Imperial College London - Cancer & Surgery

SPIB - A potential biomarker in breast cancer?

Jan-Philipp Cieslik (CID 02139878)

1 Introduction

Breast cancer is the most common cancer in women. One in every seven women in the UK will be diagnosed with breast cancer in their lifetime. Due to the high incidence, breast cancer is also accountable for the most cancer/associated deaths. To predict patient outcomes and to find new actionable targets for cancer therapy new biomarkers are required. There is already a multitude of established clinical markers like the tumour size, lymphatic node invasion status, distant metastasis status (as in the TNM classification) and molecular characterization (mainly oestrogen, progesterone and HER2 receptor status). High throughput data is becoming more readily available, allowing for in silico analysis of multiomics (e.g., genomics, proteomics, ...) of large cohorts. This essay tries to demonstrate the utility of Spi-B as a possible breast cancer biomarker. The transcription factor Spi-B is encoded by the gene SPIB and is a member of the Erythroblast Transformation Specific (ETS) group, which is defined by a common highly conserved DNA/binding domain. In the literature Spi-B is described as both a tumor suppresor and an oncogenic protein. Studies in lung cancer cells found Spi-B to be involved in recruitment of tumor associate macrophages (TAM) [1], further Spi-B was found to promotes anoikis resistance [2]. In colorectal cancer cells on the other hand, Spi-B displayed tumor suppressing characteristics by activating NF-κB and JNK signaling pathways [3].
(Introduction word count: 233)

2 Methods

2.1 Data Sources

The Cancer Genome Atlas (TCGA) dataset for breast cancer (BRCA)[4] was acquired from the Xena platform [5].

2.2 Data Pre-Processing and Normalization

Some data is downloaded after pre-processing steps have already been performed. I will outline all important data pre-processing steps to give a complete representation of the resulting data.

RNA-Seq

Transcript abundance is calculated with RSEM (RNA-Seq by Expectation Maximization) and then transformed into z-scores (distance from the median measured

in multiples of the standard deviation). The reference for median and standard deviation is generated only from diploid tumour samples.

Methylation

Methylation data from the Illumina Infinium HumanMethylation450 BeadChip (Methylation450k) were transformed into β values ranging from 0 to 1. A higher β value indicates a higher level of DNA methylation. Methlyation probes are annotated with their corresponding gene symbol. The data is then reduced to only contain one methylation probe per gene, keeping only the probe with the strongest anti-correlation with the corresponding RNA-Seq data.

Copy Number

Copy number data was obtained via a whole genome microarray. The raw data was processing using the GISTIC2 (Genomic Identification of Significant Targets in Cancer) pipeline. Finally, the resulting values were grouped by thresholds and transformed into one of five levels from deep deletion (-2) to high-level amplification (+2).

2.3 Identification of a Possible Novel Prognostic Marker

A multivariate Cox proportional hazards regression model was created with data from 1218 breast cancer patients. Overall survival was defined as the dependent variable and mRNA z-scores (as determined by RNA sequencing from the primary tumour) together with age and tumour stage (I/II as low; III/IV as high) as the independent variables. The calculated p-values were corrected with the false discovery rate (FDR) method and are shown as q-values.

2.4 Survival Analysis

The patient population was divided into SPIB-high and SPIB-low divided by the median of the SPIB mRNA expression. A logrank test and an associated Kaplan-Meier plot were generated for the two subpopulations.

2.5 Methylation and mRNA Expression Correlation

Methylation values were tested for normal distribution using Shapiro-Wilk normality test. The β values fail to show a normal distribution (e.g. cg07979271: $p < 2.2e-16$), so a non-parametric test was used for correlation. Methylation β values were correlated against the mRNA z-scores using spearman.

2.6 Copy Number and mRNA Expression

A chart with multiple box plots (one per copy number threshold) were generated displaying the corresponding mRNA z-scores. To test for dependence of mRNA expression on copy number I performed a one-way independent ANOVA (analysis of variance).
(Methods word count: 425)

3 Results

3.1 Identification of a possible novel prognostic marker

In a multivariate Cox analysis in 1218 breast cancer patients I could identify 2078 significantly (q-value ¡ 0.05) associated genes (mRNA expression) with the overall survival. SPIB was associated with a beneficial harzard ratio (HR: 0.91; q-value: 2.82%; Fig. 1). Further, patients overexpressing SPIB have an overall median survival of 130 months, while low expressing patients survive for a median of 112 months (logrank p: 0.01; Fig. 1).

3.2 Methlyation and mRNA Expression (cg19387862)

Next, I focussed on the regulation of SPIB through methylation. Through spearman correlation, a significant negative association between methylation of SPIB and mRNA expression levels could be shown ($cor = 0.55; p = 4.11 \cdot 10^{-44}$; Fig. 2).

