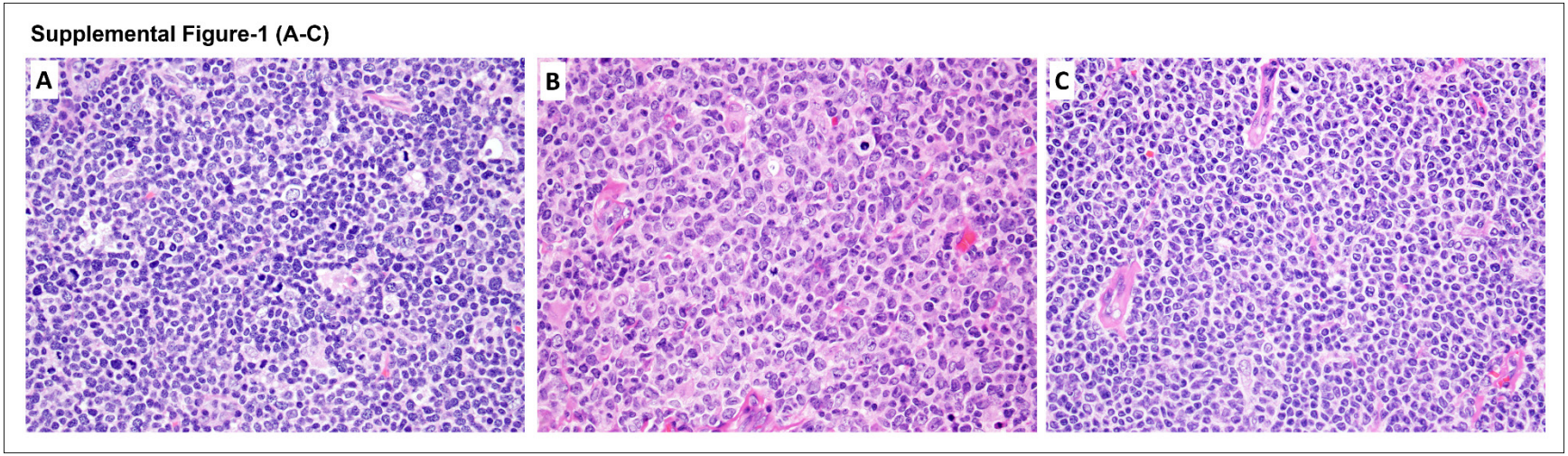
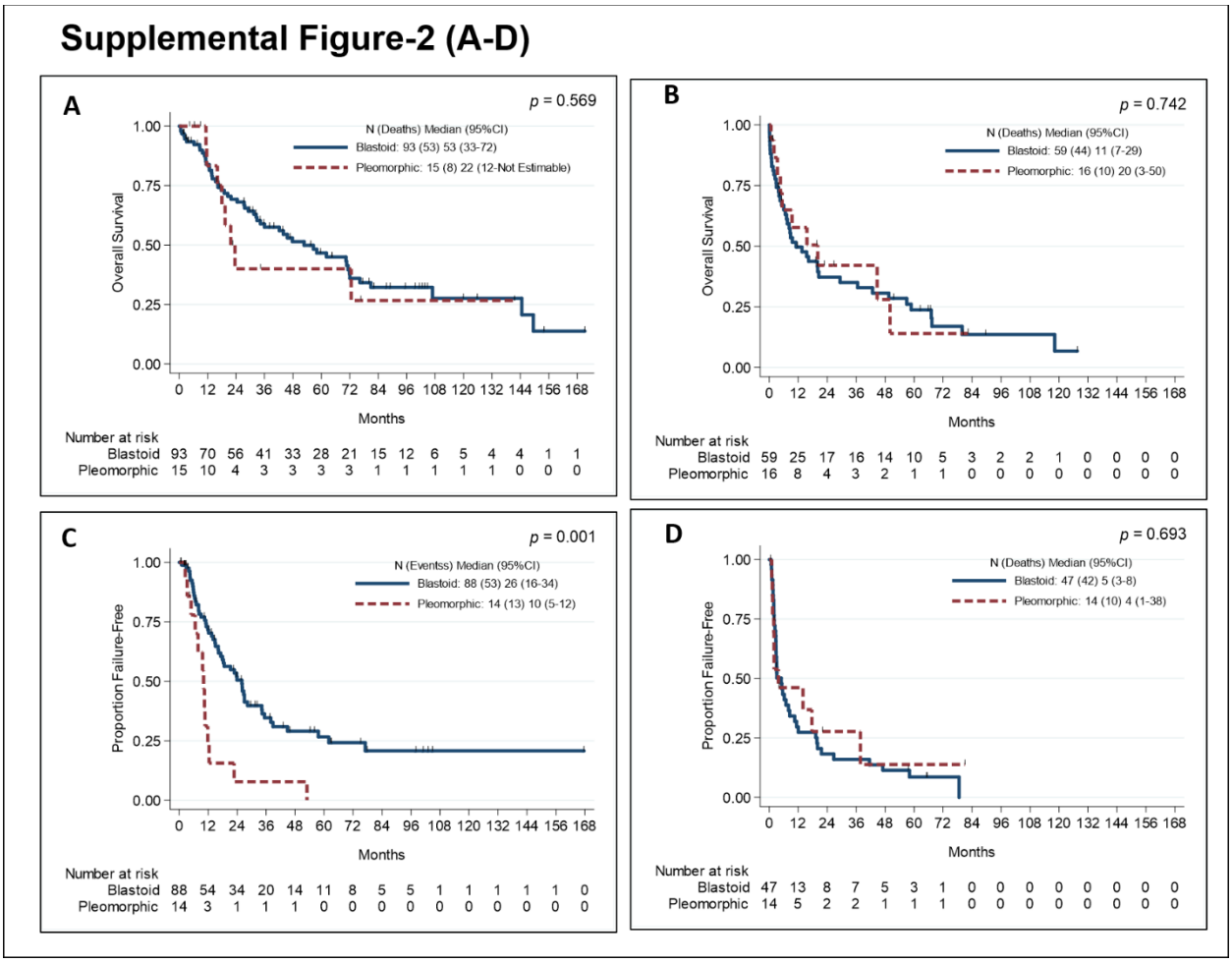
