haematologica
Journal of The Ferrata Storti Foundation

Genetic lesions in MYC and STAT3 drive oncogenic transcription factor overexpression in plasmablastic lymphoma

by Julia Garcia-Reyero, Nerea Martinez Magunacelaya, Sonia Gonzalez de Villambrosia, Sanam Loghavi, Angela Gomez Mediavilla, Raul Tonda, Sergi Beltran, Marta Gut, Ainara Pereña Gonzalez, Emanuele d' Ámore, Carlo Visco, Joseph D. Khoury, and Santiago Montes-Moreno

Haematologica 2020 [Epub ahead of print]

Citation: Julia Garcia-Reyero, Nerea Martinez Magunacelaya, Sonia Gonzalez de Villambrosia, Sanam Loghavi, Angela Gomez Mediavilla, Raul Tonda, Sergi Beltran, Marta Gut, Ainara Pereña Gonzalez, Emanuele d' Ámore, Carlo Visco, Joseph D. Khoury, and Santiago Montes-Moreno. Genetic lesions in MYC and STAT3 drive oncogenic transcription factor overexpression in plasmablastic lymphoma.
Haematologica. 2020;105:xxx
doi:10.3324/haematol.2020.251579

Genetic lesions in *MYC* and *STAT3* drive oncogenic transcription factor overexpression in plasmablastic lymphoma.

Julia Garcia-Reyero[1, 2], Nerea Martinez Magunacelaya[2], Sonia Gonzalez de Villambrosia[3], Sanam Loghavi[4], Angela Gomez Mediavilla[2], Raul Tonda[5], Sergi Beltran[5], Marta Gut[5], Ainara Pereña Gonzalez[2], Emmanuel D´Ámore[6], Carlo Visco[7], Joseph D. Khoury[4] and Santiago Montes-Moreno[1, 2].

Running title: Oncogenic somatic mutations in Plasmablastic lymphoma.

1. Anatomic Pathology Service. Hospital Universitario Marqués de Valdecilla/IDIVAL. Universidad de Cantabria. Santander, Spain

2. Translational Hematopathology Lab, IDIVAL. Centro de Investigación Biomédica en Red de Cáncer (CIBERONC). Santander, Spain

3. Cytogenetics Unit, Department of Hematology, HUMV, Santander, Spain.

4. Hematopathology Department, MD Anderson Cancer Center, Houston, TX, USA.

5. Centre Nacional d'Anàlisi Genòmica (CNAG-CRG). Barcelona Institute of Science and Technology (BIST). Universitat Pompeu Fabra (UPF), Barcelona, Spain.

6. Departments of Pathology and Hematology, San Bortolo Hospital, Vicenza, Italy.

7. Department of Medicine, Section of Hematology, University of Verona, Italy.

WC text 3292, WC abstract 242, 5 figures, 1 table, 34 references.

Author for correspondence: Santiago Montes Moreno MD, PhD.

Anatomic Pathology Service, Hospital Universitario Marqués de Valdecilla.

Translational Hematopathology Lab. IDIVAL. Universidad de Cantabria.

Avda de Valdecilla s/n, 39010, Santander, Cantabria, España.

Tlf/ FAX: (34) 942-203492. E-mail: santiago.montes@scsalud.es

Acknowledgements: This study was supported by grants from MINECO (PI16/1397, SMM, Principal Investigator) and IDIVAL (NEXTVAL 15/09, SMM, Principal Investigator). NMM was supported by Asociación Española contra el Cancer (AECC). The authors want to acknowledge the Valdecilla Tumor Biobank Unit (Tissue Node, PT13/0010/0024) for their skillful handling and processing of tissue samples and specially all the clinical colleagues and pathologists who provided with clinical data and samples for this research study.

ABSTRACT

Plasmablastic lymphoma mutational profile is undescribed. Here we performed a targeted exonic NGS analysis of 30 plasmablastic lymphoma cases with a B cell lymphoma dedicated panel and FISH for the detection of *MYC* rearrangements. A complete phenotyping of the neoplastic and microenvironment cell populations was also performed. We have identified an enrichment in recurrent genetic events in *MYC* (69% with *MYC* translocation or amplification and 3 cases with missense point mutations), *PRDM1*/Blimp1 and *STAT3* mutations. These gene mutations were more frequent in EBV positive disease. Other genetic events included mutations in *BRAF*, *EP300*, BCR (*CD79A* and *CD79B*), NOTCH pathway (*NOTCH2*, *NOTCH1* and *SGK1*) and *MYD88p*L265P. Immunohistochemical analysis showed consistent MYC expression, higher in cases with *MYC* rearrangements together with phospho-STAT3 (Tyr705) overexpression in cases with *STAT3* SH2 domain mutations. Microenvironment populations were heterogeneous and unrelated with EBV, with an enrichment of Tumor Associated Macrophages (TAM) and PD1 positive T cells. PD-L1 was expressed in all cases in the TAM population but only in 5 cases in the neoplastic cells (4 out of 14 EBV positive cases). HLA expression was absent in the majority of PBL cases. In summary, Plasmablastic lymphoma mutational profile is heterogeneous and related with EBV infection. Genetic events in *MYC*, *STAT3* and *PRDM1*/Blimp1 are more frequent in EBV positive disease. An enrichment in TAM and PD1 reactive T lymphocytes is found in the microenvironment of PBL cases, that express PD-L1 in the neoplastic cells in a fraction of cases.

Key points:

Plasmablastic Lymphomas cases show activating mutations in *MYC*, *STAT3* and *BRAF*, providing a rationale for targeted therapy.

An enrichment in TAM and PD1 reactive T lymphocytes is found in the microenvironment of PBL cases that together with PD-L1 expression by the tumor cells may suggest an option for immune-checkpoint interference.

INTRODUCTION

Plasmablastic lymphoma is an aggressive non-Hodgkin B cell lymphoma type defined as a high grade large B-cell neoplasm with plasma cell phenotype (i.e. loss of B-cell antigens with downregulation of CD20 and PAX5 expression and overexpression of PRDM1/Blimp1 and XBP1s) [1-4].

Epstein-Barr virus (EBV) infection is found in the majority of the cases but is not required for the development of a plasmablastic phenotype since clear-cut plasmablastic lymphoma can be negative for EBV [3-5]. In addition, recent evidence suggests that EBV or human immunodeficiency virus (HIV) status does not influence the gene expression profile patterns of plasmablastic lymphoma[6]. However EBV positivity in plasmablastic lymphoma has been found to be associated with increased PD-L1 protein expression as well as other immune scape markers [7,8] and decreased expression of MHCII/HLA-DR by the neoplastic cells[9]. This has been recently found to be associated with an increased antiviral cytotoxic immunity involving different immune cell populations[7].

The genetic landscape of plasmablastic lymphoma with regards to somatic mutations is poorly understood. So far, *MYC-IGH* translocations are the most commonly detected alteration, involving 60% of the cases[10,11]. Concurrent mutations in *PRDM1*/Blimp1 have been found in half of these cases[12]. Very recently, exome sequencing of a series of HIV- positive plasmablastic lymphoma cases showed somatic mutations involving components of the non-canonical NFkB pathway as well as genes involved in immune response (meeting abstract[13]) but such data remain limited.

Our aim was to characterize the genetic profile of a series of plasmablastic lymphoma cases using targeted exonic NGS and correlate the findings with EBV infection and the expression status of immune checkpoint proteins in both the neoplastic population and the microenvironment. In addition, we quantified the components of the microenvironment and searched for skewed T cell populations in this tumor. We have found that the mutational profile of PBL is related with EBV infection in the tumor cells and we identified recurrent genetic events in *MYC*, *STAT3* and *PRDM1*/Blimp1 that are more frequent in EBV positive disease. In addition, we identified PD-L1 expression on tumor cells in a subset of cases as well as an enrichment in Tumor associated macrophages (TAM) and PD1 reactive T cells in the microenvironment of PBL cases.

METHODS

Case Selection

Twenty-eight new cases were retrieved from the files of the Pathology Department of the Hospital Universitario Marqués de Valdecilla, 10 samples from the files of the University of Texas MD Anderson Cancer Center (Houston) Hematopathology Department files and 4 cases from the

department of Pathology of San Bortolo Hospital, Vicenza, Italy. Material Transfer Agreements were signed by IDIVAL and the corresponding institutions to share the material in the project. The study and sample collection were approved by the local ethics committee (CEIC Cantabria, IRB code 2016.168) and complies with the Declaration of Helsinki. All cases were diagnosed according to the WHO classification of Hematolymphoid Neoplasms[14]. Negativity for pan B cell markers (CD20), HHV-8 and ALK was required in every case to be included. The phenotype of the cases was consistent with a plasma cell differentiation program[4,15]. The clinical features of the cases were recorded and a summary is available in Supplementary Table 1.

Immunohistochemistry and in situ hybridization.

Immunohistochemical reactions were performed following conventional automated procedures. Chromogenic in situ hybridization for EBV-EBERs and Fluorescent in situ hybridization for the detection of *MYC* rearrangement were also done.

Quantification of the cellular composition of the tumor and transcription factor abundance.

Quantification of different lymphoid and histiocytic/dendritic subpopulations, identified with CD3, CD8, PD1, CD163, PD-L1 and MHCII/HLA DP/DR and absolute quantification of the number of nuclei showing expression of MYC and Phospho-STAT3 (Tyr705) were done.

Next Generation Sequencing using amplicon-based library generation.

DNA was extracted from formalin fixed paraffin embedded samples using the PicoPure™ DNA Isolation Kit (ThermoFisher Scientific) and was quantified by Qbit fluorometer (ThermoFisher Scientific). All samples subjected to NGS analysis were required to have >50% of neoplastic cells as identified by morphology (H&E).

