Dissolving heterogeneity in cancer

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG

vorgelegt von
Katharina Meyer
aus
Bonn

im Jahr 2013

Das Promotionsgesuch wurde eingereicht am:
30.01.2013

Die Arbeit wurde angeleitet von:
Prof. Rainer Spang

Unterschrift:

Contents

Summary

Cancer - this one word stands for a vast variety of complex diseases. Cancers are named according to the organ they originate from, like breast cancer, colon cancer or brain cancer. However, when analyzing the molecular phenotype of the cells of theses different cancer types one can observe that each type itself consists of many different diseases. The genetic aberrations leading to malignant growth can vary from case to case and so can the molecular mechanisms that a tumour cell uses to facilitate de-differentiation and progression. To the eye of a molecular biologist, cancers are very heterogeneous.
Modern treatment strategies target specific oncogenic mechanisms, like the specific inhibition of a signaling pathway. In view of the molecular heterogeneity of cancers, it cannot be expected that all patients will benefit from this treatment. Some cancers depend on the activity of a pathway and will hence respond to its inhibition, while others might not. To use targeted therapies effectively, we must first understand the heterogeneity of tumours.
A modern genomics based approach to classify cancers is gene expression profiling using microarrays or high throughput sequencing. Expression profiles monitor the expression of virtually all genes in a tumour simultaneously. They thus characterize its molecular phenotype at a high resolution. Importantly, the oncogenic events, like aberrations in the genome and associated perturbations of signaling pathways, shape these profiles. Hence, classifying tumours by expression profiles means classifying them by the underlying mutational changes that distinguishes them from healthy tissue.

The tumour profiling projects, covered in my thesis, are carried out in close collaboration with various experimental groups: Cell of origin expression signatures were developed in collaboration with the group of Christoph Beier from the Department of Neurology, University of Regensburg. The finding that specific genetic perturbations direct the development of different brain tumour types was discovered in collaboration with the group of Ulrike Nuber from the Department of Laboratory Medicine of Lund Strategic Research Center for Stem Cell Biology. The characterization of signaling pathways in human transformed GC B cells and their activity in lymphomas was conducted in collaboration with the group of Dieter Kube from the Department of Haematology and Oncology, University Medical Center Göttingen.

My contribution was the statistical analysis of the expression profiles. I designed experiments, analyzed the quality of expression data, identified biomarkers and developed gene expression signatures allowing for clinical distinction of different cancer types. For building gene expression signatures I added a novel statistical method to the pool of existing tools.

Gene expression signatures combine the expression levels of multiple genes to allow for highly resolved and robust molecular tumour classification. In this work I defined multiple gene expression signatures: One that allows the clinical distinction of brain tumour types and others that allow the measurement of pathway activities in lymphomas. For building gene expression signatures I developed a novel statistical method, which combines the expression of multiple co-expressed genes to a single real valued index per tumour sample.

My thesis addresses the molecular heterogeneity of two tumour types.
The first type are brain tumours. They can develop from different cells of origin. While these cells of origin are all tumour stem cells, they nevertheless differ as do the tumours they initiate. I describe how different cellular origins of glioblastomas are reflected in their expression profiles and I develop a cell of origin gene expression signature. Moreover, genomic aberrations like the mutation of oncogenes determine the tumour phenotype and thus shape its expression profile. I describe how the temporal order of these genetic events contributes to the heterogeneity of brain tumours, and analyze the characteristic expression profiles of the different subtypes.

The second tumour type are lymphomas. Lymphomas represent a heterogeneous group of cancers affecting cells of the immune system. 95% of lymphomas are B cell lymphomas and evolve during B cell differentiation. Differentiation includes crosstalk of different signaling pathways. Most malignant B cell lymphoma use key factors of these pathways to generate survival and proliferation signals. The activity of the dysregulated pathways differs between lymphoma types and thus shapes their expression profiles differently. Two aggressive forms of lymphomas are Burkitt's Lymphoma (BL) and Diffuse Large B Cell Lymphoma (DLBCL). I characterize different patterns of gene expression changes by B cell specific stimuli and associate these with different pathway activities in DLBCL and BL. Furthermore I was able to show that LEF1 is aberrantly expressed in BL cells. I identified LEF1 target genes and used these genes to build an LEF1 activity index. This index can be used to measure LEF1 activity in lymphoma samples. We showed that LEF1 activity is higher in Burkitt Lymphomas as compared to Diffuse Large B-cell lymphomas.

Introduction

Cancer has become a leading cause of death. According to the World Health Organization (WHO) cancer resulted in 7.6 million deaths worldwide in 2008 [50]. The biologist Theodor Heinrich Boveri (1862-1915) was the first one to suggest that cancer might arise from one single cell that was not able to separate chromosomes accurately [18]. Based on his studies he stated: "A malignant tumour cell is [...] a cell with a specific abnormal chromosome constitution" [19]. Today it is widely accepted that cancer is a genetic disease, caused by the deregulation of proto-oncogenes and tumour suppressor genes in a cell.

Genes expressed in a normal cell as well as in a tumour cell define its type and behavior. Processes like cell survival, apoptosis and functional responses are regulated. A major control point for this regulation is transcriptional initiation. It is a highly regulated process which depends on timing and on levels of expression of many genes. In cancer such regulations are often changed or destroyed and so proto-oncogenes and tumour suppressor genes are deregulated. Thus, a deeper understanding of gene expression and gene regulation is essential for understanding normal development and diseases like cancer.

One tool to monitor gene expression is the microarray technology [133]. Microarrays make it possible to measure gene expression of thousands of genes simultaneously. They are used to study whole genome transcription levels. Detection of transcripts is based on hybridization of two cDNA strands. Microarrays consist of thousands of spots that include different DNA sequences each interrogating a particular gene. Analyses included in this thesis all make use of oligonucleotide arrays. Probes of oligonucleotide arrays are synthesized directly on the chip. Oligonucleotides are about 25 nucleotides long. There are about 35 probes per gene on the microarray. Each array is hybridized with cDNA or RNA from one sample which is labeled beforehand with a fluorescence dye. Signal intensities are read with a laser scanner. The signal intensity of one spot is higher the more target molecules hybridize to the complementary strands at that spot [163, 113, 133]. Thus, it is possible to make inference among samples, genes and their expression levels.

Outline

The thesis is divided into four chapters. The **first Chapter** gives an overview of computational methods used throughout the thesis. Amongst others I introduce a method to summarize gene expression values to a single representative index.
Chapter 2 and **Chapter 3** include tumour profiling projects addressing brain tumours and lymphomas. Both chapters start with a biological introduction of the respective tumour type. **Chapter 2** divides into three Sections. In **Section 2.2** I describe how the temporal order of genetic events contributes to the heterogeneity of brain tumours and I analyze the characteristic expression profiles of the different subtypes. In the **Sections 2.3** and **2.4** I describe how different cellular origins of glioblastomas are reflected in their expression profiles and I develop a cell of origin gene expression signature. Furthermore, I associate the different cellular origin of glioblastoma with different functions of $TGF\beta$.
Chapter 3 contains two further sections in addition to the biological introduction. In **Section 3.2** I characterize major patterns of gene expression changes by B cell specific stimuli and associate these with different pathway activities in diffuse large B cell lymphoma and Burkitt lymphoma. In **Section 3.3** I identify LEF1 target genes, allowing the measurement of LEF1 activity in lymphoma.

Chapter 1

Methods

1.1 Summarizing gene expression profiles: Indexing

Gene and protein expression are tightly regulated processes. They are controlled by internal and external conditions so that not all proteins are produced under all circumstances. Proteins which function together in a cell often have common regulatory control elements and are thus coordinately expressed. Therefore those genes are co-regulated and co-expressed.

In microarray data co-expression of genes is reflected by strong correlation between expression levels. The standard correlation coefficient (i.e., the dot product of two normalized gene expression vectors) conforms well to the biological notion of the meaning of two genes being co-expressed. That is because the standard correlation coefficient captures similarity in shape of the gene expression, but places no emphasis on the gene expression magnitude.

I developed a method *getIndex* to summarize gene expression values of co-expressed genes to a single real valued index per sample.

The *getIndex* function uses a gene expression matrix as input, with expression values of genes in rows and samples in columns.
Additionally the function can take a *sign* vector, which assigns how a gene is meant to react to a specific signal ($+1$ = up regulation, -1 = down regulation). This information is not mandatory, but can be used if it is known a priori. If a *sign* vector is given to the method the resulting index will represent the activity of the associated signal.

The method can be divided into four main steps:

1. A gene-wise scaling of the data: All rows (genes) are normalized to mean zero and standard deviation of 1. This is done in order to make gene expression values comparable to each other.

2. A reference gene is determined, showing the highest sum of correlation coefficients to all other genes.
   If there is no *sign* vector given to the method, all gene expression values of genes anticorrelated to the reference gene are multiplied by -1. Afterwards, the gene expression values of all genes show positive correlation to the gene expression values of the reference gene.
   If there is a *sign* vector given to the function, the gene expression values of genes, that are known to be down regulated by the associated signal are multiplied by -1. Thereby the resulting gene expression values reflect the activity of the signal associated with the *sign* vector.

3. The median polish procedure is used to decompose expression levels into into several additive parts:

$$y_{ij} = \mu + g_i + s_j + r_{ij}$$

   $i \in \{1, ..., n\},\ n = Number\ of\ genes$
   $j \in \{1, ..., m\},\ m = Number\ of\ samples$

   The gene expression values y_{ij} are split into an *overall gene expression value* μ, a *gene effect* g_i, a *sample effect* s_j and r_{ij}. Where r_{ij} are the residuals of y_{ij} in the additive model.

   Median polishing is an iterative process. The procedure works by subtracting the medians of each row (gene) from each element in the row, then the medians of each column (sample) from each element in the column.

The row medians are recorded in the *gene effect* vector and column medians in the *sample effect* vector. This subtraction process is repeated until convergence.

The medians of the *gene effect* and *sample effect* vectors are then added up to give the *overall gene expression value* μ and subtracted from each entry in the *gene effect* vector and each entry in the *sample effect* vector resulting in a *gene effect* vector g and a *sample effect* vector s.

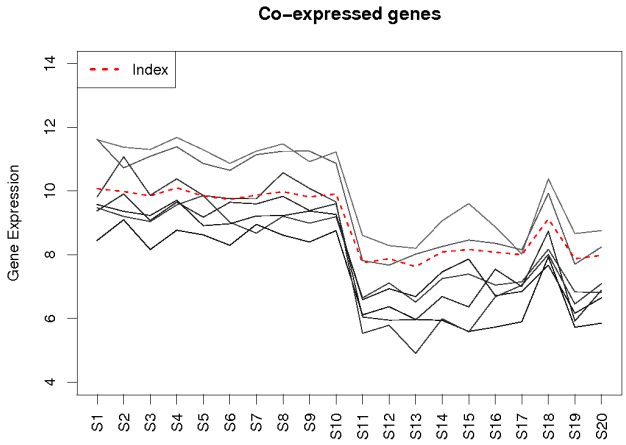
The procedure returns the *overall gene expression value* μ, the *sample effect* vector s, the *gene effect* vector g and the estimated residuals. Because the median polish procedure uses medians rather than means it is robust to outliers.

Note that the median polish procedure to decompose gene expression data is also used to compute probe-set summaries in the popular normalization package RMA for Affymetrix gene chips [84].

4. The *getIndex* method returns two variables. The first variable is the sample index s, which is the sample effect estimated by the median polish procedure.
   The second variable is the residual matrix R containing the residuals r_{ij} produced by summarizing the data with the median polish procedure.

One should keep in mind that if the data does not fit the model, information is lost during summarization. Hence data that does not fit the model should not be summarized.

An example for a co-expressed gene cluster can be seen in Figure 1.1 (a), in contrast to the gene cluster in Figure 1.2 (a).
The genes in Figure 1.1 (a) are highly correlated (Figure 1.1 (b)). The genes in Figure 1.2 (a) are chosen randomly and are mainly uncorrelated (Figure 1.2 (b)). The red line in both Figures show the resulting index, when summarizing the corresponding genes.

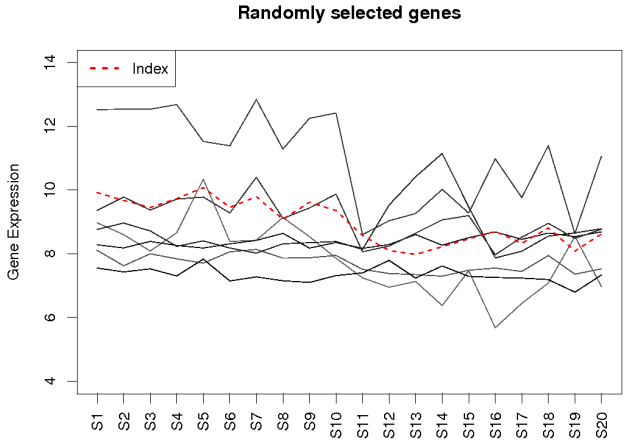


(a)

	S Gene 2	S Gene 3	S Gene 4	S Gene 5	S Gene 6	S Gene 7
S Gene 1	0,97	0,96	0,94	0,94	0,96	0,93
S Gene 2	1,00	0,96	0,96	0,95	0,97	0,95
S Gene 3		1,00	0,94	0,94	0,96	0,94
S Gene 4			1,00	0,95	0,92	0,97
S Gene 5				1,00	0,92	0,94
S Gene 6					1,00	0,91

(b)

Figure 1.1: (a) Gene expression of 7 co-expressed genes in 20 samples, and index values after summarizing the 7 genes in red. (b) Correlation coefficients of the seven selected genes (S Genes)



(a)

	R Gene 2	R Gene 3	R Gene 4	R Gene 5	R Gene 6	R Gene 7
R Gene 1	0,52	0,47	-0,29	-0,42	0,66	0,06
R Gene 2	1,00	0,82	-0,55	-0,15	0,66	-0,06
R Gene 3		1,00	-0,39	-0,21	0,78	0,10
R Gene 4			1,00	-0,02	-0,12	0,32
R Gene 5				1,00	-0,37	-0,48
R Gene 6					1,00	0,12

(b)

Figure 1.2: (a) Gene expression of 7 randomly selected genes in 20 samples, and index values after summarizing the 7 genes in red. (b) Correlation coefficients of the seven random genes (R Genes)

When summarizing gene expression values one should consider that information is lost. If the lost information is too great, then summarizing does not make sense.
How munch information is lost can be seen by looking at the residuals produced by the median polish. The residuals produced by summarizing gene expression values of the two given gene clusters in Figure 1.1 (a) and Figure 1.2 (a) are shown in Figure 1.3. As one can see, the randomly chosen genes produce very high residuals. In contrast to that the highly correlated gene cluster produces residuals located around zero. Thus by summarizing gene expression values of uncorrelated genes the loss of information is too great and one should better look at the genes separately.

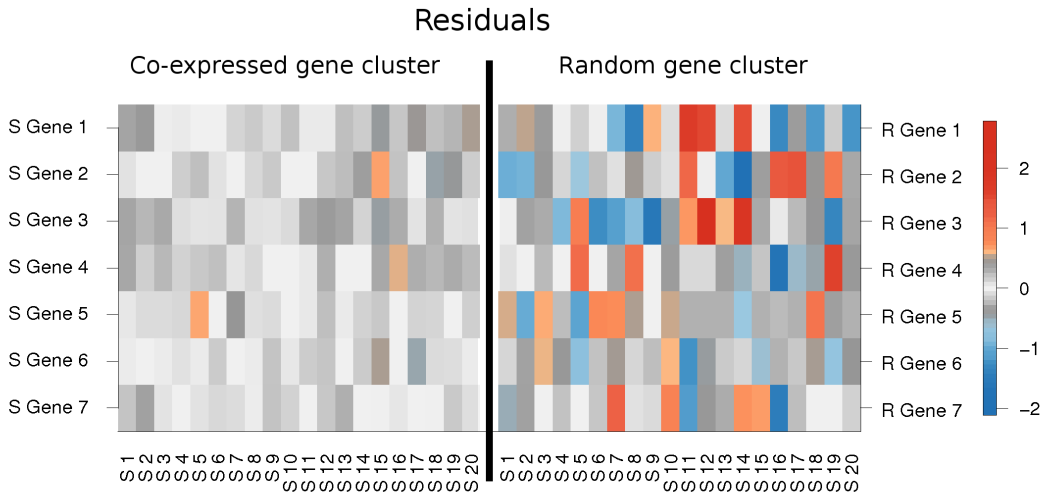


Figure 1.3: Matrix of residuals produces by the median polish summarization. The rows are the genes, where R Genes are the randomly chosen genes and S Genes are the selected highly correlated genes. S stands for Sample.

In conclusion *getIndex* provides a possibility of combining gene expression values to a single real valued index per sample. This index can be used as a representative for the summarized gene cluster. But if genes are not co-expressed, then summarization leads to a loss of too much information and can be misleading.

1.2 Explorative methods

1.2.1 Hierarchical Clustering

The goal of clustering is to partition the observed data set, $D = e_1, e_2, ..., e_N$, into a set of non-overlapping clusters. A K-cluster partition P of D can be specified as $P = P_1, P_2, ..., P_K$, such that $\bigcup_{k=1}^{K} P_k = D$, and $P_i \cap P_j = \emptyset, \forall_{i,j} : i \neq j$.
If we want to cluster N samples ($e_i \in D$) or experiments for example and the features are the genes whose expression is measured in each of the N experiments [135].

Hierarchical Clustering is a method widely used to split objects into groups also called clusters, so that objects within a cluster are more similar to one another than to objects from other clusters. Hierarchical Clustering is for example used to form descriptive statistics to ascertain whether or not the data consists of distinct subgroups, each group representing objects with substantially different properties.