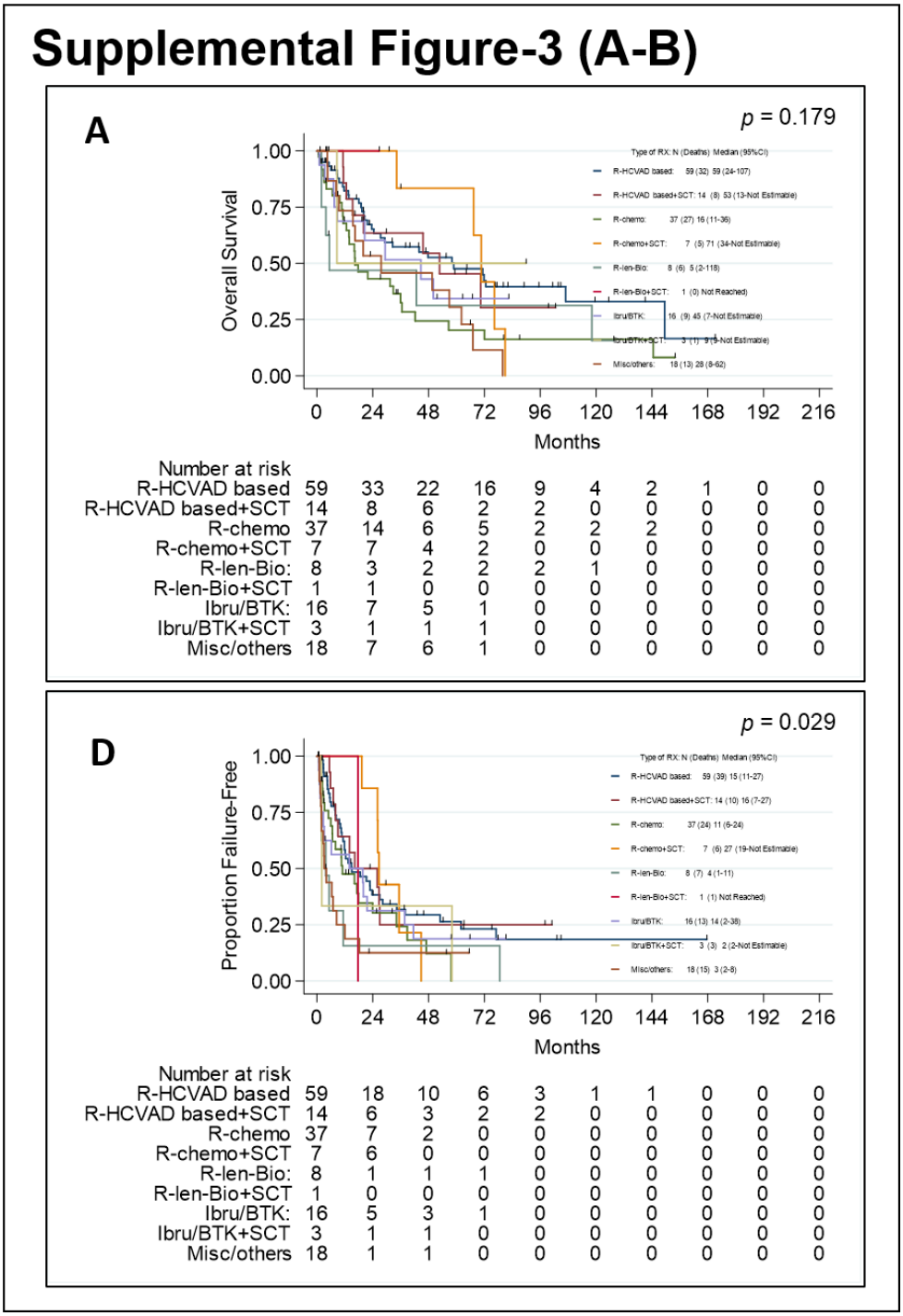
Supplemental Figure legend, tables and supplemental methods



