Integrative analysis of copy-number and expression data

Genomic DNAs from 18 DLBCL (SUDHL4, SUDHL6, SUDHL7, SUDHL8, SUDHL10, OCILy1, OCILy3, OCILy8, OCILy10, OCILy18, OCILy19, DB, FARAGE, Pfeiffer, Toledo, HT, KARPAS-422, WSUNHL), 1 MLBCL (KARPAS11106P) and 6 HL (L428, KMH2, L540, L1236, SUPHD1, HDLM2) cell lines and 21 peripheral blood lymphocyte samples from normal donors were extracted as previously described and profiled using Affymetrix SNP6.0 microarrays.[1] The above-mentioned high density (HD) SNP array includes over 900,000 SNP probes and ≈ 950,000 additional probes for copy number variation resulting in substantially improved whole genome coverage.

The inference of DNA copy number from '.cel' files was performed using a GenePattern pipeline that runs the following modules: *SNPFileCreator, CopyNumberInference, RemoveCopyNumberOutliers, DivideByNormals,* and *Quality Control*.[2] Normalized log2-ratios were segmented using the Circular Binary Segmentation (CBS) algorithm (v1.12.0) using 10,000 permutations, α=0.01, and undo splits (undo.sd=1).[3,4] Significance of copy number alterations across samples was assessed by the GISTIC algorithm[5], applied using the GenePattern platform and thresholds corresponding to t_{amp}=2.178 for amplifications and t_{del}=1.820 for deletions. These thresholds were determined based on an empirical analysis of the distribution of copy numbers (CNs) in known regions of amplification (and deletion), and by selecting the mid-point between the modes corresponding to two and three CNs (and two and one CNs). The chromosome-wide amplification in chromosome 7 and the chromosome arm-wide deletion in chromosome 6 were used for this purpose.

GISTIC defines peaks of interest with associated FDR q-values determined by multiple hypothesis testing (MHT) correction, with regions obtaining q-values below 0.25 being considered significant. Within each region a smaller peak (or peaks) is identified as the set of contiguous markers with the highest q-values. Naturally occurring copy number variants (CNVs) were removed before running GISTIC. These CNVs were compiled from SNP6.0 analysis of HapMap normals[6] the Database of Genomic Variants (DGV, http://projects.tcag.ca/variation)[7] and an automated search of profiled normals.

Transcriptional profiling was performed for all cell lines using Affymetrix U133 A&B microarrays as previously described.[8] Integrative analysis, combining DNA copy number and gene transcript data, was performed in order to assess relationships between DNA copy number change and alteration in gene transcript abundance so as to refine the list of candidate genes within alteration regions. Genes within the peak (region) of GISTIC-identified alterations were tested for difference in expression between samples with or without each lesion by a two-group t-statistic, and significance was assessed both by a permutation test procedure and by asymptotically derived p-values (using a Student's t distribution). The permutation-based p-values were "smoothed" by the Laplace rule of succession, i.e., by adding 1 to the numerator and 2 to the denominator of the ratio used to compute the empirical p-value.[9] Q-values were derived by MHT correction of nominal p-values using the FDR method applied to the union of all genes within all the peaks (regions). Genes were considered positive by integrative analysis if

differences in transcript abundance attained an FDR≤0.25 by at least one of the two procedures (permutation-based or asymptotic).

REFERENCES

[1] Takeyama, K. *et al.* Integrative analysis reveals 53BP1 copy loss and decreased expression in a subset of human diffuse large B-cell lymphomas. *Oncogene* 27, 318–322 (2007).
[2] Network, C. G. A. R. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068 (2008).
[3] Olshen, A., Venkatraman, E., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array based DNA copy number data. *Biostat* 5, 557–572 (2004).
[4] Venkatraman, E. & Olshen, A. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 23, 657–663 (2007).
[5] Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application in glioma. . *Proc Natl Acad Sci U S A* 104, 20007–20012 (2007).
[6] McCarroll, S. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* 40, 1166–1174 (2008).
[7] Iafrate, A. *et al.* Detection of large-scale variation in the human genome. *Nat Genet* 36, 949–951 (2004).
[8] Juszczynski, P. *et al.* The AP1-dependent secretion of galectin-1 by Reed Sternberg cells fosters immune privilege in classical Hodgkin lymphoma. *Proc Natl Acad Sci U S A* 104, 13134–13139 (2007).
[9] Gould, J., Getz, G., Monti, S., Reich, M. & Mesirov, J. P. Comparative gene marker selection suite. *Bioinformatics* 22, 1924–1925, doi:10.1093/bioinformatics/btl196 (2006).