3.3 Copy Number Alteration and mRNA Expression

Another way mRNA expression could be modified is through copy number alteration. As seen in Fig. 3, no significant difference between the copy number levels could be shown (one-way independent ANOVA; $p > 0.05$).
(Results word count: 146)

4 Conclusion

SPIB is an interesting gene that is currently under explored in breast cancer. It could already be demonstrated to be a viable candidate gene in lung cancer. This essay shows a possible positive assotation of SPIB with survival in breast cancer.
(Conclusion word count: 41)

5 Figures and Tables

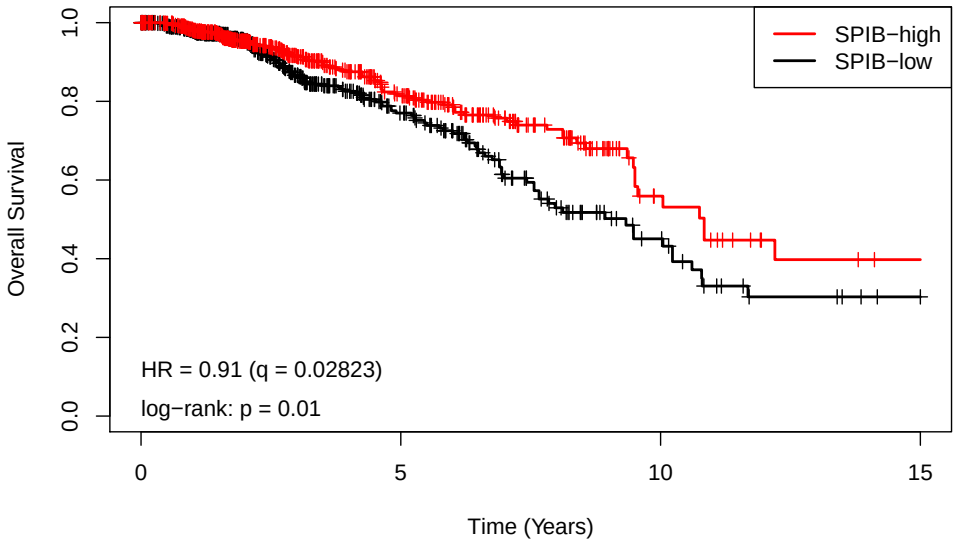

Figure 1: Survival of SPIB-high and SPIB-low breast cancer patients. Data from the TCGA BRCA dataset [4, 5].

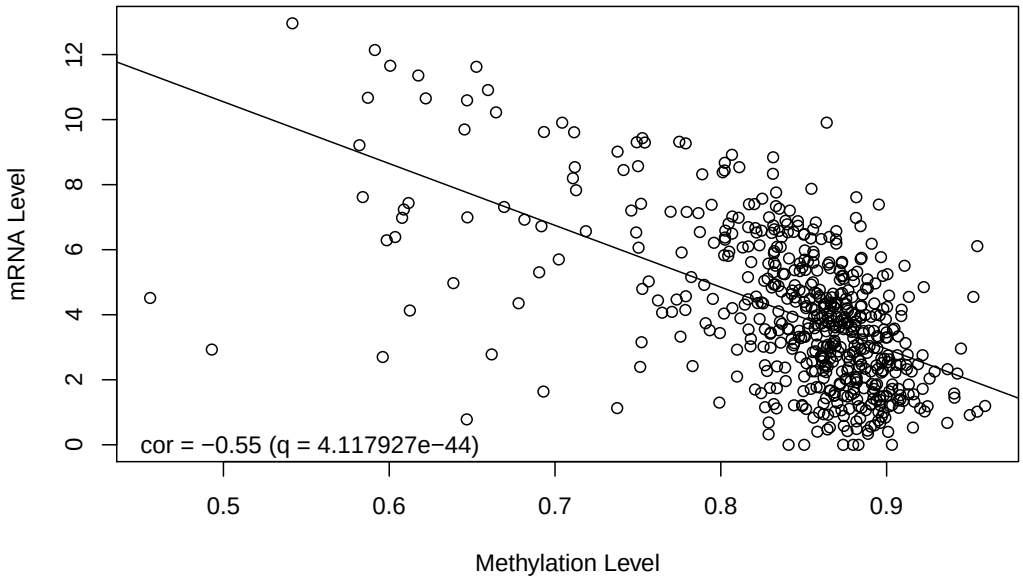

Figure 2: Scatter plot of SPIB methlyation vs mRNA expression values. Spearman correlation shown. Data from the TCGA BRCA dataset [4, 5].

Table 1: Methylation Sites

CpG	Cor	p-value	q-value	Mean (SPIB-High)	Mean (SPIB-Low)
cg07979271	-0.514	5.55e-60	9.43e-59	0.816	0.864
cg13403724	0.288	3.8e-18	3.23e-17	0.248	0.206
cg17774764	-0.281	3.07e-17	1.74e-16	0.737	0.775
cg03763616	-0.277	7.68e-17	3.26e-16	0.832	0.859
cg19387862	0.272	2.7e-16	9.2e-16	0.191	0.168
cg15690347	0.266	1.25e-15	3.54e-15	0.383	0.301
cg08201854	0.265	1.87e-15	4.55e-15	0.26	0.226
cg15007959	0.263	3.08e-15	6.55e-15	0.253	0.203
cg18254819	0.246	1.53e-13	2.9e-13	0.235	0.212
cg24092179	0.228	1e-11	1.71e-11	0.271	0.244
cg22268231	-0.147	1.22e-05	1.88e-05	0.45	0.487
cg06512885	-0.133	7.89e-05	0.000112	0.875	0.881
cg26522743	-0.0968	0.0042	0.00549	0.394	0.424
cg21152077	0.0921	0.00653	0.00792	0.464	0.448
cg13918544	0.0845	0.0125	0.0141	0.132	0.143
cg22745102	0.0705	0.0372	0.0396	0.469	0.462
cg04508467	2.02e-05	1	1	0.696	0.686

