SUPPLEMENTARY APPENDIX

IL6R-STAT3-ADAR1 (P150) interplay promotes oncogenicity in multiple myeloma with 1q21 amplification

Phaik Ju Teoh,[1,2] Tae-Hoon Chung,[1] Pamela Y.Z. Chng,[1] Sabrina H. M. Toh[1] and Wee Joo Chng[1,2,3]

[1]Cancer Science Institute of Singapore, National University of Singapore; [2]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and [3]Department of Haematology-Oncology, National University Cancer Institute of Singapore, National University Health System, Singapore

 doi:10.3324/haematol.2019.221176

Received: March 6, 2019.

Accepted: August 12, 2019.

Pre-published: August 14, 2019.

Correspondence: *WEE JOO CHNG* - csicwj@nus.edu.sg

Supplementary Materials and Methods

Patient samples, human MM cell lines (HMCLs) and reagents

Plasma cells from the patients were isolated through Ficoll-Hypaque centrifugation system and were purified with CD138-immunomagnetic beads (Stemcell Technologies). These primary cells are cultured in DMEM medium supplemented with 20% FBS, 1% Penicillin/Streptomycin, 1% L-glutamine, 100ng/mL IGF-1, 50ng/mL BAFF, 55uM beta-mecarptoethanol, 10ng/mL IL6 (by default, but adjusted according to experimental set up). All HMCLs have been tested negative for mycoplasma contamination and were cultured in RPMI-1640, supplemented with 10% FBS and 1% Penicillin/Streptomycin. Cells were passaged every 2 days to keep them at optimum condition. XG6 and XG7, the IL6-dependent cell lines were consistently cultured with 3ng/mL IL6. List of antibodies used of Western blot and immunofluoresence applications is found in Table S2.

GEP and aCGH data processing

We analyzed 9 publicly available datasets altogether. (Table S1) MMRF dataset is composed of GEP (GSE26760; 304 samples) and aCGH (GSE26849; 254 samples) data of newly diagnosed or relapsed MM patients, compiled as a resource for the MM research community[1]. HMCL dataset is composed of GEP (45 samples) and aCGH (46 samples) of representative MM cell lines[2]. UAMS (GSE2659; 559 samples) is a GEP dataset of newly diagnosed MM patients assayed by researchers in University of Arkansas Medical School[3]. APEX (GSE9782; 264 samples) is a GEP dataset of relapsed MM patients participating in phase 2 and phase 3 APEX/SUMMIT/CREST clinical trial of bortezomib[4]. HOVON (GSE19784; 320 samples) is a GEP dataset of newly diagnosed MM patients participating in HOVON-65/GMMG-HD4 clinical trial, a large multicentre, prospective, randomized phase 3 trial[5]. Mayo (GSE6477; 162 samples) is a GEP dataset encompassing normal, MGUS,

SMM, newly diagnosed and relapsed MM patients assayed by researchers in Mayo clinic[6]. Italy (GSE13591; 158 samples) is a GEP dataset encompassing normal, MGUS, MM and PCL patients assayed by Italian researchers[7]. DFCI dataset is composed of GEP (GSE4452; 65 samples) and aCGH (66 samples) of newly diagnosed patients before total therapy analyzed by the Dana-Farber Cancer Institute researchers[8]. CoMMpass (811 samples) is an RNA-seq dataset of patients participating in the ongoing Relating Clinical Outcomes in MM to Personal Assessment of Genetic Profile (CoMMpass) study, the flagship of MMRF's Personalized Medicine Initiative[19]. Of all gene expression data, 8 (MMRF, HMCL, UAMS, APEX, HOVON, Mayo, Italy, DFCI) of which were of microarray type and only one (CoMMpass) was of RNA-seq type. In addition, 3 datasets (MMRF, HMCL, DFCI) contained both aCGH and gene expression data, allowing us to connect copy number aberrations to gene expression changes for targeted regions.

All data analysis was performed using the R/Bioconductor system. GEP data were either downloaded from GEO or obtained from the source indicated in Table S1 by themselves and no further processing was applied. For aCGH data, we performed segmentation of log-ratio values according to probes' chromosomal positions by using the circular binary segmentation (CBS) algorithm[9,10] implemented in the "DNAcopy" Bioconductor package, and determined each segment's copy number aberration status as explained in the Supplementary Information of our previous study in detail[11], which can be recapitulated briefly as follows: following segmentation, we analyzed the distribution of segments' average log-ratio values, identified the crests and troughs particularly near log-ratio=0 (WT), and used the troughs to define each segment's copy number status. Although the log-ratio values should in principle be enough to delineate various copy number statuses, it is not usually the case due to noise from enormous number of probes masking true values. This complexity can be greatly reduced by the segmentation procedure and cutoff values between different copy number statuses

appear naturally at least for low copy number aberrations. For more details, please be advised to confer Teoh PJ *et al.*[11]. IL6R and ADAR1 copy number status was finally identified as the copy number of the segment harboring the indicated genes.

For Affymetrix platform datasets, signal intensities for probesets were transformed to represent expression values for genes by first gathering all intensities for probesets corresponding to specific genes, then finding probesets whose average signal intensities were larger than half of maximum average, and finally averaging intensities of chosen probesets for each sample. This transformation was inspired by the fact that even probesets of a specific gene displayed drastically divergent intensity values with tens of folds' difference among them were not rare and intensities values too small only magnified the contribution from non-specific sources.

STAT3 signature development

We gathered putative members of STAT3 signature by collecting 130 genes that were reported to be members of STAT3 signaling from diverse sources such as MSigDB, GO and literature. Despite such assertions, the relevance of the claims of STAT3 association for these genes may not be translated into myeloma. So we went through another round of selection. We estimated the correlation between expression profiles of putative member genes with STAT3 expression profile using MMRC dataset and selected those whose correlation with STAT3 expression profile was positive with $p < 0.01$ to make all subsequent index estimation straightforward, which resulted in the following 21 genes: *STAT3, JAK1, JAK2, TYK2, IL6, IL6R, OSM, SOCS3, SOCS6, MCL1, BCL6, VEGFA, NFKB1, MAPK1, HCK, SUMO3, FAS, ITGB1, PTP4A3, TJP1, NCOR2, YWHAZ*.

STAT3 signature index estimation

We first divided each gene's expression values with its global median in a dataset (gene-by-gene median-normalization) and performed log2-transformation to the

expression profile. Then, the STAT3 signature index of a sample was estimated as the median of all member genes' normalized log2-transformed expression values in that sample.

Chromatin immunoprecipitation (ChIP) Assays

Cells were cross-linked with formaldehyde, harvested and lysed in lysis buffer. STAT3 protein was captured with anti-STAT3 and negative control IgG antibody (5ug per reaction) and the binding of the protein on P110, P150 and IL6R promoter regions were quantified with qPCR.

Luciferase Assay

Cells were transfected with Luciferase plasmid encoding for specific P110 and P150 promoters (Active Motif). At 24 hours' post-infection, the cells were stimulated with IL6 (10ng/mL) or co-transfected with STAT3-overexpressing plasmids (pIRES-STAT3-CA and pIRES-STAT3-DN). 24 hours later, the cells were harvested and analyzed with the Dual Luciferase system (Promega). Luciferase signals were quantified with Tecan plate reader (Mannedorf, Switzerland).

Co-immunoprecipitation Assay (Co-IP)

STAT3 or ADAR1-transfected-293T cells or IL6-induced-H929 were harvested and lysed with RIPA buffer, a gentle lysis buffer that helps to retain protein-protein interaction. Magnetic beads (Dynabeads from Invitrogen, California, USA) were incubated with either of these antibodies: STAT3, P150, Flag and IgG (negative control), for 2 hours at room temperature, to generate antibody-beads complex. Total cellular protein was then incubated with either of these antibody complexes overnight at 4C. The protein complexes were boiled to release the proteins and potential interaction between these proteins was investigated by Western blot analysis.