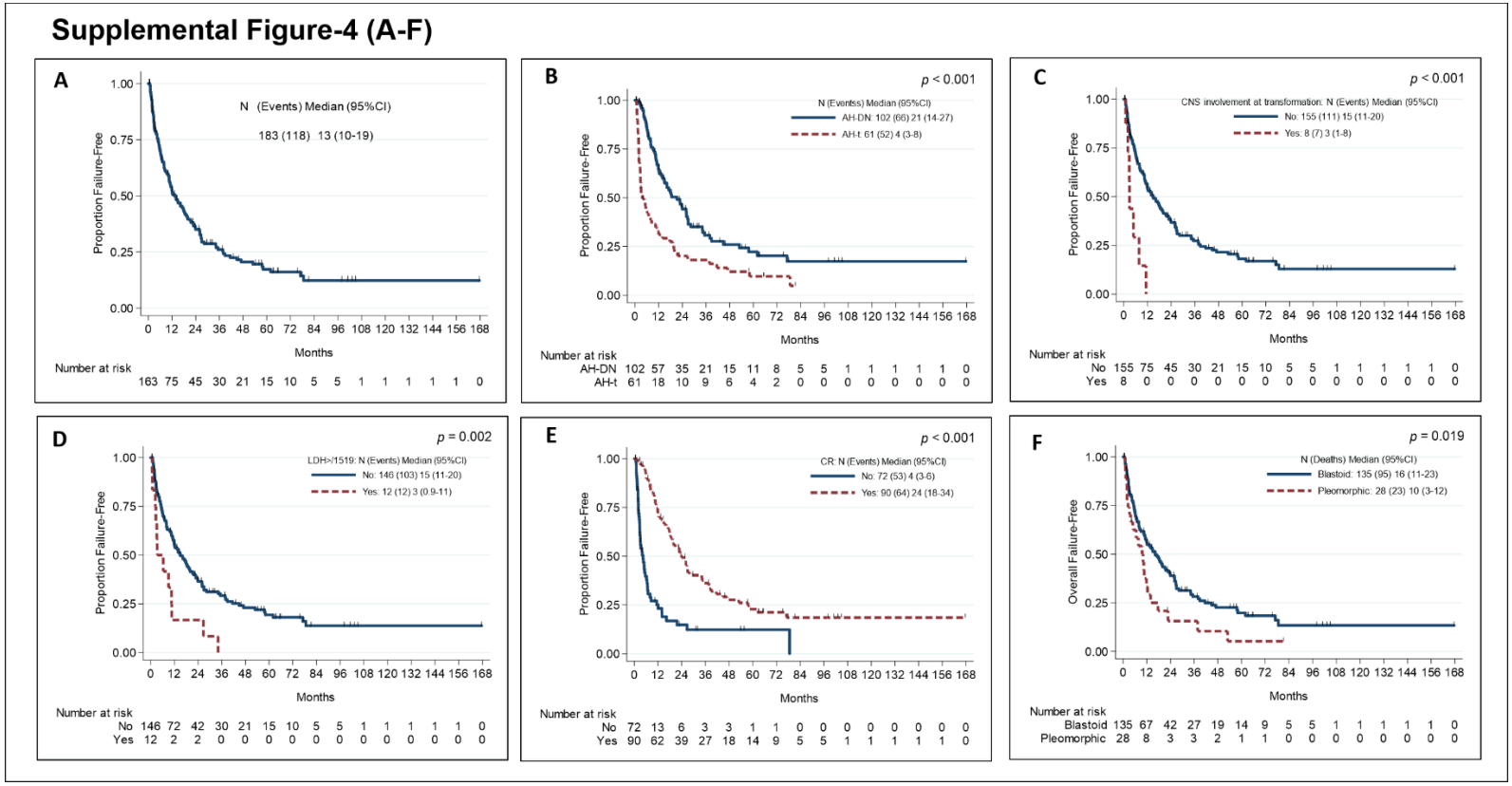
Supplemental Figure – 1 (A-C) Representative Hematoxylin and eosin (H&E) images (400X) of aggressive (AH-MCL; blastoid, pleomorphic variants) and non-aggressive MCL from lymph node biopsies – A) Blastoid morphology with homogenous pattern, round nuclei, fine thin chromatin resembling lymphoblasts is shown **B)** Pleomorphic morphology, heterogeneous distribution, small to large MCL cells with anaplastic nucleus, irregular but frequent and prominent nuclei, resembling large B cell lymphoma is shown. **C)** Non aggressive classic morphology MCL cells. Monotonous population with slightly or markedly irregular nuclear contours and moderately dispersed chromatin. This variant can have a nodular or diffuse pattern of distribution.



