Supplementary Materials and Methods

Materials and Methods

Sample collection and cell preparation

BM samples were filtered with a 100 µm strainer and red blood cells were lysed (BulkLysis, Cytognos, Spain) for 15 min at 20 °C and washed with PBS. PCs were isolated using an immuno-magnetic CD138 positive selection kit according to manufacturer's instructions (EasySepTM, STEMCELL Technologies, Vancouver, Canada). N=2 PCs samples were cryopreserved in 10% Dimethyl Sulfoxide, while N=3 samples were freshly processed. For patient P4, both $CD138^{pos}$ and $CD138^{neg}$ cell fractions were collected. Cell numbers and viability of the positive and negative fractions after PC separation were checked with Trypan Blue exclusion assay to ensure only samples with a viability >80% were used.

Single-cell library preparation and sequencing

Cells were counted and loaded into the Chromium Single Cell Chip G (PN-1000120) to yield a recovery of 9,000 single-cell transcriptomes. Subsequently, the libraries were pooled and sequenced on a NovaSeq6000 platform (Illumina Inc.) with a 150bp Paired End protocol, aiming at ~150000X and 15000 reads/cell for 5' gene expression and V(D)J enriched libraries, respectively. Raw sequences were preprocessed using Cell Ranger (version 5.0.0) pipelines using GRCh38 human reference transcriptome. In detail, mkfastq command for demultiplexing and FASTQ generation, then count and vdj commands for transcriptome and V(D)J raw data alignment and gene count matrix generation, respectively.

Preprocessing and batch correction of scRNA sequencing datasets

We used findintegrationanchors and intergratedata commands and proceeded with Seurat standard workflow to analyze the combined datasets. Batch-correction was carried out using the Harmony tool (github.com/immunogenomics/harmony) (v1.2.0)[25,26] using the "fresh/frozen" state as the variable on which compute the correction. Then the merged data were scaled and used for principal component analysis. To identify canonical marker genes that were conserved across genotypes and define the

clusters, the FindConservedMarkers function was used. To identify the DEGs between genotypes or selected groups, the FindMarkers function was used.