An appropriate metric is necessary for the method, to measure the similarity/distance between pairs of objects, and a linkage criterion, which specifies the similarity of sets as a function of the pairwise distances of objects. In the following chapters I use the euclidean distance as distance measure. The clustering algorithm takes as input the already calculated $N \times N$ distance matrix M, where N is the number of objects, and each element in the matrix M_{ij} records the proximity between the ith and jth object $dist(i, j)$ [56].

Hierarchical clustering starts with n objects and n clusters, so each object is assigned to its own cluster. Then iteratively the two most similar clusters are joined into a new cluster. That is done until only one cluster remains. Thus a hierarchical representation is built, in which the clusters at each level of the hierarchy are created by merging clusters of the previous level. At each step distances between the clusters are recomputed, depending on the linkage criterion specified. As linkage criterion I use the complete linkage method [56].

Complete linkage computes the distance between two clusters by taking the maximum distance between a pair of clusters. Thus distance $M(X, Y)$ between cluster X and cluster Y, can be described by the following expression: $M(X, Y) = max(dist(x, y))$, where $dist(x, y)$ is the distance between elements $x \in X$ and $y \in Y$. Complete linkage tends to build compact clusters of approximately equal diameters [39].

The result of a hierarchical clustering can be visualized as a dendrogram or tree whose branch lengths reflect the degree of similarity between the clusters, as assessed by a

pairwise similarity function. By not choosing a priori a specific number of clusters, hierarchical clustering provides an opportunity for a multi-resolution view of the data. This is useful in exploratory data analysis [56].

1.2.2 Consensus Clustering

A problem in cluster analysis is to validate the clustering result. Consensus clustering is a method to assess the "stability" and robustness of clusters to sampling variability [135]. The basic assumption of consensus clustering is that for two sets of data, each representing a sample of objects from the same population, the clustering result should not be radically changed. Thus, the more stable the clusters are to sampling variability, the more confident we can be that the found clusters represent real structure. Consensus clustering gives a consensus across multiple runs of a clustering algorithm and assesses the stability of the discovered clusters [135].

Before applying the method one has to select a resampling scheme and a clustering algorithm. The result of consensus clustering is a consensus matrix which quantifies the agreement among the clustering runs over the given dataset D. A consensus matrix is defined as a $N \times N$ matrix, that stores for each pair of samples (experiments) the proportion of clustering runs in which these two samples are clustered together.

The first step of the method is to resample the original dataset D H-times. This is done by drawing H-times a specific percentage of samples from the available samples. Let the obtained new datasets be denoted as $D^{(1)}, D^{(2)}, \ldots, D^{(H)}$. Then for each of these datasets a clustering is done each resulting in a $N \times N$ connectivity matrix, $M^{(h)}$. The entries of the $M^{(h)}$ matrices are defined as follows:

$$M^{(h)}(i,j) = \begin{cases} 1, & \text{if samples } i \text{ and } j \text{ belong to the same cluster} \\ 0, & \text{otherwise} \end{cases}$$

Additionally the calculation of a $N \times N$ indicator matrix, $I^{(h)}$ is needed. This matrix indicates which samples were included in the different runs. Thus the (i,j)-th entry of $I^{(h)}$ is 1 if the samples i and j are present in the dataset $D^{(h)}$, and 0 otherwise.

In the last step consensus clustering calculates the consensus matrix $\mathcal{M}$, which is defined as the normalized sum of the connectivity matrices $M^{(h)}$.

$$\mathcal{M}(i,j) = \frac{\sum_h M^{(h)}(i,j)}{\sum_h I^{(h)}(i,j)}$$

The consensus matrix $\mathcal{M}$ is symmetric and its elements are real numbers between 0 and 1. The entry (i,j) in the consensus matrix $\mathcal{M}$ represents the number of times the samples i and j are assigned to the same cluster divided by the number of times both samples were included in a set $D^{(h)}$. 0 values are recorded when a sample pair never fell into the same cluster and 1 is recorded for sample pairs that always fell into the same cluster.

If the samples of the matrix are rearranged, in such a way that samples belonging to the same cluster are next to each other, a perfect consensus clustering would transform into a block-diagonal matrix with all the values of the matrix being either 0 or 1 [135].

Based on the consensus matrix, summary statistics can be defined accounting for the stability of a given cluster and additionally for the single members of a cluster. For each cluster $k \in K$, a cluster consensus $m(k)$ is defined. Furthermore for each sample $e_i \in D$ and each cluster k a consensus $m_k(i)$ is defined. These statistics can be used to validate cluster stability and to identify the more representative samples within each cluster.

The cluster consensus $m(k)$ can be defined as the average consensus index between all pairs of samples belonging to the same cluster and can be computed as follows:

$$m(k) = \frac{1}{N_k(N_k-1)/2} \sum_{\substack{i,j \in I_k \\ i \leq j}} \mathcal{M}(i,j),$$

where $I_k^{(h)} = j : e_j \in k$ is the indicator matrix and N_k is the number of samples in cluster k. For a perfect consensus clustering the cluster consensus's $m(k)$ would be 1 for all k's. The sample consensus $m_i(k)$ measures the average consensus index between sample e_i and all other samples in cluster k.

The sample consensus can be defined for each sample e_i and each cluster k as follows:

$$m_i(k) = \frac{1}{N_k - 1\{e_i \in I_k\}} \sum_{\substack{j \in I_k \\ j \neq i}} \mathcal{M}(i,j),$$

where N_k is the number of samples in cluster k and $1\ \{condition\}$ is the indicator function that is equal to 1 when the *condition* is true, and 0 otherwise. The sample consensus $m_i(k)$ for a perfect consensus clustering and a given k, would be 1 for all samples $e_i \in k$, and 0 for the other samples.
The cluster consensus $m(k)$ and the sample consensus $m_i(k)$ can be used to quantify the stability of each cluster [135].

The consensus clustering method is implemented in the function *consensusClustering* included in the package *compdiagTools*. It returns a consensus matrix, sample consensus and cluster consensus values. In the next chapters I use the following subsample scheme: I subsample the data 100 times, each time including 80% of the available samples.

The visualization of the consensus matrix helps to assess the cluster composition and stability of each cluster. A color gradient is associated with the values ranging from 0 to 1, so that white corresponds to 1 and dark red corresponds to 0. Thus a block-diagonal matrix corresponding to a perfect consensus clustering would be visualized as a heat map including white blocks along the diagonal on a red background. Additionally the sample consensus values are shown for each sample as row names on the left hand side of the heatmap and the cluster consensus values are plotted below the heatmap for each cluster [135].

1.2.3 Guided Clustering

Guided clustering is a method to combine experimental and clinical high-throughput data [131]. It identifies clusters of genes that stand out in the experimental data and are additionally strongly correlated in the clinical data. The clinical data consists of gene expression profiles of patients with, for example, a specific disease. The experimental data consists of gene expression profiles from a perturbation experiment containing a perturbed and a control group [131].

The method is implemented in a R-package called *GuidedClustering* [131]. It needs the following variables to be specified: *Data.t, data.g, sigma, ncluster* and *label.g.*
Data.t (T_{ij}), also called clinical data, consists of expression profiles from patients.

Data.g (G_{ij}), also called the guided data, includes expression profiles from a perturbation experiment. *Sigma* specifies the global smoothing strength. *Ncluster* defines how many clusters the method should extract. *Label.g* defines the groups of the guided data (*data.g*) , where 0 represents the control samples and 1 represents the perturbed samples. How the method uses these variables and how they are defined will be explained in the following [131].

Let T_{ij} be a set of clinical expression profiles and G_{ij} a guiding dataset. The datasets include genes in rows i and samples or expression profiles in columns j. The guiding dataset includes two different types of expression profiles: Normal unperturbed celllines (controls) and celllines where a specific pathway is perturbed. Thus the samples can be labeled as perturbed (= 1) and unperturbed (= 0) resulting in a binary label vector L [131].

Guided clustering calculates a matrix of pairwise gene similarities M. The gene similarities are high if two genes are strongly correlated in the clinical data T and if their gene expression in the guided data G is strongly correlated with the label vector L.
The method starts by calculating two matrices which capture the gene similarity on both data sets separately. Then these two matrices are fused by matrix multiplication resulting in the matrix M.

The matrix A_T captures the similarity on the clinical data T. The matrix A_T is defined on the clinical data T for two genes g and $h$as follows:

$$A_T(g,h) = \exp\left(\frac{-(1-\omega)(1-max(cor(g,h),0))^2}{2\sigma^2}\right),$$

where $cor(g,h)$ is the Spearman's correlation coefficient of the expression vectors of genes g and h. The matrix A_G captures the similarity of the genes on the guiding data G. The matrix A_G is calculated as follows for one gene g:

$$A_G(g) = \exp\left(\frac{-\omega(1-cor(g,L))^2}{2\sigma^2}\right),$$

where $cor(g,L)$ is the Spearman's correlation coefficient of the expression vector of gene g and the binary label vector L. The bandwidth of the Gaussian smoothing function is specified by σ. Additionally $\omega \in 0\ldots1$ is a tuning parameter that balances the information between the dataset G and T. The method takes a vector including different values of σ as input and internally calibrates ω. The user can than choose, depending on the resulting clusters, which σ to take [131].

As already mentioned, the algorithm fuses the two different similarity matrices by matrix multiplication ($W = A_G^{1/2} * A_T * A_G^{1/2}$). Using the resulting matrix W, the number of beforehand specified (*Ncluster*) clusters are extracted. The extracted clusters include genes that are strongly correlated within the clinical data T and are differentially expressed in the perturbed dataset G.
In addition to the different clusters the method returns a consensus expression index [131]. The consensus expression index is calculated based on the model that was also used for index calculation in the first Section of this Chapter.
Guided clustering yields the opportunity to combine informations from experimental and clinical gene expression data.

1.3 The Bioconductor package Limma

Limma [176] is a Bioconductor package to find differentially expressed genes between e.g. under different conditions, in different tissues or developmental stages or at different time points. The package is designed to analyze complex experiments involving comparisons between many transcripts simultaneously. The idea is to fit a linear model to the expression values of each gene. Because empirical Bayes is used to borrow information across genes, the analysis is also stable for an experiment with a small number of samples. Within the linear model, several groups of samples can be defined and compared. It is possible to include multiple confounding effects, such as dye or batch effects. To define the linear model *limma* requires two matrices [175].

The first matrix is a *design matrix*, which defines different groups e.g. *treated*, *untreated*, *batch one* and *batch two* are defined. Each row of a *design matrix* corresponds to a sample in an experiment and each column corresponds to a coefficient to be estimated. Thus the number of coefficients is defined by the number of columns in a *design matrix*. The second matrix needed is a *contrast matrix*, which answers the question: Which comparisons are to be made? Each contrast corresponds to a comparison between the RNA targets e.g. $treated - untreated$ [176].

Mathematically a linear model can be written as:

$$E[y_j] = X\alpha_j$$

where y_j is the expression data for the gene j, X is the *design matrix* and α_j is the vector of coefficients.

The linear model is fitted to the data using the method $lmFit()$. The fitted model returned by this method includes a coefficient component with estimated values for the α_j [176].

The contrasts are:

$$\beta_j = C^T * \alpha_j$$

where C is the *contrasts matrix* [176].

The next step is to apply $contrasts.fit()$, which again returns a model including now the coefficients of interest, the estimated values for β_j. Additionally it is possible to obtain an ordered list of differentially expressed genes with logarithmic fold changes and p-values. The list contains all genes included in the analysis. The next step is to decide on a p-value threshold, genes with p-vales above this threshold are considered to be significantly differentially expressed. The decision on the p-value criterion has to be made by the individual researcher taking into account the experiment and what further analysis and experimental steps will be done [176, 175].

1.4 Gene enrichment

1.4.1 Overrepresentation analysis

Microarray experiments are used to measure the expression of several thousands of genes. First analysis steps often end in a list of differentially expressed genes. To determine common functional groups of genes an enrichment analysis for specific Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways can be carried out. GO terms are defined by the GO database [5], where genes are assigned to biological processes, molecular functions and cellular locations. KEGG [93] is a collection of pathway maps of molecular interactions and reaction networks in cells, and variants of them.

For each of these GO and KEGG terms one can compute a probability that a specific number of genes associated with one annotation term are present in a predefined genelist, e.g. of differentially expressed genes by chance. The hypergeometrical distribution is used to calculate this probability.

Lets say we have N genes and M genes are for example annotated to a specific GO term. Additionally n genes are found to be differentially expressed. Furthermore there are k genes that are differentially expressed and are also annotated to that specific GO term.

Then we can calculate the probability to find k differentially expressed genes annotated in that GO term with the following formula:

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

The probability needed is not for having a specific number k of genes but for having at most k genes. Thus the cumulative probability needs to be calculated:

$$P(X \leq k) = \sum_{y=0}^{k} \frac{\binom{M}{y}\binom{N-M}{n-y}}{\binom{N}{n}}$$

The analysis returns the GO terms or KEGG pathways that are overrepresented in a given list of genes, for example differentially expressed genes. The resulting GO terms and KEGG pathways can be used to describe the set of differentially expressed genes. I used the method *gokeggLister* implemented in the package *compdiagTools* to do such an analysis. This method returns HTML code that can be included in a HTML page. The HTML code encodes a table with all GO terms and KEGG pathways with a p-value below 0.1.

1.4.2 Gene Set Enrichment Analysis

Detecting modest changes in gene expression datasets is hard, due to a large number of variables (genes), a high variability between samples, and often a limited number of samples. The goal of Gene Set Enrichment Analysis (GSEA) is to determine modest but coordinated changes in a priori defined sets of related genes. Such a set might for instance include all genes connected to a specific pathway. [180]

GSEA considers experiments with genome wide expression profiles from samples belonging to two classes. A ranked genelist $L = g_1, \ldots, g_N$ is produced by ranking the

genes based on a score between their expression and the group distinction by using any suitable metric ($r_1 \leq \cdots \leq r_N$). For example genes can be ranked according to the fold change between two groups. Additionally an a priori defined geneset S is given. S can for example include genes encoding products in a specific pathway or sharing a specific GO category.

GSEA determines whether the genes of a geneset S are randomly distributed throughout L or can be primarily found at the top or the bottom of list L. It is expected that sets, which are related to the phenotypic distinction, will show an enrichment in either end of the list L.

The GSEA method consists of three steps. The first step is to calculate an enrichment score (ES). That is done by recording the enrichment over the ranking, by calculating the following sums:

$$P_{HIT}(S,i) = \sum_{g_j \in S, j \leq i} \frac{|r_j|^p}{N_R}$$

$$P_{MISS}(S,i) = \sum_{g_j \notin S, j \leq i} \frac{1}{N - N_H}$$

Where N_R stands for: $N_R = \sum_{g_j \in S} |r_j|^p$. The parameter p is used to control the influence of the scores. In addition, the number of genes $g_1, \ldots, g_n$ that are also in geneset S are counted, $N_H = \sum_{g_j \in S} 1$.

The ES is the maximum deviation from zero of the difference between P_{HIT} and P_{MISS}. If p is set to 0, the $ES(S)$ is reduced to the standard Kolmogorov-Smirnov statistic, when testing the ranks at genes in S against a uniform distribution.

The second step is to assess the significance of the ES. Therefore the phenotype labels are permuted 1000 times. Permuting the phenotype labels instead of the genes maintains the correlation structure of the gene expression data. For each permutation the ES is computed, which results in a distribution of random ES scores. Then the ES score for the actual data is compared with the distribution (null distribution) of the random ES scores.

The third step is to normalize the ES for the size of each geneset, yielding the normalized enrichment score (NES). In addition a false discovery rate (FDR) is calculated, corresponding to each NES, to control the proportion of false positives. Therefore the tails of the observed and null distributions for NES are compared.

GSEA is available from the following website: http://www.broadinstitute.org/gsea/. Several existing genesets including positional genesets, curated genesets, motif genesets, computational genesets and gene ontology genesets are also provided by that site. [180]

1.4.3 The Bioconductor package Ordered list

OrderedList is a Bioconductor package for comparing ranked genelists, for example two ranked genelists resulting from two knockdown experiments. The function *compareList* implemented in the package *OrderedList* searches for similarities in two ranked genelists [122].

As input two ranked genelists (G_A and G_B) of length m, starting with the strongest up-regulated genes and ending with the strongest down-regulated genes per experiment, are needed. The method computes a similarity score by comparing the two rankings of the genelists.

First the overlap between the gene lists is computed. For each rank n, $n = 1, \ldots, m$, it is counted how often genes appear in both ordered lists up to position n. The values $O_n(G_A, G_B)$ denote the size of the overlap at position n. The total overlap of position n is defined as the overlap of up-regulated genes $O_n(G_A, G_B)$ plus the overlap of down-regulated genes $O_n(f(G_A), f(G_B))$, where $f()$ refers to the flipped list with down-regulated genes on top [199]. At position n it is given as:

$$An = O_n(G_1, G_2) + O_n(f(G_1), f(G_2))$$

The weights ω are chosen to decay exponentially with rank n:

$$\omega_\alpha = exp{-\alpha_n},$$

where α is a tuning parameter. A smaller α puts more weight on genes further down the list. The similarity score S'_α is defined as the sum over all weighted overlaps [199]:

$$S'_\alpha(G_A, G_B) = \sum_1^m exp(\alpha n) A_n$$

The hole procedure is done twice, once with the original two ranked genelists and once with the second genelist reversed. Thus the reversed genelist starts with the strongest down regulated genes so that the up regulated genes in the first list are compared to the down regulated genes in the second list.

To estimate if the detected similarities are significant, *compareList* shuffles the two ranked genelists 1000 times (by default) to generate 1000 random lists. Then the method calculates similarity scores for the shuffled genelists generating a null distributions for similarity scores. From this an empirical P-values and the 95% confidence interval can be deduced [122].

CompareList yields a significant estimation for the similarity of genelists and graphically illustrates the list comparison analysis [122].

