**Supplementary Tables**

**Table S1. Characteristics of patients with hairy cell leukemia**. b.c.m. : cm below costal margin on examination; *at time of investigation; ^differential represented the percent of neutrophils, lymphocytes and hairy cells (HCs) of the white blood cells (WBC) as calculated by morphological examination of peripheral blood smear; Hb: haemoglobin; PLT: platelets; ~phenotype was determined on the peripheral blood mononucleated cell (PBMC) fraction in flow cytometry and percent of HCs was calculated as the CD103+ B-cell fraction of the lymphocyte gate; 2CdA: 2-Chloro-deoxyAdenosine; DCF: DeoxyCoFormicin; IFN: interferon-alpha.

| Characteristics | n/total (%) | Range (Median) |
|---|---|---|
| ***Clinical*** | | |
| Age | | 33-75 (47) |
| Male gender | 7/11 (64) | |
| Splenomegaly | 7/11 (64) | |
| Spleen b.c.m. | | 0-16 (5) |
| Second neoplasm* | 0/11 (0) | |
| ***FBCs and differential^*** | | |
| Hb (g/dl) | | 8-17 (11.2) |
| PLT (x109/L) | | 28-114 (79.0) |
| WBC (x109/L) | | 4-41 (10.4) |
| Neutrophils (%) | | 3-38 (9.0) |
| Lymphocytes (%) | | 47-98 (86) |
| HCs (%) | | 10-93 (58) |
| ***Phenotype~*** | | |
| HCs (%) | | 33-73 (66.8) |
| Kappa+ | 4/11 (36) | |
| CD20+ | 11/11 (100) | |
| CD11c+ | 11/11 (100) | |
| CD27+ | 11/11 (100) | |
| CD38+ | 11/11 (100) | |
| CD103+ | 11/11 (100) | |
| ***Bone Marrow histology*** | | |
| HCs (%) | | 50-95 (85) |
| DBA44+ | 11/11 (100) | |
| HCI | | 0.11-0.81 (0.44) |
| ***Molecular*** | | |
| IGHV homology | | 91.7-99.0 (97.2) |
| IGHV≤98% homology | 10/11 (91) | |
| BRAF V600E mutation | 11/11 (100) | |

**Supplementary Table legends**

**Table S2. Biologic concepts obtained grouping over 6,100 gene sets belonging to the MSigDB v5.1 GSEA collection** (1) **into 100 main biological themes.**

**Table S3. HCL gene expression signature by Basso et al**. (2). Supervised comparison by moderated t-test (limma) and Fisher's exact test (methylation only) on the HCL methylation and gene expression profiles with the post-GC B-cell subsets (MGZ, loMBC and intMBC, grouped as one pool), and with SMZL and CLL (M-CLL, U-CLL and CLL-VH3-21). Fold changes corresponding to the M-value difference (methylation) or to the $\log_2$ difference (gene expression) between HCL and post-GC B-cell averages, and between HCL and other B-cell tumor averages respectively.

**Table S4. Methylation profiles in HCL**. Supervised comparison by moderated t-test (limma) and Fisher's exact test of the HCL methylation profiles with the post-GC B-cell subsets (MGZ, loMBC and intMBC, grouped as one pool) and with SMZL and CLL (M-CLL, U-CLL and IGHV3-21+ CLL, called CLL-VH3-21, grouped as one pool). Fold changes corresponding to the M-value difference between HCL and other B-cell averages.

**Table S5. Methylation profiles in HCL and post-GC B-cells**. Supervised comparison by moderated t-test (limma) and Fisher's exact test of the HCL methylation profiles with the post-GC B-cell subsets (MGZ, loMBC and intMBC, grouped as one pool). Fold changes corresponding to the M-value difference between HCL and other B-cell averages.

**Table S6. Methylation profiles in HCL and other B-cell tumors**. Supervised comparison by moderated t-test (limma) and Fisher's exact test of the HCL methylation profiles with SMZL and CLL (M-CLL, U-CLL and IGHV3-21+ CLL, called CLL-VH3-21, grouped as one pool). Fold changes corresponding to the M-value difference between HCL and other B-cell averages.

