for approaches aimed at functional single-cell MRD studies, where residual cells after treatment could be identified based on their V(D)J sequence and further characterized for gene and surface protein expression. The simulation of a post-therapy sample that we conducted by *in silico* dilution of clonal tumor cells supports the feasibility of such a scRNA-seq in the MRD context. Overall, our studies align with the general view that the potential of NGS is currently under-utilized in translational applications in MM [20–22], while a comprehensive characterization of its heterogeneity, as well as the functional properties of residual cells after treatment, may favorably impact the outcome of patients.

4. Materials and Methods

4.1. Patients

The study was based on a series of 18 patients with PC dyscrasia admitted to our institution. Monoclonal gammopathy of undetermined significance (MGUS), smoldering MM (SMM), and MM were diagnosed according to the International Myeloma Working Group (IMWG) revised criteria [23]. The study was approved by the local Ethics Committee (Provision n. 575 dated 29 March 2018) and written informed consent was obtained from all of the patients involved in the study. The study was conducted according to good clinical practice and the ethical principles outlined in the Declaration of Helsinki. BM sampling was done at diagnosis, and repeated in a SMM patient at progression to symptomatic disease, totaling 19 samples. White blood cells were obtained from BM aspirates, after red cells lysis, and CD138⁺ PCs were isolated by an immunomagnetic method with anti-CD138 monoclonal antibodies (STEMCELL Technologies, Vancouver, BC, Canada).

4.2. DNA-Based Molecular Analysis of IGHV Rearrangement

For the purification of genomic DNA from CD138⁺ PCs, we used the AllPrep DNA/RNA/miRNA isolation kit (Qiagen, Hilden, Germany), following the manufacturer's protocol. We searched for clonal IGHV rearrangements by applying the EuroClonality-NGS Working Group standard operating procedure for two-step Ig NGS-based marker identification [Version 1.0 (11 June 2019)] using the IGH V-J set of primers. Libraries were pooled and sequenced using 250 bp paired-end runs on a MiSeq instrument (Illumina) to an average of 31.96 Mbp per sample.

Fastq files were uploaded on the Vidjil [16] web application for the analysis of high-throughput sequencing reads on IGH locus.

4.3. Single-Cell V(D)J Analysis

Fresh cells isolated via CD138⁺ magnetic bead separation were processed according to 10x Genomics Chromium Single-cell 5' Gene Expression workflow and, in parallel, Chromium Single-cell V(D)J Enrichment protocol. In particular, sample partitioning and molecular barcoding were done on the Chromium Controller (10x Genomics), where we loaded cellular suspensions together with the Single-cell 5' Gel Beads on a Single-cell 5' chip, in which gel beads in emulsion (GEM) generation took place. Each gel bead is functionalized with barcoded oligonucleotides that consists of: (i) an Illumina R1 sequence, (ii) a 16 bp 10x barcode to index GEMs, (iii) a 10 bp randomer to index molecules (unique molecular identifier, UMI), and (iv) a 13 nt template switch oligo (TSO). Reverse transcription (RT) of polyadenylated RNA transcripts took place using poly(dT) primers inside each GEM, after which cDNAs (each containing a UMI and shared 10x barcode per GEM (cell), and ending with a TSO at the 3' end) were pooled for amplification and library construction in bulk. Specifically, from the amplified cDNA of each sample, we prepared in parallel a 5' gene expression library and a V(D)J enriched library. The latter was obtained after enrichment of 10x barcoded, full-length V(D)J segments via PCR amplification with primers specific to Ig constant regions. Generated libraries were combined according to Illumina specifications and paired-end sequenced (2 × 150 bp) on Illumina NovaSeq platform to a depth of ~150,000 and 15,000 reads/cell for 5' gene expression and V(D)J enriched libraries, respectively.

