**DDX3X loss is Associated with Aggressive Phenotypes in Non-Hodgkin's Lymphomas**

Atish Kizhakeyil, Nurmahirah binte Mohammed Zaini, Zhi Sheng Poh, Brandon Han Siang Wong, Xinpeng Loh, Aik Seng Ng, Zun Siong Low, Praseetha Prasannan, Chun Gong, Michelle Guet Khim Tan, Chandramouli Nagarajan, Dachuan Huang, Pang Wan Lu, Jing Quan Lim, Sharon Barrans, Choon-Kiat Ong, Soon-Thye Lim, Wee-Joo Chng, George Follows, Daniel Hodson, Ming-Qing Du, Yeow Tee Goh, Suat Hoon Tan, Nicholas Francis Grigoropoulos, Navin Kumar Verma

**Supplementary Methods**

---

## Whole exome sequencing (WES) and analysis

Tissue sections (10 μm) prepared from formalin-fixed paraffin-embedded tissue blocks on glass slides were dewaxed using a standard xyelene/ethanol series. Tissue samples were crudely dissected using a scalpel and DNA was extracted using the QIAamp DNA micro kit (QIAGEN) according to the manufacturer's instructions. DNA concentration was measured using Qubit (Life Technologies) and 2 μg DNA from each sample was used for WES.

Exome libraries were created from template DNA and sequenced at the Wellcome Trust Sanger Institute. DNA was sheared to 100-400 bp on an ultra-sonicator (Covaris) and purified with the QIAquick system (QIAGEN). Fragmented DNA was end-repaired and re-purified. This was followed by addition of an A-tail and Minelute purification (QIAGEN). Adapters were ligated to the DNA fragments with Quick T4 DNA ligase (Agilent) and purified with Ampure magnetic beads (Beckman Coulter). DNA was quantified on a BioAnalyser (Agilent) and specimens with less than 500 ng of libraried sample underwent a further 6 PCR cycles. Libraries were hybridised to biotynilated 120mer cRNA baits designed for exome capture (SureSelect, Agilent). Hybridised DNA was captured using streptavidin magnetic beads, purified, and sequenced on the HiSeq platform (Illumina).

Reads were aligned to the human genome and unmapped reads were dropped from analysis. Taking into account the read position, orientation, and base quality, a naïve Bayesian algorithm (CaVEMan) was used to estimate the posterior likelihood of each possible nucleotide at every position. The Pindel algorithm was used to analyse insertions and deletions[1]. Where matched normal tissue was available, genomic variants were excluded. Where a library was

successfully created for a tumour sample but not its matched normal, a normal sample from a different patient was used instead. SNVs were cross-referenced with the dbSNP database but not filtered on this basis. In order to enrich the dataset for high-confidence variants, a called variant was excluded if any of the following filters applied[2]: *i*) Less than 1/3 mutant alleles were ≥25 base quality, *ii*) Coverage was less than 8 and no mutant alleles were found in the first 2/3 of a read, *iii*) More than 3% of mutant alleles that were ≥15 base quality were also found in the matched normal, *iv*) At least 2% of mutant alleles of base quality ≥20 were found in at least 2 unmatched normal samples, *v*) Mutant alleles were all on one direction of a read and only in the second half of the read. Second half of read contains the motif GGC[AT]G in sequenced orientation and the mean base quality of all bases after the motif was <20, *vi*) Mean mapping quality of the mutant allele reads was < 21, *vii*) The variant position falls within a simple repeat using the supplied normal, *viii*) Position falls within a centromeric repeat using the supplied normal, *ix*) Mutant reads were on one strand (permitted proportion on other strand: 4%), and mean mutant base quality was <21, *x*) More than 10% of reads covering the variant's position contained an indel according to mapping, *xi*) More than 80% of reads contain the mutant allele at the same read position, *xii*) The variant falls within a high sequencing depth region using the supplied bed file, *xiii*) The variant position could not be annotated against a transcript using the supplied bed file, *xiv*) The variant position has ≥3 mutant alleles present in at least 1% of unmatched normal samples in the unmatched VCF, *xv*) Coverage is ≥10 on both strands but the mutant allele is only present on one strand, and *xvi*) Tumour to normal sample mutant allele proportion is <0.2.