**Table S7. Methylation profiles in HCL and SMZL**. Supervised comparison by moderated t-test (limma) and Fisher's exact test of the HCL methylation profiles with SMZL. Fold changes corresponding to the M-value difference between HCL and other B-cell averages.

**Table S8. Methylation profiles in HCL and M-CLL**. Supervised comparison by moderated t-test (limma) and Fisher's exact test of the HCL methylation profiles with M-CLL. Fold changes corresponding to the M-value difference between HCL and other B-cell averages.

**Table S9. Methylation profiles in HCL and U-CLL**. Supervised comparison by moderated t-test (limma) and Fisher's exact test of the HCL methylation profiles with U-CLL. Fold changes corresponding to the M-value difference between HCL and other B-cell averages.

**Table S10. Methylation profiles in HCL and IGHV3-21+ CLL**. Supervised comparison by moderated t-test (limma) and Fisher's exact test of the HCL methylation profiles with IGHV3-21+ CLL (CLL-VH3-21). Fold changes corresponding to the M-value difference between HCL and other B-cell averages.

**Table S11. Gene expression changes driven by methylation in HCL (I)**. Supervised comparison by moderated t-test (limma) and Fisher's exact test (methylation only) on the HCL methylation and gene expression profiles with the post-GC B-cell subsets (MGZ, loMBC and intMBC, grouped as one pool). Fold changes corresponding to the M-value difference (methylation) or to the $\log_2$ difference (gene expression) between HCL and post-GC B-cell averages.

**Table S12 Methylation and gene expression profiles in HCL of BCR, TLR-MAPK-NFKB, BRAF signaling genes; cell adhesion, B-cell differentiation markers and methylated transcripts in cancer**. Single sample GSEA (ssGSEA) (3) was performed on the methylation and gene expression profiling data. Differentially methylated and expressed signatures were selected by supervised comparison (moderated *t-test*) on the methylation and gene expression profiles of HCL and post-GC B-cell subsets (MGZ, loMBC and intMBC, grouped as one pool). Fold changes corresponding to the difference on the ssGSEA scores between HCL and post-GC B-cell averages.

**Table S13. Gene expression changes driven by methylation in HCL (II)**. Supervised comparison by moderated t-test (limma) and Fisher's exact test (methylation only) of the HCL methylation and gene expression profiles and with SMZL and CLL (M-CLL, U-CLL and IGHV3-21+ CLL, called CLL-VH3-21). Fold changes corresponding to the M-value difference (methylation) or to the $\log_2$ difference (gene expression) between HCL and other B-cell averages.

**Table S14. Enrichment analyses on Methylation and gene expression profiles in HCL**. Supervised comparison by moderated t-test (limma) of the HCL methylation and gene expression concepts with the post-GC B-cell subsets (MGZ, loMBC and intMBC, grouped as one pool), and with SMZL and the CLL pool (M-CLL, U-CLL and IGHV3-21+ CLL, called CLL-VH3-21, grouped as one pool). Fold changes corresponding to the difference in single-sample GSEA (ssGSEA) score (methylation or gene expression) between HCL and post-GC B-cell, and between HCL and other B-cell tumors (SMZL and the CLL pool), respectively.