Identification of IGH translocations with FUSCIA

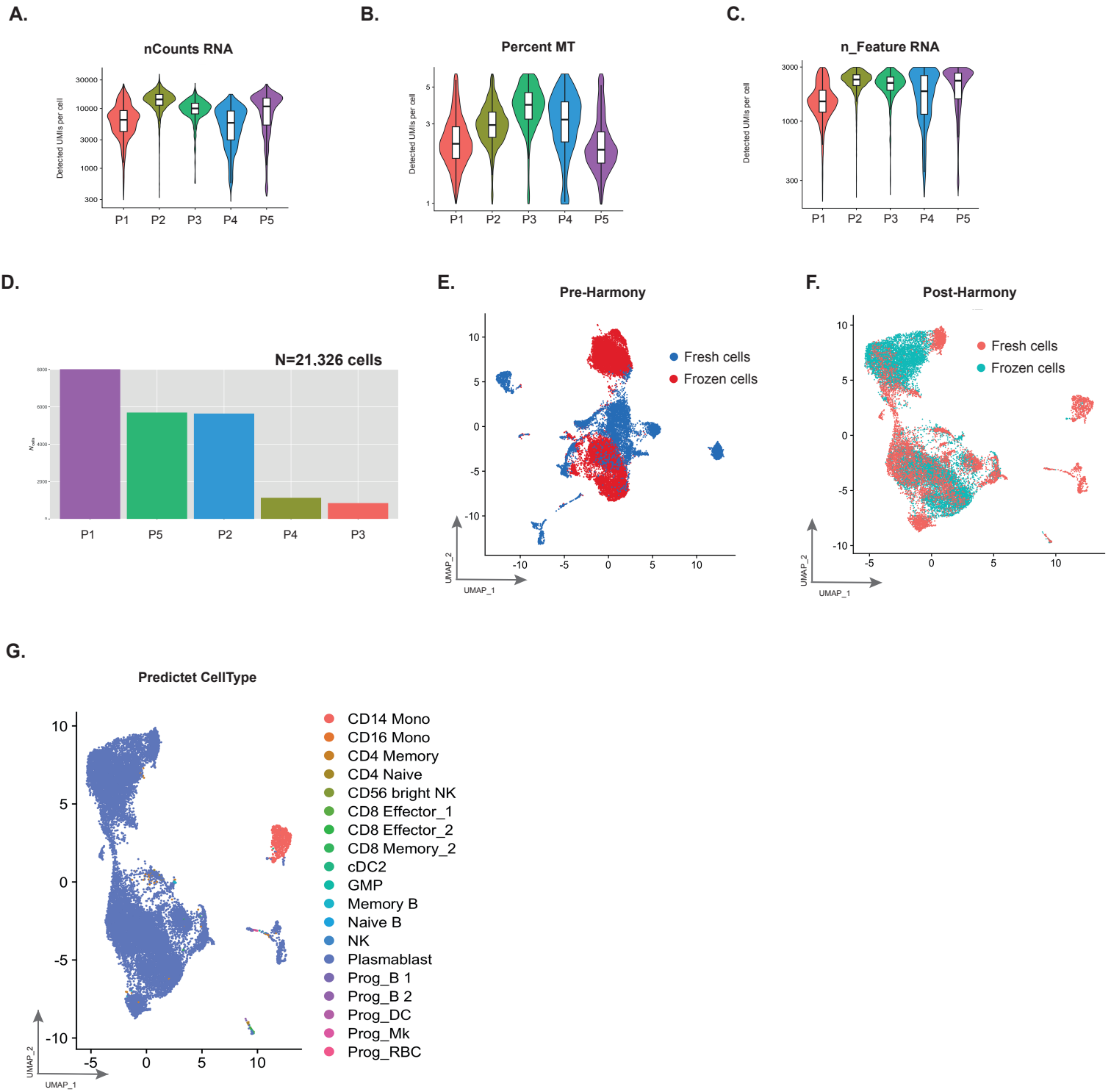
As input we used scRNA-seq.bam files and start-end coordinates of the genes, and the "discover_discordant_reads.py" function. For the analysis we investigated the main translocations involving the IGH locus in chr14 in MM (t(4;14), t(6;14), t(11;14), t(14;16), t(14;20)).
Barplots were created using the Prism 5.0 software (GraphPad Prism, RRID:SCR_002798).

inferCNV

In all samples, Ig genes were retained for this analysis. We used "denoise" and the default hidden markov model (HMM) settings, "subclusters" as analysis_mode and "ward.D2" as hclust_method. Low-probability CNVs were filtered out using "0.05" as threshold of tumor_subcluster_pval.

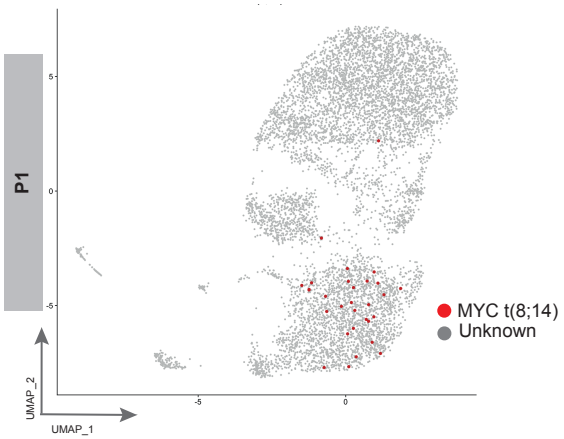
Pseudotime trajectory inference analysis

To reduce the dimensionality of our data and to filter out uninformative genes a filtering step was applied: only genes robustly expressed (at least 4 reads) in at least 10 cells were considered as potentially interesting cell-type marker genes. Quantile normalization was used to remove unwanted technical or biological artifacts from the data. Dimensionality reduction was performed using principal components analysis (PCA). After clusters identification the next step was to determine the global structure of the underlying lineages. Finally, the slingshot function was applied to obtain the trajectory inference.