### *In silico* mutational and functional analysis of DDX3X

Silent mutations and those in non-coding regions were excluded. The damaging potential of mutations was assessed using PolyPhen-2 (Polymorphism Phenotyping v2 software), which predicts potential impact of an amino acid substitution based on protein structure and function[3]. The PolyPhen-2 score ranges from 0.0 (tolerated) to 1.0 (deleterious). Variants with scores (0.0 to 0.15), (0.15 to 1.0) and (0.85 to 1.0) were predicted to be benign, probably damaging and possibly damaging respectively. Variants in known cancer genes in the census of the Catalogue of Somatic Mutations in Cancer (COSMIC) which were predicted to be "benign" were also included[4]. Variants in recurrently mutated genes were cross-referenced with COSMIC to determine if they are novel[5]. Manual pathway analysis was carried out on these variants. High

confidence SNVs and indels identified in tumour samples for which the patient's matched normal tissue were sequenced were subjected to gene functional annotation and functional classification clustering analyses using DAVID, with medium classification stringency and Bonferroni and Benjamini corrections[6].

**DDX3X stable knockdown in DLBCL cells**

The hairpin sequences of the shRNAs used are (CTL shRNA, DDX3X shRNA#1 and DDX3X shRNA#1; Supplementary information, **Table S7**). Retroviral packaging plasmids 1 μg of pHIT60 (gag-pol, kind gift of Dr Louis Staudt, National Cancer Institute, USA) and 1 μg of pHITEA6*3 (envelope) were used together with 4 μg of a retroviral construct to transfect each 10 $cm^2$ dish of HEK-293T, after mixing with 1 ml of Opti-MEM medium (Invitrogen) and 18 μl of TransIT-293 (Mirus). The viral supernatant was harvested 48 h after transfection and filtered through a 0.45 μM filter. For transduction of the lymphoma cell lines, 1-2 × $10^6$ target cells were resuspended with viral supernatant and infected by centrifugation (1500g, 90 min at 32°C) with the addition of 10 μg/ml Polybrene (INSIGHT Biotechnology) and 25 μM HEPES (ThermoFisher) in 12 or 24 well plates. Viral supernatant was replaced with fresh medium immediately after centrifugation. Cells were maintained at 37°C with 5% $CO_2$ in RPMI supplemented with 10% tetracycline-free FBS and 1% of Penicillin-Streptomycin for another 48 h and puromycin was added in (1.5 μg/ml) for the selection of transduced cells until the percentage of GFP-expressing cells reached more than 98%. Expression of the hairpins were induced by doxycycline (100 ng/ml) treatment.

**RNA sequencing (RNA-seq) analysis of DDX3X-mutated/depleted cells**

RNA samples were isolated from wild-type, mutant and DDX3X-depleted U2392, HuT78 and SNK1 cells using RNeasy® Mini Kit (QIAGEN). RNA integrity was assessed by 2100 Bioanalyzer system (Agilent). RNA-seq was performed by NovogeneAIT Genomics Singapore Pte Ltd. on Illumina HiSeq platform. The raw reads were aligned using the HISAT2 version 2.1.0 algorithm in National Supercomputing Centre (NSCC) Singapore server[7]. Subsequently, the aligned sequences were sorted using SAM tools 1.3 and downloaded into the local system for further processing on RStudio[8]. Using featureCounts, a program under the Bioconductor package "Rsubread", the mapped reads were counted for genomic features such as exons and chromosomal locations[9]. The summarised output table was uploaded into another Bioconductor

package “DESeq2”, which uses a negative binomial distribution model to test for differential expression[10]. Differentially expressed genes (DEGs) were identified using a cut-off of 2-fold change and a false discovery rate (FDR) of less than 0.05. The generated gene list was analysed through DAVID[6], GSEA[11], IPA (QIAGEN)[12] and GO[13] to identify top diseases and functions as well as pathways associated with the genes that were significantly expressed.

**Checkerboard method and combination index**

The stock dilution of vorinostat (Sigma #SML0061) and WP1066 (sc-203282) starting from 2-fold concentrations of $IC_{50}$ were prepared in cell culture medium. A total of 50 µL the first drug and the second drug of the combination was serially diluted along the ordinate and abscissa, respectively and distributed into each well of the 96-well plates. HuT78, SNK1, or U2392 cells were added into each well ($1 \times 10^4$ cells/well) and the plates were placed in the cell culture incubator (37°C, 5% $CO_2$) for 24 h. The viability of cells was determined using MTS-based assay kit (CellTiter 96® AQueous One Solution Cell Proliferation Assay, Promega) according to manufacturer’s recommendation. The combinatorial concentrations of each drug and their corresponding cell viability values were “entered as datapoints” in CompuSyn software[14,15] and “Combination Index” were determined.