## Supplementary Figures

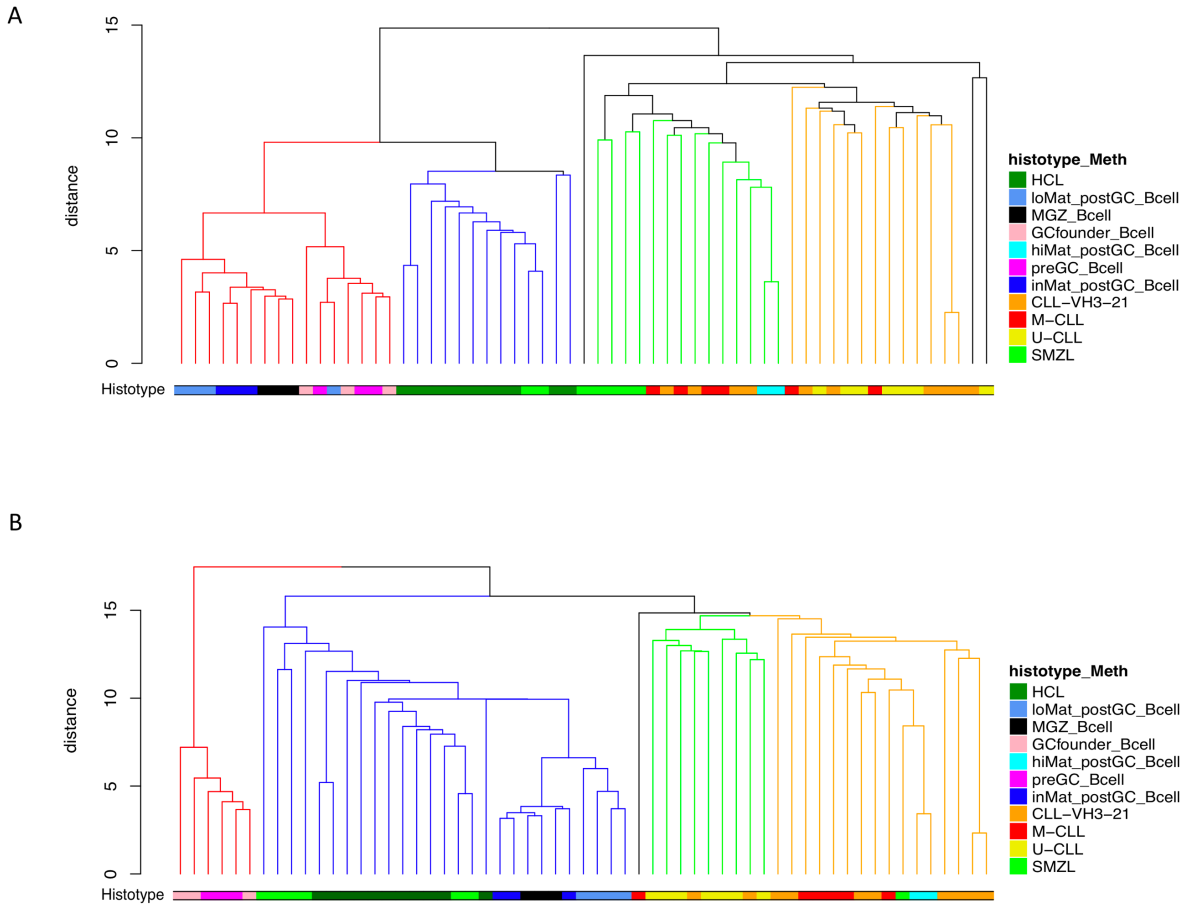


**Figure S1. (A) Unsupervised hierarchical clustering (Euclidean distance, complete linkage) using CGI-probes; (B) Unsupervised hierarchical clustering using non CGI-probes (Euclidean distance, complete linkage)**. Methylation profiling (histotype_Meth in the legend) included hairy cell leukemia (HCL, dark green); chronic lymphocytic leukemia samples included un-mutated (U-CLL, yellow), mutated (M-CLL, red) IGHV, and IGHV3-21+ (CLL-VH3-21, orange); splenic marginal zone lymphoma (SMZL, light green); naive B cells (preGC_Bcell, magenta); germinal center (GC) founder B-cells (B-cells upon antigen encountering, GCfounder_Bcell, light pink); low-, intermediate- and high-maturity memory B cells (loMat_postGC_Bcell in light blue; inMat_postGC_Bcell in blue; and hiMat_postGC_Bcell in cyan, respectively); and splenic marginal zone B-cells (MGZ_Bcell in black).