In vitro functional assays:

(a) Cell viability and growth assay

MM cell viability and growth rate was assessed by CTG (Cell Titre Glo, Promega) luminescence assay. Cells were seeded and cultured in 96-well plates in 100uL complete medium at a density of 30 x 10^4 cells/100uL/ well. The cells were incubated either for different time points or with different concentration of drugs. Then the CTG reagent was added to the cells in a 1:1 volume ratio and the mixture was further incubated for 20 minutes. Finally, the luminescence signals were read with Tecan microplate reader. Each or condition or set of cells was seeded in triplicates, and each experiment was conducted at least twice. Data presented are mean ± standard deviation of biological replicates.

(b) Colony formation assay

Colony formation assay was done to determine the rate of MM cells forming colonies as a representation of proliferation of MM cells. The cells were seeded in the methylcellulose-based medium (StemCell Technology) containing important growth factors for the optimum growth of MM cells to a final concentration of 7000 cells/ml. They were left to grow in a 24 well plate for 7-14 days depending on the colony forming capacity of the cells, after which the colonies were visualized and counted under the microscope. Each experiment was done in triplicates and repeated twice. Data presented are mean ± standard deviation of biological replicates. Images reported here are representative of technically and biologically-replicated experiments.

(c) Cell cycle analysis

Indicated cell lines were serum starved for 16 hours to synchronise them into G0 cell cycle stage. The cells were then allowed to recover into the progressive cell cycle phases by supplementing them with 10% FBS for at least 24 hours. The cells were harvested and fixed in 70% ethanol overnight, followed by the staining with propidium

iodide (PI). Their cell cycle profiles were then read on a flow cytometer (BD LSRII). Subsequently, data analysis was done with Flowjo software (Ashland, Oregan, USA).

(d) Annexin-V-FITC apoptosis assay

U266 and H929 cells were infected with lentivirus particles containing shCtr, shP150 #1 and shP150 #2 and the cells were harvested at 48 hours for the assay. Cells were stained with annexin-V-FITC dye, with PI serving as the exclusion dye, for 30 minutes at room temperature, in dark. The percentage of cells undergoing apoptosis was analysed with flow cytometer (BD LSRII) and Flowjo software.

Immunofluorescence

Cells were cytospinned to generate slides with monolayer cells. The slides were fixed with 4% formaldehyde at room temperature for 15 minutes, washed with PBS and permeabilized with 0.2% Triton-X for 10 minutes. They were then blocked with 1% BSA for 30 minutes and stained with respective primary antibodies overnight. The slides were washed with PBS and were incubated with fluorescent-tagged secondary antibodies for 1 hour at room temperature, after which they were counterstained with DAPI for 1 minute. The slides were visualized under the microscope iXLR7 for the localization signal. Merging of images and protein co-localisation analysis was done with Image J. Images reported here are representative of technically and biologically-replicated experiments.

RNA-sequencing (RNA-seq) and RNA editing analysis

Whole transcriptome sequencing was done on the Illumina Hi-Seq-4000 platform, with a sequencing depth of 100 million reads per sample, 100bp paired-end, conditions that are sufficient for the detection of A-to-I editings [12-14]. A bioinformatics pipeline adapted from a previously published method was used to identify RNA editing events[15]. For each sample, raw reads were mapped to the reference human genome (hg19) and a

splicing junction database generated from transcript annotations derived from UCSC, Refseq, Ensembl and Gencode (v19) by using Burrows-Wheeler Aligner with default parameters (bwa mem algorithm, v0.7.15-r1140)[13]. To retain high quality data, PCR duplicates were removed (samtools rmdup function, v1.4.1)[14] and the reads with mapping quality score less than 20 were discarded. Junction-mapped reads were then converted back to the genomic-based coordinates. An in-house perl script was utilized to call the variants from samtools pileup data and the sites with at least 2 supporting reads were retained. The candidate events were filtered by removing the SNPs reported in different cohorts (1000 Genomes Project[15], NHLBI GO Exome Sequencing Project (http://evs.gs.washington.edu/EVS/), dbSNP v138)[16] and excluding the sites within the first 6 bases of the reads caused by imperfect priming of random hexamer during cDNA synthesis. For the sites not located in Alu elements, the candidates within the 4 bases of a splice junction on the intronic side, and those residing in the homopolymeric regions and in the simple repeats were all removed. Candidate variants located in the reads that map to the non-unique regions of the genome by using BLAST-like alignment tool (BLAT)[17] were also excluded. Finally, only A-to-G editing sites based on the strand information were considered. To identify high confidence editing events for all the downstream analyses, the candidate sites were required to be supported by at least 20 reads and having more than 10% absolute editing frequency.

Table S1. List of publicly available patient datasets used for in silico analyses

Name	Type	Platform	N	Note
MMRC	aCGH	Agilent Human 244K	254	GSE26849 [1]
	GEP	Affymetrix HG-U133 Plus 2	304	GSE26760 [1]
HMCL	aCGH	Agilent Human 44K	46	Supplementary Data.[2]
	GEP	Affymetrix HG-U133 Plus 2	45	Supplementary Data.[2]
UAMS	GEP	Affymetrix HG-U133 Plus 2	559	GSE2658 [3]
APEX	GEP	Affymetrix HG-U133 A/B	264	GSE9782 [4]
HOVON	GEP	Affymetrix HG-U133 Plus 2	320	GSE19784 [5]
MAYO	GEP	Affymetrix HG-U133A	162	GSE6477 [6]
Italy	GEP	Affymetrix HG-U133A	158	GSE13591[7]
DFCI	aCGH	Agilent Human 22K	66	Supplementary Data[8]
	GEP	Affymetrix HG-U133 Plus 2	65	GSE4452[18]
CoMMpass	RNA-seq		811	MMRF Researcher Gateway[19]

Table S2. List of Antibodies used for Western blot and Immunofluoresence

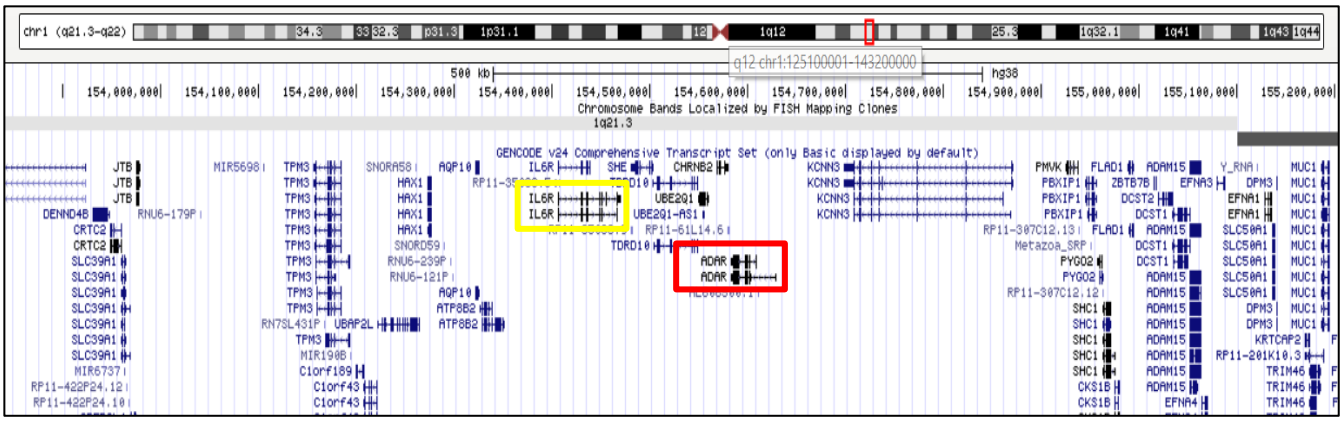
Antibody	Vendor	Catalog number
STAT3	Santa Cruz	sc-8019
Phospho-STAT3 (Y705)	Cell Signaling Technology	CST-9145
MCL1	Santa Cruz	SC-12756
JAK2	Santa Cruz	Sc-390539
BCL2	Santa Cruz	SC-509
Total ADAR1	Abcam	Ab168809
ADAR1-P150	Abcam	Ab-126745
IL6R	Santa Cruz	SC-661
GAPDH	Santa Cruz	SC-47724
Kappa	Vectorlab	CI-3060
Lambda	Vectorlab	FI-3070

Table S3. shRNA sequences that against P150 and IL6R used for the experiements. The sequences were cloned into pLKO.1 vector.