A TruSeq® Custom Amplicon Low Input Library containing exonic regions of 35 selected genes of interest was used to isolate the DNA for sequencing (Illumina). The list of genes was *CARD11, ARID1A, NOTCH1, TCF3, SMARCA4, STAT6, EP300, CREBBP, MLL2, BTK, NOTCH2, TNFRSF14, ATM, FOXO1, B2M,*
PLCG2, CD79B, TP53, STAT3, BCL2, MEF2B, CD79A, CXCR4, PTPN1, MYD88, FAT2, PRDM1, TNFAIP3, SGK1, CCND3, PIM1, EZH2, BRAF, MYC, NOTHC2. Of note, variants occurring in regions outside the coverage of our targeted design, were not explored using this approach. Details about library preparation can be found in supplementary material.

Sequencing on a HiSeq instrument (Illumina, paired end, 2x150) at the National Genomic Analysis Center (CNAG, Barcelona, Spain) was performed.

Sequencing data interpretation and reporting.

In brief only variants in which both libraries had a coverage greater or equal to 300 reads and had the same genotype were selected for downstream analysis. Subsequently only missense, frameshift, and nonsense somatic mutations with variant frequency>10% were considered (supplementary Table 2). SNPs were filtered out using VAF criteria, comparison with dbSNP and with a in house germline variants database. Finally, 34 somatic mutations (31 missense, 3 nonsense) in 14 genes were considered (Table 1).

For further details on the methods here used please see supplementary material.

RESULTS

1. **The mutational profile of plasmablastic lymphoma is heterogeneous and correlates with EBV infection in the neoplastic cells.**

After targeted NGS with a lymphoma dedicated panel, somatic missense and nonsense mutations were identified in 18 out of 30 plasmablastic lymphoma cases (60%). EBV negative cases tended to show a higher rate of mutations, as compared to EBV positive cases (87.5% vs 54%, respectively, Chi square, $p > 0.05$) (Figure 1).

Interestingly the pattern of mutations was also different between EBV positive and EBV negative cases. Recurrent somatic mutations restricted to EBV positive cases were found in *PRDM1*/Blimp1 in 6 cases and *STAT3*, in 5 cases. Notably, a recurrent *PRDM1*/Blimp1 variant, D203E, was identified in 4 out of 6 cases, involving the PR domain of the protein.

STAT3 mutations were found in 5 out of 30 cases (16%), all EBV positive. Interestingly all but one (*STAT3p*D566Y) of the mutations involved the SH2 domain of STAT3 protein (*STAT3p*Y640F, *STAT3p*M648L, *STAT3p*G618R, *STAT3p*N647I) (Figure 2) and lead to phosphoSTAT3 (Tyr705) protein overexpression (see below).

The majority (16 out of 23 cases tested, 69%) of plasmablastic lymphoma cases harbored structural abnormalities at the *MYC* locus. 14 cases showed *MYC* translocation (60%) using break apart probes. *MYC-IGH* was confirmed in 7 of 9 cases tested (77%). *MYC* was found to be amplified by FISH in 2 additional cases (Figure 1). Thus, in cases with *MYC* rearrangements, *MYC-IGH* is the most frequent alteration. Although there was a clear trend for the association between EBV positivity and *MYC* rearrangement the difference was not statistically significant (Chi square 0.06).

Furthermore, *MYC* was found to be mutated in 3 cases with all but one of the mutations involving exon 2 and consisting of transversions and transitions at C: G pairs (4 out of 7 mutations, see Table

1). Furthermore, *MYCp*79S mutation involves the WRCY consensus motif. All these features are consistent with a mechanism related with aSHM as described in early reports[16].

Common mutations in DLBCL NOS, involving BCR activation, TLR/NFkB, histone modifying genes and NOTCH pathway were found in 8 cases (Table 1 and Figure 1). Those mutations involved *CD79p*AW76*, *CD79Bp*D34N, *MYD88p*L265P, *NOTCH1p*P401L, *NOTCH2p*R2400*, *SGK1Kp*136*and *EP300p*M2010I/*EP300p*R1731H. NOTCH pathway was affected by somatic mutations in NOTCH2 (1 case), NOTCH1 (1 case) and SGK1 (2 cases). Other mutations found were *SMARCA4p*R1005Q and *TP53p*R273H. Of note two cases, both EBV negative, had mutations in BRAF gene, one case with the canonical activating *BRAFp*V600E mutation and the other with the mutation *BRAFp*G469A in the ATP binding site.

2. ***STAT3* mutations are associated with constitutive phospho-STAT3 (Tyr705) activation. MYC protein overexpression is related to *MYC* rearrangement status.**

Phospho-STAT3 (Tyr705) protein immunohistochemical expression was quantified in 20 cases with available mutational data. Mean phospho-STAT3 expression was 48 nuclei per HPF (40x) in these 20 cases. Mean expression for 2 out of 4 SH2 domain mutated cases with available IHC data was 249 nuclei per HPF (40x). Mean phosho-STAT3 expression for *STAT3* wild type cases was 28 nuclei per HPF (40X). Mean phospho-STAT3 expression for the single nonSH2 *STAT3* mutated sample was 40 nuclei per HPF (40X). Thus, *STAT3* SH2 domain mutations (*STAT3p*Y640F, *STAT3p*M648L, *STAT3p*G618R, *STAT3p*N647I) were associated with overexpression of phospho-STAT3 by IHC in the tissue samples (Figure 2B**).**

MYC protein was consistently expressed in all the cases (range 59-236 nuclei per HPF, mean 236), irrespective of the presence of *MYC* translocation, as previously reported[12,17]. However, significant differences in the level of MYC expression were quantified, according to *MYC* gene status. *MYC* translocated (14 cases) and amplified cases (2 cases) had, as expected, higher MYC protein expression than cases without *MYC* rearrangements (7 cases). Mean number of positive nuclei was 109 per HPF in non-rearranged cases vs mean 282 positive nuclei per HPF in MYC rearranged cases, Mann-Whitney test p value <0.0001 (Figure 3).

Mean MYC protein expression in 22 cases with available data was 236 nuclei per HPF, significantly higher than the mean 48 nuclei per HPF in the cases of phospho-STAT3 protein expression (Wilcoxon test p 0.001). There was no correlation between the expression levels of both proteins (Pearson non-significant). Due to the high prevalence of *MYC* translocations and amplification in PBL and the relatively low levels of phospho-STAT3 expression and absence of correlation between both

proteins, it is unlikely that STAT3 activation contributes to MYC overexpression in most PBL cases. However, one of our cases with *STAT3* SH2 domain mutations and absence of *MYC* translocation by FISH showed high levels of both phospho-STAT3 and MYC proteins, without detectable PRDM1/Blimp1 mutations, suggesting that MYC overexpression might be related with STAT3 activation by mutations in rare PBL cases.

In sum MYC protein overexpression is due to rearrangements involving *MYC* in a significant proportion of plasmablastic lymphoma cases (69% in these series). Most translocations fuse *MYC* to *IGH* and few cases may show amplifications of the *MYC* gene. Both alterations lead to MYC protein overexpression. Genetic alterations in the *MYC* regulatory domains of *PRDM1*/Blimp1 may also contribute to its overexpression[12]. In addition here we show that a fraction of plasmablastic lymphoma cases has STAT3 activation, due to somatic mutations in the *STAT3*-SH2 domain that may increase MYC expression, as previously described in DLBCL[18] (Figure 4).

3. **The Immune microenvironment in PBL is characterized by enrichment in TAM and PD1 positive T cells. PD-L1 in neoplastic cells in PBL is expressed in a fraction of cases, independently of EBV infection.**

We quantified the expression of CD163 and PD-L1 in histiocytic/dendritic cells within plasmablastic lymphoma cases. The mean expression of PDL1 was 33 nuclei per HPF (range 1.67-61) and the mean expression of CD163 was 38 nuclei per HPF (range 2-84, Figure 5). The correlation between CD163 and PD-L1 expression was significant (Pearson 0.6, p value <0.05), suggesting that PD-L1 positive cells are histiocytes in plasmablastic lymphoma cases. There was not a significant difference in the content and distribution of CD163 and PD-L1 positive histiocytes between EBV positive and EBV negative cases (Mann-Whitney test p>0.05).

CD8 positive and PD1 positive T cell subpopulations were quantified. The mean number of CD8 positive lymphocytes was 52 nuclei per HPF (range 1-117) and the mean number of PD1 positive lymphocytes was 32 nuclei per HFP (range 0-76). There was a significant difference in the distribution of CD8 and PD1 positive cell subset (Wilcoxon test, significant p<0.001) consistent with different cell populations. Pearson correlation value was however significant (Pearson 0.59, p value <0.05). There was no significant difference in the content and distribution of CD8 and PD1 positive lymphocytes between EBV positive and EBV negative cases (Mann-Whitney test p>0.05) (Figure 5).

PDL1 was expressed by tumor cells in 5 out of 24 (20%) cases evaluated (mean 59 nuclei per HPF, range 25-98). Four out of 5 PD-L1 positive cases (in the neoplastic cells) were EBV positive. 14 EBV positive PBL cases were PD-L1 negative in the tumor cells. Thus 4 out of 18 (22%) EBV positive PBL

cases were PD-L1 positive, while one out of 6 (16%) EBV negative cases were PD-L1 positive. Thus, there was no association between EBV infection by tumor cells and PDL1 expression, since most of the EBV positive cases were PD-L1 negative (p value, non-significant, Figure 5). Interestingly one case with *STAT3* SH2 mutations showed concurrent PD-L1 and Phospho-STAT3 (Tyr705) expression. For the other STAT3 SH2 mutated cases PD-L1 expression data was not available to test this association.

Consistent with previously published data[9] major histocompatibility class II (MHCII) protein/ HLA (DP,DR) was virtually absent in PBL. Only 3 cases out of 25 tested were positive (12%, mean 349 nuclei per HPF, range 284-440). Two cases showed a membranous and cytoplasmic granular pattern and the other a membranous pattern. All 3 cases were EBV positive. The other 22 cases were completely negative for HLA expression in tumor cells (Figure 5).