1.5 Classification: Prediction Analysis of Microarrays

The problem of classification by microarrays is that there is a large number of genes from which to predict a class, and most often a small number of samples. Prediction Analysis of Microarrays (PAM) is a statistical technique for classification of gene expression data using nearest shrunken centroids. Briefly, a standardized centroid for each class is computed. This is the average gene expression for each gene in each class divided by the within-class standard deviation for that gene. Then the method "shrinks" each of the class centroids toward the overall centroid for all classes by an amount called the threshold. To classify the method takes the gene expression profile of a new sample, and compares it to each of these shrunken class centroids. The predicted class for that new sample is the class whose centroid is closest, in squared distance to it.

As a byproduct the method identifies subsets of genes that best characterize each class [183]. PAM is available in the Bioconductor package *pamr* which includes many useful functions some are explained here [186].

Let us assume we have k classes $(1, 2, \ldots, k)$ and we want to build a classifier to predict all k classes. Given is the expression values x_{ij} for each gene $i = 1, 2, \ldots, p$ and for each sample $j = 1, 2, \ldots, n$. C_k denotes the indices of the n_k samples in class k.

The i-th component of the centroid for class k is the mean expression value for gene i in class k and can be calculated as follows:

$$\bar{x}_{ik} = \sum_{j \in C_k} x_{ij}/n_k.$$

The i-th component of the overall centroid is:

$$\bar{x}_{ik} = \sum_{j=1}^{n} x_{ij}/n.$$

The method shrinks the class centroids toward the overall centroids.
Let

$$d_{ik} = \frac{\bar{x}_{ik} - \bar{x}_i}{m_k s_i}$$

where $m_k = \sqrt{1/n_k - 1/n}$ and s_i is the pooled within-class standard deviation for gene i.

$$s_i^2 = \frac{1}{n-K} \sum_k \sum_{i \in C_k} (x_{ij} - \bar{x}_{ik})^2$$

Thus d_{ik} is a t-statistic for gene i, comparing class k to the other classes.

The method shrinks each d_{ik} towards zero giving new shrunken centroids and d'_{ik}, which is calculated as follows:

$$d'_{ik} = sign(d_{ik})(|d_i k| - \triangle)_+$$

The "+" means positive part ($t_+ = t$ if $t \geq 0$, and zero otherwise.). When the shrinkage parameter $\triangle$ increases, an increasing number of genes have all their $d'_{ik} = 0$ [183]. Genes are discarded if d'_{ik} becomes zero. The new shrunken class centroid $\bar{x}'_{ik}$ can then be calculated as follows:

$$\bar{x}'_{ik} = \bar{x}_i + m_k(s_i + s_0)d'_{ik}$$

The result depends on the shrinkage parameter $\triangle$.

Pamr.cv is a function to cross-validate the nearest shrunken centroid classifier produced by *pamr.train* and thus it helps to determine the right amount of shrinkage [186]. The method does a n-fold cross-validation to help choosing a shrinkage parameter $\triangle$.

That means the data is randomly divided into n approximately equal-size parts (including now different samples). The n parts are roughly balanced, that means the k classes are distributed proportionally among each of the n parts. Then the model is fitted n times, each time using 90% of the samples for training. The class labels of the remaining 10% of the samples are then predicted. The errors on all n parts are added together to compute an overall error.
The shrinkage parameter $\triangle$ should be chosen such that the cross-validated error is minimized [183].

To predict new samples the package includes the function *pamr.predict*. The function returns prediction information from a nearest shrunken centroid fit. As input it takes the trained classifier, a matrix of features of which predictions are to be made and a chosen shrinkage value. Then PAM classifies a new sample x_j to belong to the nearest shrunken centroid. To do so the distance between a class centroid and a sample is measured by the discriminant score:

$$\delta_k(x_j) = \sum_{i=1}^{p} \frac{(x_{ij} - \bar{x}'_{ik})^2}{(s_i + s_0)^2} - 2log\pi_k$$

where π_k is the prior probability of class k that is defined as the overall frequency of class k in the population, hence $\sum_{k=1}^{K} \pi_k = 1$. Usually π_k is set to n_k/n.
The prediction model can thus be defined as

$$f_{class}(x_j, \triangle) = r, \text{ where } \delta_r(x_j) = \min_k \delta_k(x_j)$$

The result of the classification depends on the chosen shrinkage parameter $\triangle$ [183].

1.6 Survival Analysis

1.6.1 Survival and hazard

Survival analysis attempts to answer different questions, for example: what is the fraction of a population which will survive past a certain time point? How do particular influences or characteristics (confounding factors) increase or decrease the survival time? Survival data is often censored. That means that for some patients the event of interest did not occur during the study period.

Two related functions are generally used to describe survival data: The survival probability $S(t)$ and the hazard $h(t)$.

The survival probability $S(t)$ is defined as the probability that a patient survives from the starting point of the study to a specified future time t.
The survival probability can be estimated from observed survival times by using the Kaplan Meier (KM or product-limit) method [94].

The hazard $h(t)$ is defined as the instantaneous event rate for a patient who has already survived to time t. The hazard reflects the incident event rate, while survival reflects the cumulative non-occurrence of the event [27].

1.6.2 Kaplan-Meier survival estimate

The survival probability can be estimated from observed survival times by using the KM method [94].
Suppose that in k patients events occur at distinct times $t_1 < t_2 < t_3 < \cdots < t_k$ (all in the period of follow-up). The events are assumed to occur independently of one another. Thus the probabilities of surviving from one interval to the next can be multiplied to calculate the cumulative survival probability. The probability of being alive at time t_j, $S(t_j)$, can be calculated as follows:

$$S(t_j) = S(t_{j-1})(1 - \frac{d_j}{n_j})$$

where $S(t_{j-1})$ stands for the probability of being alive at time t_{j-i}, n_j is the number of patients to be alive just before t_j and d_j is the number of events at t_j. Additionally $t_0 = 0$ and $S(0) = 1$. The value of $S(t)$ is constant between the events. Hence the estimated probability is a step function that changes values only at the times of events. The estimator uses the information from each patient for as long as they are known to be event-free [27].
The KM survival curve is a plot of the KM survival probability against time. The plot provides a useful summary of the data and can be used to compare survival times of different patient groups. The R-package *survival* provides the function *survfit* which can be used to create survival curves given either a formula (e.g. KM) or a previously fitted model.

1.6.3 The log-rank test

Let us assume two groups of patients are given, differentiating by a certain characteristic. To see if this characteristic makes a difference in survival one can plot for each of these groups a KM survival curve. The question to ask now is, if these two KM curves are "statistically equivalent". If that is the case then there is no evidence to indicate that the true survival curves are different. To test the "statistically equivalence" of two KM curves a log-rank test can be used [104].
The log-rank test is a large-sample χ^2 test, with the null hypothesis that two given groups have identical survival and hazard functions. The log-rank statistic is computed as follows: Let $j = 1, ..., J$ be the distinct times of observed events in either group. For each time j, the N_{1j} and N_{2j} are the number of patients still alive at the start of period j in the respective group. Let $N_j = N_{1j} + N_{2j}$ and let O_{1j} and O_{2j} be the observed number of events in the groups at time j. The observations at time j are defined as $O_j = O_{1j} + O_{2j}$. The log-rank statistic compares the observations O_{1j} to its expectation E_{1j} under the null hypothesis and is defined as:

$$Z = \frac{\sum_{j=1}^{J}(O_{1j} - E_{1j})}{\sqrt{\sum_{j=1}^{J} V_j}},$$

where $E_{1j} = O_j \frac{N_{1j}}{N_j}$ and the variance $V_j = \frac{O_j(N_{1j}/N_j)(1-N_{1j}/N_j)(N_j-O_j)}{N_j-1}$.
Thus the null hypothesis being tested is that there is no overall difference between the two survival curves. Under the null hypothesis the log-rank statistic is approximately χ^2 distributed with one degree of freedom [104].

1.6.4 Cox proportional hazards model

The Cox proportional hazards (PH) model [32] is a multivariate approach for analyzing survival data in medical research. It is a regression model, which describes the relation between the hazard function and a set of covariates.
The Cox model is defined as:

$$h(t) = h_0(t) \times exp(b_1x_1 + b_2x_2 + \cdots + b_px_p),$$

where the hazard function $h(t)$ is determined by a set of p covariates $(x_1, x_2, \ldots, x_p)$ and by the baseline hazard h_0. The underlying hazard function corresponds to the probability of reaching an event (dying) when all the x_i are equal to zero [20].

The aim of the analysis is to estimate the impact of the covariates x_i on survival. This is done by a maximum likelihood approach. The impact of the covariates is measured by the size of the estimated coefficient b_i. The baseline hazard function h_0 is estimated non parametrically, hence the survival times are not assumed to follow a particular statistical distribution. The covariates $(x_1, x_2, \ldots, x_p)$ act as factors on the hazard $h(t)$ at any time point t. Thus the hazard of the event in any group is a constant multiple of the hazard in any other. This proportionality implies that the quantities $exp(b_i)$ are called hazard ratios [20].

If b_i is greater than zero (the hazard ratio is greater than one), the value of the ith covariate increases, the hazard of the event increases and thus the length of survival decreases. In other words a covariate i with a positive coefficient b_i says that the hazard is higher, and thus the prognosis worse with higher values of that covariate [104].

The function *coxph* provided by the R-package *survival* fits a Cox proportional hazards regression model. The function takes as input a formula, that includes a response variable on the left of a "$\sim$" operator, and on the right the covariates. The response variable on the left side is a survival object, as returned by the *surv* function and includes the survival informations. In addition a $data.frame$ needs to be provided including the values for the covariates named in the formula on the right side [20].
The function returns a fitted model yielding coefficient estimates, standard errors and hazard ratios. Furthermore an appropriate test of significance for each coefficient is carried out by dividing the regression estimate b by its standard error $SE(b)$ (Wald statistic) [104].

In summary a Cox model is a statistical approach for exploring the relationship between the survival of a patient and several explanatory variables. A Cox model estimates the effect of each explanatory variable on survival after adjustment for other explanatory variables. In addition, it allows to estimate the hazard of death for a patient, given their prognostic variables [104].

Chapter 2

Brain tumours

Synopsis

First a brief introduction to brain tumours is given. The second Section describes how the temporal order of genetic events contributes to the heterogeneity of brain tumours and I analyze the characteristic expression profiles of subtypes. In the last two Sections I describe how different cellular origins of glioblastomas are reflected in their expression profiles. I develop a cell of origin gene expression signature differentiating between two types of glioblastoma cancer stem celllines. In addtion, I connect the two types of glioblastoma cancer celllines to different $TGF\beta$ responsiveness.

2.1 Biological introduction

Brain tumours are cells within the brain or the central spinal canal which are growing abnormally. There are more than 120 types of brain tumours. They are sub-grouped by the World Health Organization (WHO) classification system [124]. The first concept of grading brain tumours with a systematic classification scheme was introduced in 1926 by Baily and Cushing [33]. The WHO classifications follows a histological principle: Brain tumours are classified according to the type of normal tissue they most closely resemble; their presumed cell of origin. For most brain tumours the actual cell of origin is still unknown. Hence the concept of classifying the brain tumours according to their presumed cell of origin is critical. Nevertheless, there are still no other diagnostic approaches which do better. The WHO classification of brain tumours by morphology is the "gold standard" [184].

Some tumour types are additionally assigned a grade, which ranges from Grade I (least malignant) to Grade IV (most malignant). The grading estimates tumour malignancy and patient's prognosis [124, 6].

Glioblastoma Multiform (GBM) are graded by the WHO as grade IV. GBM are the most common and deadly brain tumours in adults [123]. Patients have a short live expectancy despite aggressive therapeutic approaches [42]. Glioblastomas are tumours that form from malignant glial astrocytes: A type of glial cell in the brain and central nervous system. They are the most malignant astrocytic gliomas. Glioblastoma multiform are associated with a high degree of vascularisation and angiogenesis [156]. The vast majority of glioblastomas develop in the cerebral hemispheres. They consist of cells with various morphologies. In addition glioblastoma multiform tumours consist of cells that have lost their structural and functional differentiation as normal cells. These cells are called anaplastic. The histological diagnosis of glioblastoma is usually made on the basis of two distinctive secondary features. The first one is microvascular proliferation, by which vascular cells and other vascular elements, such as vascular smooth muscle cells and fibroblasts, divide. The second feature in glioblastoma is necrosis: The presence of dying cells in living tissue. The presence of necrosis and/or micro-vascular proliferation is essential for diagnosis because these features distinguish glioblastomas from anaplastic astrocytomas [184].
GBM contain subpopulations of cells having stem cell-like properties, such as self-renewal and the ability to differentiate into specialized cell types. These cells are called cancer stem cells or tumour initiating cells [174, 83, 61]. Histological different GBM exist and they resemble different types of cancer stem cells [10].

Central nervous system primitive neuroectodermal tumour (WHO grade IV) (CNS PNET) is a malignant embryonal tumour. It got its name from its cells that did not develop properly. Thus the neuroepithelial cells appear primitive: They are small, they are round, they are densely packed and they are undifferentiated. Additionally these cells show a high proliferation rate. Immature neuroepithelial cells still have the capacity to differentiate into various neural celllineages. The PNET usually occurs in children and young adults. Boys are more often affected than girls. PNETs are often diagnosed wrongly [184]. One reason for this is that the immunophenotype and the histology of PNETs are quite similar to medulloblastomas. Additionally PNETs are far less common than cerebellar medulloblastoma. In contrast to medulloblastomas, PNETs do not show an alteration of chromosome 17 or mutations of parts of the hedgehog pathway, which are frequent features in classic medulloblastomas [184]. Additionally, in PNET the formation of rosette-like cell structures can be observed. They tend to be large and to spread throughout the CNS [6].

Atypical Teratoid/Rhabdoid Tumour (WHO grade IV) (AT/RT) is a rare, highly malignant embryonal CNS tumour. AT/RTs can occur in any area of the CNS but often involve the cerebellum. These tumours are associated with a loss of both copies of INI1, a tumour suppressor gene on chromosome region 11q11.2. They predominantly affect newborns. Prognosis of patients is dismal [6]. In patients less than 18 years old, AT/RT accounts for approximately 2% of all brain tumours. Adults are rarely affected. AT/RT contain rhabdoid cells: Cells that are rod-shaped. They also contain regions with or without primitive neuroectodermal tumour-like areas and/or cells differentiating along epithelial, mesenchymal, glial or neuronal lineages. They sometimes contain hemorrhagic (profuse bleeding) or necrotic areas. Even areas composed of small undifferentiated cells are not uncommon in AT/RT [184].

2.2 Specific genetic perturbations direct the development of different brain tumour types from postnatal neural stem/progenitor cells

This project was done in close collaboration with the group of Ulrike Nuber from the department of laboratory medicine of the Lund Strategic Research Center for Stem Cell Biology. The group of Ulrike Nuber did all the experiments. The microarray data were produced and normalized by Falk Hertwig. I performed the statistical analysis and evaluation of the data. The results are published in Cancer Research: "Definition of Genetic Events Directing the Development of Distinct Types of Brain Tumours from Postnatal Neural Stem/Progenitor Cells" by F. Hertwig, K. Meyer, S. Braun, S. Ek, R. Spang, C. Pfenninger, I. Artner, G. Prost, X. Chen, J. Biegel, A. Judkins, E. Englund and U. Nuber [76].

2.2.1 Motivation

Many of the morphological defined tumour types are still heterogeneous. They include more than one subgroup. Gene expression profiling can help to understand the underlying mechanisms leading to different histological phenotypes. In most cases, it is still unknown how cell-intrinsic and cell-extrinsic factors contribute to the development of distinct brain tumour types. Furthermore, it is not clear if the mere accumulation of genetic events or the order of them determines tumour phenotypes and to which extent determined tumour types are stable or can be changed into another distinct type.

2.2.2 Over-expression of multiple genes leads to oncogenic transformation of $p53^{-/-}$ postnatal NSC/NPCs

Neural stem cells were genetically perturbed with ten different oncogene combinations and with the empty vector. The neuronal stem cells were $p53^{-/-}$, which means that the gene $TP53$, encoding for the tumour suppressor $p53$ is deleted on both alleles. $p53$ is an important regulator of the cell cycle.
Thus cells that do not express any $p53$, like the neuronal stem cells which we used, are more susceptible to tumours. The oncogenes used for perturbing the $p53^{-/-}$ neuronal stem cells were: $HRAS$, MYC, $Bmi1$ and $Ezh2$.

All possible double perturbations and single perturbations were conducted. We transplanted the different perturbed $p53^{-/-}$ neuronal stem cells into three primary mice.

Oncogene $Bmi1$ alone, $Ezh2$ alone, the combination of $Bmi1 + Ezh2$ and the empty vector control did not give rise to a tumour upon transplantation into mice. All other perturbations led to tumour growth (in at least two out of the three mice).

2.2.3 Microarray hybridization and preprocessing

Isolated total RNA was analyzed on the Aligent 2100 Bionanalyzer (Agilent Technologies, Santa Clara, CA, USA). A generation of probes, hybridization to Affymetrix Gene 1.0 ST Array, washing and scanning were performed as suggested by the Affymetrix standard protocol. 300 ng RNA was used for the first strand cDNA synthesis. Microarray data was normalized by the robust multi-array average (RMA) [85] using the expression console software. For the analysis of the microarray data I used R [182] and Bioconductor [64].

2.2.4 Over-expression of HRAS or MYC alone or in combination leads to the development of three brain tumour types

According to histological analysis three tumour types developed: High grade gliomas, PNETs and different epithelioid tumours with abundant giant rhabdoid cells, named AT/RT-like in the following. AT/RT-like tumours developed from cells over-expressing $HRAS+MYC$. Gliomas developed from cells over-expressing $HRAS+Bmi1$ $HRAS+Ezh2$, or only $HRAS$. The PNETs developed from cells perturbed by three different combinations: $MYC + Bmi1$, $MYC + Ezh2$, and MYC on its own.