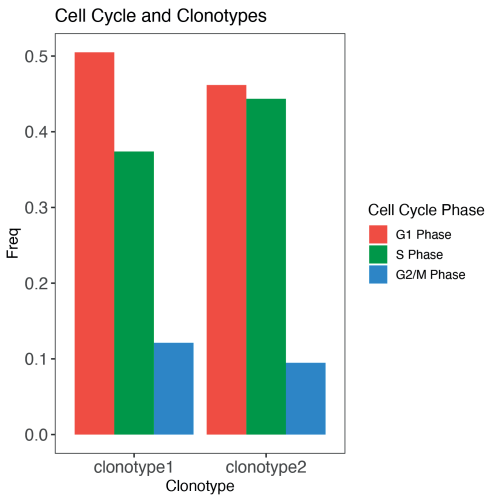

A.
nCounts RNA
Detected UMIs per cell
30000
10000
3000
1000
300
P1
P2
P3
P4
P5
B.
Percent MT
Detected UMIs per cell
5
3
1
P1
P2
P3
P4
P5
C.
n_Feature RNA
Detected UMIs per cell
3000
1000
300
P1
P2
P3
P4
P5
D.
N=21.326 cells
P1
P5
P2
P4
P3
E.
Pre-Harmony
Fresh cells
Frozen cells
UMAP_2
UMAP_1
F.
Post-Harmony
Fresh cells
Frozen cells
UMAP_2
UMAP_1
G.
Predictet CellType
CD14 Mono
CD16 Mono
CD4 Memory
CD4 Naive
CD56 bright NK
CD8 Effector_1
CD8 Effector_2
CD8 Memory_2
cDC2
GMP
Memory B
Naive B
NK
Plasmablast
Prog_B 1
Prog_B 2
Prog_DC
Prog_Mk
Prog_RBC
UMAP_2
UMAP_1

Supplementary Figure 2

A.

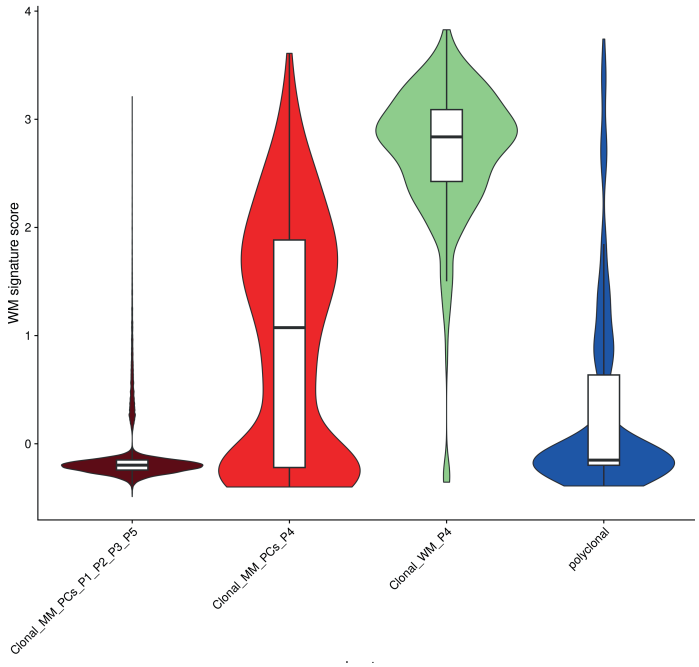


B.



Supplementary Figure 3

A.



Supplementary Figure Legends

Supplementary Figure 1. **Quality control of single cell transcriptome data. A-C.** Violin plot of quality metrics, as nCounts_RNA, percent MT and n-feature RNA per patient. **D.** Bar plot of cell counts by patients. **E,F**. UMAP embedding of five patients before and after batch effect correction, by Harmony pipeline. Colored by fresh vs. frozen processing. **G.** UMAP visualization of "*predicted cell type*" annotation after the process of mapping datasets of human BMNC to annotated references in Seurat.

Supplementary Figure 2. P1 clonotypes genomic inference and cell cycle analysis

A. t(8;14) fusion transcripts mapped on P1 UMAP . Cells characterized by fusion transcripts are highlighted by red dots. **B.** Barplot depicting the relative frequency (y-axis) of cell cycle phases: G1 phase (in red), S phase (in green), G2/M phase (in blue) with respect to the two dominant clonotypes of P1 (x-axis).