DISCUSSION

In this study we characterized the genetic profile of a series of plasmablastic lymphoma cases using targeted exonic NGS and correlate with EBV infection and the expression of immune checkpoint proteins in both the neoplastic population and tumor microenvironment. Our results show that genetic abnormalities (including translocations, amplifications and point mutations) in the MYC gene are the most common genetic event in plasmablastic lymphoma. In addition to previously described translocations, involving *IGH* and *MYC*[10,11], here we demonstrate that few cases may have *MYC* amplification, confirming our previous observations[12]. Both *MYC* translocation and amplification lead to a significantly increased MYC protein overexpression. Interestingly we have also identified *MYC* point mutations, mainly consisting of transversions and transitions at C: G pairs and involving exon 2 and, in the case of *MYCp*79S mutation the WRCY consensus motif. All these features are consistent with a mechanism related with aSHM [16]. The oncogenic effect of these point mutations remains however unclear.

In addition, we have found, that 16% of our cases (5 cases) carry recurrent somatic mutations in the oncogene *STAT3*, preferentially involving the SH2 domain of the protein. Interestingly these mutations were restricted to EBV positive PBL. Here we demonstrate that these mutations lead to an increased expression of phospho-STAT3 (Tyr705).

STAT3 mutations and phospho-STAT3 overexpression have been found very rarely in DLBCL NOS (6% according to [19]). In cases of ALK positive large B cell lymphoma that show commonly a plasmablastic phenotype, phospho-STAT3 expression has been found to be associated with the presence of ALK rearrangements and overexpression[20]. Importantly, STAT3 activation, due to somatic mutations in the *STAT3*-SH2 domain may contribute to MYC overexpression, as previously described in DLBCL[18]. In

addition, one case in our series showed concurrent *STAT3* SH2 mutations and Phospho-STAT3 (Tyr705) expression and PD-L1 overexpression, confirming previous results in other lymphoma types suggesting that STAT3 activation triggers PD-L1 overexpression[21].

STAT3 somatic mutations in plasmablastic lymphoma have not been previously described so far and may have therapeutic implications for the clinical testing of STAT3 inhibitors in these patients.

Interestingly the pattern of somatic mutations in EBV negative disease was more heterogeneous. Mutations involving BCR activation, TLR/NFkB, histone modifying genes and NOTCH pathway were found in 8 cases (Table 1 and Figure 1). *MYD88p*L265P mutation, involving the TIR domain of the *MYD88* gene has been previously described in ABC-type DLBCL, Primary Central Nervous System Lymphoma and other immune privileged site DLBCL[22,23] and LPL/WM[24] and leads to downstream activation of IRAK4/IRAK1/TRAF6 complex and NFkB activation. The pattern of mutations in *CD79A/B* in PBL cases was distinct than that found in DLBCL NOS. Mutations in *CD79A/B* were found located outside the ITAM domains related with constitutive BCR activation in ABC-type DLBCL[25]. NOTCH pathway was mutated in *NOTCH2, NOTCH1* and *SGK1*. *NOTCH2p*R2400* is a nonsense mutation that truncates the PEST domain of the NOTCH2 protein and has been already described in B-NHL, including DLBCL NOS[26]. PEST domain truncating mutations have been found present in multiple tumor types and functional studies suggest that this class of mutations can be targeted with Notch inhibitors including gamma secretase inhibitors[27]. *NOTCH1p*P401L has been already reported in CLL in a previous study[28] and lies within the Ca binding EGF-like domains repeat. Mutations in *SGK1*, involved *SGK1p*S451F, *SGK1p*A380V point mutations and *SGK1p*K136* truncating mutations. These mutations have not been previously described in DLBCL NOS cases[26]. SGK1 has been suggested to be a negative regulator of NOTCH signaling enhancing NOTCH protein degradation and reducing its activation by the gamma-secretase[29]. Other mutations found were *SMARCA4p*R1005Q and *TP53p*R273H.

Of note MAPK/ERK pathway activating mutations, involving *BRAF* (*BRAFp*V600E, *BRAFp*G469A) were found in two cases, both EBV negative. *BRAF* mutations have been observed, rarely, in related neoplasms such as multiple myeloma. Previous studies have found *BRAF* mutations in 4% of multiple myeloma cases[30] and associated with aggressive clinical features, plasmablastic phenotype and clonal evolution[31,32], with obvious clinical implications for targeted therapy.

In addition to the genetic profile of the cases we explored further the composition of the tumor microenvironment and the expression of immune-checkpoint markers in both the neoplastic and other lymphoid and histiocytic/dendritic populations. Our results confirm previous studies showing an enrichment in Tumor Associated Macrophages that express CD163 and PD-L1. In addition PBL

cases show a significant population of CD8 positive T cells, irrespective of the almost absent expression of MHCII/HLA by the neoplastic cells[9]. Importantly, together with CD8 positive T cells, PBL shows a distinct population of PD1 positive T cells. In our cases EBV does not influence the immune populations, with regards to the content of TAM, CD8 positive and PD1 positive T cells quantified in the tissue. Furthermore, in our series, PD-L1 expression by the neoplastic cells was found in 20% of the cases analyzed, similar to previously published series[8] and there was no association between EBV infection by tumor cells and PD-L1 expression, since PD-L1 was found in both EBV positive and negative variants and most of the EBV positive cases were PD-L1 negative. These findings are in agreement with previously published data in plasmablastic lymphoma, with variable expression of PD-L1 by the neoplastic population, ranging 20-44%[8,33]. In our series, however, we do not confirm an association between EBV infection and PD-L1 expression as suggested by others[8]. These differences might be due to a combination of factors, including different clones used for the detection of PD-L1 expression (22C3 clone in this study, SP142 in others[8]) and different quantification and statistical methods used. In addition another biological factor related to the uncommon PD-L1 expression in PBL cases could be related to the usual latency pattern found in these cases, since PD-L1 expression in EBV positive PTLD has been highly associated with EBV latency patterns 2 and 3[34] while PBL cases use to disclose EBV latency pattern 1[7]. Notably one of our cases points to STAT3 activation as a potential cause for PD-L1 overexpression in PBL cases. Collectively our results on the microenvironment and immune-checkpoint expression in PBL may indicate a potential for immune checkpoint interference in patients with plasmablastic lymphoma.

In summary, here we have found that the mutational profile of PBL is related with EBV infection in the tumor cells and identified recurrent genetic events in *MYC*, *STAT3* and *PRDM1*/Blimp1 that are associated with EBV positive disease. Both *MYC* genetic alterations (including translocations and amplification) and SH2 domain *STAT3* mutations lead to *MYC* and Phospho-STAT3 (Tyr705) protein overexpression, respectively. Other somatic mutations including *BRAFp*V600E, *MYD88p*L265P, *NOTCH2p*R2400* and *TP53p*R273H, appear in EBV negative disease, suggesting an overlapping mutational profile with both multiple myeloma and DLBCL NOS. Furthermore, the tumor microenvironment in PBL is characterized by an enrichment in PD-L1 positive TAM and PD1 reactive T lymphocytes with expression of PD-L1 by the neoplastic tumor cells in a fraction of cases. Novel molecular targets derived from the present study include MYC and STAT3 activation, MAPK/ERK and NOTCH2 pathway mutations and immune-checkpoint interference.

Authorship contributions: JGR performed research, analyzed data and approved the paper, NMM performed research, analyzed data and approved the paper, SGV analyzed data and approved the paper, SL performed research, provided clinical data and approved the paper, AGM performed research and approved the paper, RT analyzed data and approved the paper, SB analyzed data and approved the paper, MG analyzed data and approved the paper, AGP performed research and approved the paper, EDA performed research, provided clinical data and approved the paper, CV provided clinical data and approved the paper, JK provided clinical data and approved the paper, SMM designed research, performed research, analyzed data, wrote the paper and approved the paper.

The authors have no potential conflict of interest to disclose related to the content of this work.

REFERENCES

1. Swerdlow SH CE HN, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. 2008.

2. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. Blood. 2011;117(19):5019-5032.

3. Delecluse HJ, Anagnostopoulos I, Dallenbach F, et al. Plasmablastic lymphomas of the oral cavity: a new entity associated with the human immunodeficiency virus infection. Blood. 1997;89(4):1413-1420.

4. Montes-Moreno S, Gonzalez-Medina AR, Rodriguez-Pinilla SM, et al. Aggressive large B-cell lymphoma with plasma cell differentiation: immunohistochemical characterization of plasmablastic lymphoma and diffuse large B-cell lymphoma with partial plasmablastic phenotype. Haematologica. 2010;95(8):1342-1349.

5. Colomo L, Loong F, Rives S, et al. Diffuse large B-cell lymphomas with plasmablastic differentiation represent a heterogeneous group of disease entities. Am J Surg Pathol. 2004;28(6):736-747.

6. Chapman J, Gentles AJ, Sujoy V, et al. Gene expression analysis of plasmablastic lymphoma identifies downregulation of B-cell receptor signaling and additional unique transcriptional programs. Leukemia. 2015;29(11):2270-2273.

7. Gravelle P, Péricart S, Tosolini M, et al. EBV infection determines the immune hallmarks of plasmablastic lymphoma. Oncoimmunology. 2018;7(10):e1486950.

8. Laurent C, Fabiani B, Do C, et al. Immune-checkpoint expression in Epstein-Barr virus positive and negative plasmablastic lymphoma: a clinical and pathological study in 82 patients. Haematologica. 2016;101(8):976-984.

9. Schmelz M, Montes-Moreno S, Piris M, Wilkinson ST, Rimsza LM. Lack and/or aberrant localization of major histocompatibility class II (MHCII) protein in plasmablastic lymphoma. Haematologica. 2012;97(10):1614-1616.

10. Valera A, Balagué O, Colomo L, et al. IG/MYC rearrangements are the main cytogenetic alteration in plasmablastic lymphomas. Am J Surg Pathol. 2010;34(11):1686-1694.

11. Taddesse-Heath L, Meloni-Ehrig A, Scheerle J, Kelly JC, Jaffe ES. Plasmablastic lymphoma with MYC translocation: evidence for a common pathway in the generation of plasmablastic features. Mod Pathol. 2010;23(7):991-999.

12. Montes-Moreno S, Martinez-Magunacelaya N, Zecchini-Barrese T, et al. Plasmablastic lymphoma phenotype is determined by genetic alterations in MYC and PRDM1. Mod Pathol. 2017;30(1):85-94.