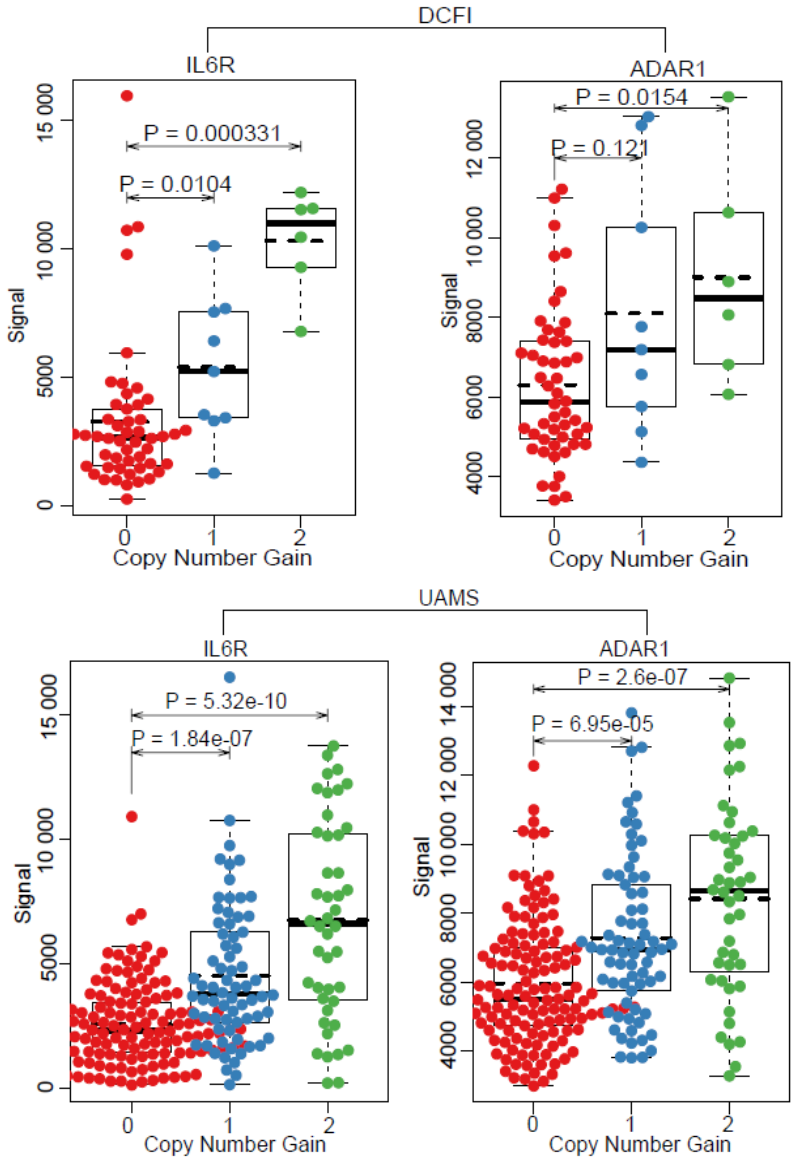
Target gene	shRNA sequences
shP150 #1	CCGGAGTTTCCTGCTTAAGCAAATACTCGAGTATTTGCTTAAGC AGGAAACTTTTTTG
shP150 #2	CCGGTGATTGCCTTTCCTCACATTTCTCGAGAAATGTGAGGAA AGGCAATCATTTTTG
shIL6R	CCGGTATCGGGCTGAACGGTCAAAGCTCGAGCTTTGACCGTTC AGCCCGATATTTTTG

Figure S1

(A)


chr1 (q21.3-q22)
q12 chr1:125100001-143200000
hg38
Chromosome Bands Localized by FISH Mapping Clones
1q21.3
GENCODE v24 Comprehensive Transcript Set (only Basic displayed by default)
IL6R
UBE2Q1
UBE2Q1-AS1
ADAR
ADAR1

(B)


DCFI
IL6R
P = 0.000331
P = 0.0104
ADAR1
P = 0.0154
P = 0.121
Signal
Copy Number Gain
UAMS
IL6R
P = 5.32e-10
P = 1.84e-07
ADAR1
P = 2.6e-07
P = 6.95e-05
Signal
Copy Number Gain

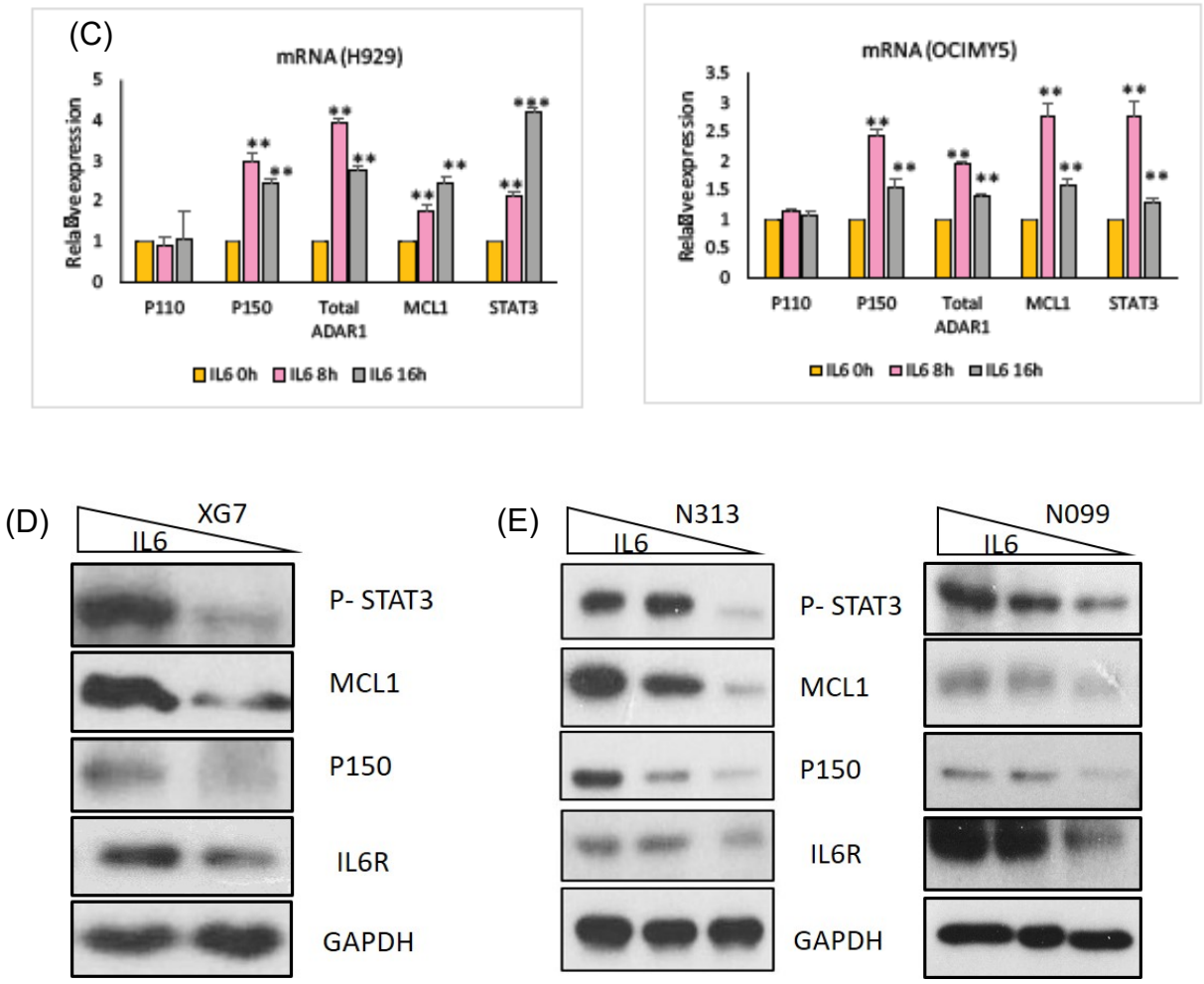
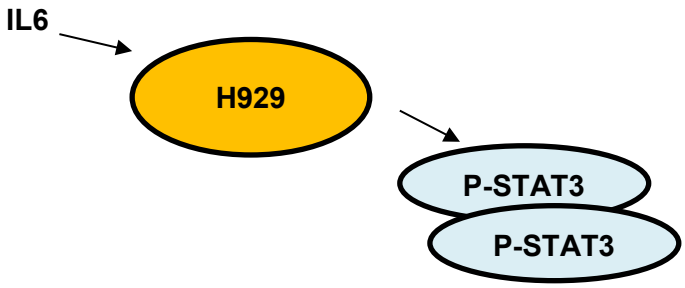