A

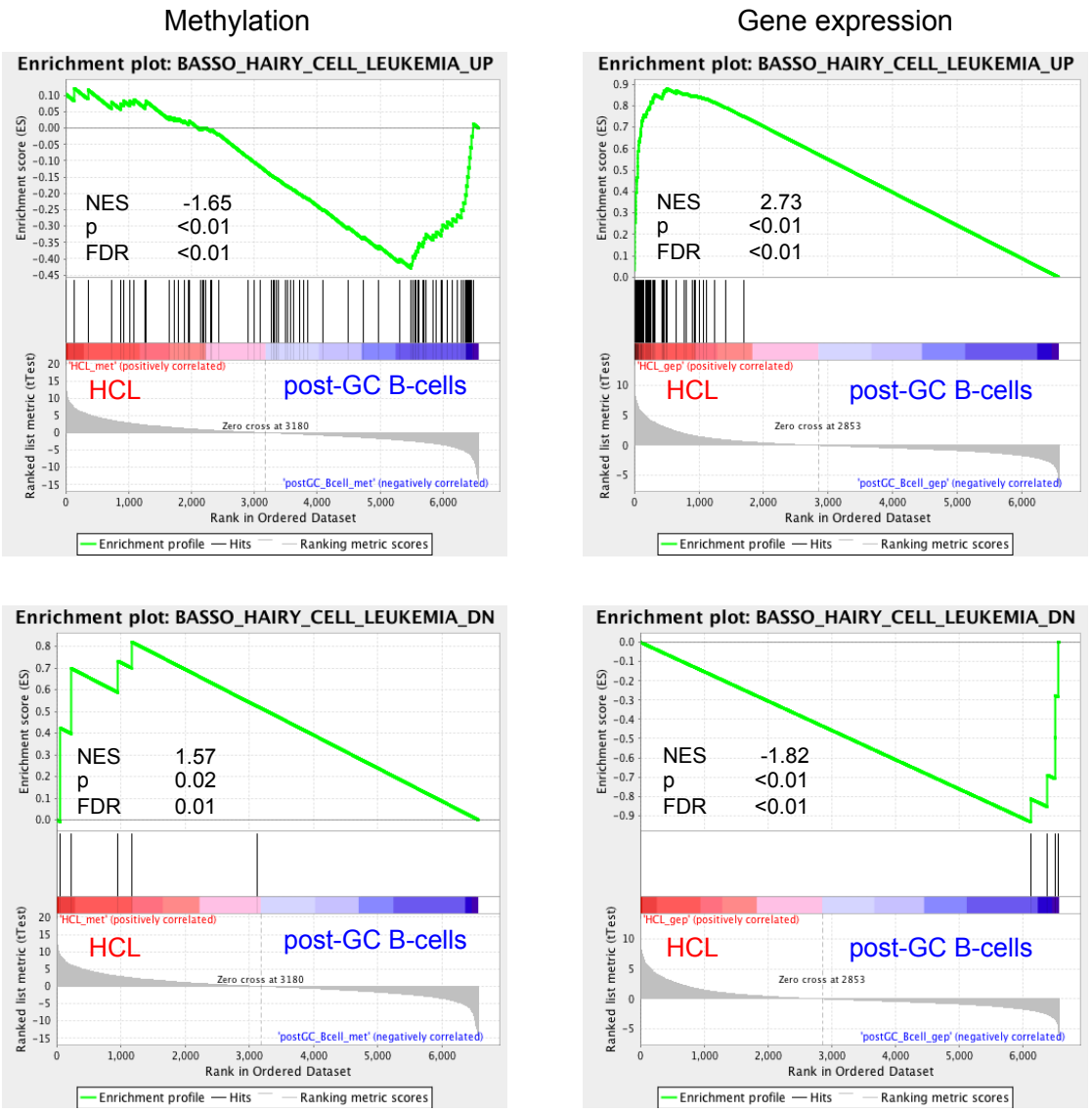


**Figure S2A. Methylation importantly contributes to the HCL gene expression signature (I).** Gene expression signatures of HCL (described by *Basso et al* (2)) were analyzed by gene set enrichment analyses (GSEA) (1) of the methylation (left panel) and gene expression (right panel) profiles of HCL (red) and post-GC B-cells (blue). NES: normalized enrichment score, p: nominal p-value, FDR: false discovery rate.