13. Munevver C, Rong HR, Chineke I, et al. Genetic Analysis of Plasmablastic Lymphomas in HIV (+) Patients Reveals Novel Driver Regulators of the Noncanonical NF-Kb Pathway. Blood. 2018;132(Suppl 1):1565.

14. Swerdlow SH CE, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J (Eds). WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (revised 4th edition). IARC. Lyon 2017.

15. Montes-Moreno S, Martinez-Magunacelaya N, Zecchini-Barrese T, et al. Plasmablastic lymphoma phenotype is determined by genetic alterations in MYC and PRDM1. Mod Pathol. 2017;30(1):85-94.

16. Pasqualucci L, Neumeister P, Goossens T, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001;412(6844):341-346.

17. Loghavi S, Alayed K, Aladily TN, et al. Stage, age, and EBV status impact outcomes of plasmablastic lymphoma patients: a clinicopathologic analysis of 61 patients. J Hematol Oncol. 2015;8:65.

18. Sarosiek KA, Malumbres R, Nechushtan H, Gentles AJ, Avisar E, Lossos IS. Novel IL-21 signaling pathway up-regulates c-Myc and induces apoptosis of diffuse large B-cell lymphomas. Blood. 2010;115(3):570-580.

19. Ohgami RS, Ma L, Monabati A, Zehnder JL, Arber DA. STAT3 mutations are present in aggressive B-cell lymphomas including a subset of diffuse large B-cell lymphomas with CD30 expression. Haematologica. 2014;99(7):e105-107.

20. Valera A, Colomo L, Martinez A, et al. ALK-positive large B-cell lymphomas express a terminal B-cell differentiation program and activated STAT3 but lack MYC rearrangements. Mod Pathol. 2013;26(10):1329-1337.

21. Tabanelli V, Corsini C, Fiori S, et al. Recurrent PDL1 expression and PDL1 (CD274) copy number alterations in breast implant-associated anaplastic large cell lymphomas. Hum Pathol. 2019;90:60-69.

22. Ngo VN, Young RM, Schmitz R, et al. Oncogenically active MYD88 mutations in human lymphoma. Nature. 2011;470(7332):115-119.

23. Chapuy B, Roemer MG, Stewart C, et al. Targetable genetic features of primary testicular and primary central nervous system lymphomas. Blood. 2016;127(7):869-881.

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

25. Davis RE, Ngo VN, Lenz G, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. Nature. 2010;463(7277):88-92.

26. Karube K, Enjuanes A, Dlouhy I, et al. Integrating genomic alterations in diffuse large B-cell lymphoma identifies new relevant pathways and potential therapeutic targets. Leukemia. 2018;32(3):675-684.

27. Wang K, Zhang Q, Li D, et al. PEST domain mutations in Notch receptors comprise an oncogenic driver segment in triple-negative breast cancer sensitive to a γ-secretase inhibitor. Clin Cancer Res. 2015;21(6):1487-1496.

28. Sutton LA, Ljungström V, Mansouri L, et al. Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. Haematologica. 2015;100(3):370-376.

29. Mo JS, Ann EJ, Yoon JH, et al. Serum- and glucocorticoid-inducible kinase 1 (SGK1) controls Notch1 signaling by downregulation of protein stability through Fbw7 ubiquitin ligase. J Cell Sci. 2011;124(Pt 1):100-112.

30. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011;471(7339):467-472.

31. Bohn OL, Hsu K, Hyman DM, Pignataro DS, Giralt S, Teruya-Feldstein J. BRAF V600E mutation and clonal evolution in a patient with relapsed refractory myeloma with plasmablastic differentiation. Clin Lymphoma Myeloma Leuk. 2014;14(2):e65-68.

32. Andrulis M, Lehners N, Capper D, et al. Targeting the BRAF V600E mutation in multiple myeloma. Cancer Discov. 2013;3(8):862-869.

33. Chen BJ, Chapuy B, Ouyang J, et al. PD-L1 expression is characteristic of a subset of aggressive B-cell lymphomas and virus-associated malignancies. Clin Cancer Res. 2013;19(13):3462-3473.

34. Veloza L, Teixido C, Castrejon N, et al. Clinicopathological evaluation of the programmed cell death 1 (PD1)/programmed cell death-ligand 1 (PD-L1) axis in post-transplant lymphoproliferative

disorders: association with Epstein-Barr virus, PD-L1 copy number alterations, and outcome. Histopathology. 2019;75(6):799-812.

TABLE 1

ID	Gene	Location chrom	Domain	Allele	cDNA_position	Codons	AA Change		Consequence*	Existing_variation
4	**STAT3**	17	---	A	2009	Gac/Tac	566	D/Y	deleterious	COSM220689
4	**EP300**	22	---	A	7249	atG/atA	2010	M/I	tolerated	---
11	**MYC**	8	---	G	578	aCc/aGc	23	T/S	tolerated	---
11	**MYC**	8	---	A	775	Tac/Aac	89	Y/N	tolerated	---
11	**MYC**	8	---	C	899	tTc/tCc	130	F/S	deleterious	COSM4171775
11	**MYC**	8	---	G	945	atC/atG	145	I/M	deleterious	---
14	**STAT3**	17	SH2	A	2255	Atg/Ttg	648	M/L	tolerated	---
14	**STAT3**	17	SH2	A	2232	tAc/tTc	640	Y/F	probably_damaging	COSM1155743
17	**STAT3**	17	SH2	G	2165	Ggc/Cgc	618	G/R	deleterious	COSM1166777
17	**PRDM1**	6	PR	G	843	gaC/gaG	203	D/E	neutral	rs811925*, COSM4160094
28	**PRDM1**	6	PR	G	843	gaC/gaG	203	D/E	neutral	rs811925*, COSM4160094
7	**MYC**	8	---	T	1085	tAc/tTc	192	Y/F	probably_damaging	---
7	**CD79B**	17	---	T	175	Gac/Aac	34	D/N	tolerated	---
8	**SMARCA4**	19	---	A	3295	cGa/cAa	1005	R/Q	deleterious	---
8	**PRDM1**	6	Ac	A	2546	gGc/gAc	771	G/D	tolerated	---
2	**STAT3**	17	SH2	A	2253	aAc/aTc	647	N/I	deleterious	COSM1155744
2	**NOTCH1**	9	EGF-like	A	1278	cCc/cTc	401	P/L	deleterious	COSM4745915
5	**STAT3**	17	SH2	A	2232	tAc/tTc	640	Y/F	probably_damaging	COSM1155743
10	**PRDM1**	6	Pro-rich	A	1295	aGc/aAc	354	S/N	tolerated	rs143040512,COSM4406870
10	**CD79A**	19	---	A	413	tgG/tgA	76	W/*	---	COSM5493940
26	**PRDM1**	6	PR	G	843	gaC/gaG	203	D/E	neutral	rs811925*, COSM4160094
27	**PRDM1**	6	PR	G	843	gaC/gaG	203	D/E	neutral	rs811925*, COSM4160094
3	**ARID1A**	1	---	A	762	Ggg/Agg	131	G/R	deleterious	---
3	**ARID1A**	1	---	C	6526	tGc/tCc	2052	C/S	deleterious	---
3	**MYD88**	3	TIR	C	794	cTg/cCg	265	L/P	deleterious	COSM85940
15	**BRAF**	7	ATP binding site	G	1467	gGa/gCa	469	G/A	deleterious	COSM460
18	**SGK1**	6	---	A	1950	tCc/tTc	451	S/F	deleterious	---
18	**SGK1**	6	---	A	1737	gCt/gTt	380	A/V	tolerated	---
9	**NOTCH2**	1	PEST	A	7418	Cga/Tga	240	R/*	deleterious	COSM36210

							0			
1	**MYC**	8	---	T	747	agC/agT	79	S/	---	---
1	**EP300**	22	---	A	6411	cGc/cAc	1731	R/H	deleterious	---
1	**BRAF**	7	STKc_Raf	T	1860	gTg/gAg	600	V/E	deleterious	COSM476
1	**SGK1**	6	---	A	1004	Aag/Tag	136	K/*	deleterious	---
13	**TP53**	17	---	T	1008	cGt/cAt	273	R/H	possibly damaging	COSM10660

Summary of the mutations found in 18 out of 30 cases (60%) analyzed by targeted exonic next generation sequencing. Gene name, exonic location, cDNA position, single nucleotide change observed, and amino acid change predicted, together with predicted consequence using 3 different algorithms is shown. In addition, the dbSNP and the COSMIC id is provided when available.

Figure Legends

FIGURE 1

Summary of the mutations found in 18 out of 30 cases (60%) analyzed by targeted exonic next generation sequencing. EBV positivity by the tumor cells (EBV) and HIV infection by the patient are shown, together with the status of *MYC* gene as determined by interphase FISH. The pattern of somatic mutations is heterogeneous with a trend to a higher rate of mutations in EBV positive disease. Mutations (including translocations, amplifications and point mutations) in *MYC* gene are the most common genetic event in plasmablastic lymphoma. Previously undescribed mutations in plasmablastic lymphoma such as *STAT3* (16% of cases), *BRAF*, *MYD88*, *NOTCH2* and *TP53* were also identified (see detail in Table 1).

FIGURE 2

STAT3 mutations were found in 5 cases (16%), all EBV positive. Interestingly all but one (*STAT3p*D566Y) of the mutations involved the SH2 domain of STAT3 protein (A. Mean expression for SH2 domain mutated cases (2 cases with available mutational and IHC data) was 249 nuclei per HPF (40x). Mean Phosho-STAT3 expression for STAT3 wild type cases was 28 nuclei per HPF (40X) (B). Thus, STAT3 SH2 domain mutations lead to PhosphoSTAT3 (Tyr705) protein overexpression. Representative microphotographs are shown in C.

