STIMULUS REPORT

blood advances

Comparative genomics of *CXCR4*^{MUT} and *CXCR4*^{WT} single cells in Waldenström's macroglobulinemia

Cristina Jiménez,[1] Lian Xu,[1] Nickolas Tsakmaklis,[1] Maria G. Demos,[1] Amanda Kofides,[1] Gloria G. Chan,[1] Maria Luisa Guerrera,[1] Jiaji G. Chen,[1] Xia Liu,[1] Manit Munshi,[1] Christopher J. Patterson,[1] Guang Yang,[1,2] Jorge J. Castillo,[1,2] Steven P. Treon,[1,2] and Zachary R. Hunter[1,2]

[1]Bing Center for Waldenström's Macroglobulinemia, Dana-Farber Cancer Institute, Boston, MA; and [2]Department of Medicine, Harvard Medical School, Boston, MA

Key Points

- Single-cell whole-genome amplification can be used to interrogate the genomic architecture of Waldenström's macroglobulinemia.
- The mutational signature of *CXCR4*^{MUT} cells may be associated with alterations in DNA repairing genes and tumor suppressors.

Introduction

MYD88 and *CXCR4* somatic mutations are the most common alterations in Waldenström's macroglobulinemia (WM), affecting 95% to 97% and 30% to 40% of patients, respectively.[1] *CXCR4* mutations occur in the C-terminal domain and are often subclonal to mutated *MYD88* with a median allele burden of 35%, suggesting that these mutations occur after acquisition of the *MYD88* mutation. These activating mutations in *CXCR4* are typically nonsense or frameshift mutations that impact clinical presentation. The nonsense variants in particular are associated with higher bone marrow disease burden, high serum immunoglobulin M levels, symptomatic hyperviscosity, and earlier time to first treatment. *CXCR4* mutations also impact ibrutinib response, including depth and time to major response and progression-free survival.[2,3] Differences in response rates are minor when using a combination of ibrutinib and rituximab, but delayed responses are still observed in *CXCR4*^{MUT} patients.[4,5] The acquisition of *BTK*^{Cys481} mutations that underlie ibrutinib resistance also appears more common in *CXCR4*-mutated patients, suggesting an underlying genomic predisposition. As such, we sought to examine the *CXCR4* mutant subclone at a single-cell level to identify the alterations that may explain these unique features.

Methods

We selected 1 untreated patient with the clinicopathologic diagnosis of WM, with mutated *MYD88* and *CXCR4* for this proof-of-concept study. Single-cell sorting was performed on CD19^{+} immunoglobulin M^{+} bone marrow mononuclear cells, followed by whole-genome amplification (WGA) of the DNA on each isolated cell with the REPLI-g single cell kit (Qiagen, Valencia, CA). *MYD88* and *CXCR4* mutational status was assessed by Sanger sequencing.[6] A total of 22 single cells, distributed in 13 *MYD88*^{MUT}/*CXCR4*^{MUT} and 9 *MYD88*^{MUT}/*CXCR4*^{WT}, together with the bulk tumor CD19^{+} fraction and the bulk CD19^{-} germline sample (non–whole-genome amplified) from the same patient, were sent for whole-genome sequencing (WGS) to the Broad Institute of Massachusetts Institute of Technology and Harvard (Cambridge, MA). Data were analyzed following the Genome Analysis Toolkit Best Practice Guidelines (Broad Institute).[7] Sequencing reads were aligned to the human reference genome GRCh37/HG19 using Burrows-Wheeler Aligner. Small variants and indels were called using Strelka[8] and Ensembl Variant Effect Predictor,[9] and copy number alterations (CNA) were analyzed using Control-FREEC (Boeva Laboratory, Institut Curie, Paris, France)[10] and Genome Analysis Toolkit Copy Number Variation (Broad Institute).[11] Further analyses, including nonnegative matrix factorization,[12] Fisher's exact test, and differential gene expression analysis with voom from the edgeR/limma Bioconductor packages,[13] were conducted in R (R Foundation for Statistical Computing, Vienna, Austria). The pipeline followed is detailed in previous studies of the group.[1,14] All validations were carried out by Sanger sequencing.