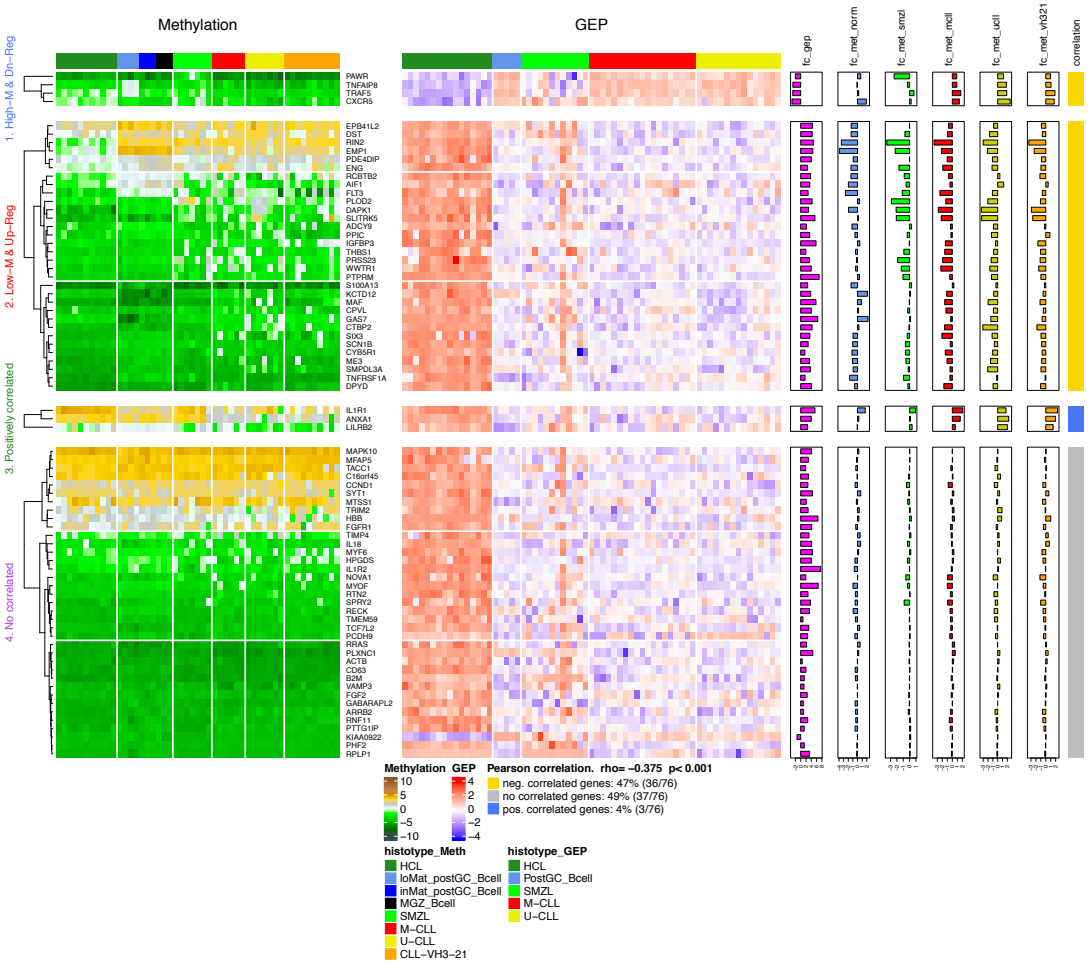


**Figure S2B. Methylation importantly contributes to the HCL gene expression signature (II).** Integration of promoter methylation data (heatmap on right) with the published gene expression profile (GEP, heatmap on left) (2). 36 out of 76 differentially transcribed genes (47%) inversely correlated with the methylation status. Hierarchical clustering (on rows, Euclidean distance, complete linkage) based on the methylation profile. Pearson's correlation was performed between methylation and GEP (rho= -0.375, p< 0.001), *fc_gep* means fold change in gene expression; *fc_met_norm, fc_met_smzl, fc_met_mcll, fc_met_ucll* and *fc_met_vh321* means differences in M-values in HCL compared to post-GC B-cells, SMZL, M-CLL, U-CLL or CLL-VH3-21 respectively. Methylation profiling (histotype_Meth in the legend) included hairy cell leukemia (HCL, dark green); chronic lymphocytic leukemia samples included un-mutated (U-CLL, yellow), mutated (M-CLL, red) IGHV, and IGHV3-21+ (CLL-VH3-21, orange); splenic marginal zone lymphoma (SMZL, light green); low- and intermediate-maturity memory B cells (loMat_postGC_Bcell in light blue and inMat_postGC_Bcell in blue, respectively), and splenic marginal zone B-cells (MGZ_Bcell in black). Gene expression profiling (histotype_GEP in the legend) included hairy cell leukemia (HCL, dark green); chronic lymphocytic leukemia samples included un-mutated (U-CLL, yellow), mutated (M-CLL, red) IGHV; splenic marginal zone lymphoma (SMZL, light green); and memory B cells (postGC_Bcell, light blue).