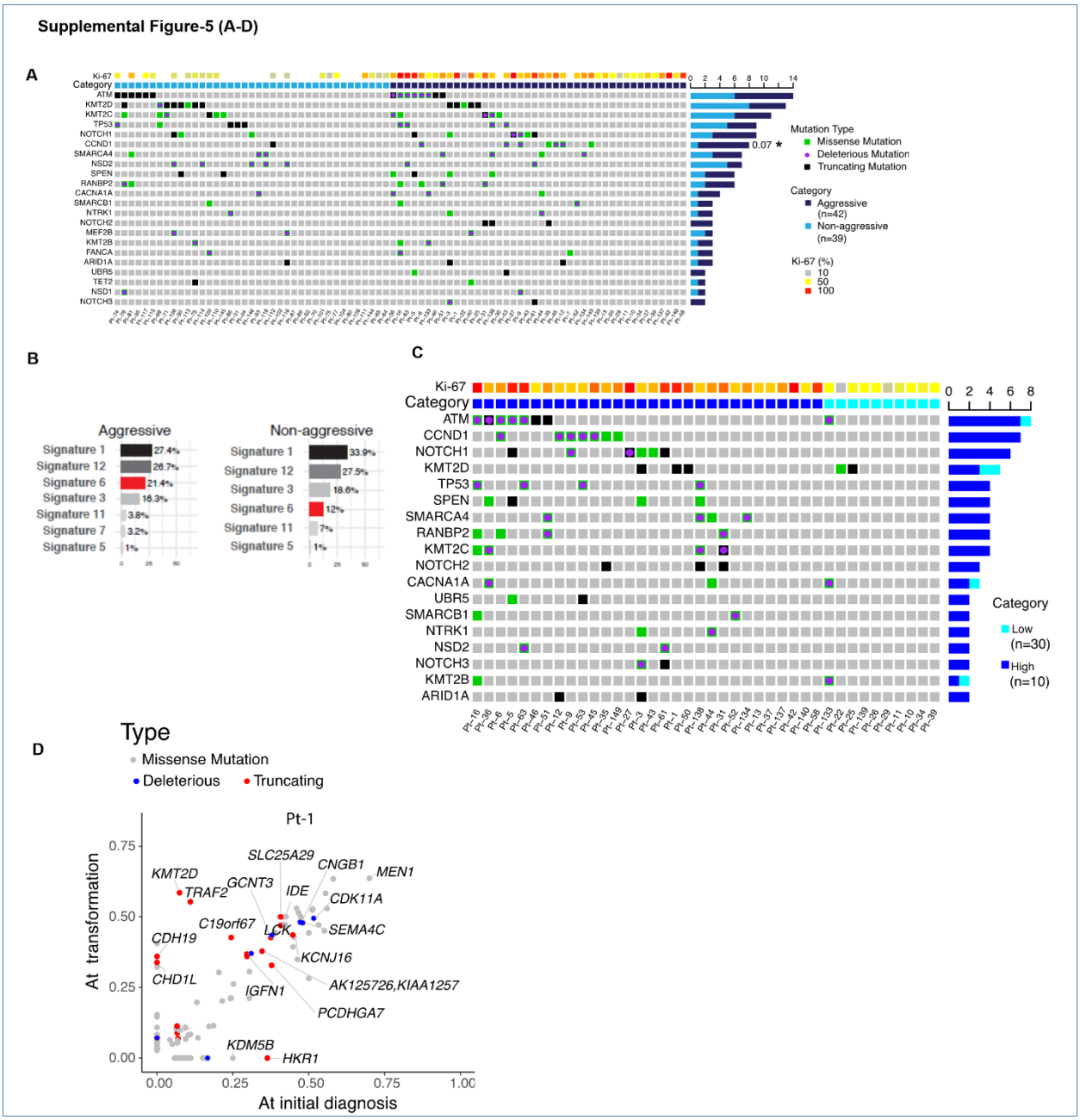
Supplemental Figure – 2 (A-D) Overall and failure free survival (FFS) in *de novo*-MCL (AH-DN) and transformed MCL (AH-t) according to histologic variant – blastoid vs pleomorphic A) With in AH-DN category, the median survival was not significantly different in blastoid vs pleomorphic (53 vs 22 month respectively; p=0.56) **B)** With in AH-t category, the median survival was not significantly different in blastoid vs pleomorphic (11 vs 20 months respectively; p=0.74) **C)** Median FFS within AH-DN category was significantly shorter in pleomorphic variant compared to blastoid variant MCL (10 vs 26 months respectively; p=0.001) **D)** With in AH-t category, the median FFS was not significantly different in blastoid vs pleomorphic (5 vs 4 months respectively; p=0.69).