Figure S1. (A) Screenshot of the 1q21 portion of Chromosome 1 from the UCSC genome browser. IL6R and ADAR1 are located at close proximity in the genome. Yellow box- IL6R, Red box—ADAR1. **(B)** Gene expression level of IL6R and ADAR1 in DCFI and UAMS patient datasets according to 1q21 status. 0- no copy number gain (WT), 1- one copy gain, 2- two or more than two copies gain. **(C)** RT-qPCR analysis of IL6/STAT3 pathway factors and ADAR1 mRNA expression in cells with 1q21(amp) and 1q21(WT) upon IL6 stimulation (10ng/mL) at different time points. **(D**) XG7 (IL6-dependent cell line) was cultured in IL6-enriched (3ng/mL) and IL6-deprieved (2ng/mL) medium for 24 hours and cells were harvested for Western blot analysis. **(E)** CD138+ plasma cells retrieved from two newly diagnosed patients (N313 and N099), was cultured in vitro in the presence of 10ng/m/, 5ng/ml and 2.5ng/ml for 24 hours and cells were harvested for Western blot analysis.

Figure S2

(A)



Protein Pulldown: (a) Anti-STAT3 antibody
(b) IgG negative control antibody

(B)

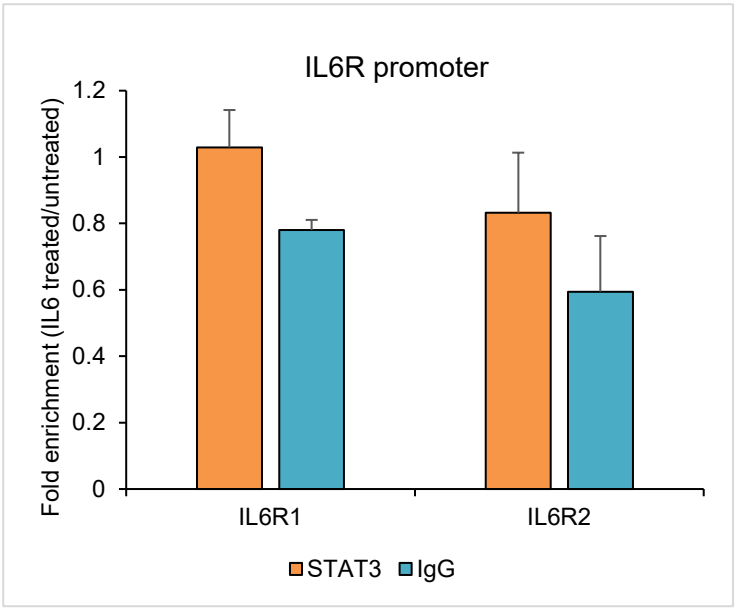


Figure S2. (A) Schematic diagram depicting H929 being treated with IL6 for 6 hours to stimulate the accumulation of phospho-STAT3 protein. These proteins were pulled down with specific anti-STAT3 or IgG negative control antibodies for ChiP-qPCR analysis. **(B)** Relative STAT3 enrichment (IL6 treated/untreated) on IL6R promoters. IL6R1 and IL6R2 are primers encompassing 2 independent regions of the IL6R promoter.