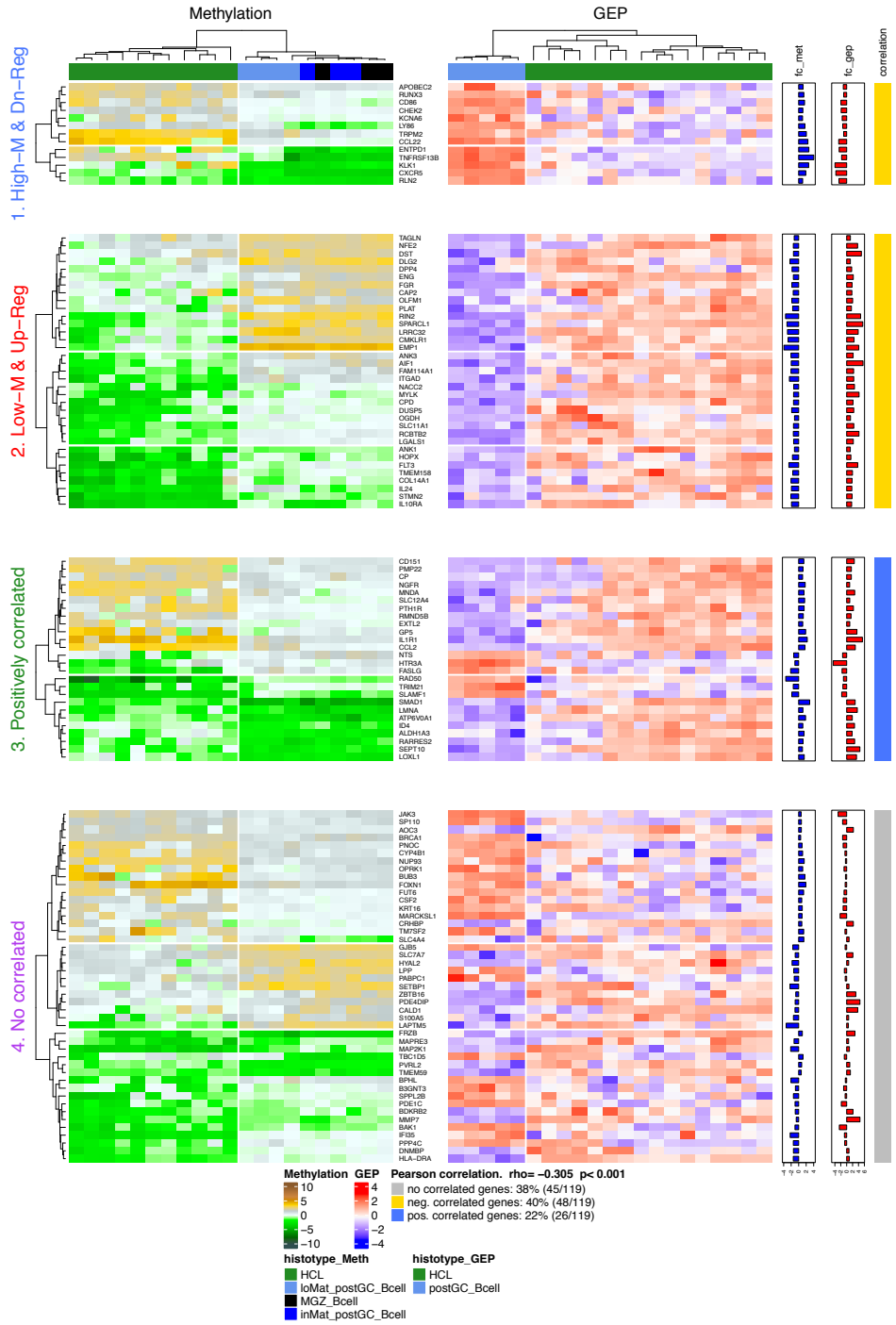


**Figure S3. Integration of methylation (heatmap on left) and gene expression (heatmap on right) profiles of HCL and post-GC (MGZ_Bcell, loMBC_postGC and inMBC_postGC) B-cells**. Supervised hierarchical clustering (Euclidean distance, complete linkage) of methylation profiles. Pearson's correlation was performed between methylation and GEP (rho= -0.305, p< 0.001), *fc_met* means fold change in methylation, *fc_gep* means fold change in gene expression. Methylation profiling (histotype_Meth in the legend) included hairy cell leukemia (HCL, dark green); low-, intermediate- and high-maturity memory B cells (loMat_postGC_Bcell in light blue; and inMat_postGC_Bcell in blue, respectively); and splenic marginal zone B-cells (MGZ_Bcell in black). Gene expression profiling (histotype_GEP in the legend) included hairy cell leukemia (HCL, dark green); and memory B cells (postGC_Bcell, light blue).