The different histological features of the three tumour types are shown in Figure 2.1. Gliomas showed features like abundant spindle-formed cells, necrosis and single giant cells. These giant cells contained nuclei in differing numbers and of different sizes and shapes. The PNETs consisted of round small cells having a large nuclear to cytoplasmic ratio. Sometimes these cells formed rosette-like structures. The AT/RT-like tumour consisted of cells of varying size with a round shape and prominent nucleoli. Several cells had a nucleus peripherally located and a large cytoplasm.[76]

The stability of the developed tumours and their self-renewal capacity was shown by performing further serial transplantations using FACS-isolated tumour cells [76].

2.2.5 Clustering highlights the three histologically different tumour types

I performed an unsupervised clustering, including the microarray gene expression data of the 26 tumour sphere samples (13 CNS PNET, 7 glioma and 6 AT/RT-like tumours). I calculated pairwise euclidean distances between all samples based on the 5000 variable genes across all the samples. Complete linkage was used to build the hierarchical clustering. The resulting clustering tree shows three groups in accordance to the histological classification (Figure 2.2).
Thus, the main differences in the data is induced by the histological tumour types. The glioma phenotype is induced by MYC and the PNET phenotype by $HRAS$.

2.2.6 Identification of different gene expression patterns for the three tumour types

I looked for genes specifically regulated in the three different tumour types. Thus I searched for genes whose expression is differentially regulated in one of the tumour types compared to the others (Figure 2.3).
I correlated the 5000 most variable genes across all samples $j = 1, \ldots, 26$ to binary vectors $(b1, b2, b3)$ of length 26, where each entry represents a tumour sample. Thus for each gene $i = 1, \ldots, 5000$ three Pearson correlations (denoted as cor) were calculated, as follows:

$$cp_i = |cor(X_i, b1)|, \text{ where } b1_j = \begin{cases} 1 & \text{if } j \text{ is a PNET sample} \\ 0 & \text{otherwise} \end{cases}$$

$$ca_i = |cor(X_i, b2)|, \text{ where } b2_j = \begin{cases} 1 & \text{if } j \text{ is a AT/RT-like sample} \\ 0 & \text{otherwise} \end{cases}$$

$$cg_i = |cor(X_i, b3)|, \text{ where } b3_j = \begin{cases} 1 & \text{if } j \text{ is a Glioma sample} \\ 0 & \text{otherwise} \end{cases}$$

The Pearson correlation coefficients of the i genes and the different labels were summarized in a correlation vector:

$$v = (cp_1,\ cp_2, \ldots,\ cp_{5000},\ ca_1,\ ca_2, \ldots,\ ca_{5000},\ cg_1,\ cg_2, \ldots,\ cg_{5000})$$

From these vector v the 1000 genes showing the highest correlation coefficients were chosen for further analysis. The expression values of these genes are shown in a heatmap in Figure 2.3.
Additionally I tested if these genes were differentially expressed between tumour types. I fitted the following linear model using the R package *limma* [176]:

$$E[y_j] = X\alpha_j,$$

where y_j is the expression data for gene j, X is the *design matrix* and α_j is a vector of coefficients for each gene j to be estimated.

The design matrix X determines how the coefficients are to be estimated. I used the following design matrix X to model the data:

$$X = \begin{array}{c} \\ S1_{Glioma} \\ \vdots \\ S6_{Glioma} \\ S1_{PNET} \\ \vdots \\ S13_{PNET} \\ S1_{AT/RT} \\ \vdots \\ S7_{AT/RT} \end{array} \begin{array}{c} \begin{array}{ccc} Glioma & PNET & AT/RT \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

The model included three different tumour types, thus 3 different coefficients for each gene j were estimated. To validate that the previously found genes are differentially expressed between tumour types, the contrast β_j were computed as follows:

$$\beta_j = C^T * \alpha_j,$$

where α_j is a vector containing the 3 estimated coefficients for each gene j and C is the following contrast matrix:

$$C = \begin{array}{c} \\ Glioma \\ PNET \\ AT/RT \end{array} \begin{array}{c} \begin{array}{ccc} C1 & C2 & C3 \end{array} \\ \begin{pmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{pmatrix} \end{array}$$

Thereby defining how the coefficients α_j should be combined to calculate the β_j. It is then tested for every gene j if the contrasts β_j are unequal zero, yielding a p-value (adjusted after Benjamini Hochberg [12]) for each gene and each contrast estimated. All previously selected, highly correlated 1000 genes showed an adjusted p-value smaller than 10^{-6}.

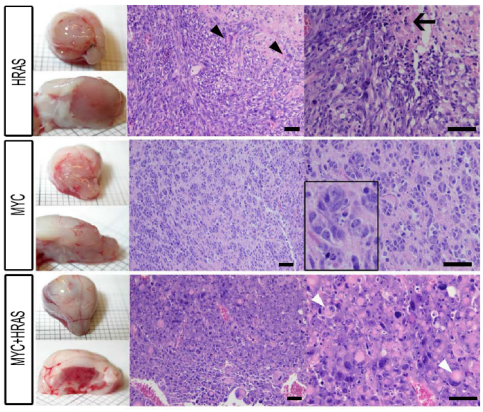


Figure 2.1: Representative macroscopic and microscopic images of Hematoxylin/eosin-stained paraffin sections of the three different brain tumour types (Scale bars: 50 μm). $HRAS$ over expression: Glioma features, necrosis (arrow), single giant cells containing multiple nuclei (arrow heads). MYC over expression: PNET features, small cells forming rosette like structure (magnified section). $MYC + HRAS$ over expression: AT/TR-like features, cells having eosinophilic cytoplasmic inclusions (arrow head). (Figure taken from a publication by Nuber et al [76].)

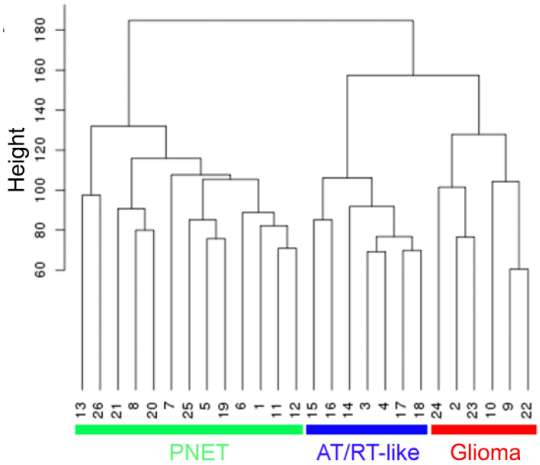


Figure 2.2: Unsupervised clustering of the gene expression data from 26 tumours which developed after primary, secondary and tertiary transplantation. The clustering was based on euclidean distances. The distances were calculated between all samples based on the 5000 most variable genes across all the samples. The labels are the internal array sample numbers, which can be found in the appendix table A1. (Figure included in publication by Nuber et al [76].

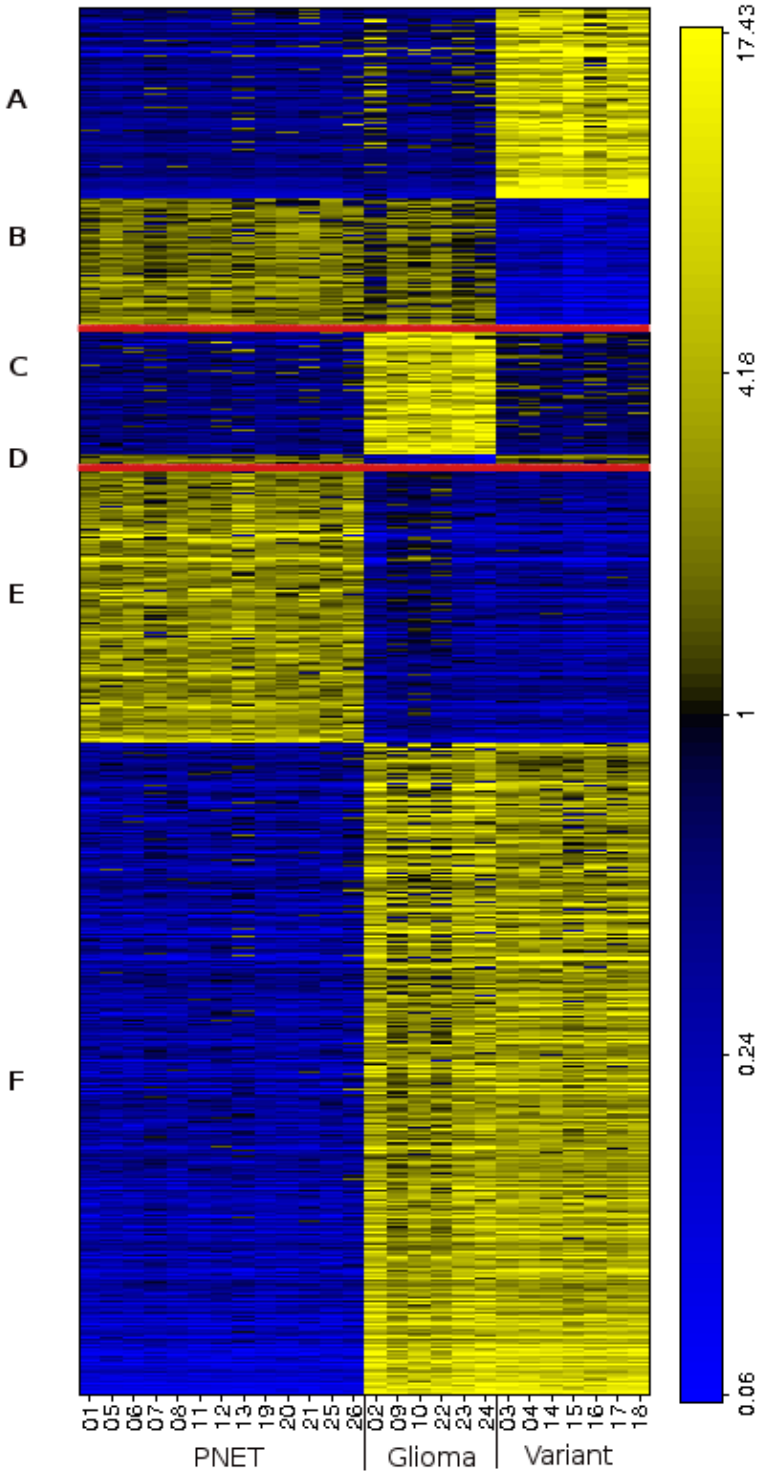


Figure 2.3: Heatmap of genes reflecting the three different tumour types. Genes were plotted along the y-axis and tumour samples along the x-axis. Expression levels: low (blue), high (yellow) (Figure included in publication by Nuber et al [76].)

2.2.7 Histological AT/RT-like tumour are related to ER-stress

AT/RT tumours are associated with a loss of both copies of the tumour suppressor gene $SMARCB1$ ($SNF5/INI1/BAF47$). But the AT/RT-like samples resembling these phenotype did not show a reduced *Smarcb*1 RNA level in comparison to the other two tumour types.
Thus we took a closer look at geneset A shown in Figure 2.3 to get a better idea of the AT/RT-like phenotype. This geneset includes genes up-regulated in the phenotype AT/RT-like in comparison to the other two phenotypes. Using a whole genome rVISTA analysis of transcription factor binding sites [54], conserved binding site motives were found 2500 bp upstream of genes from group A. Only one conserved binding site motif was significantly over-represented with an adjusted p-value smaller 0.005, namely NFE2. The sequence motif NFE2 is named after a binding site motif of the hematopoietic transcription factor $NF\text{-}E2$. But this motif is also known to bind five other transcription factors: $NRF1$-3 and $BACH1$-2. These transcription factors form together the Cap 'n' collar (CNC) family of transcription regulators [137]. Several of the genes in group A were found to be targets of the transcription factor CNC and of NF-κB. These two transcription factors are both known to play a role in the unfolded protein response (UPR) [75]. The UPR pathway is activated by endoplasmic reticulum (ER) stress. That occurs when the amount of unfolded proteins exceeds the folding capacity of the endoplasmic reticulum. From geneset A, including 134 genes, 43 genes were associated with ER stress ($32,1\%$ genes of geneset A). Triggers for the UPR pathway in tumours are for example hypoxia, glucose and amino acid starvation.
To investigate if the human AT/RT tumours can be related to ER stress, a list of previously identified genes, being up-regulated in human AT/RT and up-regulated upon deletion of *Smarcb*1 in mouse embryonic fibroblasts [86], was used. It was possible to show a relation between 35 of these genes (30.7%) to ER stress.
We were able to show that upon $SMARCB1$ reduction the phosphorylation of eIF2α increases. $EIF2\alpha$ is a central component of one of three UPR branches. Dephosphorylation of eIF2α is done by a catalytic sub-unit of protein phosphatase-1 ($PP1c$). $SMARCB1$ binds to that catalytic sub-unit of $PP1$ ($PP1c$) and additionally to the $PP1$ regulatory sub-unit 15. $SMARCB1$ increases $PP1c$ in active solution [196]. An explanation for the increase of eIF2α phosphorylation in $SMARCB1$ knockdown cells is a decreased activity of $PP1c$. But phosphorylation of eIF2α is known to enhance apoptosis and proteasome inhibition. Thus, it was tested if in AT/RT the over-activation of eIF2α could be of therapeutic use. And indeed, the treatment of $MCF7\text{-}SMARCB1$ knockdown cells with proteasome inhibitor Bortezomib leads to an increased number of apoptotic cells (determined by FACS analysis).

Such an approach could be a possible treatment for AT/RT and related malignancies with $SMARCB1$ loss-of-function. A schematic representation of the connection between $SMARCB1$ and eIF2α phosphorylation can be found in Figure 2.4.

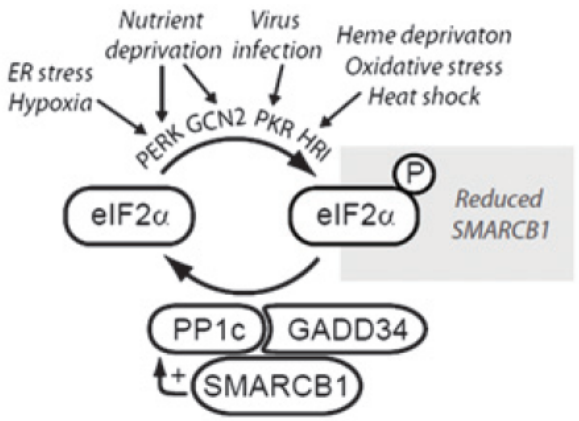


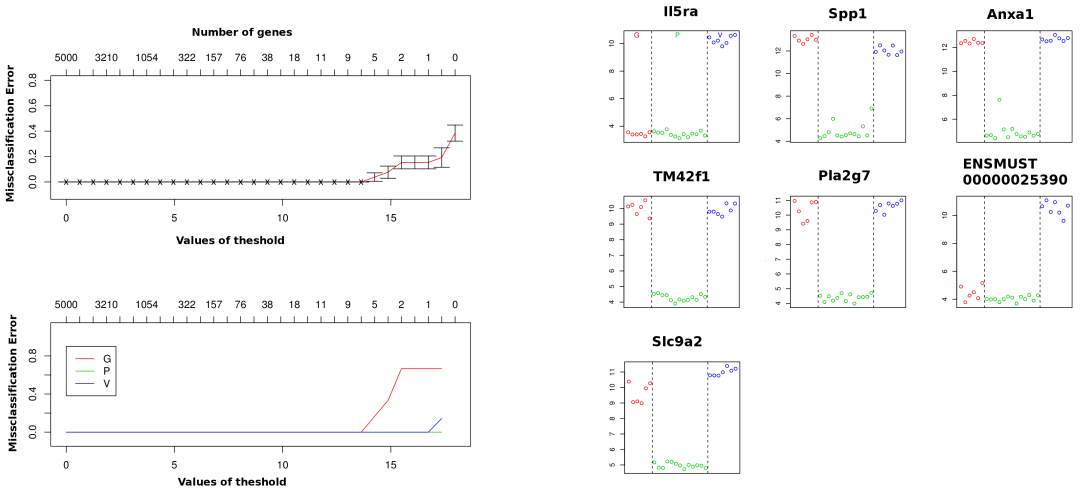
Figure 2.4: Schematic representation of the connection between SMARCB1 and eIF2α phosphorylation. (Figure included in publication by Nuber et al [76].)

2.2.8 Defining a molecular signature differentiating between three tumour types

I did a supervised classification to investigate how many genes were necessary to classify and predict the three tumour types.
Thus I trained a shrunken centroid classifier on the 5000 most variable genes across all the samples using the method implemented in the R package *pamr* [183]. As input the method got the gene expression profiles including the 5000 preselected genes and the association of samples to the 3 defined classes: PNET, AT/RT-like and Gliomas. A 10-fold cross-validation was done, splitting the data into 10 equal-sized parts, which were roughly balanced. Then the model was fitted 100 times, each time only 90% of the samples were used, the remaining 10% were used to test the resulting classifier. The errors of all 100 runs were added together to compute the misclassification error. This whole procedure is automatically done with different choices of shrinkage values $\triangle$.

In Figure 2.5 the misclassification errors are plotted against the different choices of $\triangle$, shrinkage values. The higher the shrinkage the fewer genes are included in the classifier. On top of the figures it is shown how many genes were still included in the classifier after shrinkage with the corresponding $\triangle$.

I chose a $\triangle$ shrinkage value of 13.7, where no misclassification was done in the cross-validation. That shrinkage ($\triangle = 13.7$) resulted in a classifier of seven genes. The gene expression values of these seven genes within the different tumour samples are shown in Figure 2.5 (b).



(a) (b)

Figure 2.5: (a) PAM classifier, (b) Gene expression values of the seven classifier genes plotted against the 26 tumour sphere samples. (Red: Glioma samples, Green = CNS PNET samples, Blue: AT/RT-like samples) (Figures included in publication by Nuber et al [76].)