Figure S3

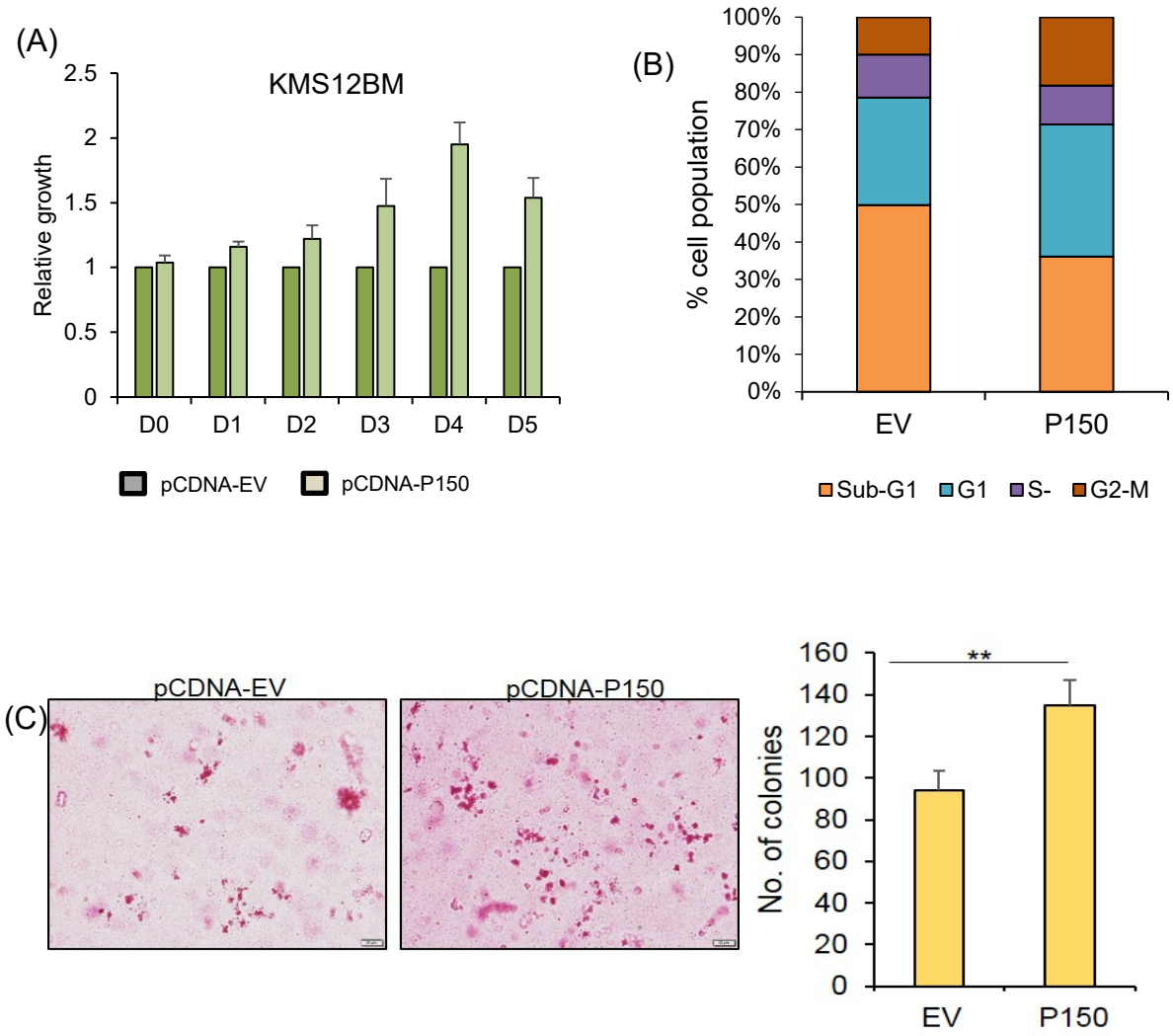


Figure S3. (A) Daily cell growth of transfected KMS12BM cells was monitored with CTG assay. Relative growth rate was represented as the growth of P150-overexpressed cells (pCDNA-P150) against control cells (pCDNA-EV). **(B)** Cell cycle analysis of control and P150-overexpressed-KMS12BM. There was an increase in S-phase and G2/M phases cell population, indicating the cells were in a more proliferative state. **(C)** Colony formation assay detects a significantly higher number of colonies formed in P150-overexpressed-KMS12BM as compared to its control empty vector cells. Left: Representative photographs of the soft agar. Darkened pink dots are the formed colonies. Right: Quantitative values of the number of colonies formed. **$p<0.05$.

Figure S4

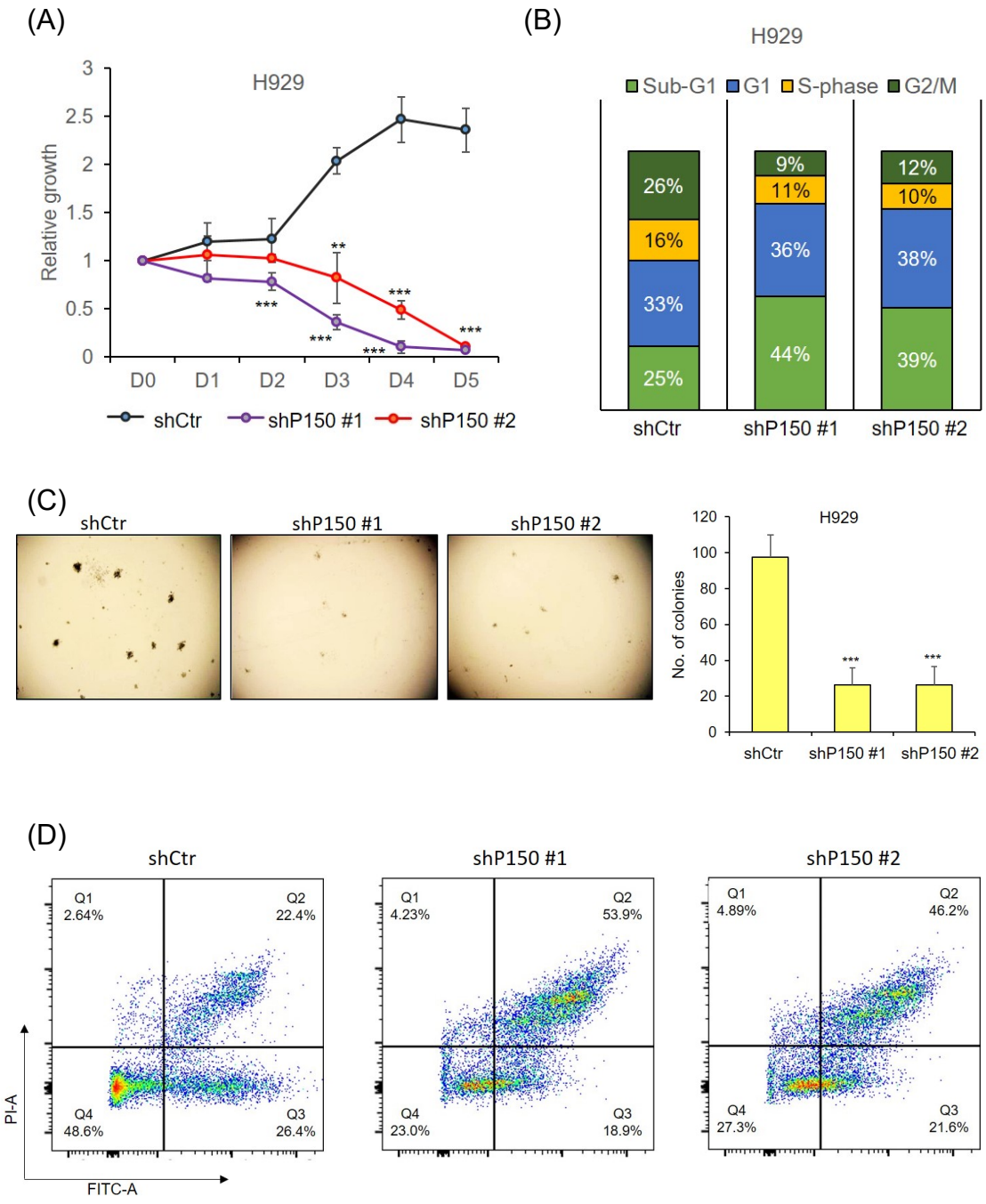
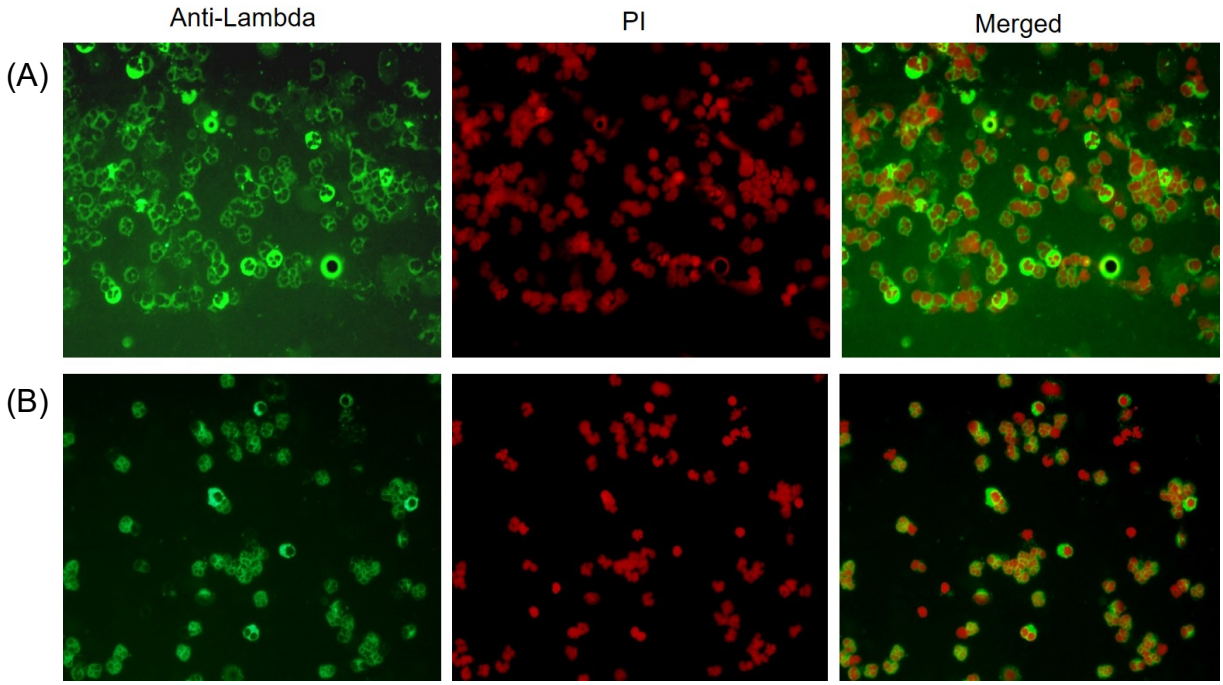