Submitted 2 March 2020; accepted 4 August 2020; published online 21 September 2020. DOI 10.1182/bloodadvances.2020001775.

The data set has been submitted to the National Institutes of Health and will be made available by the authors upon request in the interim by contacting the corresponding author at zachary_hunter@dfci.harvard.edu.

The full-text version of this article contains a data supplement.

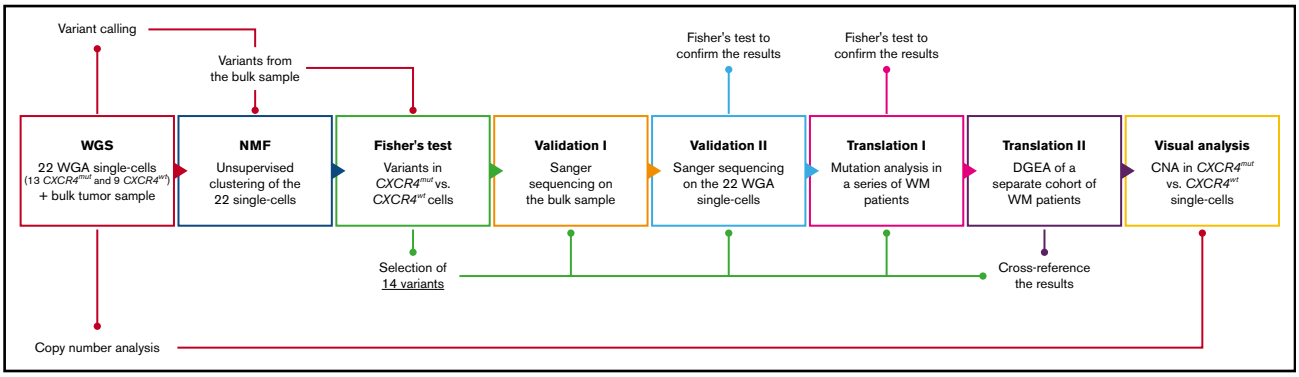

Figure 1. Workflow of the single-cell WGS analysis. Schematic representation of the steps followed for the analysis of the study. WGS was performed in 22 whole-genome amplified single-cells (13 *MYD88*^{mut}/*CXCR4*^{mut} and 9 *MYD88*^{mut}/*CXCR4*^{wt}) and the bulk tumor sample from a patient with WM. Variants called in the bulk sample were used as a matrix for the unsupervised clustering of the 22 cells. As cells did not cluster together according to *CXCR4* mutation, we intentionally searched for variants significantly present in one group vs the other, by applying the Fisher's exact test, and selected a list of 14. We validated them by Sanger sequencing on the bulk sample and on the 22 single cells. Next, we looked for mutations on these genes in a series of patients with whole-genome sequencing data to see whether they were predominant in *CXCR4*-mutated (mut) or wild-type (wt) cases. In addition, differential gene expression analysis based on *CXCR4* was conducted in an independent set of patients, and results were cross-referenced with our list of variants. Finally, copy number alterations were compared between both groups. DGEA, differential gene expression analysis; NMF, nonnegative matrix factorization.

Results and discussion

Because this was the first experiment of this kind carried out in WM, and considering the potential limitations of the WGA (allele dropout, false positives, and sequence-dependent bias),[15] we were able to establish an appropriate workflow to carry out single-cell interrogation (Figure 1). Briefly, we first performed an unsupervised clustering of the 22 single cells based on the variants of the bulk sample. Then, we applied the Fisher's exact test to identify variants significantly different in the single cells according to *CXCR4* status and validated them by Sanger sequencing on the bulk sample and on 22 single cells. Next, we searched for mutations on these genes in a series of patients to see whether they were predominant in *CXCR4*-mutated or wild-type cases, and we conducted differential gene expression analysis based on *CXCR4* in an independent set of patients. Finally, copy number alterations were compared between both groups of cells.