Table S1. Integrative analysis of DNA copy number and transcript abundance of genes within the 9p24 amplification peak

Gene Symbol	Score	P-value	FDR	Fold Change
C9ORF46	−3.39	0.0014	0.0324	2.44
CD274	−2.37	0.015	0.12	2.73
PDCD1LG2	−2.04	0.032	0.182	1.67
KIAA1432	−1.23	0.14	0.375	1.1
RLN1	−1.19	0.16	0.402	1.39
RLN2	−0.76	0.22	0.488	1.12
INSL4	−0.65	0.3	0.579	1.07

C90RF46, CD274 (PD-L1) and PDCD1LG2 (PD-L2) have most significant association between DNA copy number and transcript abundance.

Figure S1. PD-L2 expression in HL and MLBCL cell lines – comparison of cell surface immunostaining by flow cytometry (as in Fig. 2, main manuscript) (top panel) and quantitative immunohistochemistry (bottom panel)
PD-L2 levels measured by quantitative immunohistochemistry are in line with those determined by flow cytometry. After confirming that PD-L2 quantitative IHC and flow cytometry gave comparable results in the above-mentioned cell lines, we evaluated PD-L2 expression in primary MLBCLs by quantitative IHC (Fig. S2).

Figure S2. Quantitative immunohistochemistry of PD-L2 in representative primary MLBCLs with known 9p24.1 copy numbers
(Top panel) qPCR analysis of PD-1 ligand (*PD-L1*)/9p24.1 copy numbers in a series of 41 primary MLBCLs (as in Fig. 4A, main manuscript). (Bottom panel) PD-L2 quantitative immunohistochemistry in representative primary MLBCLs from this series. (Left) PD-L2 quantitative immunohistochemistry. (Right) Quantitative analysis of PD-L2 immunostaining in 10 tumor-involved non-sclerotic regions of each tumor.

Figure S3. Chemical inhibition and siRNA-mediated knockdown of JAK2 decrease PD-1 ligand transcription
(A) Western analysis of phosphoJAK2 in HL cell lines (L428 and SUPHD1) treated with the increasing doses (2.5–10 μm) of the specific JAK2 inhibitor, SD-1029, or control or siRNA JAK2. siRNA-mediated knockdown of JAK2 was obtained by transfecting cells with 75 mol of JAK2 ON-TARGETplus SMARTpool siRNA (Dharmacon) or corresponding scrambled oligonucleotides. Transfections were performed as described for luciferase constructs (main manuscript). (B) RT-qPCR analysis of PD-L1 transcript abundance in the cell lines treated with SD-1029, or control or siRNA JAK2 in A. Data are means (+/− SD) of triplicate RT-qPCR reactions from a representative experiment (experiment performed 3 times). Although siRNA-mediated JAK2 knockdown was less effective at decreasing JAK2 activity (phosphoJAK2) than chemical inhibition, JAK2 depletion also decreased PD-L1 expression in the HL cell lines. Similar to JAK2 chemical inhibition, JAK2 knockdown was more effective at inhibiting PD-L1 expression in SUPHD1 (which has high-level 9p24.1 amplification).

Figure S4. Chemical inhibition of JAK2 decreases PD-L2 ligand transcription
(A) RT-qPCR analysis of PD-L2 transcript abundance in HL cell lines treated with increasing doses (2.5–10 μM) of the JAK2 inhibitor, SD-1029 (as in Fig. 6A, main manuscript or Fig. S3). Data are means (+/− SD) of triplicate RT-qPCR reactions from a representative experiment. (B) The PD-L2 promoter regulatory module including a predicted ISRE/IRF module and other degenerate STAT-binding sites. The spacing of the ISRE and IRF1 elements in the predicted PD-L2 module was larger than that in other known ISRE/IRF1 modules (<100bp) and the PD-L1 module (63bp, Fig. 6C, main manuscript). The ISRE/IRF1 spacing and lower numbers of degenerate STAT-binding sites in the PD-L2 5′ regulatory region likely explain the more moderate effect of chemical JAK2 inhibition on PD-L2 transcript abundance (compared with PD-L1). (C) Analysis of pGL3-PD-L2p luciferase activity in L428 and SUHD1 HL cells treated with SD-1029 or vehicle. To assess the effect of JAK2 on transcription mediated by the 5′ PD-L2 regulatory element, this sequence was cloned into the pGL3 luciferase vector and transfected into