Supplementary Figure 3. Waldenstrom gene signature inference

A. Violin plot showing a WM signature score inferred across the entire dataset including: clonal PC from MM patients P1, P2, P3, P5 (in brown), clonal PCs from MM patient P4 (in red), clonal WM cells (PCs and Memory B) from patient P4 (in green), and polyclonal cells from all patients (in blue).

SUPPLEMENTARY TABLE 1. Clonotype 1 V(D)J genes and CDR3 sequences.

Patient	Clonotype 1				
	Gene C	Gene V	Gene J	CDR3	
P1	IGHG1	IGHV5-51	IGHJ4	TGTGCGAGAGGCGTAGTGGCTACGATTACCGCTCTTGACTACTGG	CARGVVATITALDYW
	IGKC	IGKV1D-33	IGKJ3	TGTCAACAGTATGATAATCTCCCTATATTCACTTTC	CQQYDNLPIFTF
P4	IGHA2	IGHV5-10-1	IGHJ6	TGCATGCAAAATCTACAAACTCCTGGGTTT	CATSSYDAAHCSGPTCHHYYMDVW
	IGKC	IGKV2D-28	IGKJ2	TGTGCGACCTCCTCCTACGATGCCGCCCATTGTAGTGGTCCCACCTGCCACCACTACTACATGGATGTCTGG	CMQNLQTPGF

CDR: Complementarity-determining region

SUPPLEMENTARY TABLE 2. Clonotype 2 V(D)J genes and CDR3 sequences.

Patient	Clonotype 2				
	Gene C	Gene V	Gene J	CDR3	
P1	-	-	-	-	-
	IGKC	IGKV1D-33	IGKJ3	TGTCAACAGTATGATAATCTCCCTATATTCACTTTC	CQQYDNLPIFTF
P4	IGHM	IGHV3-48	IGKJ6	TGTGCGAGAAGTCCTTATCACTATTACGGTCTGGACGTCTGG	CARSPYHYYGLDVW
	IGKC	IGKV3-20	IGKJ1	TGTCAGCACTATGGTAGTTCACCTATGTGGACGTTC	CQHYGSSPMWTF

CDR: Complementarity-determining region

Supplementary Table 3: WM signature gene list

	feature	group	avgExpr	logFC	statistic	auc	pval	padj	pct_in	pct_out
1	CD52	WM	2.30824159847504	1.09695403839427	14962680	0.746261741230514	2.72837743058084e-189	6.04799425036856e-187	89.7203999745239	56.8519968676586
2	RALGPS2	WM	2.0057830243136	1.00119535884933	15370866.5	0.766619990436992	5.90759831712257e-222	1.98414745296449e-219	86.2874976116171	56.7736883320282
3	JCHAIN	WM	1.68804751484783	1.49201896120123	14789954.5	0.737647079125536	5.08843340347824e-199	1.25328114727669e-196	54.3213808037705	12.1378230227095
4	CD74	WM	3.34707561752407	1.02796442316893	13308898.5	0.66377960154666	1.17125385112166e-84	7.41805260508964e-83	98.0701866123177	80.5794831636648
5	MS4A1	WM	2.54411909532941	1.31911880937416	16344861	0.815197840896866	0	0	91.4336666454366	62.9600626468285
6	ITM2B	WM	1.64865805407441	1.01462734612338	15435718.5	0.769854475598894	1.48987965997745e-231	5.79406358293336e-229	73.9507037768295	45.4189506656226
7	SYNE2	WM	1.24765396203244	1.02202291197789	14804463	0.738370688697661	1.05785715513833e-195	2.52145371590875e-193	57.2766065855678	16.6014095536413
8	SERF2	WM	1.98926557830106	1.18454664092789	16369150	0.816409251649	0	0	86.9880899305777	48.9428347689898
9	CYBA	WM	2.16578277616501	1.04804656441763	15660540.5	0.781067443943263	6.17949029360003e-246	2.79552574159657e-243	89.4146869626138	61.3938919342208
10	CD79B	WM	2.36550737323366	1.71194718322398	17855854	0.890558422501707	0	0	90.2872428507738	41.5035238841034
11	POU2F2	WM	1.69216424625106	1.11356281394611	15510775	0.773597908886291	5.40882082977949e-238	2.17995147879495e-235	74.6576651168715	36.0219263899765