Our first approach was to perform an unsupervised clustering of the 22 single cells based on the variants of the bulk sample by using nonnegative matrix factorization,[12] but the samples did not cluster according to the *CXCR4* status, possibly because of biased amplification or allele dropout of the variants in the single cells. Therefore, we decided to find variants from the bulk sample that were enriched in *CXCR4*-mutated vs *CXCR4* wild-type single cells. Fifty-three single nucleotide variants and 10 indels corresponding to 59 genes were identified (supplemental Table 1). Most of these variants (48 of 63; 76%) were predominant in *CXCR4*^{MUT} cells. We selected the variants belonging to genes that are expressed in WM or in healthy donors B cells[16] and located near transcribed genes, ending up with 14 mutations (Table 1). Among the affected genes were *MACROD2* and *CCSER1*, which are associated with chromosome instability,[17,18] and *UVRAG*, which is involved in DNA damage repair.[19] There were also tumor suppressors, such as *BTG2*[20] and *DAB2*,[21] a regulator of the cell cycle (*SCAPER*),[22] and genes responsible for posttranslational protein modifications (*LNX1* and *DCUN1D4*). The remaining genes (*TMEM14B*, transmembrane protein; *LRMP*, lymphocyte protein; *SPON1*, cell adhesion protein; *OSGEPL1*, endopeptidase; *VTA1*, protein involved in vesicle trafficking; *EXOC6B*, part of the exocyst complex) have a less known role.

We picked up half of the variants for validation in the bulk tumor sample by Sanger sequencing and confirmed all of them. Validation in the 22 single cells was also performed, although in this case, results were hampered by the biased amplification and allele drop out associated with the WGA process.[15] However, *LRMP* kept the significantly different distribution observed in the single-cell WGS results, being detected almost exclusively in *CXCR4*^{MUT} cells (supplemental Table 2). This gene encodes a lymphoid-restricted membrane protein involved in antigen receptor assembly and trafficking during lymphocyte development.[23]

The following step was to translate these findings to a cohort of WM patients. We searched for mutations in the 14 genes in a series of 46 patients from a previous WGS study[14] to see whether they were predominant in *CXCR4*-mutated or *CXCR4* wild-type WM. Characteristics of this cohort are summarized in supplemental Table 3. Results did not show any significant difference in the distribution of the mutations on these genes between both groups of patients from this cohort (supplemental Table 4). As most mutations were noncoding, we hypothesized that they may influence gene expression. Using our previously published RNASeq cohort of 57 WM patients, we checked to make sure the candidate gene was expressed in healthy donor B cells and/or WM.[16] Then we used differential gene expression analysis between the *CXCR4*^{MUT} and *CXCR4*^{WT} cohorts to look for evidence of transcriptional dysregulation. Interestingly, from our list of 14 genes, *LRMP* was significant for differential expression (adjusted $P = .003$). We then analyzed a larger cohort of patients (n = 284) from an ongoing project, and *LRMP* was within the top 10 most differentially expressed genes (sorted by adjusted P value), with a fold change of −0.6 (supplemental Table 5). Finally, we analyzed the CNA to look for alterations specific to the *CXCR4*^{MUT} subclone, but this patient did