HL cell lines with low (L428) or high (SUPHD1) 9p24.1 copy numbers. PD-L2 promoter-driven luciferase activity was increased in both L428 and SUPHD1 (empty vector [control] vs pGL3-PD-L2 [control]). Of note, SUPHD1 exhibited much higher PD-L2 promoter-driven luciferase expression than L428 (SUPHD1 vs. L428 pGL-3PD-L1p [control], compare y-axes). In addition, treatment with the specific chemical JAK2 inhibitor, SD-1029, resulted in a marked decrease in PD-L2 driven luciferase expression in SUPHD1 and a more modest decrease in L428 (compare pGL3-PD-L2p [control] vs pGL3-PD-L2 [SD-1029]). Data are means (+/− SD) of triplicate measurements from a representative experiment. The experiments in A and B were performed 3 times.

Figure S1

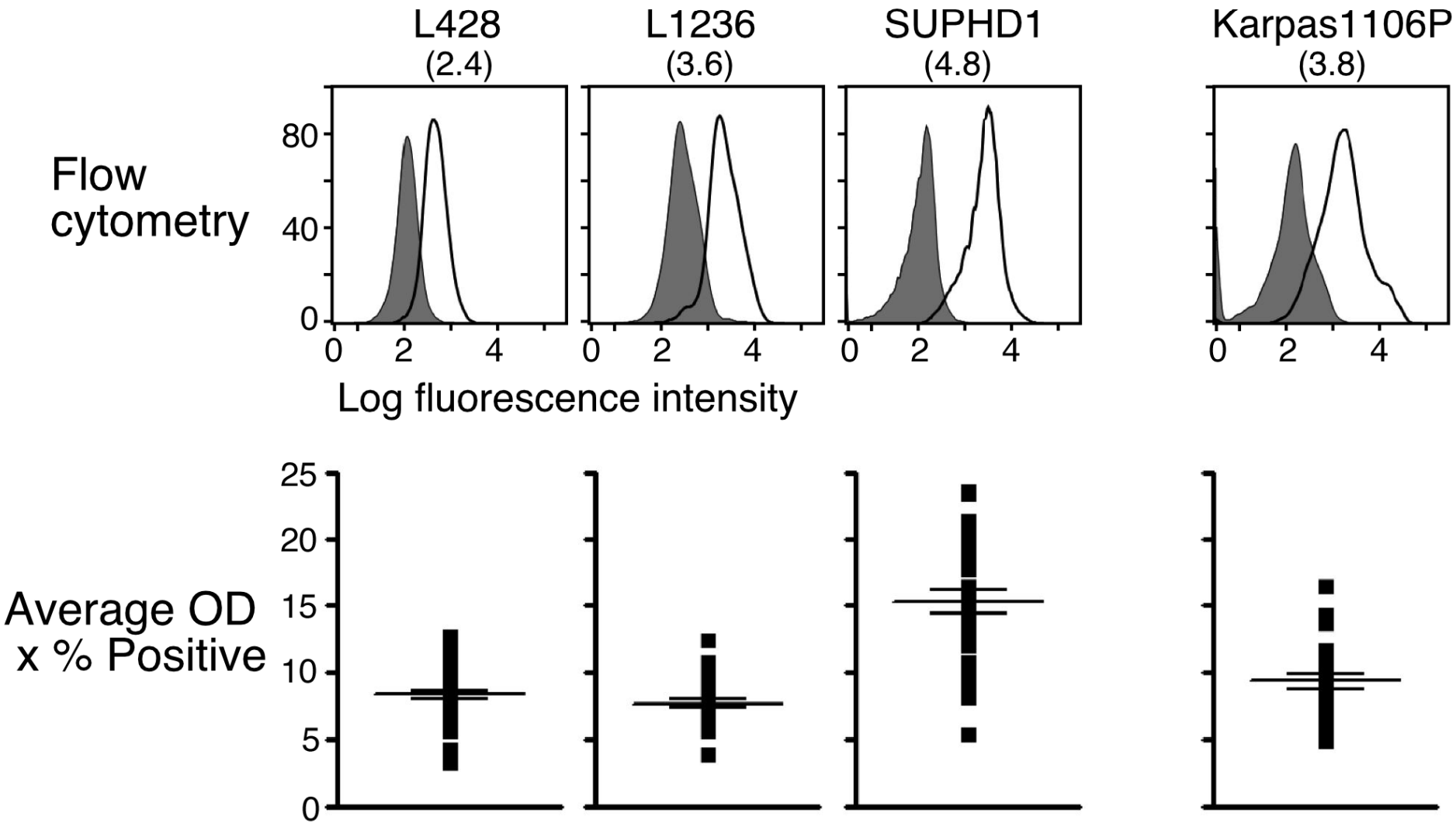

L428
(2.4)
L1236
(3.6)
SUPHD1
(4.8)
Karpas1106P
(3.8)
Flow cytometry
80
40
0
0
2
4
Log fluorescence intensity
Average OD x % Positive
25
20
15
10
5
0