Figure S4. (A) Daily cell growth of infected H929 cells was monitored with CTG assay over a course of 5 days. Growth of shCtr cells was sustainable up to day 4 and dropped slightly on day 5 whereas the growth of shP150 cells (both shRNAs) was largely compromised. The cells could hardly grow on day 1 and 2 and its viability dropped from day 3 onwards. **(B)** Cell cycle analysis of control and P150-knockdown-H929.

There was a reduction in S-phase and G2/M phases cell population, indicating the cells were cycling in slower manner. **(C)** Colony formation assay detects a significantly lesser number of colonies formed in H929-shP150 cells as compared to its shCtr cells. *Left:* Representative photographs of the soft agar. Dark coloured dots are the formed colonies. *Right:* Quantitative values of the number of colonies formed. ***$p<0.001$. **(D)** Annexin-V-FITC apoptosis assay showed increased amount of annexin-V positive cells (Q2 and Q3) upon 48 hours of P150 knockdown.

Figure S5



Immunofluorescence staining of patient samples to check for the purity of CD138+ plasma cells. Cells were cytospinned onto a clear glass slide, and were processed according to standard IF protocol. They were incubated with FITC-tagged-anti-kappa and anti-lambda antibody for the specific detection of these immunoglobulin light chains, and were counterstained with propidium iodide (PI). Up to 95% of the cells showed high expression of the cytoplasmic immunoglobulin proteins, indicating high purity of CD138 positivity. (A) Patient N291, (B) Patient N292.

Figure S6

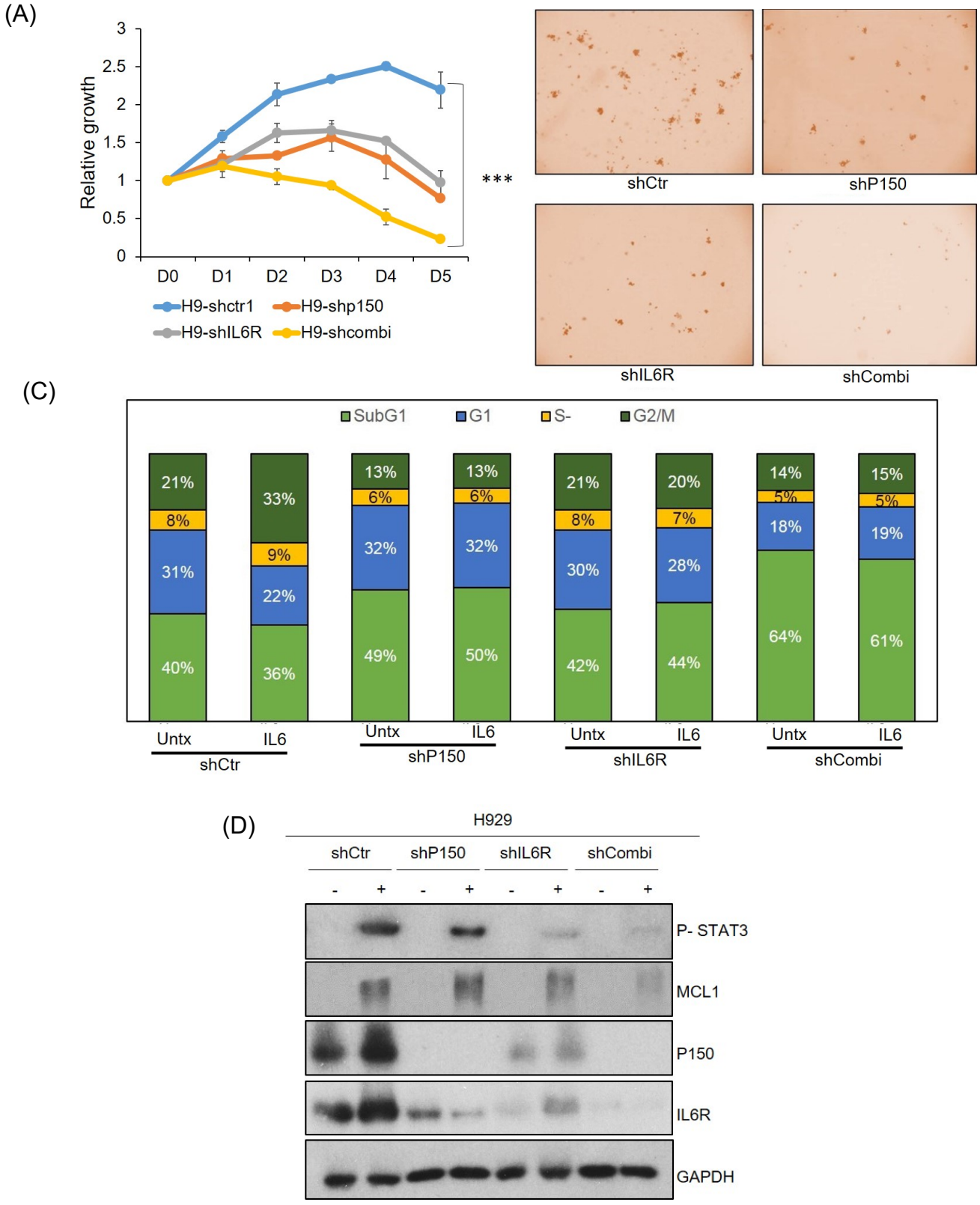


Figure S6. (A) Daily cell growth of H929 cells was monitored with CTG assay after infected with either shCtrr, shP150, shIL6R or combination of the latter 2 plasmids.

Relative growth rate of shCtr and shP150 cells from day 1 to day 5 was calculated by normalising their luminescence values of D1-D5 against D0. Growth of cells with shCombi was highly compromised as compared to the shCtr cells and those that have single loss of either P150 or IL6R. *** $p<0.001$ **(B)** Colony formation assay detects a significantly lesser number of colonies formed in H929-shCombi cells as compared to the shCtr, shP150 and shIL6R cells. **(C)** H929 cells with manipulated levels of P150 and IL6R were stimulated with IL6 for 24 hours and the cells were fixed and stained with propidium iodide for cell cycle analysis. **(D)** At 48 hours' post-lentivirus infection of the respective shRNAs, H929 cells were treated with IL6 (10ng/mL) for 8 hours and STAT3 pathway protein expression profile was checked with Western blot.