2.2.9 The CNS PNET phenotype induced by *MYC* is unstable and can be converted into the AT/RT-like tumour type

It was possible to specify the developing tumour type by the expression of different specific oncogenes in postnatal neural stem cells. Now we examined if an established tumour phenotype can be converted into another one by a consecutive perturbation. Thus, we generated Glioma and CNS PNET tumours by transplanting 500,000 neural stem cells ($p53^{-/-}$) over-expressing *HRAS* or *MYC*. Cells from two different glioma and two different PNET were isolated. After that, the glioma cells were transduced with viral vectors over-expressing *MYC* and the PNET cells with vectors over-expressing *HRAS*. Glioma and PNET cells now over-expressing *HRAS* **and** *MYC* were picked and transplanted into mice. The resulting daughter tumours expressed both *MYC* and *HRAS*. The daughter tumours, which developed from glioma cells, retained most of their original features.

CNS PNET daughter tumours converted into AT/RT-like tumours. These converted tumours consisted of cells with a rhabdoid morphology and were positive for AT/RT-like tumour markers. Microarray gene expression profiles were done from FACS-purified tumour sphere cells from one glioma mother tumour and one CNS PNET mother tumour, and from four daughter tumours.

I used the seven gene classifier, which had been trained beforehand, to predict the new daughter tumours. I used the *pamr.predict* function implemented in the bioconductor package PAM. The function returned the prediction information, from a nearest shrunken centroid fit. For input it took the 7 gene classifier, the matrix of gene expression values of the daughter tumours and the already chosen shrinkage value of 13.5. Then PAM calculates the distances to the different class centroids by using the gene expression values of the seven genes and classifies new samples to the nearest shrunken centroid. It returned the posterior class probabilities, given in the table 2.1.

		Glioma	PNET	AT/RT
Daughter Tumour	48	0.031	**0.967**	0.001
Daughter Tumour	49	**0.967**	0.006	0.027
Daughter Tumour	50	0.000	0.000	**1.000**
Daughter Tumour	51	0.000	0.000	**1.000**
Daughter Tumour	52	**0.954**	0.021	0.025
Daughter Tumour	53	**0.935**	0.045	0.019

Table 2.1: Table of posterior class probabilities

The daughter tumours of the $HRAS$ induced gliomas, which developed upon MYC over expression were classified as gliomas. In contrast to that, the daughter tumours of the MYC induced CNS PNET, which developed upon $HRAS$ over expression were assigned to the AT/RT-like group. The expression values of the 7 classifier genes in the daughter tumours can be seen in the heatmap shown in Figure 2.6)

Thus, it was possible to show the shift from the phenotype of CNS PNET to the phenotype of AT/RT-like tumours by over expression of $HRAS$. However the phenotype of the glioma cells could not be changed, even if both daughter tumours now also expressed MYC and $HRAS$.
That means that certain *in vivo* tumour phenotypes are plastic and can be converted by an additional genetic event. Furthermore, the results show that the order of the genetic events plays an important role and directs the development of different *in vivo* tumour types (Figure 2.7).

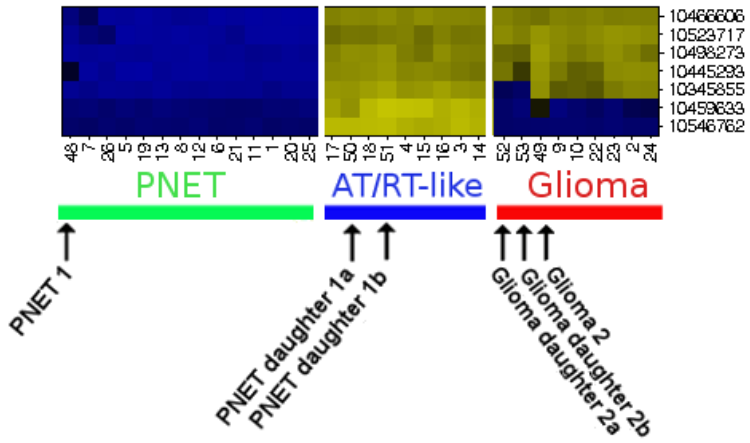


Figure 2.6: Gene expression profile of classifier genes. Rows corresponding to the seven classifier genes. Columns corresponding to sphere-cultured tumours cells samples. (Figure included in publication by Nuber et al [76].)

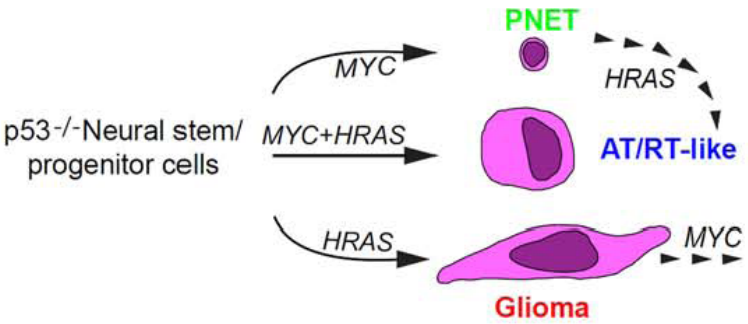


Figure 2.7: Schematic overview of consecutive perturbations (Figure included in publication by Nuber et al [76].)

2.2.10 Discussion

Several insights in the origin of brain tumours, and the role of genetic alterations in specifying tumour phenotypes were found. We showed the involvement of the UPR in tumours with loss of $SMARCB1$. We demonstrated that three different brain tumour types can develop from the same pool of $p53^{-/-}$ postnatal NSC/NPCs. Additionally we could demonstrate that the development of the three different brain tumour types is determined by the type and order of genetic perturbations. All three associated human tumour types occur in younger patients. An explanation for this could be that the cell-of-origin NSC/NPCs were extracted from young (four-week old) mice. These cells show a higher proliferative rate than NPC/NPCs from older animals.
$HRAS$ mutations are not commonly known in human gliomas, but an increase in copy numbers of RAS/RAF genes and an increase of RAS activity was found [142, 90].

MYC is known to be up-regulated in human AT/RTs [127] and in human CNS PNETs MYC and additionally $MYCN$ is often amplified [8].
The majority of AT/RT cases are connected to an inactivation of the tumour suppressor gene $SMARCB1$ [148]. There was no evidence for loss-of-function of $SMARCB1$ in the murine AT/RT-like tumours, but other similarities were found between the gene expression pattern of these tumours to the human AT/RTs and to mouse fibroblasts having no *Smarcb*1. The results suggest that different genetic events can lead to the development of similar tumour phenotypes.

In addition to already known phenotypic and genetic changes, which can appear during the development of the same tumour type, the co-existence or sequential occurrence of two different malignant tumours within the same tissue has been reported [49]. Here it was possible to show that an established *in vivo* tumour type (CNS PNET) can be transformed into another tumour type (AT/RT-like tumour) by the over-expression of a single gene. It was previously described that the insertion of the v-Ha-ras gene into a small cell lungcancer cellline which already over expressed MYC, induces *in vitro* a transition to a large cell undifferentiated lungcarcinoma, whereas no transition appeared when v-Ha-ras was inserted in small cell lungcancer celllines without MYC over expression [128]. Here we could show that a single gene led to the transition of one *in vivo* tumour type to another. Previous studies found that the order of expressing transcription factors determines the specification of normal hematopoietic cells [87]. In the case of lung neoplasia it was shown that the relative order of K-ras and APC gene mutations direct the grade, but not the tumour type [91].

Within this study, it was possible to give a proof-of-principle that the order of genetic events can direct the development of an *in vivo* tumour type. The only difference is that the system used is based on the over expression of MYC and $HRAS$ under the CMV promoter and that an *ex vivo* step is included.
MYC is known to promote cellular reprogramming in induced pluripotent stem cells which have broad developmental potential [141]. An explanation that MYC could not turn glioma cells over-expressing $HRAS$ into the AT/RT-like phenotype could be, that the binding sites of genes required for the development of the AT/RT-like tumour type are inaccessible in the glioma cells.

Furthermore, we were able to identify unique gene expression profiles per tumour type by that identifying genes that were only regulated in the murine AT/RT-like tumours. It was possible to associate these genes to an up-regulation of the UPR. Given this we were able to highlight that $SMARCB1$, a tumour suppressor primarily known for its role in chromatin remodeling, has an additional function associated to the UPR. The UPR helps to cope with conditions that lead to an increased number of misfolded proteins. This can occur in rapidly growing tumours during hypoxia and nutrient deprivation. The activation of the UPR pathway can lead to different results - an increased survival or cell death. This result depends on different branches of the UPR and the dynamics of the response. The murine AT/RT-like tumours and their expression profiles are associated to an up-regulation of the UPR. Moreover, eosinophilic cytoplasmic inclusions, another feature of AT/RT, can be found in cells that are under ER stress [197] and in some of the tumours resembling the AT/RT-like phenotype.
Due to the rapid growth of these tumour-spheres and due to the lack of glucose, the UPR is activated. Another reason for the activation of UPR is the over-expression of MYC and $HRAS$. MYC enhances protein translation, and thereby increases the ER protein load. Additionally, the mutated $HRAS$, that was used in the study, (V12 $HRAS$) particularly activates the UPR pathway [40].

The analysis of previously published gene expression profiles from human AT/RTs lacking $SMARCB1$ and Smarcb1-deficient mouse embryonic fibroblasts [159] also shows that genes related to the UPR are activated. It was shown that the $SMARCB1$ protein activates $PP1c$-$GADD34$ in solution [196]. Here it was also possible to show that reduced levels of the protein $SMARCB1$ are associated with an elevated cellular sensitivity to $eIF2\alpha$ phosphorylation, which is a central UPR mechanism. Sustained $eIF2\alpha$ phosphorylation confers cytoprotection against hypoxia, oxidative stress, and long term glucose deficiency, all of which represent typical *in vivo* conditions of rapidly-growing tumours [14, 72, 106, 138, 195].

Importantly, the application of the proteasome inhibitor bortezomib results in higher apoptosis rate of cells with reduced $SMARCB1$ levels. Bortezomib leads to an increase of ER stress and synergizes with the chemical inhibitor of the $GADD34$-$PP1c$ complex [169]. Thus, future studies should further investigate whether such an approach could be used as a novel targeted strategy for treating AT/RTs and other related malignancies.

In conclusion, the results indicate that an established *in vivo* tumour type (here CNS PNET) shifts to another tumour type (here AT/RT-like) by the over expression of a single gene (here $HRAS$). Additionally, we were able to show that the order of genetic changes can direct the specification of a tumour type. $HRAS$ over expression after MYC over expression leads to AT/RT-like tumours, whereas the reverse order (first the over expression of $HRAS$ and second MYC over expression) results in the persistence of the glioma phenotype. We defined a gene signature differentiating between the tumour types and using the signature, confirmed the shift between tumour types.
Thus, not only the cell-of-origin, genetic and epigenetic events can influence the specification of tumour types but also the order of genetic events. Furthermore, we identified tumour type specific expressed genes. These genes indicate that AT/RT and AT/RT-like tumours are associated with the activation of the UPR. We could demonstrate that the involvement of $SMARCB1$ in regulating the UPR could be therapeutically used.

2.3 Gene expression data of different glioblastoma-derived CSC indicate different cells of origin

This project was done in close collaboration with the group of Christoph Beier from the department of Neurology, University of Regensburg. I performed the statistical analysis together with Claudio Lottaz. Parts of this work are published in the article: "Transcriptional Profiles of $CD133^+$ and $CD133$ Glioblastoma-Derived Cancer Stem Celllines Suggest Different Cells of Origin" by C. Lottaz, D. Beier, K. Meyer, P. Kumar, A. Hermann, J. Schwarz, M. Junker, P. Oefner, U. Bogdahn, J. Wischhusen, R. Spang, A. Storch and C. Beier in Cancer Research [121].

2.3.1 Motivation

Many of the morphological defined tumour types are still heterogeneous. They include more than one subgroup. Additionally, in most cases it is still unknown which cell types give rise to different tumour types.
Growing evidence suggests that glioblastoma multiform develop from different cancer stem cells (CSC) or tumour initiating cells (TIC). In the following we are going to compare gene expression profiles of different possible founder cells with 17 GBM CSC lines. We address the question, whether it is possible to connect the GMB CSC lines to their corresponding founder cell.
It is already known that a subgroup of astrocytic GBM lead to $CD133^-$ CSC lines which differ in their gene expression profiles and their phenotypic properties from $CD133^+$ CSC lines. Thus, the histologically different entities of GBM might have developed from different GBM CSCs. One hypothesis is that different cells of origin exist [10] and give rise to different tumour subtypes. Postulated founder cell populations include mature astrocytes [189], unipotent "restricted" progenitor cells and astrocytic progenitor cells combined with mesenchymal stem cells (MSC) [179, 16]. Furthermore, adult neural stem cells (aNSC) from the sub-ventricular zone (SVZ) of the adult brain and fetal neural stem cells (fNSC) may be potential populations of founder cells. aNSC do not express the GBM marker $CD133$. But aNSC share many properties with their malignant counterparts [61, 174].

Recently, Phillips and colleagues determined a 35-gene signature that divided GBM into three subgroups. They connected the three types to different stages of neurogenesis, correlated with patient prognosis and described a pattern of disease progression.

Due to the physiological functions of the signature genes and their gene expression pattern, the three groups were named: "proneural", "mesenchymal" and "proliferative" [154]. Whether these three subtypes develop due to different CSC origins is still unclear.

We address this question by analyzing gene expression profiles of different founder cell populations and 17 GBM CSC lines. The CSC lines had been shown to preserve the phenotype of the original tumour and hardly acquire new mutations when propagated *in vitro* [115].

2.3.2 Microarray preprocessing and quality control

The microarray data consisted of different chip types. It was merged by matching probe sets using the matchprobes extension from the Bioconductor [64] extension for R [182]. The publicly available data was jointly normalized with our data. Gene expression values were background corrected and normalized on probe level using the variance stabilization method by Huber et al [80]. The normalized probe intensities were summarized into gene expression levels by using an additive model described in Irizarry and colleagues [84] fitted by the median polish procedure [187].

2.3.3 Identification of two different groups of GBM CSC lines

We compared 17 GBM CSC lines with different founder cell populations. As founder cell population we used gene expression profiles from three human astrocytes, three human neurons, two fNSC lines, three aNSC lines and from three MSC lines. First I did a hierarchical clustering. I calculated pairwise euclidean distances between samples based on the 500 most variable genes across all samples. Then complete linkage was used to cluster the samples. The clustering revealed three groups and is shown in Figure 2.8. The GBM CSC lines only fall in two of the three groups. One of these groups also contains the fNSC lines and the other contains the aNSC lines. The third cluster is formed by the rest of possible founder cell populations: Neurons, astrocytes and MSC. In the clustering image (Figure 2.8) the correlation between $CD133$ expression and transcriptional similarity to either aNSC or fNSC lines can be observed. All of the $CD133^{+}$ neurosphere like growing GBM CSC lines clustered together with the fNSC. In contrast to that the two $CD133^{-/+}$ semi-adherent growing and all adherently growing GBM CSC lines, except two $CD133^{-}$ clustered with aNSC [121].

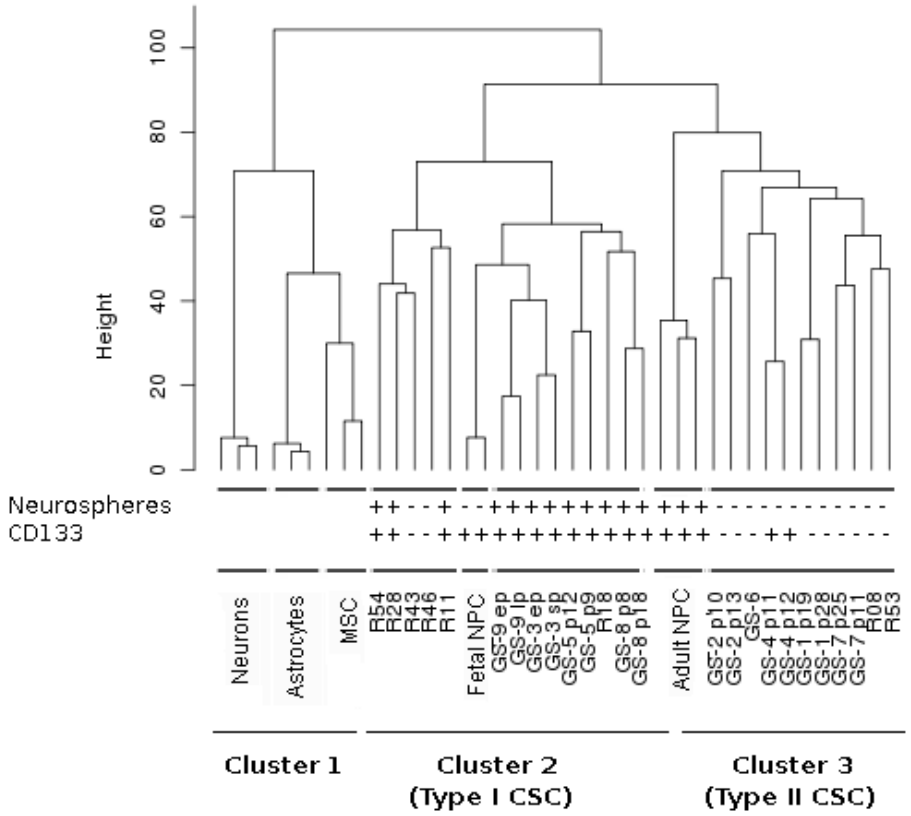


Figure 2.8: Unsupervised complete linkage clustering based on euclidean distances between samples (lp: late passage; ep: early passage; GS: Glioblastoma Sample; R: cellline from Regensburg; neurosphere: $+/-$ neurosphere formation, $CD133$: $+/-$ $CD133$ expression)

I used the consensus clustering method [135] to validate the robustness of the three clusters. The method took the gene expression matrix for input, including gene expression values of the 500 most variable genes in rows and the samples to be clustered in columns. Then the data was resampled 100-times by repeatedly drawing 80% from the given samples. After that a clustering was done for each of these datasets resulting in a connectivity matrix. Consensus clustering returns a consensus matrix and additionally sample consensus values for each sample and cluster consensus values. The consensus matrix includes real numbers between 0 and 1. The entry (i, j) in a consensus matrix represents the number of times the samples i and j are assigned to the same cluster divided by the number of times both samples were included in a resampled dataset. The result of the consensus clustering run is visualized in Figure 2.9, where a color gradient is associated with the values ranging from 0 to 1, so that beige corresponds to 1 and dark red corresponds to 0.