Figure S2

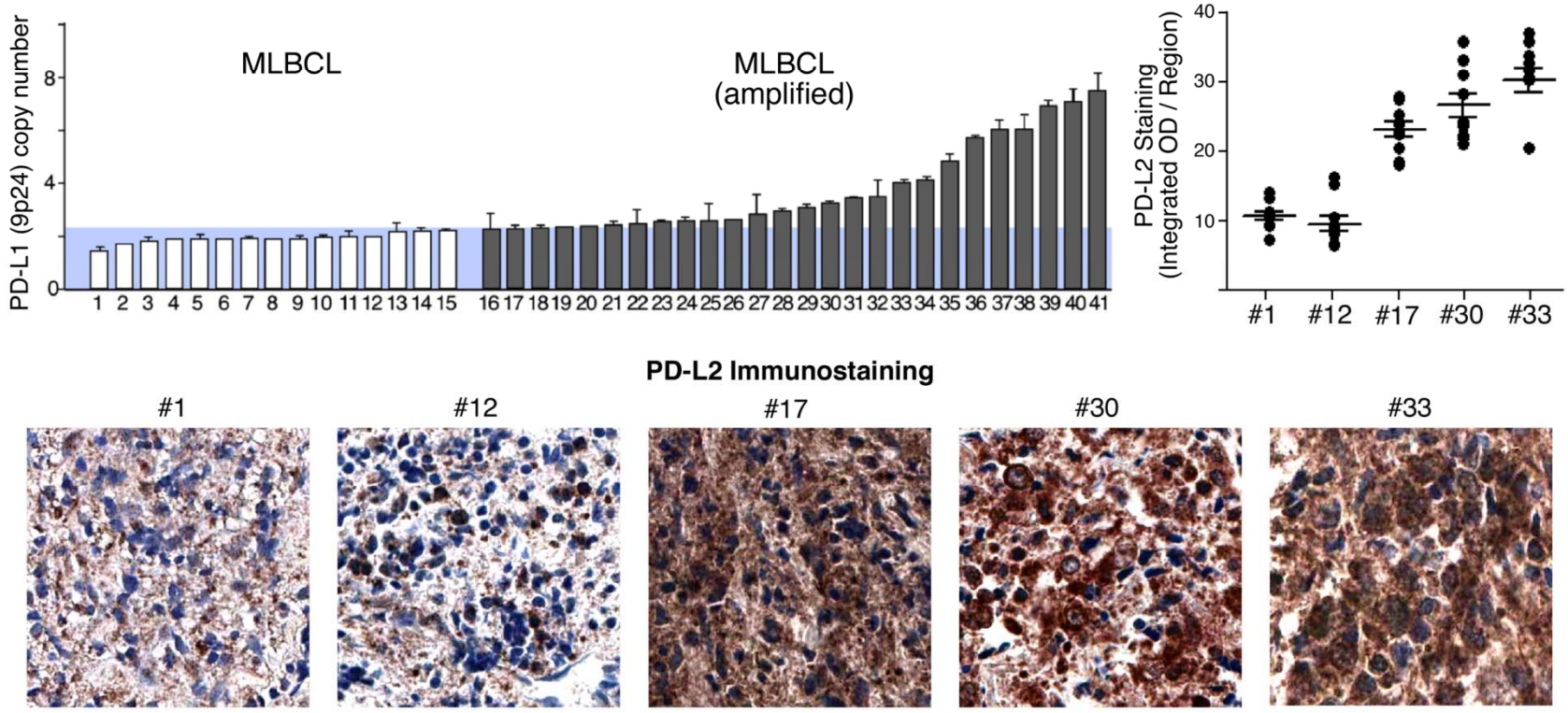

PD-L1 (9p24) copy number
MLBCL
MLBCL
(amplified)
0
4
8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41
PD-L2 Staining
(Integrated OD / Region)
10
20
30
40
#1
#12
#17
#30
#33
PD-L2 Immunostaining
#1
#12
#17
#30
#33