Supplemental Figure – 3 A) Median survival is shown according to the type of first line treatment given after the diagnosis of AH-MCL. Treatment type is divided based on whether patients received stem cell transplantation (SCT). The differences were not statistically significant in various treatment groups (p=0.107). **B)** Median FFS is shown according to the type of first line treatment given after the diagnosis of AH-MCL. Treatment type is divided based on whether patients received stem cell transplantation (SCT). The differences were statistically significant among various treatment groups (p=0.02).



Supplemental Figure – 4 (A-F) Failure free survival (FFS) in aggressive histology (AH) patients with MCL including denovo-MCL (AH-DN) and transformed MCL (AH-t) A) Median FFS after diagnosis was 13 months. This includes all patients with AH-MCL **B)** Median FFS was significantly longer in AH-DN (21 months) vs 4 months in AH-t ($p<0.001$). **C)** Median FFS was significantly inferior in patients with central nervous system (CNS) involvement at diagnosis of AH-MCL (3 months) compared to those without CNS involvement (15 months), $p<0.001$. **D)** Median survival was significantly longer in patients with low serum LDH levels (<1519; 15 months) versus high LDH levels (≥1519; 3 months) $p=0.002$. Cut off point of 1519 IU/L was based on classification and regression tree analysis (CART) **E)** Median FFS was significantly longer in patients who achieved complete remission (CR) after first line treatment after the diagnosis of AH-MCL (24 months), compared to those patients who did not achieve CR (4 months) after first line therapy. $p<0.001$. **F)** Median FFS was significantly shorter in patients with pleomorphic variant MCL compared to blastoid variant MCL ($P=0.01$)



Supplemental Figure – 5 (A-D) Genomic profile (Oncoprint, and mutation signature) in aggressive histology MCL and non-aggressive (classic variant) and Ki-67% categories. A) Oncoprint comparing AH-MCL vs non-aggressive MCL. Each column represents a patient tumor sample. The histograms on the left side show the accumulated counts of somatic alterations for each specific gene by their group. Patient category and Ki-67% are annotated on the top tracks. Color bar of Ki-67 (%) represent Ki-67 protein percentage gradually increased from low (grey) to high (red). **B)** Mutation signature and total mutation burden is shown between AH-MCL and non-aggressive groups. Mutation signature 6 associated with defective DNA damage control was differentially observed in AH-MCL group and mutational burden was