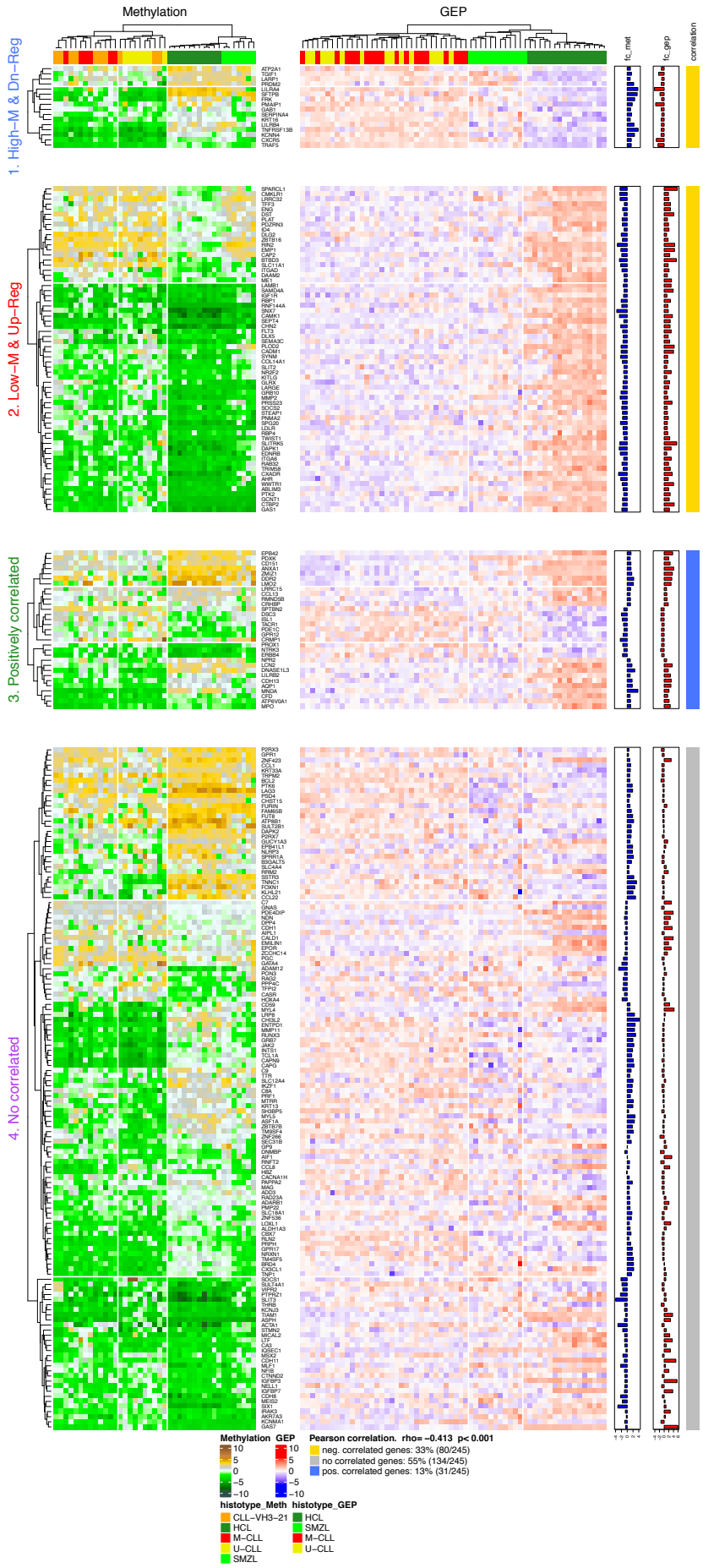
Methylation
GEP
fc_met
fc_gep
correlation
1. High−M & Dn−Reg
2. Low−M & Up−Reg
3. Positively correlated
4. No correlated
Methylation GEP
10
5
0
−5
−10
histotype_Meth
CLL−VH3−21
HCL
M−CLL
U−CLL
SMZL
Pearson correlation. rho= −0.413 p< 0.001
neg. correlated genes: 33% (80/245)
no correlated genes: 55% (134/245)
pos. correlated genes: 13% (31/245)
histotype_GEP
HCL
SMZL
M−CLL
U−CLL

**Figure S4. Integration of methylation (heatmap on left) and gene expression (heatmap on right) profiles of HCL, SMZL and the CLL pool (M-CLL, U-CLL and CLL-VH3-21)**. Supervised hierarchical clustering (Euclidean distance, complete linkage) of methylation profiles. Pearson's correlation was performed between methylation and GEP (rho= -0.413, p< 0.001), *fc_met* means fold change in methylation; *fc_gep* means fold change in gene expression. Methylation profiling (histotype_Meth in the legend) included hairy cell leukemia (HCL, dark green); chronic lymphocytic leukemia samples included un-mutated (U-CLL, yellow), mutated (M-CLL, red) IGHV, and IGHV3-21+ (CLL-VH3-21, orange); and splenic marginal zone lymphoma (SMZL, light green). Gene expression profiling (histotype_GEP in the legend) included hairy cell leukemia (HCL, dark green); chronic lymphocytic leukemia samples included un-mutated (U-CLL, yellow), mutated (M-CLL, red) IGHV; and splenic marginal zone lymphoma (SMZL, light green).

## References

1. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005 Oct 25;102(43):15545-50.

2. Basso K, Liso A, Tiacci E, Benedetti R, Pulsoni A, Foa R, et al. Gene expression profiling of hairy cell leukemia reveals a phenotype related to memory B cells with altered expression of chemokine and adhesion receptors. J Exp Med. 2004 Jan 5;199(1):59-68.

3. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013 Jan 16;14:7.