The consensus clustering showed that all three clusters are robust against sampling variability, all having a cluster consensus greater than 0.7

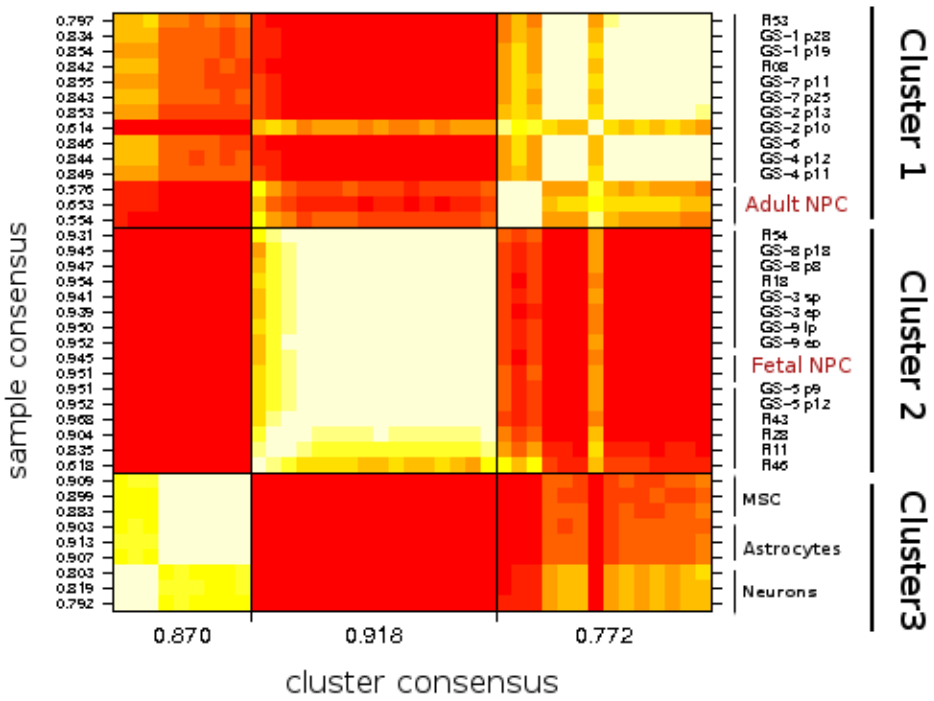


Figure 2.9: Consensus clustering result: The counts are visualized in a color-coded consensus matrix (dark red = 0 pairs that never clustered together, beige = 1 pairs that always clustered together). Additionally the sample consensus values are shown for each sample as row names on the left hand side of the heatmap and the cluster consensus values are plotted below the heatmap for each cluster

2.3.4 Associating uniquely expressed genes with the two different groups of GBM CSC lines

To get an impression of the differences between the three clusters $(C1, C2, C3)$ I searched for genes uniquely expressed by one of them. Thus I searched for genes whose expression is differentially regulated in one of the clusters compared to the other two.

I correlated the gene expression values for each gene i given in all samples $j = 1, \ldots, 39$ to binary label vectors $(b1, b2, b3)$ of length 39, where each entry j represents a sample included in one of the three clusters.

Thus for each gene i three Pearson correlations (denoted as cor) were calculated, as follows:

$$x_i = |cor(X_i, b1)|, \text{ where } b1_j = \begin{cases} 1 & \text{if } j \text{ is a sample of cluster 1} \\ 0 & \text{otherwise} \end{cases}$$

$$y_i = |cor(X_i, b2)|, \text{ where } b2_j = \begin{cases} 1 & \text{if } j \text{ is a sample of cluster 2} \\ 0 & \text{otherwise} \end{cases}$$

$$z_i = |cor(X_i, b3)|, \text{ where } b3_j = \begin{cases} 1 & \text{if } j \text{ is a sample of cluster 3} \\ 0 & \text{otherwise} \end{cases}$$

This resulted in 3 vectors (x, y, z) containing for each gene i the Pearson correlation coefficients to one of the binary label vectors $(b1, b2, b3)$. From each of these vectors I selected the 100 genes with the highest correlation coefficients, resulting in 300 genes highly correlated with one of the group label vectors. Thus I found an individual transcriptional profile per group (Figure 2.10).

We compared the transcriptional profiles of the 17 GBM CSC lines. These profiles revealed focal adhesion proteins and genes associated with the extracellular matrix. Furthermore many genes associated with $TGF\beta$ activity and $TGF\beta$/bone morphogenetic protein (BMP) signaling pathway were found up-regulated in type II GBM CSC lines compared to type I [121]. Following this line, the different types of $TGF\beta$ responsiveness were checked. It was found that all type II GBM CSC lines were responsive, and show a reduced migration and proliferation rate after treatment with $TGF\beta$. The type I GBM CSC lines displayed no altered growth, clonogenicity and migration after $TGF\beta$ treatment. In Section 2.4 I describe the role of $TGF\beta$ in more detail.

From recent comparisons of fNSC and aNSC one already knows which genes are expected as differentially expressed [130]. Among others we found the following genes: DCX, $CD44$, $PDGFR\alpha$, $EGFR$ and $CD133$. All of these markers were low or negatively regulated in aNSC and had a high expression in fNSC. The exception was $CD44$, which was down-regulated in fNSC and highly expressed in aNSC.

In the following, these markers were used to show the relationship between the two types of GBM CSC lines resulting from the clustering and the fNSC/aNSC lines. Both types of GBM CSC lines (aNSC, fNSC) express nestin as was also previously described by Beier and colleagues [10]. The type I CSC lines all express $CD133$, $PDGFR\alpha$ and $EGFR$ similar to the fNSC. They also contain DCX-positive cells like fNSC. In contrast to that the type II CSC lines express $CD44$ but not $PDGFR\alpha$, $CD133$ and they do not contain DCX-positive cells. But they also show an expression of $EGFR$. The expression of these markers vary within the different types of GBM CSC lines. But still the similar marker expression of the putative founder cells and their corresponding GBM CSC type underscores the close relationship [121].
In the following I refer to the GBM CSC lines in cluster 2 as type I CSC lines and to the GBM CSC lines in cluster 3 as type II CSC lines (Figure 2.10).

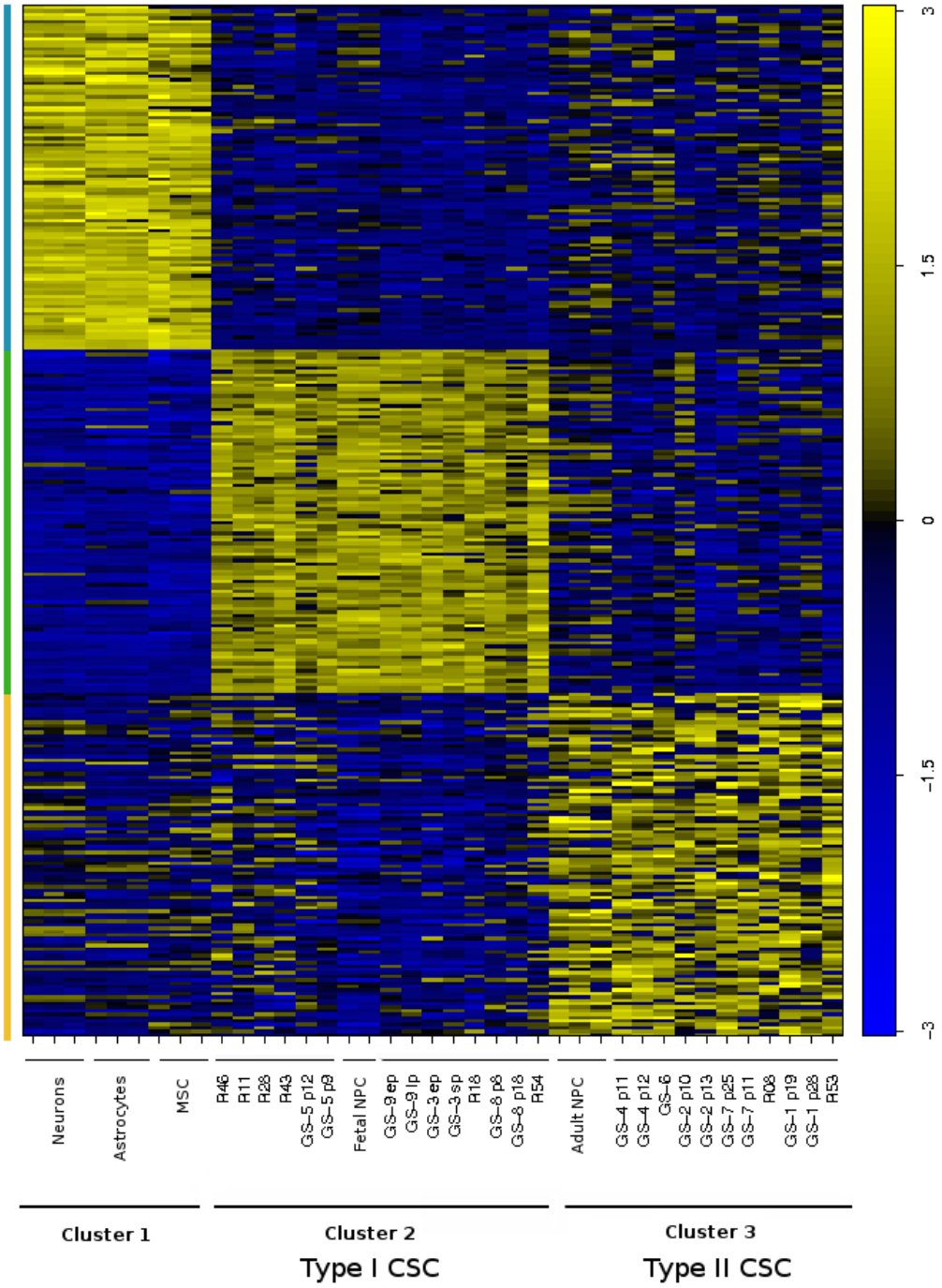


Figure 2.10: Heatmap of 300 genes. For each cluster 100 over expressed genes were found by correlation. Rows reflect genes and columns reflect samples. Expression levels: low (blue), high (yellow)

2.3.5 The difference between the two types of GBM CSC lines is independent of growth pattern

The clustering not only reflected the $CD133$ expression but also the growth pattern of the GBM CSC lines (Figure 2.8). In the following, we show that the growth pattern is not the variable driving the gene expression differences between the GBM CSC types. I included four further gene expression profiles from type I and type II CSC lines in the analysis (Type I: R18 and R54 ; Type II: R8 and R53). These CSC lines were cultured as a semi-confluent monolayer on laminin (Figure 2.12 (a)). Furthermore, I included CSC lines published by Pollard and colleagues, which had always been grown on laminin [157]. I calculated euclidean distances between all the samples based on the 2000 most variable genes across all the samples. Here, "all the samples" stands for all GBM CSC gene expression profiles grouped in cluster 2 and cluster 3 (Figure 2.8) together with the newly included gene expression profiles. I used complete linkage to do an unsupervised hierarchical clustering. The clustering result is visualized in Figure 2.11. Irrespectively of the growth pattern the new CSC lines are grouped with type I and type II GBM CSC lines.

To show that the growth pattern does not influence the transcriptional profiles, I used a linear model approach implemented in the R package *limma* [176] to identify genes that were differentially expressed between the celllines grown on laminin (L) and the corresponding celllines not grown on laminin ($nonL$). The linear model used was:

$$E[y_j] = X\alpha_j,$$

where y_j is the expression data for gene j, X is the *design matrix* and α_j is a vector of coefficients for each gene j to be estimated. The following design matrix X and contrast matrix C were used:

$$X = \begin{array}{c} \\ R54_L \\ R18_L \\ R08_L \\ R53_L \\ R54 \\ R18 \\ R08 \\ R53 \end{array} \begin{array}{c} \begin{array}{ccccc} L & nonL & P1 & P2 & P3 \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array} \qquad C = \begin{array}{c} \\ L \\ nonL \\ P1 \\ P2 \\ P3 \end{array} \begin{array}{c} \begin{array}{c} L - nonL \end{array} \\ \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{array}$$

The design matrix determines which and how the 5 coefficients for each gene j are to be estimated. The model took the two different growth patterns and the different celllines (pairs: $P1$,$P2$,$P3$) into account. The fourth pairing of the celllines, $R53$ and $R53_L$, was implicitly modeled, such that $P4$ is determined by not being $P1$, $P2$ or $P3$.
The contrast β_j were computed as follows:

$$\beta_j = C^T * \alpha_j,$$

where α_j is a vector containing the 5 estimated coefficients and C is the contrast matrix. The result yields the opportunity to see if there are any genes differentially expressed between the two growth pattern groups (L and $nonL$). No genes were found to be significantly differentially expressed between the two groups, all having p-values (adjusted after Benjamini Hochberg [12]) larger than 0.05.

Furthermore, the growth pattern of the putative founder cells were contrary to the growth pattern of the associated GBM CSC lines (Figure 2.8 and 2.12 (b)). Thus, the growth pattern did not drive the differences of the GBM CSC line profiles.

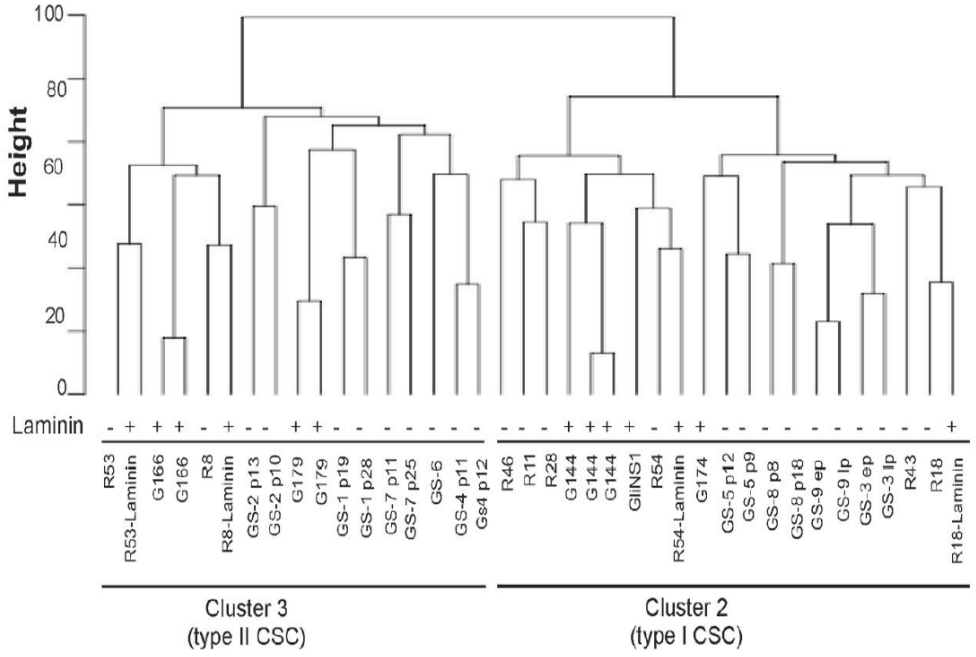
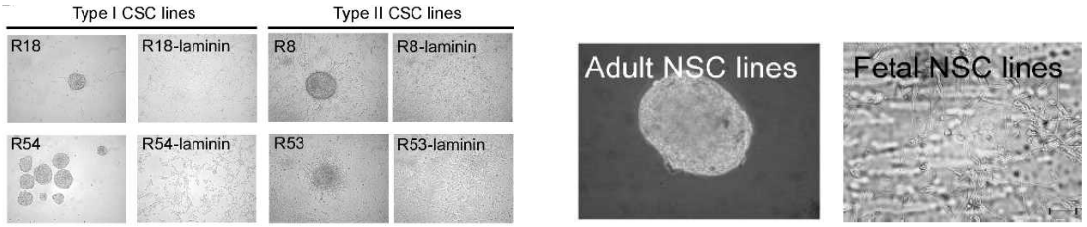


Figure 2.11: Unsupervised clustering using complete linkage. Clustering used euclidean distances calculated between all samples based on the 2000 most variable genes across all samples.



(a) (b)

Figure 2.12: (a) Growth pattern of CSC lines after seven days with or without laminin (b) Growth pattern of fNSC and aNSC

2.3.6 Defining a 24-gene-signature differentiating between the two types of GBM CSCs

We connected the two types of GBM CSC lines to $TGF\beta$ responsiveness. This might be relevant for patient selection benefiting from an anti-$TGF\beta$ treatment. But an actual patient-selection, based on the GBM subtype requires a molecularly defined gene signature. Thus, I used shrunken centroid classification, as implemented in the R package *pamr* [183], to train a classifier which discriminates between GBM CSC line type I and type II.

Thus I trained a shrunken centroid classifier on the 5000 most variable genes across all the samples. For input the method got the gene expression profiles including the 5000 preselected genes and the association of samples with the 2 defined classes: GBM CSC line type I and type II. A 10-fold cross-validation was done, splitting the data into 10 equal-sized parts, which were roughly balanced. Then the model was fitted 100 times, each time only 90% of the samples were used, the remaining 10% were used to test the resulting classifier. The errors of all 100 runs were added together to compute the misclassification error. This was done for different choices of shrinkage $\triangle$.
In Figure 2.13 the misclassification errors is plotted against the different choices of $\triangle$, shrinkage values. The higher the shrinkage the fewer genes are included in the classifier. On top of the figures the number of genes still included in the classifier after shrinkage with the corresponding $\triangle$ is shown.