significantly higher in AH-MCL. All somatic substitutions identified are included to decipher mutational signatures. Increased contribution of signature 6 to aggressive tumor when compared with non-aggressive tumor. **C)** Oncoprint showing pattern of somatic mutations in high (n=30) and low Ki-67% (n=10) categories. Differences in the two groups were statistically significant with distinct somatic mutation profile in patients with high Ki-67%. The histograms on the left side show the accumulated counts of somatic alterations for each specific gene by their group. Almost all of the somatic mutations were predominantly observed in patients with high Ki-67% group. **D)** The variant allelic fractions of somatic mutations in paired baseline-progression (BP) and (CP) tumors from Patients 1 (Pt-1). Each dot represents a mutation and the dots are colored by their mutation types. CDH19 and CHD1L truncating mutation emerged at BP tumors. KMT2D and TRAF2 mutations are also dominant at BP tumors.

Supplemental Table- 1: Univariate and multivariate analyses of factors predicting for failure free survival (FFS) after first line therapy for aggressive histology MCL (AH-MCL)

Supplemental Table- 2: Summary of predominant somatic mutations in aggressive histology MCL and its subsets

Supplemental Table- 1: Univariate and multivariate analyses of factors predicting for failure free survival (FFS) after first line therapy for aggressive histology MCL (AH-MCL)

			$Univariate			#Multivariate		
	N	Events	HR	95% CI HR	P-value	HR	95% CI HR	P-value
Hemoglobin (g/dL)	160	116	0.91	0.84-0.99	**0.02**			
Ki-67%	145	103	1.01	1.00-1.02	**0.008**			
AH-MCL Category					**<0.001**			
***AH-DN**	102	66	1.00	1.00-1.00				
AH-t	61	52	2.20	1.52-3.17		4.36	2.00-9.52	**<0.001**
AH-MCL subcategory					**<0.001**			
Blastoid-DN	88	53	1.00	1.00-1.00	-			
Pleomorphic-DN	14	13	2.54	1.37-4.71	**0.003**			
AH-t	61	52	2.50	1.70-3.68	**<0.001**			
PS-ECOG, n (%)					**<0.001**			
***0**	37	30	1.00	1.00-1.00				
1	96	65	0.81	0.52-1.25		1.57	0.83-2.99	0.168
2	21	16	1.52	0.83-2.80		3.19	1.35-7.51	**0.008**
3	4	3	2.76	0.83-9.24		1.08	0.25-4.71	0.921
4	3	2	14.6	3.28-65.21		4.88	0.77-31.05	0.093
Histology Type					**0.019**			
***Blastoid**	135	95	1.00	1.00-1.00				
Pleomorphic	28	23	1.72	1.09-2.72		1.38	0.78-2.43	0.27
CNS involvement					**<0.001**			
***No**	155	111	1.00	1.00-1.00				
Yes	8	7	3.88	1.76-8.55		5.82	2.08-16.31	**<0.001**
LDH (>ULN)					**0.002**			
No	86	61	1.00	1.00-1.00				
Yes	71	54	1.80	1.23-2.63				
Age≥72 years					**0.006**			
***No**	127	90	1.00	1.00-1.00				
Yes	36	28	1.80	1.18-2.77		1.5	0.91-2.66	0.104
Ki-67≥50%					**0.004**			
***No**	31	22	1.00	1.00-1.00				
Yes	114	81	1.99	1.23-3.22		**1.69**	**1.01-2.82**	**0.04**
LDH≥1519 (IU/L)					**0.002**			
***No**	146	103	1.00	1.00-1.00				
Yes	12	12	2.51	1.37-4.61		**3.57**	**1.63-7.79**	0.001
β2M ≥4(mg/L)					0.08			
No	65	42	1.00	1.00-1.00				
Yes	37	25	1.54	0.93-2.54				
Platelet count≥63 ($\times 10^9$/L)					**0.02**			
***No**	13	12	1.00	1.00-1.00				
Yes	147	104	0.50	0.27-0.91		1.01	0.34-3.01	0.99

Response to first line treatment					**<0.001**			
No CR	72	53	1.00	1.00-1.00				
CR	90	64	0.31	0.21-0.45				
First line treatments					**0.005**			
***R-chemotherapy +/- SCT**	44	30	1.00	1.00-1.00				
R-HCVAD based +/- SCT	73	49	0.74	0.47-1.18		0.93	0.52-1.66	0.812
R-lenalidomide +/- SCT	9	8	1.70	0.77-3.75		1.12	0.37-3.32	0.844
Ibrutinib/BTKi +/- SCT	19	16	1.03	0.56-1.90		0.56	0.22-1.40	0.216
Miscellaneous	18	15	2.07	1.11-3.86		0.94	0.36-2.49	0.904