We used the Cell Ranger software (10x Genomics) to process scRNA-seq data. In particular, *mkfastq* command allowed raw data demultiplexing, then *count* and *vdj* were used for transcriptome and V(D)J raw data alignment to the reference genome and gene count matrix generation, respectively. Output cloupe files and vloupe files (these latter containing clonotypes and CDR3 sequences deriving from paired clonotype calling) were overlaid for an integrated analysis by Loupe Browser and Loupe V(D)J Browser. In particular, Loupe VDJ Browser was used to explore the clonality and diversity of the B cell receptor repertoire at the single-cell level. For the performance evaluation of this immunoinformatics analysis, we generated an in silico dilution of reads from 5 barcodes associated with the dominant

clonotype of a patient's sample into reads from 3180 barcodes of 13 different samples. In parallel, samples were analyzed through the Seurat pipeline of analysis [24]. In detail, we initially performed a quality check assessment removing those cells with less than 200 and more than 3000 expressed features. The first threshold allowed us to remove possible cell-free mRNA, the second one to exclude doublets. Moreover, we also removed died/dying cells by filtering out cells expressing more than 5% of mitochondrial genes. Then, to better define the possible transcriptomic differences between each sample, we removed immunoglobulin related genes from the expression matrix. Finally, the 19 samples were integrated through the *findintegrationanchors* and *intergratedata* Seurat functions [24]. Once the integration was completed, the whole data set was processed using the Doubletfinder tool [25]: only cells defined as "singlets" were retained for further analysis. The entire dataset was Log normalized and scaled, and principal components analysis was performed. Automated cell assignment was performed anchoring our dataset to the already published bone marrow annotated atlas by Hao et al. [24]. The Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) was performed by the UMAP-learn algorithm [26].

Author Contributions: Investigation, A.M. (Akihiro Maeda), A.M. (Antonio Matera), S.F., S.P., L.P., F.C., F.T., E.T., G.F., V.T., E.G. and D.D.; formal analysis: A.M. (Antonio Matera), M.C.D.V., F.L., M.L. and A.M. (Alessio Marella); writing—original draft preparation, M.L., M.C.D.V. and A.M. (Antonio Matera); writing—review and editing, M.L., M.C.D.V., E.G., D.D., A.N. and N.B.; supervision, N.B. All authors have read and agreed to the published version of the manuscript.

Funding: N.B. was funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 817997). N.B. was funded by Associazione Italiana Ricerca sul Cancro (IG25739). M.C.D.V. was funded by Umberto Veronesi Foundation. M.C.D.V. was funded by Pfizer Global Medical Grants (grant tracking No. 75340503). This study was (partially) funded by Italian Ministry of Health—Current research IRCCS.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by Comitato Etico Milano Area 2 (Provision n. 575 dated 29 March 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available upon request to the corresponding author.

Conflicts of Interest: N.B. serves on Advisory Board and Speakers Bureau for Celgene and Janssen, and on Speakers Bureau for Takeda, Amgen. The remaining Authors declare no competing financial interest.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer Statistics, 2017. *CA Cancer J. Clin.* **2017**, *67*, 7–30. [[CrossRef](#)]
2. Bolli, N.; Biancon, G.; Moarii, M.; Gimondi, S.; Li, Y.; de Philippis, C.; Maura, F.; Sathiaselan, V.; Tai, Y.T.; Mudie, L.; et al. Analysis of the Genomic Landscape of Multiple Myeloma Highlights Novel Prognostic Markers and Disease Subgroups. *Leukemia* **2018**, *32*, 2604–2616. [[CrossRef](#)]
3. Morgan, G.J.; Walker, B.A.; Davies, F.E. The Genetic Architecture of Multiple Myeloma. *Nat. Rev. Cancer* **2012**, *12*, 335–348. [[CrossRef](#)]
4. Bolli, N.; Maura, F.; Minvielle, S.; Gloznic, D.; Szalat, R.; Fullam, A.; Martincorena, I.; Dawson, K.J.; Samur, M.K.; Zamora, J.; et al. Genomic Patterns of Progression in Smoldering Multiple Myeloma. *Nat. Commun.* **2018**, *9*, 3363. [[CrossRef](#)] [[PubMed](#)]
5. Oben, B.; Froyen, G.; Maclachlan, K.H.; Leongamornlert, D.; Abascal, F.; Zheng-Lin, B.; Yellapantula, V.; Derkach, A.; Geerdens, E.; Diamond, B.T.; et al. Whole-Genome Sequencing Reveals Progressive versus Stable Myeloma Precursor Conditions as Two Distinct Entities. *Nat. Commun.* **2021**, *12*, 1861. [[CrossRef](#)]
6. Bolli, N.; Avet-Loiseau, H.; Wedge, D.C.; Van Loo, P.; Alexandrov, L.B.; Martincorena, I.; Dawson, K.J.; Iorio, F.; Nik-Zainal, S.; Bignell, G.R.; et al. Heterogeneity of Genomic Evolution and Mutational Profiles in Multiple Myeloma. *Nat. Commun.* **2014**, *5*, 2997. [[CrossRef](#)] [[PubMed](#)]
7. Misund, K.; Bruinink, D.H.O.; Coward, E.; Hoogenboezem, R.M.; Rustad, E.H.; Sanders, M.A.; Rye, M.; Sponaas, A.-M.; van der Holt, B.; Zweegman, S.; et al. Clonal Evolution after Treatment Pressure in Multiple Myeloma: Heterogenous Genomic Aberrations and Transcriptomic Convergence. *Leukemia* **2022**, *36*, 1887–1897. [[CrossRef](#)] [[PubMed](#)]
8. Ziccheddu, B.; Biancon, G.; Bagnoli, F.; De Philippis, C.; Maura, F.; Rustad, E.H.; Dugo, M.; Devecchi, A.; De Cecco, L.; Sensi, M.; et al. Integrative Analysis of the Genomic and Transcriptomic Landscape of Double-Refractory Multiple Myeloma. *Blood Adv.* **2020**, *4*, 830–844. [[CrossRef](#)]