FIGURE 3

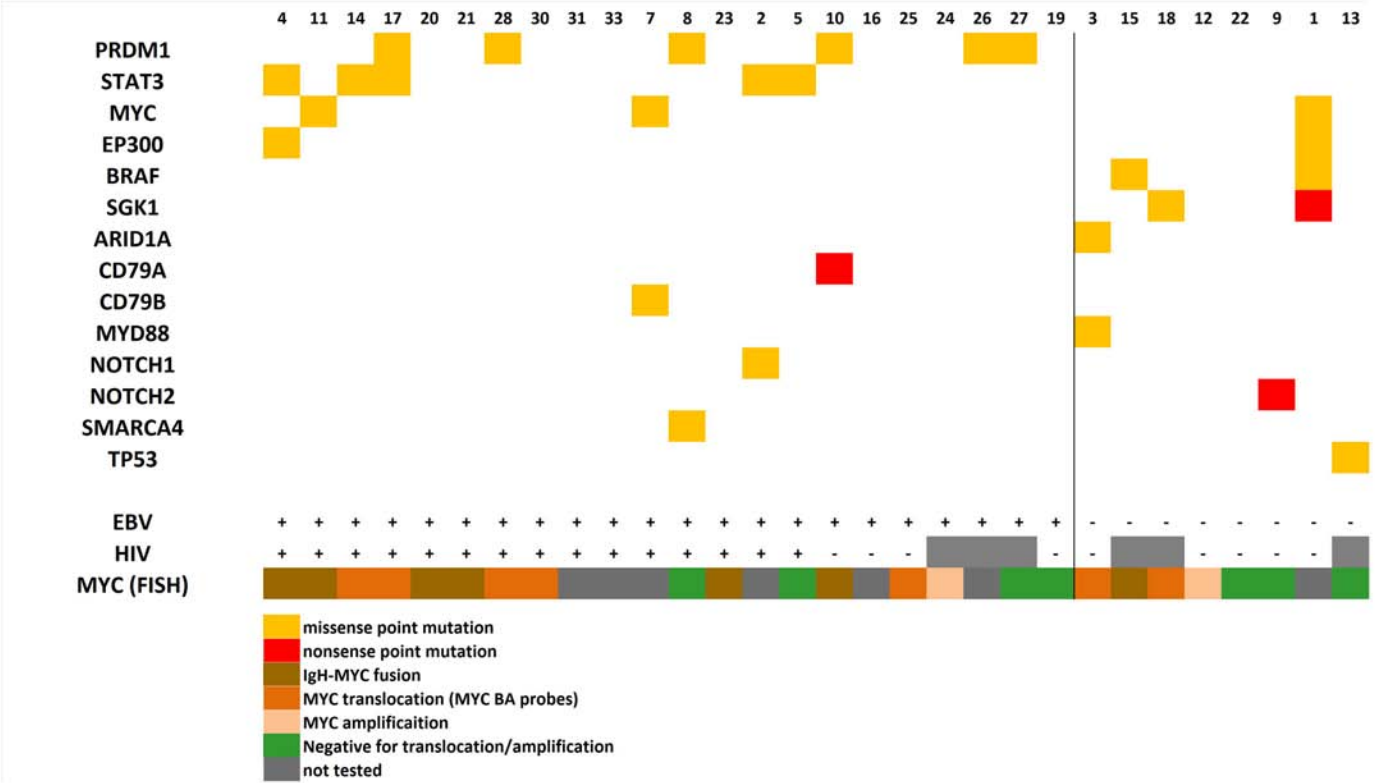
MYC protein was consistently expressed in the cases. Mean MYC protein expression in 22 cases with available data was 236 nuclei per HPF, significantly higher than the mean 48 nuclei per HPF in the cases of Phospho-STAT3 protein expression (Wilcoxon test p 0.001) (A). *MYC* translocated (14 cases) and amplified cases (2 cases) had higher MYC protein expression than cases without *MYC* rearrangements (7 cases) (Mann-Whitney test p value <0.0001) (B). Representative microphotographs are shown in C.

FIGURE 4

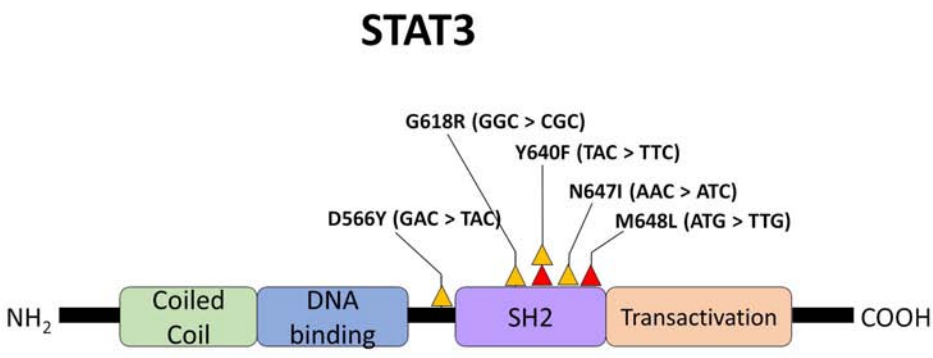
MYC protein overexpression is due to rearrangements involving *MYC* in a significant proportion of plasmablastic lymphoma cases (69% in these series). Most translocations fuse *MYC* to *IGH* and few cases may show amplifications of the *MYC* gene. In addition, here we show that a fraction of plasmablastic lymphoma cases has STAT3 activation, due to somatic mutations in the STAT3-SH2 domain that may increase MYC expression.

FIGURE 5

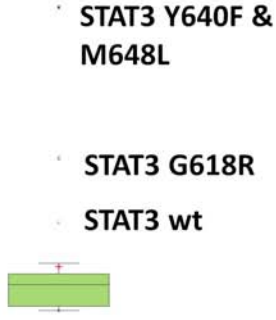
A. Scatter gram illustrating the mean and range of expression values after quantification of the IHC expression of CD8, PD1 in lymphocytes and PD-L1 and CD163 in histocyte/dendritic cell populations. B. Representative image of a case with a mean of 36 PD-L1 positive non neoplastic cells. C The same case showed a mean of 37 CD163 positive histiocytes. D. Mean expression of CD8 positive cells in this representative case was 53. E. PD1 identified a different T cell subpopulation (mean of 36 PD1 positive cells in this representative example, case n25). F. PD-L1 expression by neoplastic cells was identified in 5 out of 24 cases evaluated (20%). G. MHCII protein/ HLA (DP, DR) was, in most cases, restricted to histiocyte and endothelial cell populations. H. MHCII protein/ HLA (DP, DR) expression was identified in the neoplastic cells in 3 out of 25 cases tested (12%). 2 out of 3 cases showed a cytoplasmic granular and membranous staining (as shown in the figure) and 1 case had a membranous pattern.


4 11 14 17 20 21 28 30 31 33 7 8 23 2 5 10 16 25 24 26 27 19 3 15 18 12 22 9 1 13
PRDM1
STAT3
MYC
EP300
BRAF
SGK1
ARID1A
CD79A
CD79B
MYD88
NOTCH1
NOTCH2
SMARCA4
TP53
EBV
HIV
MYC (FISH)
missense point mutation
nonsense point mutation
IgH-MYC fusion
MYC translocation (MYC BA probes)
MYC amplificaition
Negative for translocation/amplification
not tested

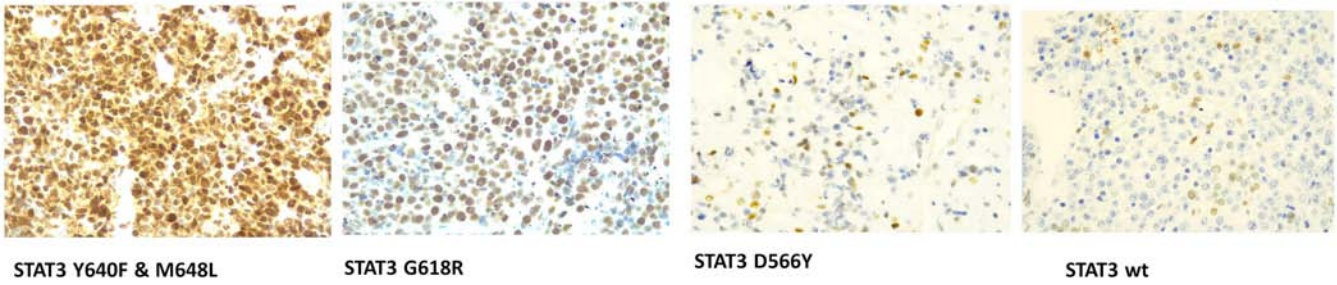
A


STAT3
G618R (GGC > CGC)
Y640F (TAC > TTC)
N647I (AAC > ATC)
D566Y (GAC > TAC)
M648L (ATG > TTG)
NH2
Coiled Coil
DNA binding
SH2
Transactivation
COOH