Figure S7

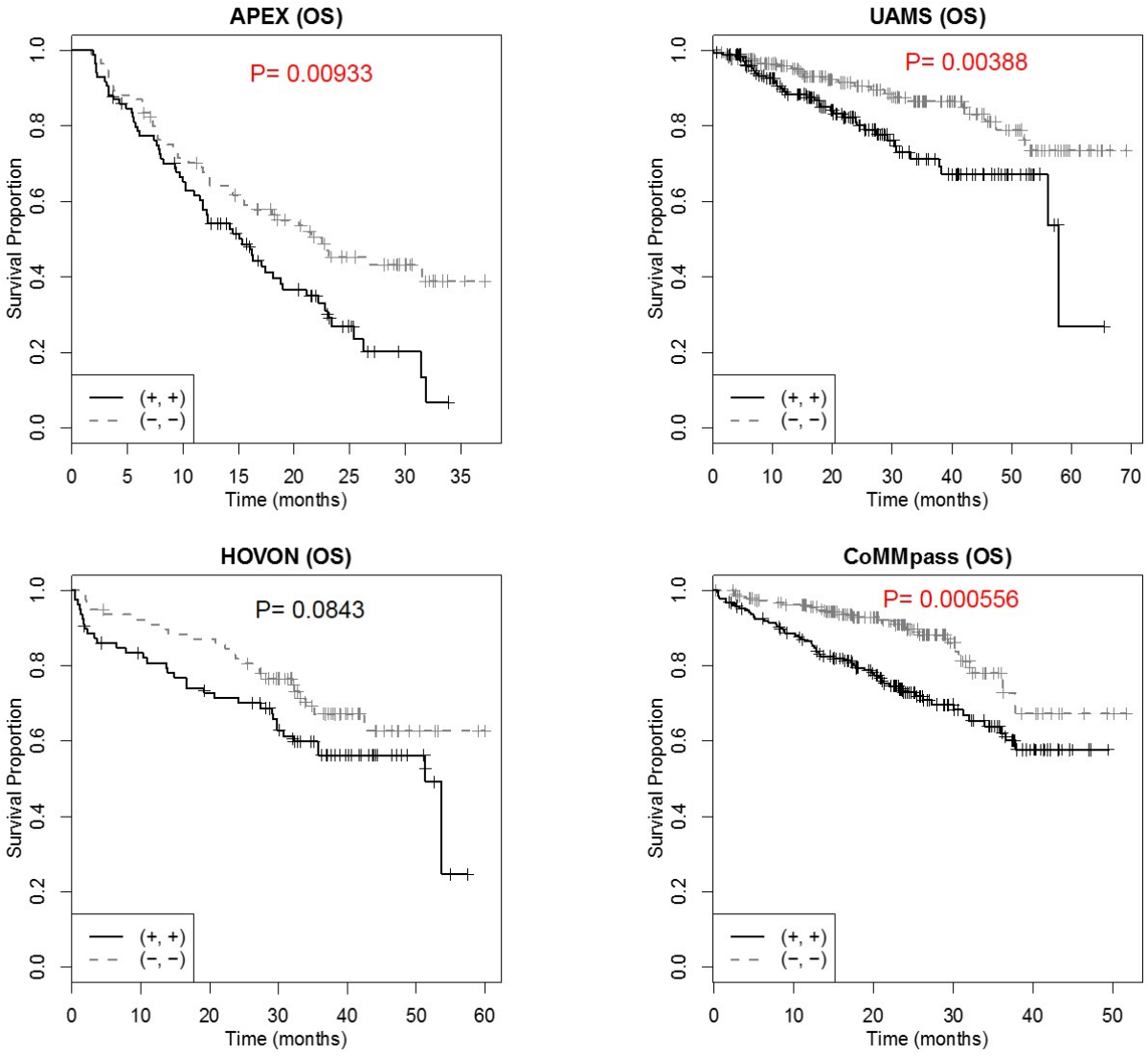
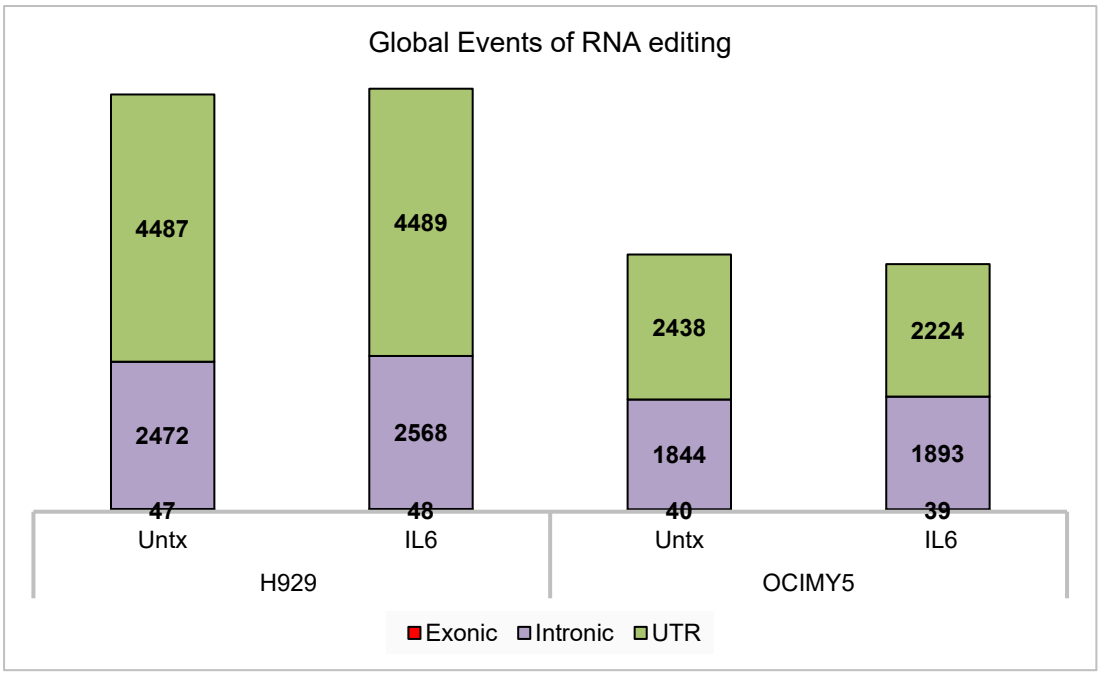


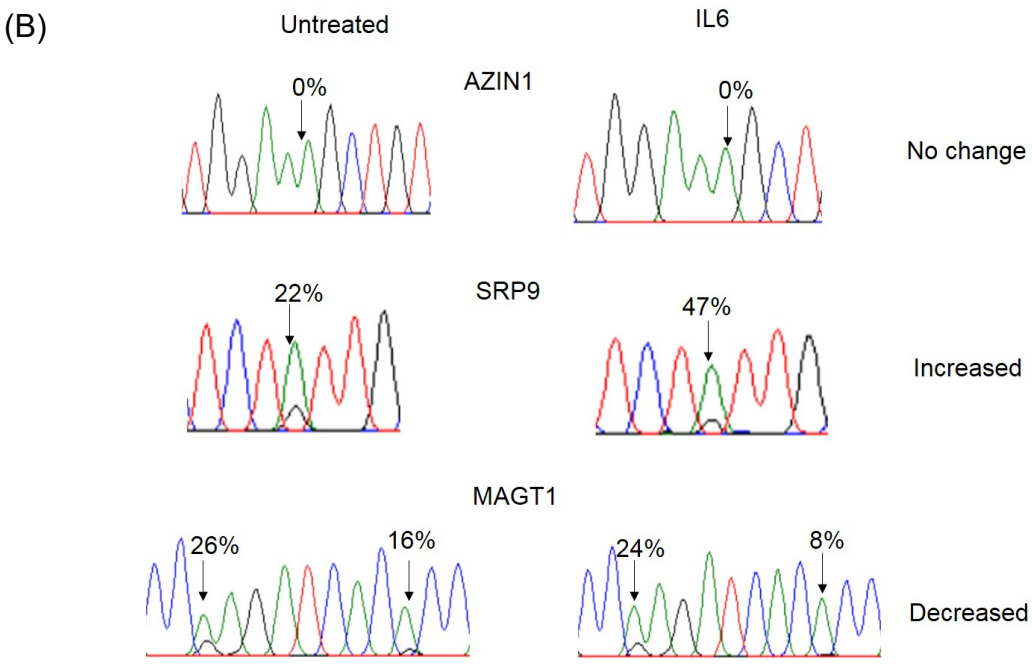
Figure S7. Prognostic difference between high ADAR1 + high IL6R and low ADAR1 + low IL6R patient groups. Significant difference was observed for OS in diverse datasets analyzed in this study. (+,+) represents high expression of both ADAR1 and IL6R and (-,-) represents low expression of both genes.