9. Rustad, E.H.; Misund, K.; Bernard, E.; Coward, E.; Yellapantula, V.D.; Hultcrantz, M.; Ho, C.; Kazandjian, D.; Korde, N.; Mailankody, S.; et al. Stability and Uniqueness of Clonal Immunoglobulin CDR3 Sequences for MRD Tracking in Multiple Myeloma. *Am. J. Hematol.* **2019**, *94*, 1364–1373. [[CrossRef](#)]
10. Ledergor, G.; Weiner, A.; Zada, M.; Wang, S.Y.; Cohen, Y.C.; Gatt, M.E.; Snir, N.; Magen, H.; Koren-Michowitz, M.; Herzog-Tzarfati, K.; et al. Single Cell Dissection of Plasma Cell Heterogeneity in Symptomatic and Asymptomatic Myeloma. *Nat. Med.* **2018**, *24*, 1867–1876. [[CrossRef](#)]
11. Liu, R.; Gao, Q.; Foltz, S.M.; Fowles, J.S.; Yao, L.; Wang, J.T.; Cao, S.; Sun, H.; Wendl, M.C.; Sethuraman, S.; et al. Co-Evolution of Tumor and Immune Cells during Progression of Multiple Myeloma. *Nat. Commun.* **2021**, *12*, 2559. [[CrossRef](#)] [[PubMed](#)]
12. Zavidij, O.; Haradhvala, N.J.; Mouhieddine, T.H.; Sklavenitis-Pistofidis, R.; Cai, S.; Reidy, M.; Rahmat, M.; Flaifel, A.; Ferland, B.; Su, N.K.; et al. Single-Cell RNA Sequencing Reveals Compromised Immune Microenvironment in Precursor Stages of Multiple Myeloma. *Nat. Cancer* **2020**, *1*, 493–506. [[CrossRef](#)] [[PubMed](#)]
13. Frede, J.; Anand, P.; Sotudeh, N.; Pinto, R.A.; Nair, M.S.; Stuart, H.; Yee, A.J.; Vijaykumar, T.; Waldschmidt, J.M.; Potdar, S.; et al. Dynamic Transcriptional Reprogramming Leads to Immunotherapeutic Vulnerabilities in Myeloma. *Nat. Cell Biol.* **2021**, *23*, 1199–1211. [[CrossRef](#)]
14. Rashid, N.U.; Sperling, A.S.; Bolli, N.; Wedge, D.C.; Van Loo, P.; Tai, Y.T.; Shamma, M.A.; Fulciniti, M.; Samur, M.K.; Richardson, P.G.; et al. Differential and Limited Expression of Mutant Alleles in Multiple Myeloma. *Blood* **2014**, *124*, 3110–3117. [[CrossRef](#)]
15. Brüggemann, M.; Kotrova, M.; Knecht, H.; Bartram, J.; Boudjogrha, M.; Bystry, V.; Fazio, G.; Froňková, E.; Giraud, M.; Grioni, A.; et al. Standardized Next-Generation Sequencing of Immunoglobulin and T-Cell Receptor Gene Recombinations for MRD Marker Identification in Acute Lymphoblastic Leukaemia; a EuroClonality-NGS Validation Study. *Leukemia* **2019**, *33*, 2241–2253. [[CrossRef](#)]
16. Duez, M.; Giraud, M.; Herbert, R.; Rocher, T.; Salson, M.; Thonier, F. Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing. *PLoS ONE* **2016**, *11*, e0166126. [[CrossRef](#)] [[PubMed](#)]