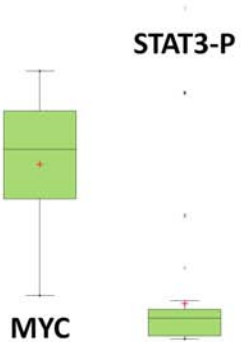
B


STAT3 Y640F & M648L
STAT3 G618R
STAT3 wt

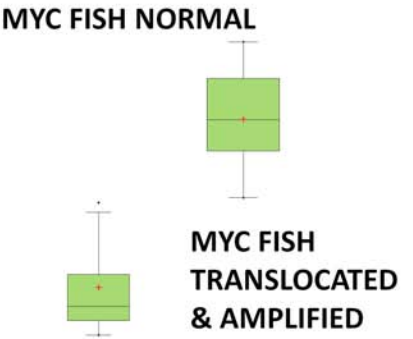
C


STAT3 Y640F & M648L
STAT3 G618R
STAT3 D566Y
STAT3 wt

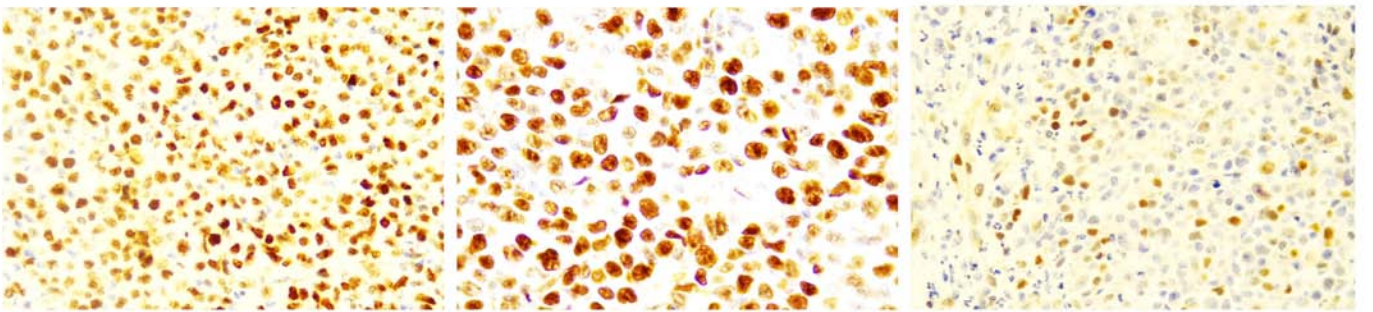
A



B



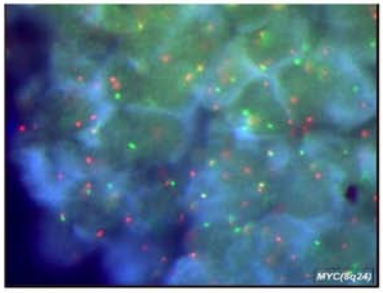
C

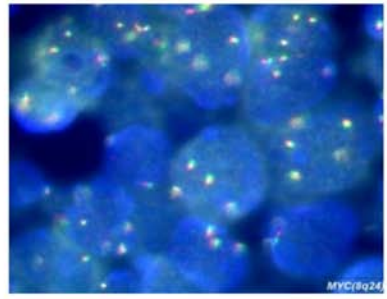


MYC-IgH POSITIVE

MYC-AMPLIFICATION POSITIVE

MYC-NON REARRANGED



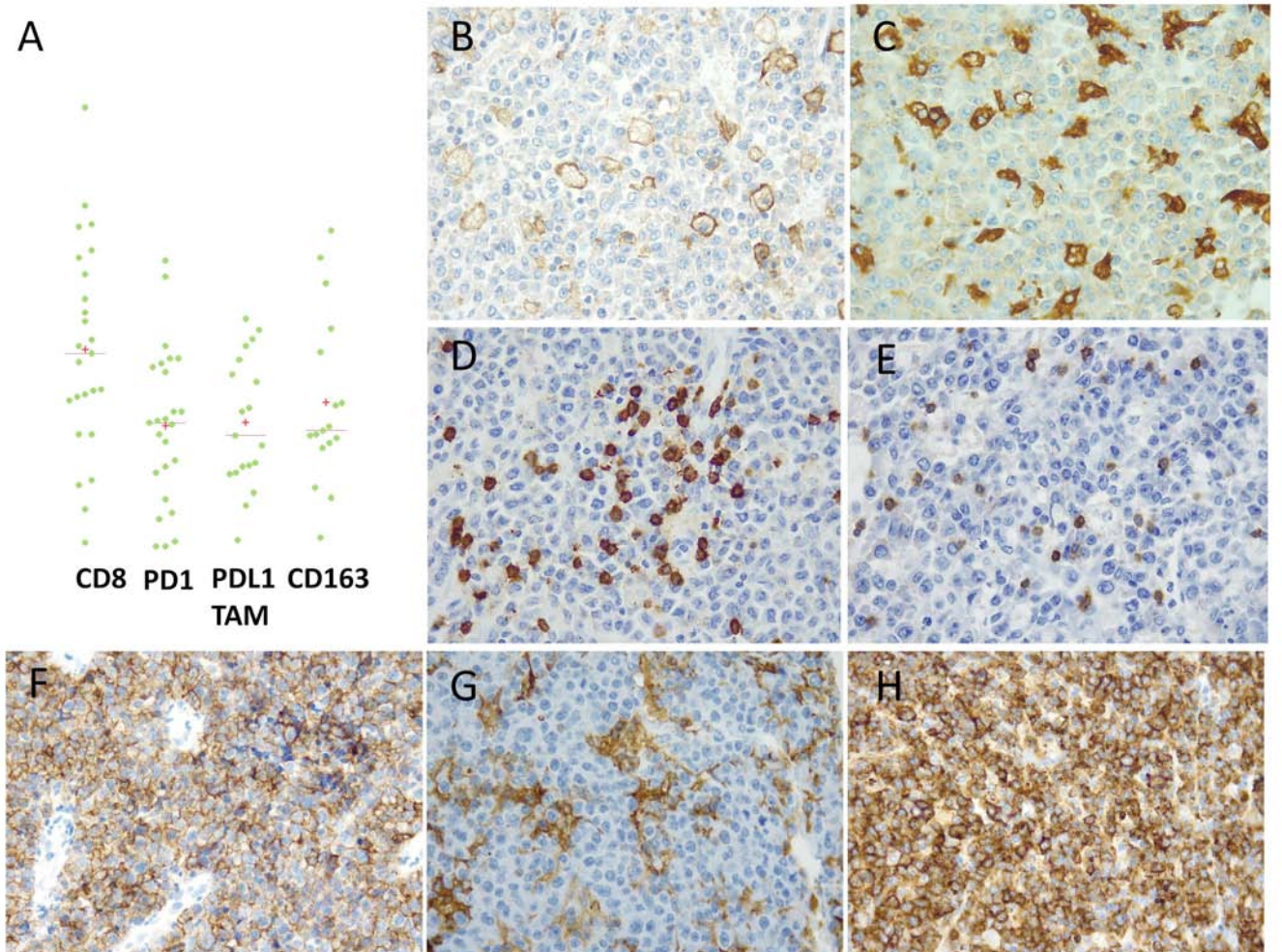


MYC translocation (60%)

MYC amplification (9%)

↑MYC PROTEIN

↑STAT3-P ← STAT3 SH2 mutations (13%)


A
B
C
D
E
F
G
H
CD8
PD1
PDL1
CD163
TAM

Supplementary Data

Supplementary Methods.

Case selection and sample collection

Twenty-eight new cases diagnosed as plasmablastic lymphoma were retrieved from the files of the Pathology Department of the Hospital Universitario Marqués de Valdecilla. Available Clinical data were retrieved. 10 samples were retrieved from the files of the University of Texas MD Anderson Cancer Center (Houston) Hematopathology Department files and 4 cases were received from the department of Pathology of San Bortolo Hospital, Vicenza, Italy. Material Transfer Agreements were signed by IDIVAL and the corresponding institutions to share the material in the project. The study and sample collection were approved by the local ethics committee (CEIC Cantabria, IRB code 2016.168) and complies with the Declaration of Helsinki. Formalin-fixed and paraffin-embedded tissue was available to construct two tissue microarrays, following conventional protocols. Whole sections were used in the rest of the cases to perform the immunohistochemical, chromogenic in situ hybridization and fluorescent in situ hybridization analysis. All cases were diagnosed according to the WHO classification of Hematolymphoid Neoplasms. Specifically, negativity for pan B cell markers (CD20), HHV-8 and ALK was required in every case to be included. The phenotype of the cases was consistent with a plasma cell differentiation program as described previously. The clinical features of the cases including age, gender, anatomic site of the biopsy, HIV status, previous transplantation or other source for immunesupression, CRAB features, M component and bone marrow biopsy infiltration were recorded and a summary is available in Supplementary Table 1.

Immunohistochemistry and in situ hybridization.

Immunohistochemical reactions were performed following conventional automated procedures (DAKO, Autostainer and Omnis automated platforms).

Primary antibodies against CD20 (DAKO, RTU), PAX-5 (DAKO, RTU), CD138 (DAKO, RTU), CD38 (Leica, 1:200), IRF4/MUM-1 (DAKO, RTU), Blimp-1 (CNIO, 1:5), Kappa (DAKO, RTU), Lambda (DAKO, RTU), BCL-6 (DAKO, RTU), BCL-2 (DAKO, RTU), CD10 (DAKO, RTU), KI67 (DAKO, RTU), CMYC (Abcam, 1:50), HHV-8 (Novus Biologicals, 1:10), EBV-LMP1 (DAKO, RTU), CD30 (DAKO, RTU), ALK (DAKO, RTU), p53 (DAKO, RTU), Phospho-STAT3 (Tyr705) (Clone EP2147Y, Millipore 1:100), PD-L1 (clone 22C3, DAKO), PD1 (clone NAT105, CNIO 1:5), CD3 (DAKO, RTU), CD4 (DAKO, RTU), CD8 (DAKO, RTU), CD163 (DAKO, RTU) and HLA-DP/DR (clone JS76 CNIO, 1:400) were used.

EBV-EBER was considered positive when ≥80% of the large atypical cells were positive. Fluorescent in situ hybridization for the detection of *MYC* rearrangement was done using a dual color break apart rearrangement probe set specific for the *MYC* gene locus on chromosome 8q24 (Abbot Molecular). At least 10% of cells with a break apart signal were required for a case to be regarded as positive for *MYC* rearrangements. At least 15% of cells with extra-copies of the *MYC* gene were required to identify a case as positive for copy number gains of the *MYC* gene.

Quantification of the cellular composition of the tumor and transcription factor abundance.

Quantification of different lymphoid and histiocytic/dendritic subpopulations, identified with CD3, CD8, PD1, CD163, PD-L1 and MHCII/HLA DP/DR and absolute quantification of the number of nuclei showing expression of MYC and Phospho-STAT3 (Tyr705) was done according to the following method:

Microphotographs of 3 representative high-power fields (40x magnification) were acquired for each case using a Leica photomicroscope (DM2000LED) with an attached camera (LEICA ICC50 W). Hotspots with higher positivity were chosen. All the above referenced immunohistochemistry markers were quantified visually averaging the number of positive cells quantified in triplicate in tissue sections. For nuclear markers we counted the positive and negative nuclei. For cytoplasmic and membrane bound markers we counted separately the positive cells with circumferential staining and the negative nuclei. For specific markers such as PD-L1, the expression was quantified in both neoplastic cells and microenvironment cells.