Abbreviations and footnotes: *PS-ECOG – performance status, eastern cooperative oncology group; AH-DN aggressive histology de novo (at the time of initial diagnosis); AH-t aggressive histology at the time of transformation (these patients had classic phase at initial diagnosis); LDH lactate dehydrogenase; WBC white blood cell count; CNS central nervous system; R-HCVAD - rituximab hyper fractionated cyclophosphamide, vincristine, adriamycin and dexamethasone alternating with high dose methotrexate, cytosine arabinoside; R-CHOP-rituximab, cyclophosphamide, adriamycin, vincristine and prednisolone; CR complete remission;* ****** *reference in multivariate analysis;***#***Variables with more than 25% of missing values were excluded from this analysis;***$** *factors not significant in univariate analysis are not shown – age at transformation, WBC count, LDH, b2microglobin, platelet count, absolute monocyte count, absolute lymphocyte count, gender type, B symptoms, leukemic phase, bone marrow involvement, SOX-11, light chain type.*

Supplemental Table- 2: Summary of predominant somatic mutations in aggressive histology MCL and its subsets

AH-MCL type	Predominant somatic gene mutations
AH-DN	*ATM, CCND1, NOTCH1, TP53, NOTCH2, FAT1, UBR5, NSD2, SMARCA4, RANBP2*
AH-t	*KMT2B, KMT2D, CACNA1A*
High (≥50%)	*CCND1, NOTCH1, TP53,SPEN, NOTCH2, FAT1,RANBP2, KMT2C, NTRK1, UBR5, NSD2, ARID1A, SMARCA4*
Low (<50%)	*CARD11*

Whole-exome sequencing (WES)

Briefly, indexed libraries were prepared from 500 ng of Biorupter Ultrasonicator (Diagenode, Denville, NJ, USA)-sheared, genomic DNA using the KAPA Hyper Library Preparation Kit (KAPABiosystems, Wilmington, MA, USA). The indexed libraries were prepared for capture with 6 cycles of preligation-mediated PCR amplification. Following amplification and reaction cleanup, the libraries were quantified fluorometrically using the Qubit™ dsDNA HS Assay (ThermoFisher, Waltham, MA, USA) and assessed for size distribution using the Fragment Analyzer (Advanced Analytical, Ames, IA, USA). Library concentrations were normalized, and the libraries were multiplexed 6 libraries/pool.

Each multiplexed library pool was hybridized to a probe pool from the SeqCap EZ Human Exome Enrichment Kit v3.0 (Roche-NimbleGen, Madison, WI, USA). The enriched libraries were amplified with 8 cycles of post-capture PCR, then assessed for exon target enrichment by qPCR. The exon-enriched libraries were then assessed for size distribution using the Fragment Analyzer (Advanced Analytical) and quantified by qPCR using the KAPA Library Quantification Kit (KAPABiosystems). Sequencing was performed on the HiSeq4000 Sequencer (Illumina, San Diego, CA, USA.), one capture (6 samples) per lane using the 76 bp paired-end configuration.

WES data processing and genotyping quality check

Raw output of the Illumina exome sequencing data was processed using Illumina's Consensus Assessment of Sequence And Variation (CASAVA) tool (v1.8.2) (http://support.illumina.com/sequencing/sequencing_software/casava.html) for demultiplexing and conversion to FASTQ format. The FASTQ files were aligned to the human reference genome

(hg19) using BWA (v0.7.5) [1] allowing up to 3 mismatches (2 mismatches must be in the first 40 seed regions) for a 76-base sequencing run. The aligned BAM files were then subjected to mark duplication, realignment and base recalibration using Picard (v1.112) and GATK (v3.1-1) software tools [2]. The generated BAM files were then used for downstream analysis. Genotyping quality check was performed to rule out any possible sample swapping or contamination. Briefly, germline SNPs were called using Platypus (v0.8.1) [3]. Samples from the same patient were confirmed/identified by the percentage of genotyping-identity between them, which was defined by the fraction of identical germline alleles among the overlapping SNPs between the two samples. All samples in this study passed quality check, and no sample swapping or contamination was detected.