17. Bashford-Rogers, R.J.M.; Nicolaou, K.A.; Bartram, J.; Goulden, N.J.; Loizou, L.; Koumas, L.; Chi, J.; Hubank, M.; Kellam, P.; Costeas, P.A.; et al. Eye on the B-ALL: B-Cell Receptor Repertoires Reveal Persistence of Numerous B-Lymphoblastic Leukemia Subclones from Diagnosis to Relapse. *Leukemia* **2016**, *30*, 2312–2321. [[CrossRef](#)]
18. Hultcrantz, M.; Rustad, E.H.; Yellapantula, V.; Arcila, M.; Ho, C.; Syed, M.H.; Papaemmanuil, E.; Zhang, Y.; Maura, F.; Landgren, O. Baseline VDJ Clonotype Detection Using a Targeted Sequencing NGS Assay: Allowing for Subsequent MRD Assessment. *Blood Cancer J.* **2020**, *10*, 76. [[CrossRef](#)]
19. van der Velden, V.H.J.; Bastian, L.; Brüggemann, M.; Hartmann, A.M.; Darzentas, N. Immunoglobulin/T-Cell Receptor Gene Rearrangement Analysis Using RNA-Seq. *Methods Mol. Biol.* **2022**, *2453*, 61–77. [[CrossRef](#)]
20. Bolli, N.; Genuardi, E.; Ziccheddu, B.; Martello, M.; Oliva, S.; Terragna, C. Next-Generation Sequencing for Clinical Management of Multiple Myeloma: Ready for Prime Time? *Front. Oncol.* **2020**, *10*, 189. [[CrossRef](#)]
21. Yellapantula, V.; Hultcrantz, M.; Rustad, E.H.; Wasserman, E.; Londono, D.; Cimera, R.; Ciardiello, A.; Landau, H.; Akhlaghi, T.; Mailankody, S.; et al. Comprehensive Detection of Recurring Genomic Abnormalities: A Targeted Sequencing Approach for Multiple Myeloma. *Blood Cancer J.* **2019**, *9*, 101. [[CrossRef](#)]
22. Maura, F.; Bolli, N.; Rustad, E.H.; Hultcrantz, M.; Munshi, N.; Landgren, O. Moving From Cancer Burden to Cancer Genomics for Smoldering Myeloma: A Review. *JAMA Oncol.* **2019**, *6*, 425. [[CrossRef](#)]
23. Rajkumar, S.V.; Dimopoulos, M.A.; Palumbo, A.; Blade, J.; Merlini, G.; Mateos, M.V.; Kumar, S.; Hillengass, J.; Kastritis, E.; Richardson, P.; et al. International Myeloma Working Group Updated Criteria for the Diagnosis of Multiple Myeloma. *Lancet Oncol.* **2014**, *15*, e538–e548. [[CrossRef](#)] [[PubMed](#)]
24. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M.; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e21. [[CrossRef](#)] [[PubMed](#)]
25. McGinnis, C.S.; Murrow, L.M.; Gartner, Z.J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **2019**, *8*, 329–337.e4. [[CrossRef](#)] [[PubMed](#)]
26. Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.-A.; Kwok, I.W.H.; Ng, L.G.; Ginhoux, F.; Newell, E.W. Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP. *Nat. Biotechnol.* **2018**, *37*, 38–44. [[CrossRef](#)] [[PubMed](#)]