Nuclear atypia was used to differentiate the neoplastic cells from others (i.e macrophages). MHCII/HLA DP/DR was quantified in neoplastic cells and the pattern of expression was recorded (membrane or cytoplasmic staining).

Next Generation Sequencing using amplicon-based library generation.

A TruSeq® Custom Amplicon Low Input Library containing exonic regions of 35 selected genes of interest was used to isolate the DNA for sequencing (Illumina). The list of genes was *CARD11, ARID1A, NOTCH1, TCF3, SMARCA4, STAT6, EP300, CREBBP, MLL2, BTK, NOTCH2, TNFRSF14, ATM, FOXO1, B2M, PLCG2, CD79B, TP53, STAT3, BCL2, MEF2B, CD79A, CXCR4, PTPN1, MYD88, FAT2, PRDM1, TNFAIP3, SGK1, CCND3, PIM1, EZH2, BRAF, MYC, NOTHC2*.
Of note, variants occurring in regions outside the coverage of our targeted design, were not explored using this approach.

Depending on the availability of DNA, between 6ng and 400ng of DNA were used for libraries preparation. Two different libraries of FFPE samples (one for each DNA strand) were prepared and required per protocol in order to eliminate false C-T mutations that commonly appear during formalin fixation. After library preparation and quantification by Qubit (ThermoFisher Scientific) libraries from 30 out of 42 samples met the quality control criteria and were pooled for sequencing on a HiSeq instrument (Illumina, paired end, 2x150) at the National Genomic Analysis Center (CNAG, Barcelona, Spain).

Sequencing data interpretation and reporting.

Reads were mapped to human genome build hg19 with decoy sequences (hs37d5) using the GEM toolkit (version 3). The Genome Analysis Tool Kit (GATK) was used for local realignment and base quality score recalibration. Variant calling was done using HaplotypeCaller from GATK following the recommended best practices. Functional annotations were added using SnpEff with the GRCh37.75 database. Variants were annotated with SnpSift using population frequencies, conservation scores and deleteriousness predictions from dbNSFP. Other sources of annotations, such as gnomAD and Clinvar were also used.

Only variants with a coverage ≥300 reads were selected for downstream analysis. Both DNA strands were available in all 30 cases. Only variants in which both libraries had a coverage greater or equal to 300 reads and had the same genotype were selected for downstream analysis.

Subsequently only missense, frameshift, and nonsense somatic mutations with variant frequency>10% were considered (supplementary Table 2). SNPs were filtered out based on the comparison of the VAF of the variant with the estimation of the amount of neoplastic cells by morphology and IHC, after search in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and after comparison with a germline variants database collected from an in house analysis of 89 germline DNA from patients with DLBCL. The COSMIC (http://cancer.sanger.ac.uk/cosmic) database was also checked in every case and the COSMIC Id was annotated. Three algorithms were used to predict the functional consequences of the variants found, including SIFT (http://sift.bii.astar.edu.sg/), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/) and Condel (http://bg.upf.edu/fannsdb/). Selected variants were visualized using the Integrative Genomics Viewer (IGV).

Finally, 34 somatic mutations (31 missense, 3 nonsense) in 14 genes were considered (Table 1 in the main text).

Statistical analysis.

XLSTAT Biomed software (version 19.4) was used for statistical analysis. Descriptive statistics were performed.

Supplementary Table 1.

The clinical features of the cases including age, gender, anatomic site of the biopsy, HIV status, previous transplantation or other source for immunesupression, CRAB features, M component and bone marrow biopsy infiltration are shown.

case ID	Age	Sex	Anatomic Site	HIV	EBV in tumor cells	Transplant	Other immune deficiency	CRAB	M Component	Bone Marrow Biopsy
2	41	M	Maxillary sinus	pos	pos	no	none	no	---	negative
4	57	M	Maxillary sinus	pos	pos	no	none	no	---	---
5	50	M	Oral cavity	pos	pos	no	none	no	---	---
7	38	M	Rectum	pos	pos	no	none	no	---	negative
8	36	M	Lymph node	pos	pos	no	none	---	---	---
11	33	F	Perineal region	pos	pos	no	none	no	negative	negative
14	56	M	Oral cavity	pos	pos	no	none	bone lesions	negative	negative
17	30	M	Rectum	pos	pos	no	none	no	negative	negative
20	29	M	Lymph node	pos	pos	no	none	---	---	---
21	36	M	Soft tissue	pos	pos	no	none	no	positive	negative
23	44	M	Perineal region	pos	pos	no	none	no	negative	negative
28	53	M	Nasopahringeal area	pos	pos	no	none	no	---	negative
30	37	F	Ovary	pos	pos	no	none	no	negative	negative
31	50	M	Lymph node	pos	pos	no	none	no	negative	negative
33	37	M	Oral cavity	pos	pos	no	none	no	negative	negative
24	66	M	Lymph node	neg	pos	no	none	no	---	---
25	62	F	GI tract	neg	pos	Renal Transplant	none	no	---	---
1	66	M	Lymph node	neg	neg	alloSCT (for relapsed CLL)	none	no	negative	negative
3	64	M	Lymph node	neg	neg	no	none	no	---	---

9	75	M	Lymph node	neg	neg	no	none	no	negative	negative
10	58	M	GI tract	neg	pos	no	none	no	---	negative
12	92	F	Nasal region	neg	neg	no	none	no	---	---
16	62	M	Maxillary sinus	neg	pos	no	none	no	---	negative
19	79	F	Lymph node	neg	pos	no	none	---	---	---
22	76	M	Skin	neg	neg	no	none	no	---	negative
13	75	M	Nasal region	---	neg	---	---	---	---	---
15	82	M	Oral cavity	---	neg	---	---	no	---	---
18	56	F	Small intestine	---	neg	no	none	no	negative	---
26	43	M	Paranasal region	---	pos	---	---	no	---	---
27	76	M	Paranasal region	---	pos	---	---	no	---	---

Supplementary Table 2.

Complete list of variants identified after quality control and annotation. Only missense, frameshift, and nonsense somatic mutations with variant frequency>10% are depicted. This list includes SNPs that were filtered out in a subsequent step.

POS: Genomic position, REF: reference base, ALT: altered base, QUAL: quality score, DP: number of reads for the allele, AF: Variant Allele Frequency.

Id	Gene	CHR	POS	REF	ALT	QUAL	Annotation	DP	AF	cDNA_pos	CDS_pos	Protein_pos	AA	Codons	Existing_variation
8	ARID1A	1	27100181	CGCA	C	2181,4	disruptive_inframe_deletion	597	0,23	4349-4351	3978-3980	1326-1327	PQ/P	ccGCAg/ccg	COSM298325
8	EP300	22	41546030	C	G	10951	missense_variant	917	0,49	3864	2645	882	P/R	cCa/cGa	COSM4385247
8	PRDM1	6	1,07E+08	G	A	9400,4	missense_variant	2090	0,30	2546	2312	771	G/D	gGc/gAc	rs80257572
8	SMARCA4	19	11135047	G	A	436,44	missense_variant	543	0,16	3295	3014	1005	R/Q	cGa/cAa	-
3	ARID1A	1	27023285	G	A	4731,4	missense_variant	380	0,57	762	391	131	G/R	Ggg/Agg	-
3	ARID1A	1	27106544	G	C	4235,4	missense_variant	1915	0,21	6526	6155	2052	C/S	tGc/tCc	-
3	TNFRSF14	1	2488153	A	G	4635,4	missense_variant	786	0,36	349	50	17	K/R	aAa/aGa	rs4870
3	MYD88	3	38182641	T	C	100	missense_variant	250	0,24		794	265	L/P	cTg/cCg	COSM85940
9	EZH2	7	1,49E+08	C	G	19555	missense_variant	4508	0,28	675	553	185	D/H	Gac/Cac	COSM3762469
9	NOTCH2	1	1,2E+08	G	A	1776,4	stop_gained	1784	0,16	7418	7198	2400	R/*	Cga/Tga	COSM36210
9	TNFRSF14	1	2488153	A	G	21798	missense_variant	2336	0,49	349	50	17	K/R	aAa/aGa	rs4870
9	TP53	17	7579472	G	C	48088	missense_variant	1754	0,99	405	215	72	P/R	cCc/cGc	COSM250061
4	EP300	22	41573745	G	A	2213,4	missense_variant	758	0,24	7249	6030	2010	M/I	atG/atA	-
4	STAT3	17	40475330	C	A	2579,4	missense_variant	1239	0,21	2009	1696	566	D/Y	Gac/Tac	COSM220689
10	CD79A	19	42383208	G	A	1076,4	stop_gained	778	0,17	413	228	76	W/*	tgG/tgA	COSM5493940
10	PRDM1	6	1,07E+08	G	A	6990,4	missense_variant	2130	0,22	1295	1061	354	S/N	aGc/aAc	rs143040512
11	MYC	8	1,29E+08	C	G	21143	missense_variant	1598	0,59	578	68	23	T/S	aCc/aGc	-
11	MYC	8	1,29E+08	T	A	7615,4	missense_variant	796	0,48	775	265	89	Y/N	Tac/Aac	-
11	MYC	8	1,29E+08	T	C	5362,4	missense_variant	890	0,37	899	389	130	F/S	tTc/tCc	COSM4171775
11	MYC	8	1,29E+08	C	G	4032,4	missense_variant	1315	0,25	945	435	145	I/M	atC/atG	-