**Tumour xenograft studies**

The immune-deficient NOD.Cg-Prkdc$^{scid}$Il2rg$^{tm1Wjll}$ (NSG) female mice (6-7 weeks, Jackson Laboratory) were kept under standard laboratory conditions in accordance with the National Advisory Committee for Laboratory Animal Research (NACLAR) guidelines. All experiments were approved by the SingHealth Institutional Animal Care and Use Committee (IACUC No: 2016/SHS/1231). Wild-type U2392 (DDX3X-WT) and DDX3X-R475C U2392 cells were resuspended in 0.2 mL Matrigel (in phosphate buffered saline, diluted 1:1) and subcutaneously implanted into the left flank of mice, 10 mice each[16]. Tumour-bearing mice of DDX3X-WT and DDX3X-R475C groups (average tumour size 150 $mm^3$) were randomly assigned into two subgroups (n = 5) for the drug treatment. Mice under the first subgroup were intratumorally injected with WP1066 (40 mg/kg) and mice under the second subgroup were injected with vehicle [20 parts DMSO (Sigma #D8418) and 80 parts PEG-300 (Sigma #91462)[17]], 2 doses/week for 7 weeks in DDX3X-WT and for 8 weeks in DDX3X-R475C (n=5 under each treatment groups). Tumour size was recorded by external measurement of the length (L) and

width (W) of the tumours using a Vernier calliper. Tumour volume (TV) was calculated by using the equation: $TV = (L \times W^2)/2$. The investigators were not blinded during experimental mice allocation and outcome assessments.

The number of mice under each treatment subgroup (n = 5) was calculated with the following assumptions: $n=1+2C(s/d)^2$[18]; where, "s" is standard deviation (in our experience, s ≈ 10%), "d" is the tumour delay difference between treatment group and vehicle group (d = 30%) and "C" is the constant determined by α and β (α is the probability that a positive finding is due to change alone and 1 - β is the chance of obtaining a false-negative result). In our experiments, C = 10.51 when α = 0.05 and 1 - β = 0.9. To counter unforeseen experimental situations, such as the variation of tumour formation in mice and treatment unrelated death, 40% more mice are usually taken in the experiment.

## References


1. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25,** 2865-2871 (2009).
2. Papaemmanuil, E. et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122,** 3616-27 (2013).
3. Adzhubei, I. A., et al. A method and server for predicting damaging missense mutations. *Nat Methods* **7,** 248-249 (2010).
4. Futreal, P. A. et al. A census of human cancer genes. *Nat Rev Cancer* **4,** 177-183 (2004).
5. Bamford, S. et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* **91,** 355-358 (2004).
6. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4,** 44-57 (2009).
7. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12,** 357-360 (2015).

8. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078-2079 (2009).

9. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41(10)**, e108 (2013).

10. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15(12),** 550 (2014).

11. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. **102,** 15545-15550 (2005).

12. Krämer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30,** 523-530 (2014).

13. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25,** 25-9 (2000).

14. Zhang, N., Fu, J. N. & Chou, T. C. Synergistic combination of microtubule targeting anticancer fludelone with cytoprotective panaxytriol derived from panax ginseng against MX-1 cells in vitro: experimental design and data analysis using the combination index method. *Am J Cancer Res* **6,** 97-104 (2015).

15. Chou, T. C. Drug combination studies and their synergy quantification using the Chou-Talalay method. *Cancer Res* **70,** 440-446 (2010).

16. Shin, D. Y., Kim, A., Kang, H. J., Park, S., Kim, D. W. & Lee, S. S. Histone deacetylase inhibitor romidepsin induces efficient tumor cell lysis via selective down-regulation of LMP1 and c-myc expression in EBV-positive diffuse large B-cell lymphoma. *Cancer Lett* **364,** 89-97 (2015).

17. Horiguchi, A. et al. STAT3 inhibitor WP1066 as a novel therapeutic agent for renal cell carcinoma. *Br J Cancer* **102,** 1592-1599 (2010).

18. Snedecor, G. W. & Cochran, W. G. Statistical Methods. 8th Ed. Ames: Iowa State Press; 1989.