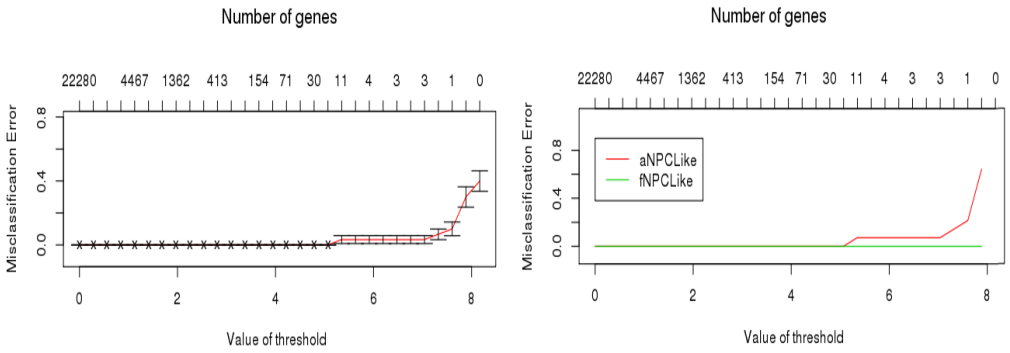


Figure 2.13: PAM results of training a classifier discriminating between GBM CSC line type I and type II. The misclassification errors is plotted against the different choices of $\triangle$, shrinkage values.

I used a $\triangle$ shrinkage value of 4.8 to identify 29 probe sets (representing 24 genes, see Figure 2.14) discriminating the GBM CSC lines with minimal cross-validation error. In the following I refer to this signature as the "24-gene-signature".

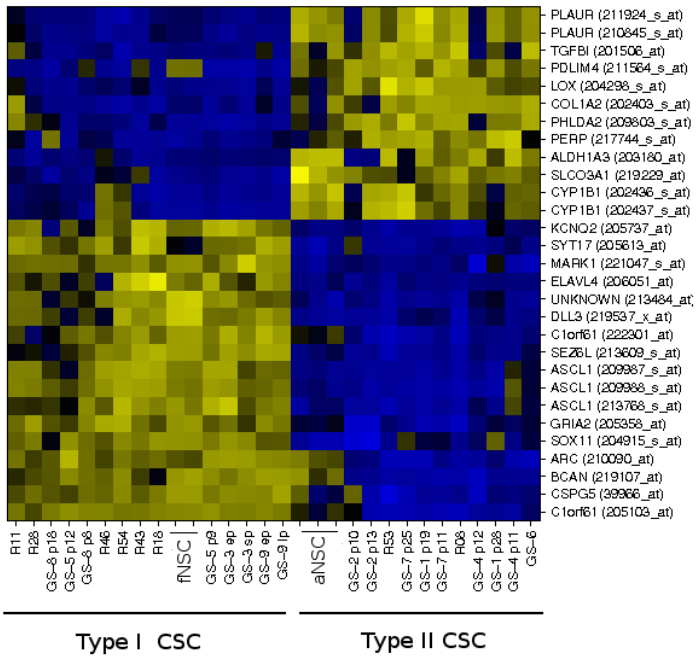


Figure 2.14: Heatmap of gene expression values of the 24-gene-signature within the training data. Rows correspond to transcripts, columns correspond to samples (low expression = blue, high expression = yellow).

2.3.7 Prediction of CSC and fetal Neural Stem Cell lines using the 24-gene-signature

To test the 24-gene-signature, I classified gene expression profiles of six recently characterized CSC and fNSC lines [117]. Pollard and colleagues cultured these CSC lines on laminin. All six CSC lines showed the same adherent growth pattern [157].

I used the *pamr.predict* function implemented in the bioconductor package PAM. The function returns prediction information, from a nearest shrunken centroid fit. It took the 24-gene-signature, the matrix of gene expression values of the gene expression profiles to be predicted and the already chosen shrinkage value of 4.8 for input. Then PAM calculates the distances to the different class centroids by using the gene expression values of the 24-gene-signature (Figure 2.15) and classifies the new expression profiles to the nearest shrunken centroid. Additionally the method returns the posterior class probabilities (table in Figure 2.2).

All published features agree with the resulting classification (Figure 2.15). All fNSC lines are recognized correctly by the gene signature and fall into the cluster of type I GMB CSC lines. The $CD133^+$ CSC lines except for $GS179$ are classified as type I. $GS179$ is classified as type II, but this cellline expresses $GFAP\delta$ filaments which are specific for the SVZ. These filaments were not detected in the analyzed type I CSC lines, so the classification into the group of type II is reasonable. The $CD133^-$ CSC line $GS166$ is classified as expected as type II. Thus, all CSC lines were correctly classified irrespective of culture-condition and growth pattern.

	CSC type II	CSC type I
G166-NS B (GSM379855)	1.000	0.000
G174-NS (GSM379856)	0.954	0.046
GS-179-NS A (GSM379857)	1.000	0.000
GS-179-NS B (GSM379858)	0.966	0.034
G144-GliNS1 (GSM379870)	0.290	0.710
GLiNS2 (GSM379871)	0.000	1.000
G144-NS A (GSM379872)	0.005	0.995
G144-NS B (GSM379873)	0.002	0.998
G166-NS A (GSM379874)	1.000	0.000
Fetal NSC (GSM379865)	0.000	1.000
Fetal NSC (GSM379866)	0.001	0.999
Fetal NSC (GSM379867)	0.002	0.998
Fetal NSC (GSM379868)	0.011	0.989
Fetal NSC (GSM379869)	0.000	1.000

Table 2.2: Table of posterior probabilities for the predicted samples published by Pollard and colleagues [157]

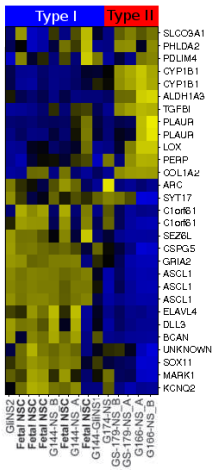
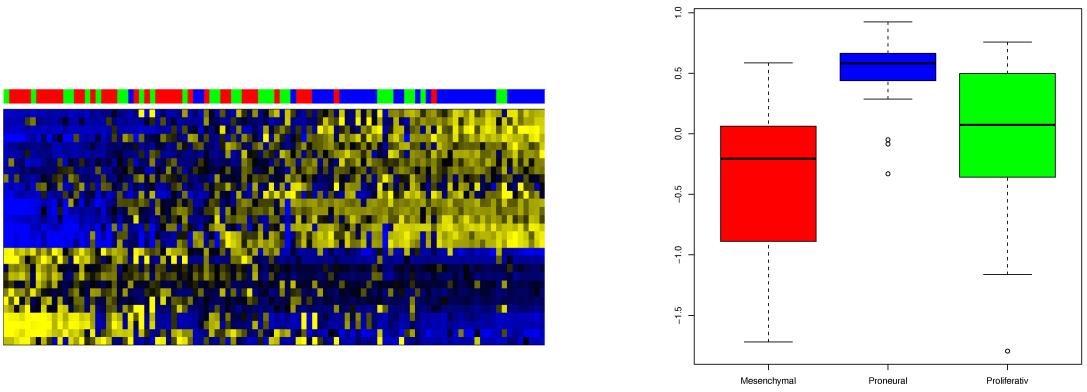


Figure 2.15: Tabel including the posterior probabilities for the predicted samples published by Pollard and colleagues and heatmap of gene expression values of the 24-gene-signature in six different CSC lines and four fNSC lines (Samples published by Pollard and colleagues [157])(low expression = blue, high expression = yellow).

2.3.8 Comparison of our classification system with the subgroups related to the 35-gene-signature

I compared our classification with the subgroups related to the 35-gene-signature established by Phillips and colleagues [154]. The three different GBM subtypes characterized by Phillips et al are referred to as proneural, mesenchymal and proliferative. I calculated an GBM CSC index representing the expression of the 24-gene-signature on 100 high-grade glioma samples from Phillips et al, by using the *getIndex* method explained in Section 1.1. The GBM CSC index reflects the expression of the 24-gene-signature within these samples. A low index represents a type II-like expression profile and a high index represents a type I-like expression profile.

The 24-gene-signature clearly distinguishes the mesenchymal from the proneural subtype (Figure 2.16). The proneural subtype is associated with an high CSC Index suggesting a connection between type I GBM CSC lines. The mesenchymal subtype is associated with a low CSC index indicating that these GBMs might be driven by type II GBM CSCs. The subgroup referred to as proliferative is ambiguous and presents a continuum between the other two types of GBM CSC lines. One reason for this could be that in some GBMs there is not only one type of CSCs present which drives the tumour but a mixture of GBM CSCs.



(a) (b)

Figure 2.16: (a) Heatmap of gene expression values of the 24-gene-signature in 100 high-grade glioma samples. The samples are ordered according to the calculated GBM CSC index based on the 24-gene-signature (low expression = blue, high expression = yellow). The color bar above the heatmap represents the three, by Phillips and colleagues defined, subgroups [154](red = mesenchymal, green = proliferative, blue= proneural). (b) Boxplots of GBM CSC index values of samples grouped into the three subgroups characterized by Phillips and colleagues. Calculation of the GBM CSC index was based on the 24-gene-signatur (red = mesenchymal ,green = proliferative, blue= proneural).

2.3.9 Discussion

Phillips and colleagues characterized three molecular subgroups of GBM associated with different stages of neurogenesis [154],[115]. Conversely, Beier and colleagues [10] and Guenther and colleagues [68] described two types of CSC lines which were derived from newly diagnosed GBM and cultured in medium.
This study now uses these three reports and suggests that a reason for the heterogeneity of GBM is that different cells of origin lead to different types of GBM CSC. The collected data suggests that, together with a recent study by Kenny and colleagues [96], neither the culture conditions ($+/-ECM$ proteins) nor the growth patterns have a dominant effect on the transcriptional profile and with that on the group assignment of CSC lines. In fact, the assignment to one of the groups describes a profound biological difference as also indicated by the different age of patients. The 24-gene signature derived from *in vitro* CSC lines can now be used to classify primary GBM samples directly *ex vivo* without culturing.

The cell of origin of GBM is still vague. In line with a previous report [115], it was shown that all GBM CSC lines do not cluster with astrocytes but with NSC lines.
Thus, most likely GBMs originate from cells that have preserved features of NSC or acquired these features, for example by de-differentiation. However, such a statement cannot be made for approximately 60% of the GBM that did not lead to CSC lines on propagation in NSC medium [61, 10, 147]. The limitation of this study is that it was not possible to observe the transformation of a putative founder cell into a GBM CSC. Thus, based on the results, it was only possible to generate hypotheses which need further validation. Only new conditional *in vivo* experiments will allow final conclusions.
Nevertheless, our data suggests that $CD133$ negative type II GBM CSC may be derived from $CD133$ negative aNSC located at the SVZ/ hippocampus [152]. But the type II CSC lines are still a heterogeneous group. A reason could be that different molecular mechanisms may be involved in the genesis. In the future, larger series will be needed to identify possible subgroups within the type II CSCs.
In contrast, at this moment it is not possible to connect the $CD133$ positive type I GBM CSC to a specific cell of origin, which has preserved or acquired a phenotype like the fNSC. Putative candidate cells could be among others the $CD133$ positive radial glia-like ependymal cells already published by Coskun and colleagues [31]. Another possibility could be the reprogramming of aNSC into pluripotent embryonal stem cells, which was already shown to be possible by activation of the transcription factor $Oct4$ [99].

Thus, the similarities between type I CSC lines and fNSC lines could also have occurred due to the reacquisition of fNSC features by a aNSC. Therapies for GBM should aim at eliminating the CSCs and not only aim at inhibiting certain pathways [11, 155]. To target the CSCs as effectively as possible it is important to know which kind of CSC is responsible for the tumour. In parts, this may be possible by grouping of GBM based on $CD133$ expression. Nevertheless, we suggest a 24-gene-signature which is more accurate than the expression of $CD133$, and thus, should be evaluated further for its potential.

In conclusion, the existence of two distinct types of GBM CSC lines was shown. During transition from the respective NSC to the related CSC, the loss of the appropriate differentiation potential may be the most critical event. But although additional alterations, giving rise to an accelerated proliferation, are required for the malignant phenotype (particularly in type II GBM CSC), the collected data provides a deeper insight into the development of GBM, their heterogeneity, and the according variable prognosis of patients. In future, this may lead to a more personalized and improved therapy targeting GBM CSCs.

2.4 The glioblastoma CSC subtype $TGF\beta$-dependently determines the degree of immune infiltration

This project was done in collaboration with the group of Christoph Beier from the department of Neurology, University of Regensburg. Together with Claudio Lottaz I carried out the statistical analysis and evaluation of the microarray data. Parts of this work are published in the article: "The Cancer Stem Cell Subtype Determines Immune Infiltration of Glioblastoma" by C. Lottaz, D. Beier, K. Meyer, P. Kumar, A. Hermann, J. Schwarz, M. Junker, P. Oefner, U. Bogdahn, J. Wischhusen, R. Spang, A. Storch and C. Beier in Stem Cell and Development [9].

2.4.1 Motivation

GBM are heterogeneous and consist of distinct molecular subtypes [117, 82]. So far only two types of GBM subtypes have been identified by transcriptional [154], genetic [192] and proteomic studies [22]: the proneural subgroup of GBM and the mesenchymal subgroup. But additional subtypes are likely to exist. In Section 2.3 I describe two types of different CSC lines $CD133^+$ and $CD133^-$ contributing to the heterogeneity of GBM [121]. I have already shown a connection between different GBM CSCs and the proneural and mesenchymal phenotype. In addition, I derived a 24-gene-signature differentiating between the "proneural-like CSC" lines, being $CD133^+$ and the "mesenchymal-like CSC" lines being $CD133^-$. It was observed that patients suffering from mesenchymal GBM have more infiltrating immune cells and show a longer survival in comparison to patients with proneural GBM [160]. The underlying mechanism of the variable immune infiltration remains unknown.
In this study we show that $TGF\beta$ regulates proliferation, migration, and tumourigenicity of mesenchymal cancer stem cells (CSC). In contrast, proneural CSC resist $TGF\beta$ due to $TGFR2$ deficiency. *In vivo* mesenchymal glioblastoma shows a substantially increased infiltration of immune cells while proneural glioblastoma display minor immune infiltration. On a functional level, proneural CSC lines cause a significantly stronger $TGF\beta$-dependent suppression of $NKG2D$ expression on $CD8+$ and $CD68+$ cells *in vitro* which provide a mechanistic explanation for the reduced immune infiltration of proneural glioblastoma. Thus, the molecular subtype of CSC $TGF\beta$-dependently contributes to the degree of immune infiltration.

2.4.2 Microarray hybridization and preprocessing

Acquisition of microarray data was described in [10]. For the analysis I used R [182] and Bioconductor [64]. Expression values were corrected and normalized using the robust multi-array average (RMA) [85] method implemented in the Affymetrix package.

2.4.3 Proneural GBM show a reduced immune infiltration

We stained microglia with markers for cytotoxic T-lymphocytes $CD8^+$ and $CD68^+$ in a series of GBMs to show the varying degrees of immune infiltration. Double staining with the proneural marker genes $Olig2$ and $CD8/CD68$ revealed an anti-correlation of the proneural marker expression and the immune infiltration markers (Figure 2.17). Thus, as expected the proneural GBM show reduced immune infiltration.

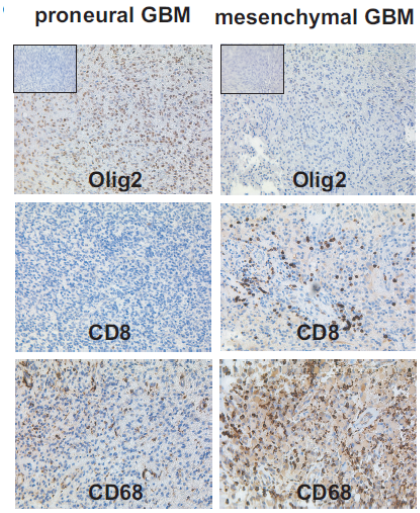
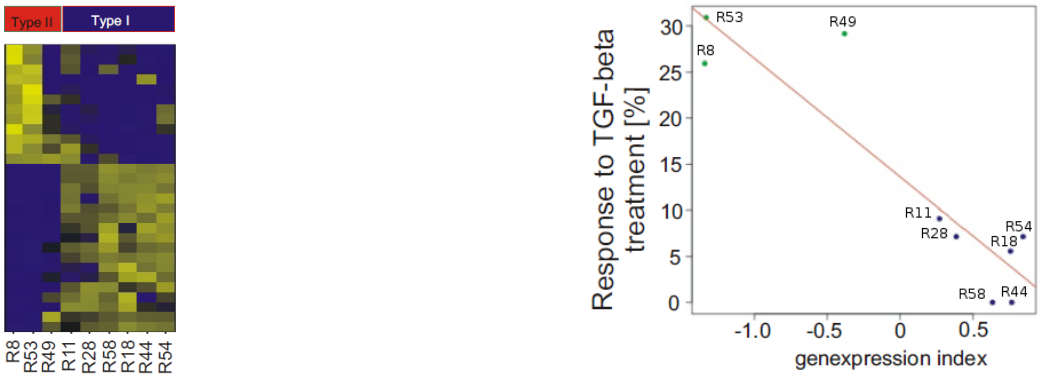


Figure 2.17: Olig2, DLL3, NeuN, YKL-40, CD44, VEGF, CD8, and CD68 expression in GBM (n=19) was quantified by a blinded rater (CPB). The expression of markers for proneural GBM (DLL3, Olig2, NeuN) correlated with the amount of infiltrating CD8, and CD68 cells. Representative images of Olig2 and $CD8^+$ and $CD68^+$ staining are given. (Figure taken from publication by Beier et al [9].)