Table 1. List of 14 variants selected

Gene	Chromosome	Position	Ref	Var	Consequence	Protein position	AA change	No. single cells (n = 22)	*P*	Differentially expressed in *CXCR4*^{mut} vs *CXCR4*^{wt} patients*
SCAPER	15	76646031	C	A	Intron variant	NA	NA	7	.017	No
TMEM14B	6	10748058	C	T	5′ UTR variant	NA	NA	9	.03	No
MACROD2	20	15161087	G	T	Intron variant	NA	NA	6	.05	Yes
LNX1	4	54385495	G	A	Intron variant	NA	NA	5	.05	Yes
CCSER1	4	92040821	T	G	Intron variant	NA	NA	8	.074	No
LRMP	12	25227005	G	C	Intron variant	NA	NA	8	.07	No
DCUN1D4	4	52756917	G	T	Intron variant	NA	NA	8	.074	No
UVRAG	11	75783330	G	T	Intron variant	NA	NA	8	.07	Yes
SPON1	11	14043635	C	T	Intron variant	NA	NA	11	.080	No
BTG2	1	203275298	T	G	Intron variant	NA	NA	10	.09	No
OSGEPL1	2	190620347	A	G	Intron variant	NA	NA	10	.09	No
DAB2	5	39407706	T	C	Intron variant	NA	NA	10	.099	No
VTA1	6	142509639	C	T	Intron variant	NA	NA	10	.099	Yes
EXOC6B	2	72611453	T	TA	Intron variant	NA	NA	8	.074	No

We used the Fisher's exact test in the whole-genome data to find mutations whose presence differed significantly (or close, $P < .1$) between both groups of cells (*CXCR4* mutant vs wild type). From the total of 63 significant variants, we selected 14 based on the expression of the gene in WM and/or normal B cells according to data from a previous work.[16]

AA, amino acid; NA, not applicable; Ref, reference; UTR, untranslated region; Var, variant.

*Data from a cohort of 56 patients evaluated by RNASeq.[16]

not present any CNA of sufficient size to analyze at the single-cell level.

In summary, this is the first single-cell study carried out in WM to characterize the clonal diversity of the disease. Our results have highlighted several alterations associated with the *CXCR4*^{MUT} clone in DNA repairing genes and tumor suppressors, suggesting that *CXCR4* mutations could be related to the alteration of certain mechanisms rather than to specific genes. The findings may therefore help guide future studies to determine the role of CXCR4 in the clonal evolution of WM. Different treatment approaches may be needed for these patients according to the underlying pathogenic mechanisms associated with the presence of CXCR4 mutation. The analysis tools and workflow provided in this paper will help set up the basis for future studies on single-cell interrogation in WM.

Acknowledgments

The authors thank the Orszag Family Fund for WM Research, Peter S. Bing, the International Waldenström's Macroglobulinemia Foundation, and the Leukemia and Lymphoma Society. Z.R.H. was supported by an American Society of Hematology Scholar Award. C.J. is supported by a grant from the Fundación Española de Hematología y Hemoterapia-Fundación CRIS.

Authorship

Contribution: C.J., N.T., M.G.D., A.K., and L.X. performed the experiments; C.J., G.G.C. and Z.R.H. performed the data analysis; C.J. and Z.R.H. wrote the paper; S.P.T. and Z.R.H. conceived and designed the experiments; X.L., M.M., M.L.G., J.G.C., C.J.P., and G.Y. prepared samples; and J.J.C. and S.P.T. provided patient care, obtained consent, and were responsible for sample collection.

Conflict-of-interest disclosure: J.J.C. has received honoraria and/or research funds from AbbVie, Beigene, Janssen, Millennium, Pharmacyclics, and TG Therapeutics. Z.R.H. has received honoraria from Janssen and Pharmacyclics. S.P.T. has received honoraria and/or research funds from Beigene, Janssen, Pharmacyclics, and Bristol-Myers Squibb. The remaining authors declare no competing financial interests.

ORCID profiles: C.J., 0000-0002-5543-5154; G.G.C., 0000-0003-4541-9713; G.Y., 0000-0003-3049-4200; J.J.C., 0000-0001-9490-7532; S.P.T., 0000-0001-6393-6154; Z.R.H., 0000-0002-1689-1691.

Correspondence: Zachary R. Hunter, Bing Center for Waldenström's Macroglobulinemia, Dana-Farber Cancer Institute, Harvard Medical School, Mayer 544, 450 Brookline Ave, Boston, MA 02215; e-mail: zachary_hunter@dfci.harvard.edu.