Figure S8

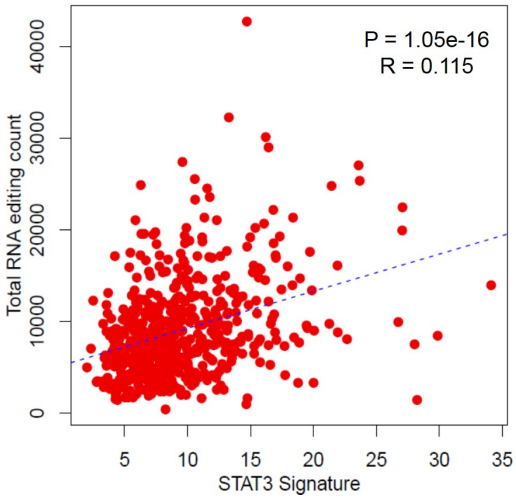
(A)


Global Events of RNA editing
4487
2472
47
4489
2568
48
2438
1844
40
2224
1893
39
Untx
IL6
Untx
IL6
H929
OCIMY5
Exonic
Intronic
UTR

(B)


Untreated
IL6
AZIN1
0%
0%
No change
SRP9
22%
47%
Increased
MAGT1
26%
16%
24%
8%
Decreased

(C)



(D)

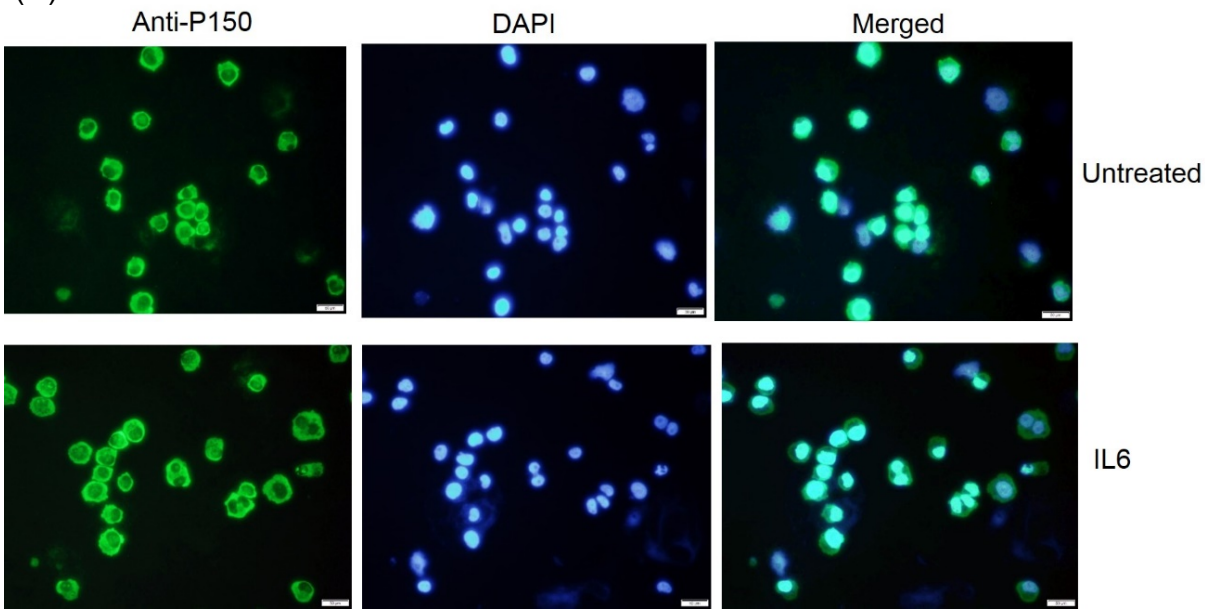


Figure S8. (A) The number of A-to-I editing events at the global level according to the genomic region detected by whole transcriptome sequencing (Hi-seq 4000, Illumina). The numbers indicated on the bar are the number of editing events within the particular region. Exonic region accounts for the least number of events, thus, the red bar was hardly visible. H929 and OCIMY5 were treated with IL6 for 24 hours before they were sent for RNA-sequencing. **(B)** Sanger Sequencing of known ADAR1 targets. Percentage indicated on the picture represents the editing frequency of the known editing sites of the gene. Green peak reprewsents A nucleotide, black peak represents

G(I) nucleotide. Editing frequency was calculated with the previously reported formula.(16) **(C)** Correlation analysis of total RNA editing and STAT3 signature in CoMMpass patients. **(D)** Immunofluorescence analysis of untreated and IL6-treated-H929 for the localisation of P150 protein. P150 was found predominantly at cytoplasm, and did not change much even upon IL6 stimulation.

References:

1. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011;471(7339):467-472.
2. Keats JJ, Fonseca R, Chesi M, et al. Promiscuous mutations activate the noncanonical NF-kappaB pathway in multiple myeloma. *Cancer Cell*. 2007;12(2):131-144.
3. Zhan F, Huang Y, Colla S, et al. The molecular classification of multiple myeloma. *Blood*. 2006;108(6):2020-2028.
4. Mulligan G, Mitsiades C, Bryant B, et al. Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. *Blood*. 2007;109(8):3177-3188.
5. Broyl A, Hose D, Lokhorst H, et al. Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. *Blood*. 2010;116(14):2543-2553.
6. Chng WJ, Kumar S, Vanwier S, et al. Molecular dissection of hyperdiploid multiple myeloma by gene expression profiling. *Cancer Res*. 2007;67(7):2982-2989.
7. Agnelli L, Mosca L, Fabris S, et al. A SNP microarray and FISH-based procedure to detect allelic imbalances in multiple myeloma: an integrated genomics approach reveals a wide gene dosage effect. *Genes Chromosomes Cancer*. 2009;48(7):603-614.
8. Carrasco DR, Tonon G, Huang Y, et al. High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. *Cancer Cell*. 2006;9(4):313-325.
9. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004;5(4):557-572.
10. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. 2007;23(6):657-663.
11. Teoh PJ, An O, Chung TH, et al. Aberrant hyperediting of the myeloma transcriptome by ADAR1 confers oncogenicity and is a marker of poor prognosis. *Blood*. 2018;132(12):1304-1317.
12. Ramaswami G, Zhang R, Piskol R, et al. Identifying RNA editing sites using RNA sequencing data alone. *Nat Methods*. 2013;10(2):128-132.
13. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595.
14. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
15. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
16. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-311.
17. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res*. 2002;12(4):656-664.
18. Fonseca R, Bergsagel PL, Drach J, et al. International Myeloma Working Group molecular classification of multiple myeloma: spotlight review. *Leukemia*. 2009;23(12):2210-2221.

19. https://research.themmrf.org