22	NOTCH2	1	1,21E+08	G	C	7845,4	missense_variant	880	0,47	277	57	19	C/W	tgC/tgG	rs11810554
23	EP300	22	41548008	A	G	9499,4	missense_variant	1380	0,39	4208	2989	997	I/V	Att/Gtt	rs20551
12	CARD11	7	2962848	G	A	6196,4	missense_variant	602	0,54	2464	2060	687	A/V	gCg/gTg	rs41493047
12	EP300	22	41548008	A	G	8543,4	missense_variant	658	0,65	4208	2989	997	I/V	Att/Gtt	rs20551
12	TNFRSF14	1	2488153	A	G	44442	missense_variant	1795	1,00	349	50	17	K/R	aAa/aGa	rs4870
13	ATM	11	1,08E+08	G	A	31006	missense_variant	5357	0,35	5942	5557	1853	D/N	Gat/Aat	COSM41596
13	EZH2	7	1,49E+08	C	G	170330	missense_variant	5215	1,00	675	553	185	D/H	Gac/Cac	COSM3762469
13	NOTCH2	1	1,21E+08	G	C	8499,4	missense_variant	2632	0,26	277	57	19	C/W	tgC/tgG	rs11810554
13	TNFRSF14	1	2488153	A	G	70550	missense_variant	6973	0,52	349	50	17	K/R	aAa/aGa	rs4870
13	TP53	17	7577120	C	T	98089	missense_variant	7897	0,59	1008	818	273	R/H	cGt/cAt	COSM10660
13	TP53	17	7579472	G	C	91137	missense_variant	3400	1,00	405	215	72	P/R	cCc/cGc	COSM250061
24	NOTCH2	1	1,21E+08	G	C	9698,4	missense_variant	932	0,53	277	57	19	C/W	tgC/tgG	COSM132738
24	TNFRSF14	1	2488153	A	G	55725	missense_variant	2190	0,98	349	50	17	K/R	aAa/aGa	rs4870
14	STAT3	17	40474459	T	A	13963	missense_variant	1179	0,59	2255	1942	648	M/L	Atg/Ttg	-
14	STAT3	17	40474482	T	A	13870	missense_variant	1179	0,59	2232	1919	640	Y/F	tAc/tTc	COSM1155743
26	ATM	11	1,08E+08	G	A	48248	missense_variant	5546	0,47	5942	5557	1853	D/N	Gat/Aat	COSM41596
26	NOTCH2	1	1,21E+08	G	C	3762,4	missense_variant	1080	0,26	277	57	19	C/W	tgC/tgG	rs11810554
26	PRDM1	6	1,07E+08	C	G	34258	missense_variant	2422	0,58	843	609	203	D/E	gaC/gaG	rs811925
26	TNFRSF14	1	2488153	A	G	48412	missense_variant	4358	0,55	349	50	17	K/R	aAa/aGa	rs4870
27	EP300	22	41548008	A	G	20536	missense_variant	1638	0,58	4208	2989	997	I/V	Att/Gtt	rs20551
27	NOTCH2	1	1,21E+08	G	C	3165,4	missense_variant	577	0,35	277	57	19	C/W	tgC/tgG	COSM132738
27	PRDM1	6	1,07E+08	C	G	31926	missense_variant	3046	0,53	843	609	203	D/E	gaC/gaG	rs811925
27	TP53	17	7579472	G	C	40713	missense_variant	1503	1,00	405	215	72	P/R	cCc/cGc	COSM250061
15	BRAF	7	1,4E+08	C	G	27862	missense_variant	2185	0,62	1467	1406	469	G/A	gGa/gCa	COSM459
15	EP300	22	41548008	A	G	5628,4	missense_variant	655	0,45	4208	2989	997	I/V	Att/Gtt	rs20551
15	NOTCH2	1	1,21E+08	G	C	3181,4	missense_variant	288	0,54	277	57	19	C/W	tgC/tgG	rs11810554
5	ATM	11	1,08E+08	A	C	61148	missense_variant	6156	0,51	4747	4362	1454	K/N	aaA/aaC	COSM22501
5	ATM	11	1,08E+08	G	A	54734	missense_variant	5588	0,51	5942	5557	1853	D/N	Gat/Aat	COSM41596

5	EZH2	7	1,49E+08	C	G	58716	missense_variant	5226	0,48	675	553	185	D/H	Gac/Cac	COSM3762469
5	NOTCH2	1	1,21E+08	G	C	2979,4	missense_variant	689	0,30	277	57	19	C/W	tgC/tgG	COSM132738
5	STAT3	17	40474482	T	A	29044	missense_variant	5011	0,35	2232	1919	640	Y/F	tAc/tTc	COSM1155743
16	ATM	11	1,08E+08	T	C	45184	missense_variant	4755	0,48	2504	2119	707	S/P	Tct/Cct	COSM41595
16	NOTCH2	1	1,21E+08	G	C	18815	missense_variant	1928	0,51	277	57	19	C/W	tgC/tgG	rs11810554
16	TP53	17	7579472	G	C	66471	missense_variant	2194	1,00	405	215	72	P/R	cCc/cGc	COSM250061
28	PRDM1	6	1,07E+08	C	G	13852	missense_variant	3293	0,26	843	609	203	D/E	gaC/gaG	rs811925
28	TNFRSF14	1	2488153	A	G	24340	missense_variant	2063	0,60	349	50	17	K/R	aAa/aGa	rs4870
17	ATM	11	1,08E+08	T	A	43593	missense_variant	4493	0,51	763	378	126	D/E	gaT/gaA	COSM22498
17	NOTCH2	1	1,21E+08	G	C	1699,4	missense_variant	844	0,21	277	57	19	C/W	tgC/tgG	rs11810554
17	PRDM1	6	1,07E+08	C	G	25161	missense_variant	2936	0,46	843	609	203	D/E	gaC/gaG	rs811925
17	STAT3	17	40475058	C	G	21942	missense_variant	1498	0,69	2165	1852	618	G/R	Ggc/Cgc	COSM1166777
17	TNFAIP3	6	1,38E+08	T	G	158181	missense_variant	5806	0,99	446	380	127	F/C	tTc/tGc	COSM1685340
17	TNFRSF14	1	2488153	A	G	24580	missense_variant	2386	0,51	349	50	17	K/R	aAa/aGa	rs4870
30	NOTCH2	1	1,21E+08	G	C	6521,4	missense_variant	699	0,48	277	57	19	C/W	tgC/tgG	rs11810554
18	ATM	11	1,08E+08	A	G	115074	missense_variant	4261	1,00	6333	5948	1983	N/S	aAt/aGt	rs659243
18	NOTCH2	1	1,21E+08	G	C	2996,4	missense_variant	1148	0,23	277	57	19	C/W	tgC/tgG	rs11810554
18	SGK1	6	1,34E+08	G	A	30903	missense_variant	5213	0,36	1950	1352	451	S/F	tCc/tTc	-
18	SGK1	6	1,34E+08	G	A	40653	missense_variant	6831	0,36	1737	1139	380	A/V	gCt/gTt	-
18	SGK1	6	1,34E+08	C	T	36890	5_prime_UTR_premature_start_codon_gain_variant	6373	0,35	#N/D	#N/D	#N/D	#N/D	#N/D	#N/D
18	TNFRSF14	1	2488153	A	G	45495	missense_variant	5114	0,47	349	50	17	K/R	aAa/aGa	rs4870
18	TP53	17	7579472	G	C	34947	missense_variant	1764	0,86	405	215	72	P/R	cCc/cGc	COSM250061
7	CD79B	17	62008716	C	T	2934,4	missense_variant	963	0,24	175	100	34	D/N	Gac/Aac	-
7	KMT2D	12	49426460	A	G	25817	missense_variant	1772	0,66	12028	12028	4010	S/P	Tct/Cct	rs80132640
7	KMT2D	12	49448463	C	T	14800	missense_variant	1052	0,57	248	248	83	R/Q	cGg/cAg	rs55865069
7	MYC	8	1,29E+08	A	T	2614,4	missense_variant	794	0,26	1085	575	192	Y/F	tAc/tTc	-
31	ATM	11	1,08E+08	A	G	10575	missense_variant	378	1,00	6333	5948	1983	N/S	aAt/aGt	rs659243

31	NOTCH2	1	1,21E+08	G	C	5897,4	missense_variant	348	0,76	277	57	19	C/W	tgC/tgG	rs11810554
31	TNFRSF14	1	2488153	A	G	11911	missense_variant	1217	0,51	349	50	17	K/R	aAa/aGa	rs4870
2	ATM	11	1,08E+08	C	G	16840	missense_variant	1591	0,55	2021	1636	546	L/V	Ctg/Gtg	rs2227924
2	ATM	11	1,08E+08	C	T	24400	missense_variant	3367	0,41	2999	2614	872	P/S	Cct/Tct	rs3218673
2	NOTCH1	9	1,39E+08	G	A	13195	missense_variant	938	0,38	1278	1202	401	P/L	cCc/cTc	COSM4745915
2	STAT3	17	40474461	T	A	5089,4	missense_variant	1930	0,23	2253	1940	647	N/I	aAc/aTc	COSM1155744
2	TNFAIP3	6	1,38E+08	T	G	9057,4	missense_variant	1327	0,39	446	380	127	F/C	tTc/tGc	COSM1685340
33	EP300	22	41546030	C	G	4277,4	missense_variant	538	0,35	3864	2645	882	P/R	cCa/cGa	COSM4385247
1	ATM	11	1,08E+08	G	A	62638	missense_variant	6032	0,50	5942	5557	1853	D/N	Gat/Aat	COSM41596
1	BRAF	7	1,4E+08	A	T	61284	missense_variant	5236	0,58	1860	1799	600	V/E	gTg/gAg	COSM476
1	EP300	22	41572907	G	A	93605	missense_variant	#####	0,45	6411	5192	1731	R/H	cGc/cAc	-
1	MYC	8	1,29E+08	C	T	32277	missense_variant	4424	0,41	747	237	79	S	agC/agT	-
1	NOTCH2	1	1,21E+08	G	C	5784,4	missense_variant	1536	0,31	277	57	19	C/W	tgC/tgG	rs11810554
1	SGK1	6	1,34E+08	G	T	3083,4	missense_variant	395	0,43	1018	420	140	N/K	aaC/aaA	COSM1581714
1	SGK1	6	1,34E+08	T	A	2868,4	stop_gained	395	0,42	1004	406	136	K/*	Aag/Tag	-
1	SGK1	6	1,34E+08	C	T	2884,4	missense_variant	395	0,42	978	380	127	R/K	aGg/aAg	COSM220583
1	TNFAIP3	6	1,38E+08	T	G	61438	missense_variant	5765	0,48	446	380	127	F/C	tTc/tGc	COSM1685340
1	TNFRSF14	1	2488153	A	G	110496	missense_variant	4029	0,97	349	50	17	K/R	aAa/aGa	rs4870