Figure S3

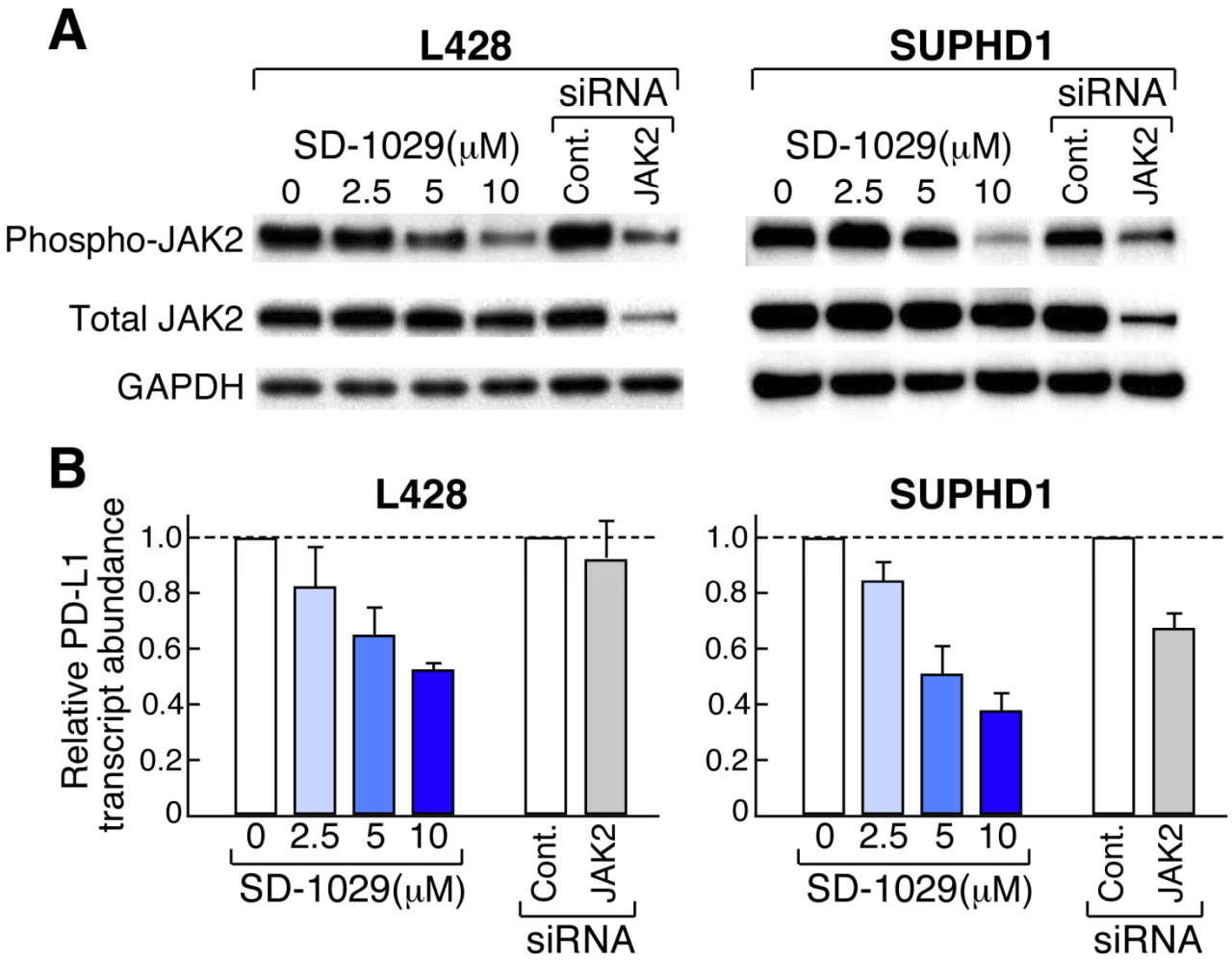
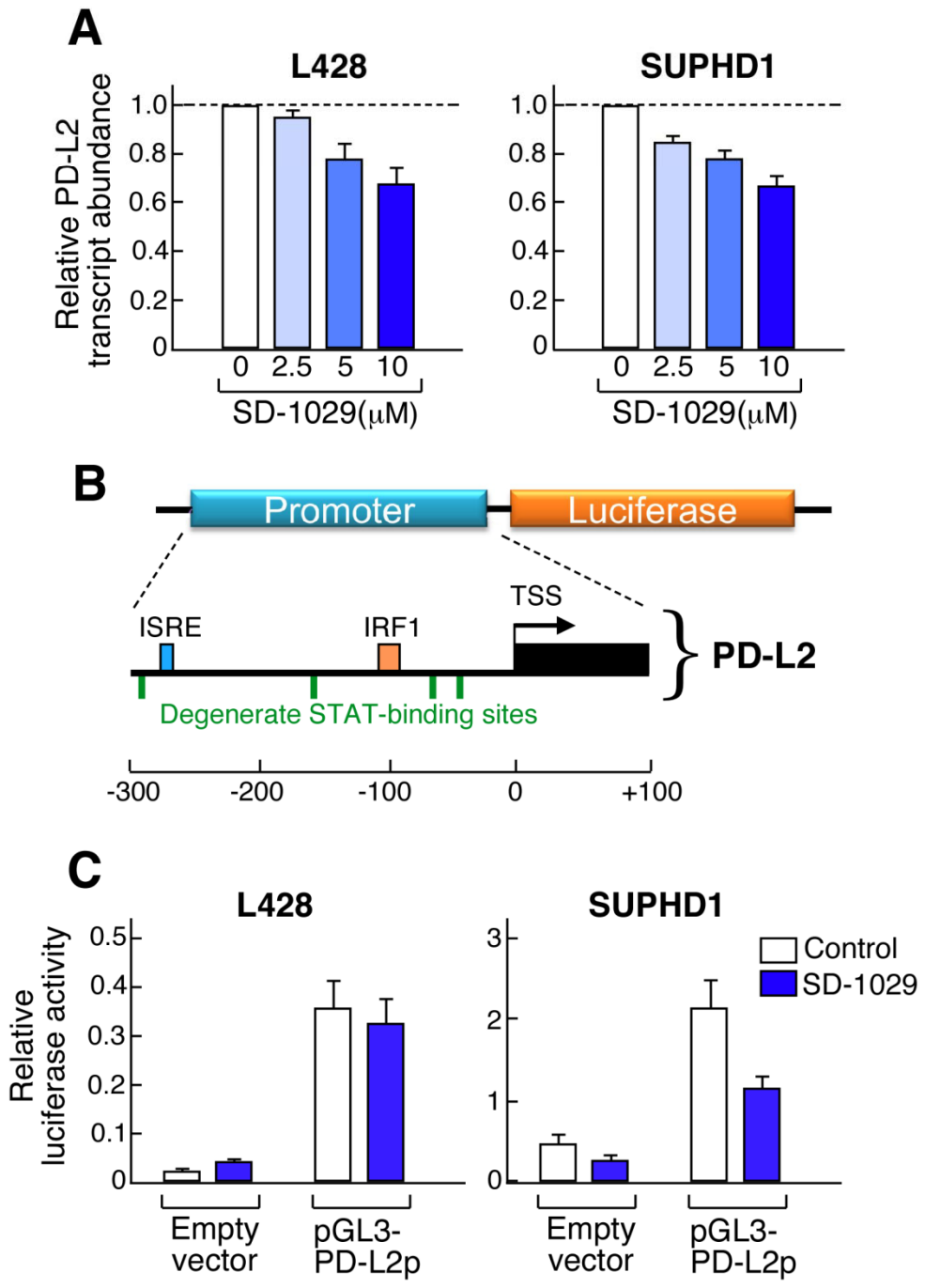

A
L428
SUPHD1
siRNA
SD-1029(μM)
0 2.5 5 10
Cont.
JAK2
Phospho-JAK2
Total JAK2
GAPDH
B
L428
SUPHD1
Relative PD-L1 transcript abundance
1.0
0.8
0.6
0.4
0.2
0
0 2.5 5 10
SD-1029(μM)
Cont.
JAK2
siRNA

Figure S4


A
L428
SUPHD1
Relative PD-L2 transcript abundance
1.0
0.8
0.6
0.4
0.2
0
0 2.5 5 10
SD-1029(μM)
B
Promoter
Luciferase
TSS
ISRE
IRF1
PD-L2
Degenerate STAT-binding sites
-300 -200 -100 0 +100
C
L428
SUPHD1
Relative luciferase activity
0.5
0.4
0.3
0.2
0.1
0
3
2
1
Control
SD-1029
Empty vector
pGL3-PD-L2p