Somatic mutation calling, filtering, and functional annotation

MuTect (v1.1.4) [4] was applied to identify somatic point mutations, and Pindel (v0.2.4) [5] was applied to identify small insertion and deletions (Indels). The MuTect and Pindel outputs were then run through our pipeline for filtering and annotation. Briefly, only MuTect calls marked as "KEEP" were selected and taken into the next step. For both substitutions and Indels, mutations with a low variant allelic fraction (VAF < 0.01) or had a low total read coverage (< 20 reads for tumor samples; <10 reads for germline sample), were removed. In addition, Indels that had an immediate repeat region within 25 base pairs downstream towards it 3' region were also removed. After that, common variants reported by the ExAc (the Exome Aggregation Consortium, http://exac.broadinstitute.org), Phase-3 1000 Genome Project (http://phase3browser.1000genomes.org/Homo_sapiens/Info/Index), or the NHLBI GO Exome Sequencing Project (ESP6500, http://evs.gs.washington.edu/EVS/) with a population minor allele

frequency greater than 0.5% were removed. The intronic mutations, mutations at 3' or 5' UTR or UTR flanking regions, silent mutations, small in-frame insertions and deletions were also removed.

To evaluate the probability of a missense mutation being functionally deleterious, dbNSFP (v3.0) [6] was applied to add prediction scores for all missense mutations from twelve commonly used functional prediction algorithms: Polyphen-2 [7], SIFT [8], MutationTaster [9], Mutation Assessor [10], LRT [11], FATHMM-MKL [12] and DANN [13], PROVEAN[14], WEST3 [15], CADD [16], GERP++ [17], MetaSVM and MetaLR [18]. A missense mutation that was called as "deleterious" by five or more algorithms were defined as a "deleterious" mutation.

DNA copy number analysis

DNA copy number analysis was conducted using an in-house application ExomeLyzer [19] followed by CBS segmentation as described previously [20]. The segmentation files were loaded to Nexus for visualization. R package was used to identify copy number gains (Log2 copy ratio ≥ 0.5) and losses (Log2 copy ratio ≤ -0.5) and the burdens of DNA copy number gains and losses were calculated using approach as previously described (Roh W, et al. Science Translational Medicine. 2017. PMID: 28251903).

References -

1. Li H and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754-1760. DOI: 10.1093/bioinformatics/btp324.
2. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43: 491-498. DOI: 10.1038/ng.806.
3. Rimmer A, Phan H, Mathieson I, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet* 2014; 46: 912-918. DOI: 10.1038/ng.3036.
4. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; 31: 213-219. DOI: 10.1038/nbt.2514.

5. Ye K, Schulz MH, Long Q, et al. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009; 25: 2865-2871. DOI: 10.1093/bioinformatics/btp394.
6. Liu X, Wu C, Li C, et al. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* 2016; 37: 235-241. DOI: 10.1002/humu.22932.
7. Adzhubei I, Jordan DM and Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; Chapter 7: Unit7 20. DOI: 10.1002/0471142905.hg0720s76.
8. Kumar P, Henikoff S and Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; 4: 1073-1081. DOI: 10.1038/nprot.2009.86.
9. Schwarz JM, Cooper DN, Schuelke M, et al. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014; 11: 361-362. DOI: 10.1038/nmeth.2890.
10. Reva B, Antipin Y and Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011; 39: e118. DOI: 10.1093/nar/gkr407.
11. Chun S and Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res* 2009; 19: 1553-1561. DOI: 10.1101/gr.092619.109.
12. Shihab HA, Rogers MF, Gough J, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* 2015; 31: 1536-1543. DOI: 10.1093/bioinformatics/btv009.
13. Quang D, Chen Y and Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 2015; 31: 761-763. DOI: 10.1093/bioinformatics/btu703.
14. Choi Y, Sims GE, Murphy S, et al. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012; 7: e46688. DOI: 10.1371/journal.pone.0046688.
15. Carter H, Douville C, Stenson PD, et al. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* 2013; 14 Suppl 3: S3. DOI: 10.1186/1471-2164-14-S3-S3.
16. Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; 46: 310-315. DOI: 10.1038/ng.2892.
17. Davydov EV, Goode DL, Sirota M, et al. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 2010; 6: e1001025. DOI: 10.1371/journal.pcbi.1001025.
18. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* 2015; 24: 2125-2137. DOI: 10.1093/hmg/ddu733.
19. Zhang J, Fujimoto J, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014; 346: 256-259. DOI: 10.1126/science.1256930.
20. Olshen AB, Venkatraman ES, Lucito R, et al. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004; 5: 557-572. DOI: 10.1093/biostatistics/kxh008.