Spearman correlation of methylation of CpG islands and the mRNA expression of Spi-B.

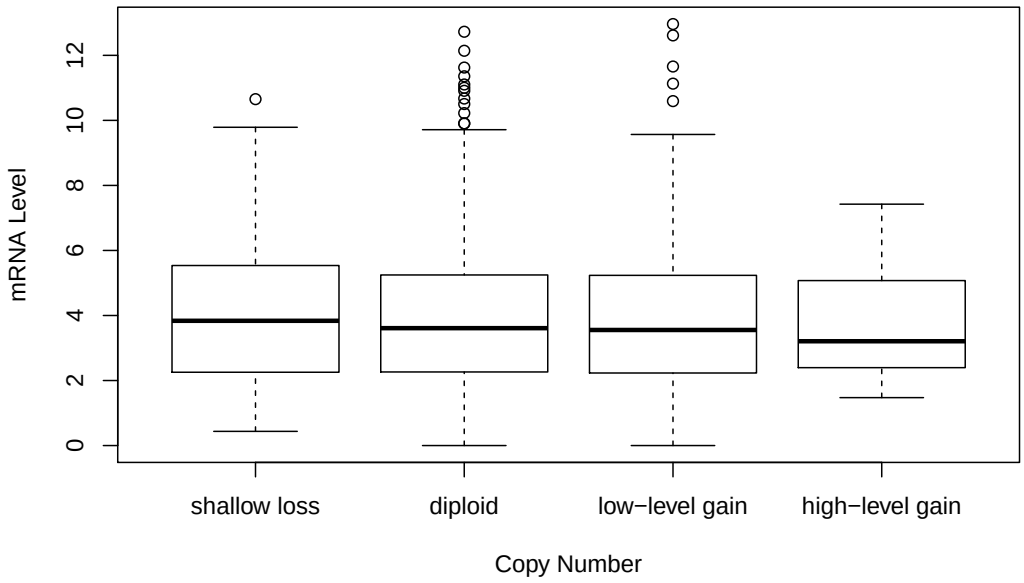

Figure 3: mRNA expression of SPIB compared under different copy number alteration levels. Data from the TCGA BRCA dataset [4, 5].

References

[1] Q. Huang, J. Liu, S. Wu, X. Zhang, Z. Xiao, Z. Liu, and W. Du, "Spi-B Promotes the Recruitment of Tumor-Associated Macrophages via Enhancing CCL4 Expression in Lung Cancer," *Frontiers in Oncology*, vol. 11, p. 1877, 2021.

[2] H. Zhang, G. Wang, R. Zhou, X. Li, Y. Sun, Y. Li, W. Du, X. Yan, J. Yang, X. Chang, Z. Liu, and Z. Ma, "SPIB promotes anoikis resistance via elevated autolysosomal process in lung cancer cells," *The FEBS Journal*, vol. 287, no. 21, pp. 4696–4709, 2020.

[3] X. Zhao, L. Li, S. Yuan, Q. Zhang, X. Jiang, and T. Luo, "SPIB acts as a tumor suppressor by activating the NFkB and JNK signaling pathways through MAP4K1 in colorectal cancer cells," *Cellular Signalling*, vol. 88, p. 110148, Dec. 2021.

[4] G. Ciriello, M. L. Gatza, A. H. Beck, M. D. Wilkerson, S. K. Rhie, A. Pastore, H. Zhang, M. McLellan, C. Yau, C. Kandoth, R. Bowlby, H. Shen, S. Hayat, R. Fieldhouse, S. C. Lester, G. M. K. Tse, R. E. Factor, L. C. Collins, K. H. Allison, Y.-Y. Chen, K. Jensen, N. B. Johnson, S. Oesterreich, G. B. Mills, A. D. Cherniack, G. Robertson, C. Benz, C. Sander, P. W. Laird, K. A. Hoadley, T. A. King, TCGA Research Network, and C. M. Perou, "Comprehensive Molecular

Portraits of Invasive Lobular Breast Cancer," *Cell*, vol. 163, pp. 506–519, Oct. 2015.

[5] M. Goldman, B. Craft, M. Hastie, K. Repecka, A. Kamath, F. McDade, D. Rogers, A. Brooks, J. Zhu, and D. Haussler, "The UCSC Xena platform for public and private cancer genomics data visualization and interpretatio," tech. rep., 2018. Type: article.