References

1. Hunter ZR, Xu L, Yang G, et al. The genomic landscape of Waldenstrom macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood*. 2014;123(11):1637-1646.

2. Treon SP, Tripsas CK, Meid K, et al. Ibrutinib in previously treated Waldenström's macroglobulinemia. *N Engl J Med*. 2015;372(15):1430-1440.

3. Treon SP, Gustine J, Meid K, et al. Ibrutinib monotherapy in symptomatic, treatment-naïve patients with Waldenström macroglobulinemia. *J Clin Oncol*. 2018;36(27):2755-2761.

4. Dimopoulos MA, Trotman J, Tedeschi A, et al; iNNOVATE Study Group and the European Consortium for Waldenström's Macroglobulinemia. Ibrutinib for patients with rituximab-refractory Waldenström's macroglobulinaemia (iNNOVATE): an open-label substudy of an international, multicentre, phase 3 trial. *Lancet Oncol*. 2017;18(2):241-250.

5. Dimopoulos MA, Tedeschi A, Trotman J, et al; iNNOVATE Study Group and the European Consortium for Waldenström's Macroglobulinemia. Phase 3 trial of ibrutinib plus rituximab in Waldenstrom's macroglobulinemia. *N Engl J Med*. 2018;378(25):2399-2410.

6. Xu L, Hunter ZR, Tsakmaklis N, et al. Clonal architecture of CXCR4 WHIM-like mutations in Waldenström macroglobulinaemia. *Br J Haematol*. 2016; 172(5):735-744.

7. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma*. 2013;43:11.10.1-11.10.33.

8. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-1817.

9. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics*. 2010;26(16):2069-2070.

10. Boeva V, Popova T, Bleakley K, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*. 2012;28(3):423-425.

11. Babadi M, Benjamin DI, Lee SK, et al GATK CNV: copy-number variation discovery from coverage data [abstract]. *Cancer Res*. 2017;77(13 suppl). Abstract 3580.

12. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*. 2010;11(1):367.

13. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5(10):R80.

14. Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenström's macroglobulinemia. *N Engl J Med*. 2012;367(9):826-833.

15. Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu Rev Genomics Hum Genet*. 2015;16(1):79-102.

16. Hunter ZR, Xu L, Yang G, et al. Transcriptome sequencing reveals a profile that corresponds to genomic variants in Waldenström macroglobulinemia. *Blood*. 2016;128(6):827-838.

17. Sakthianandeswaren A, Parsons MJ, Mouradov D, et al. MACROD2 haploinsufficiency impairs catalytic activity of PARP1 and promotes chromosome instability and growth of intestinal tumors. *Cancer Discov*. 2018;8(8):988-1005.

18. Patel K, Scrimieri F, Ghosh S, et al. FAM190A deficiency creates a cell division defect. *Am J Pathol*. 2013;183(1):296-303.

19. Jo YK, Park NY, Shin JH, et al. Up-regulation of UVRAG by HDAC1 inhibition attenuates 5FU-induced cell death in HCT116 colorectal cancer cells. *Anticancer Res*. 2018;38(1):271-277.

20. Mao B, Zhang Z, Wang G. BTG2: a rising star of tumor suppressors (review). *Int J Oncol*. 2015;46(2):459-464.

21. Tong JH, Ng DC, Chau SL, et al. Putative tumour-suppressor gene DAB2 is frequently down regulated by promoter hypermethylation in nasopharyngeal carcinoma. *BMC Cancer*. 2010;10(1):253.

22. Tsang WY, Dynlacht BD. Double identity of SCAPER: a substrate and regulator of cyclin A/Cdk2. *Cell Cycle*. 2008;7(6):702-705.

23. Behrens TW, Jagadeesh J, Scherle P, Kearns G, Yewdell J, Staudt LM. Jaw1, A lymphoid-restricted membrane protein localized to the endoplasmic reticulum. *J Immunol*. 1994;153(2):682-690.