2.4.4 $TGF\beta$ responsiveness is determined by the expression of $TGFR2$

I analyzed gene expression profiles of nine CSC lines, three mesenchymal-like and six proneural-like CSC lines. The assignment to the two subtypes of GBM and their appropriate CSC is given in Section 2.3 using the 24-gene-signature (Figure 2.18 (a)). Here we show that the effects of $TGF\beta$ vary between the two groups.

We tested whether $SMAD2$ was phosphorylated after treatment with $TGF\beta$. All mesenchymal celllines and one proneural cellline showed a time-dependent $TGF\beta$-induced $SMAD2$ phosphorylation (appendix Figure A1). Although there is no autocrine $SMAD2$ phosphorylation in any of the CSC lines *in vivo*, it was possible to detect $TGF\beta$ mRNA and unprocessed $TGF\beta$ *in vitro* and the expression of $TGF\beta$ 1 and 2 *in vivo*. All celllines which are not responsive to $TGF\beta$ show a significantly decreased expression of the $TGF\beta$ receptor $TGFR2$. More precisely, five resistant CSC lines show a lower expression of $TGFR2$ than four, which show a $SMAD2$ phosphorylation after $TGF\beta$ treatment (appendix Figure A2). This could explain the molecular mechanism leading to different resistances of the CSC lines to $TGF\beta$.



(a) (b)

Figure 2.18: (a) Heatmap of the 24-gene-signature in 9 CSC lines (low expression = blue, high expression = yellow). The color bar above the heatmap represents the GBM CSC types (red= mesenchymal-like, blue=proneural-like) (b) The gene expression index calculated based on the $TGF\beta$ target genes plotted against the relative change of proliferation after $TGF\beta$ treatment. The samples represented by blue dots were classified as proneural-like, the green samples as mesenchymal-like.(Figures taken from publication by Beier et al [9].)

2.4.5 Mesenchymal-like CSC lines are responsive to $TGF\beta$ treatment

The effects of $TGF\beta$ 1, 2 and the combination of both were the same in all experiments. In the mesenchymal-like CSC lines we observed that $TGF\beta$ has an influence on migration, proliferation and tumourgenicity. But it did not affect colongenicity or the proportion of $CD133^+$ cells and it did not alter the differentiation profile. In contrast to that $TGF\beta$ incubation in proneural-like CSC lines lacking $SMAD2$ phosphorylation and the expression of $TGFR2$ did neither affect migration nor stem cell properties. An exception was the proneural-like CSC line R11 that was resistant to $TGF\beta$ treatment despite $SMAD2$ phosphorylation.

In order to understand this discrepancy the $TGF\beta$-LIF mediated signaling pathway [151] was tested to see whether it might be operational in this GBM CSC line. Until now the medium used for the experiments was routinely supplemented with LIF. This fact obscured the LIF-depending growth-promoting effects triggered by $TGF\beta$. The experiment was reproduced using LIF-free medium. It was possible to observe an increased proportion of clonogenic cells confirming the presence of the described $TGF\beta$-LIF-$STAT3$ mediated signaling pathway in the CSC line R11. In all other celllines the $TGF\beta$ mediated effects were independent of LIF. Thus, R11 was dependent on the $TGF\beta$-LIF-$STAT3$ signaling pathway and hence resisted $TGF\beta$ treatment as long as LIF was available (appendix Figure A2).

2.4.6 Defining $TGF\beta$ target genes and a $TGF\beta$ response index

To identify GBM that respond to $TGF\beta$, I used gene expression data to define $TGF\beta$ target genes. I compared transcriptional profiles of $TGF\beta$ responsive CSC lines (R8, R53) with and without $TGF\beta$ treatment. I used a linear model approach implemented in the R package *limma* [176] to identify genes that were differentially expressed between the celllines treated with $TGF\beta$ and without treatment. The linear model used was:

$$E[y_j] = X\alpha_j,$$

where y_j is the expression data for gene j, X is the *design matrix* and α_j is a vector of coefficients for each gene j to be estimated.
The following design matrix X and contrast matrix C were used:

$$X = \begin{array}{c} \\ R8 \\ R53 \\ R8_{TGF\beta} \\ R53_{TGF\beta} \end{array} \begin{array}{c} \begin{array}{ccc} TGF\beta & nonTGF\beta & R8 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$

$$C = \begin{array}{c} \\ TGF\beta \\ nonTGF\beta \\ R8 \end{array} \begin{array}{c} \begin{array}{c} TGF\beta - nonTGF\beta \end{array} \\ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \end{array}$$

The design matrix determines how the 3 coefficients for each gene j were estimated. The model took the two different treatments and the different cellline pairings into account. The second cellline $R53$ was implicitly modeled, in a way that being cellline $R53$ is determined by not being cellline $R8$.
The contrast β_j were computed as follows:

$$\beta_j = C^T * \alpha_j,$$

where α_j is a vector containing the 3 estimated coefficients and C is the given contrast matrix. The result was a list of differentially expressed genes between the $TGF\beta$ treated and untreated gene expression profiles of the CSC lines. From that list I selected 100 genes showing the highest log fold change, between control and $TGF\beta$ treated samples. All the genes showed an absolute log fold change higher than 1.47.

I calculated a $TGF\beta$-response index on the gene expression profiles of 9 CSC lines based on the 100 $TGF\beta$ target genes, by using the method getIndex described in Section 1.1. This $TGF\beta$ index was used as representative for $TGF\beta$ signaling. We could show that the relative change of proliferation after $TGF\beta$ treatment is highly correlated to the $TGF\beta$ response index. Thus we were able to observe an increase in proliferation when treating the celllines with $TGF\beta$ (Figure 2.18 (b)).

2.4.7 Associating $TGF\beta$ response with the type of GBM CSC line

For further analysis, I used expression profiles of 80 well-characterized GBM samples published by Murat and colleagues [140].

I calculated pairwise euclidean distances between all the samples only using the gene expression values of the 100 $TGF\beta$ target genes. Then I clustered the 80 GBM samples using the complete linkage method. The clustering resulted in two clusters. To validate the robustness of the found clusters I used the consensus clustering method [135]. The method took the gene expression matrix, with the 100 $TGF\beta$ target genes in rows and the samples to be clustered in columns for input. Then the data was resampled 100-times by repeatedly drawing 80% from the given samples. After that a clustering was done for each of these dataset resulting in a connectivity matrix. Consensus clustering returned a consensus matrix and additionally sample consensus values for each sample and cluster consensus values.

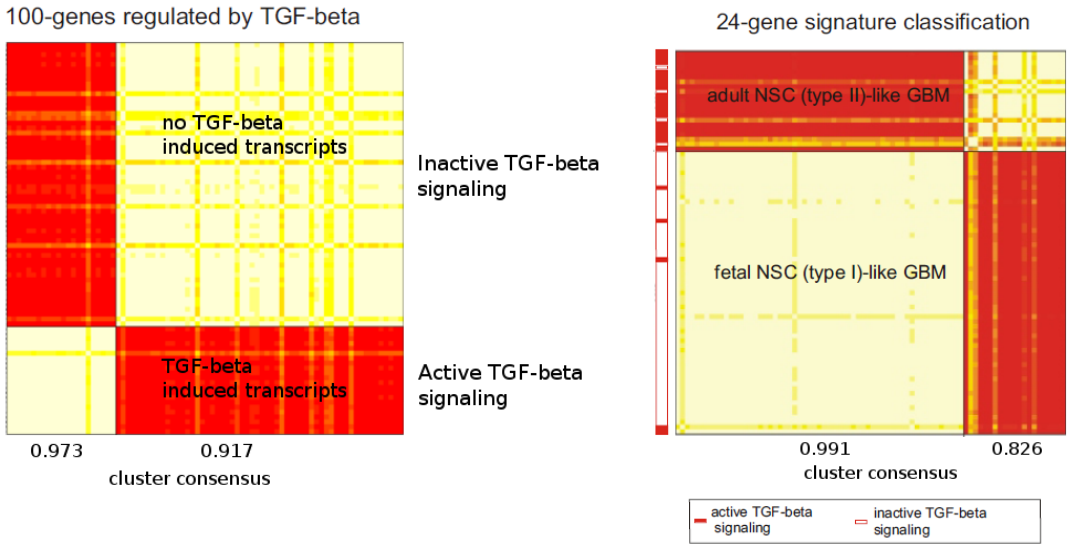
The consensus matrix includes real numbers between 0 and 1. The entry (i, j) in a consensus matrix represents the number of times the samples i and j are assigned to the same cluster divided by the number of times both samples were included in a re-sampled dataset. The result of the consensus clustering based on the expression values of the 100 $TGF\beta$ taget genes is visualized in Figure 2.19 (a), where the color gradient is associated with the values ranging from 0 to 1, so that beige corresponds to 1 and dark red corresponds to 0.

The two resulting clusters were defined as a $TGF\beta$ responsive cluster characterized by expression patterns typical for activated $TGF\beta$ signaling and a $TGF\beta$ unresponsive cluster. The $TGF\beta$ responsive cluster comprised of only 22 of the 80 GBM samples indicating that $TGF\beta$ signaling is only active in a subgroup of GBM.

To connect these two groups of different $TGF\beta$ responsive GBMs to the previously defined subgroups (proneural-like and mesenchymal-like GBM), I used the 24-gene-signature defined in Section 2.3. Again I clustered the 80 GBM samples, but now the pairwise euclidean distances were calculated based on the 24-gene-signature. Two clusters were found. To validate the robustness of the newly found clusters I followed the same analysis steps as above, the only difference was that instead of using the $TGF\beta$ target gene expression, the expression of the 24-gene-signature was used. I obtained two clusters that turned out to be highly stable with respect to sub-sampling. The result of the consensus cluster analysis, on the basis of 24-gene-signature, is shown in Figure 2.19 (b).
Many samples classified as mesenchymal-like are $TGF\beta$ responsive or more precisely do express $TGF\beta$ target genes and many samples classified as proneural-like are unresponsive to $TGF\beta$ treatment.

To validate the connection between $TGF\beta$ responsiveness and the two different types of GBM CSC lines I calculated two indices for the 80 GBM samples with the getIndex method described in Section 1.1. The first index was the $TGF\beta$ response index, which was calculated based on the gene expression of the $TGF\beta$ targets.
The second index was based on the gene expression of the 24-gene-signature. The two calculated indices were highly anti-correlated ($r = -0.885$). Thus, most $TGF\beta$ responsive GBM samples are similar to mesenchymal-like CSC lines and conversely $TGF\beta$ unresponsive GBM samples mainly display a proneural-like CSC expression. The heatmap in Figure 2.20 displays the gene expression of the 24-gene-signature in the 80 GBM samples and shows the connection of the 80 samples to $TGF\beta$ responsiveness.

All this leads to the conclusion that the 24-gene-signature not only indicates different cells of origin but also distinguished GBM with and without active $TGF\beta$ signaling and with that, $TGF\beta$ responsiveness respectively *in vivo* and *in vitro*. These results might help to identify patients who may benefit from $TGF\beta$ targeting therapies.



(a) (b)

Figure 2.19: Consensus clustering results of 80 GBM samples, (a) based on 100 $TGF\beta$ target genes, discriminating GBM samples after the activity of $TGF\beta$ signaling; (b) based on the 24-gene signature discriminating proneural-like and mesenchymal-like GBM samples. Side bar indicates in which $TGF\beta$ responsiveness clusters the corresponding sample fell in (a) The counts of how often samples clustered together are visualized in a color-coded consensus matrix (Dark red = pairs that were never clustered together, beige = pairs that always clustered together)

2.4.8 $TGF\beta$ mediates immunosuppression in proneural-like CSC

The presence of GBM with different responsiveness to $TGF\beta$ suggests that $TGF\beta$ has different predominant functions depending on the GBM subgroup. Given the almost complete $TGF\beta$ resistance and the expression of $TGF\beta$ in proneural like CSC an immunosuppresive function of $TGF\beta$ in this subtype can be assumed. The $TGF\beta$-induced down-regulation of $NKG2D$ receptor expression on $CD8^+$ T-cells and NK-cells *in vitro* was investigated [57]. The $TGF\beta$ signaling was blocked by $TGFR1$ kinase inhibitor SD-208 [188] to differentiate between $TGF\beta$ dependent and independent effects.

Peripheral blood mononuclear cells (PBMC) cultured with supernatant CSC lines remained without effect.
Thus, in the absence of immune cells CSC lines do not secrete active $TGF\beta$. But co-culturing of PBMC with CSC lines revealed an increased $TGF\beta$ dependent down-regulation of $NKG2D$ on NK-cells and of CD8 T-cells. The increase of $TGF\beta$ dependent down-regulation of $NKG2D$ is more dominant in the proneural-like CSC lines. Thus, the $TGF\beta$ dependent immuno-suppression is enhanced by proneural-like CSC lines which are therefore resistant to $TGF\beta$ treatment.

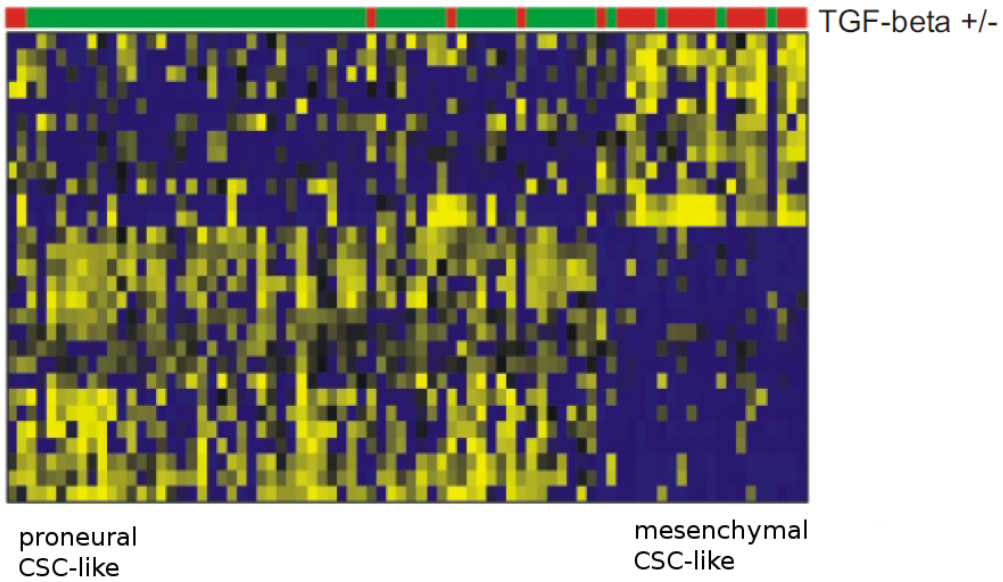


Figure 2.20: Heatmap of the 24-gene-signature expression within 80 GBM samples (published by Murat and colleagues [140]). Samples (columns) are ordered according to 24-gene-signature-index. (low gene expression = blue, high gene expression = yellow). Samples expressing $TGF\beta$ regulated transcripts are labeled in red.

Additionally, I compared the expression of a geneset associated with innate immune response between the $TGF\beta$ responsive and unresponsive subgroups. The geneset was published by Murat and colleagues [140] and consisted of 134 genes. In line with the assumption that $TGF\beta$ resistant GBM show an increased immuno-suppression, I found that the geneset has on average higher expression levels in GBM samples assigned to the proneural-like CSC subgroup (Figure 2.21).

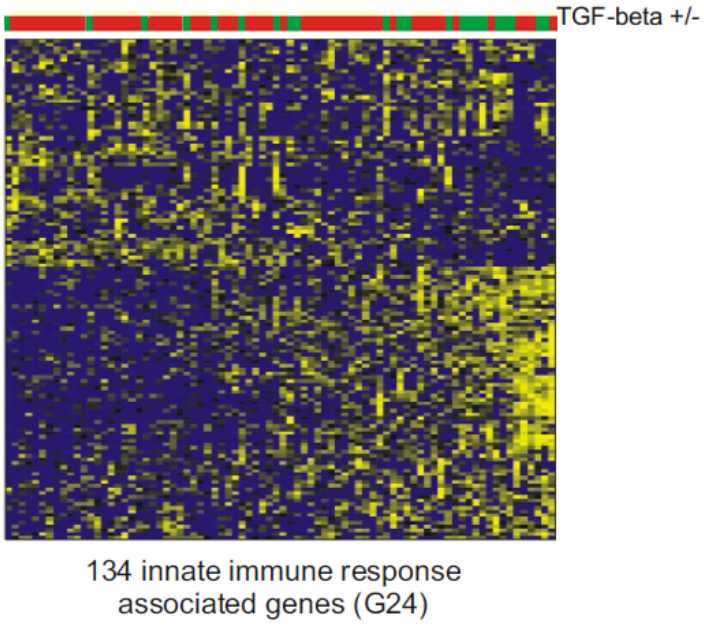


Figure 2.21: Heatmap of 134 genes associated with innate immune response published by Murat and colleagues [140]. Samples (columns) were sorted according to an index calculated on the shown genes, genes (rows) were clustered (low gene expression = blue, high gene expression = yellow). Samples expressing of $TGF\beta$ regulated transcripts are labeled in red.

2.4.9 Discussion

We were able to reveal a relationship between GBM and proneural-like respectively mesenchymal-like CSCs indicating a profound biological difference in the function of $TGF\beta$ in these tumours. Tumours expressing transcripts of proneural-like CSC also express $TGF\beta$ *in vivo*. $TGF\beta$ is so far the most potent immuno-suppressive cytokine known. Thus, it may suppress the function and invasion of immune cells in proneural GBM [160, 172]. It was possible to substantiate this hypothesis using functional assays which revealed a significantly stronger $TGF\beta$ dependent suppression of $NKG2D$ receptor expression on NK- and T-cells by proneural-like CSC lines in comparison to mesenchymal controls [9].
The different responsiveness of the CSC to $TGF\beta$ and the immunosuppression effects provides a possible explanation for the refractory phenotype of proneural GBM in an otherwise highly promising vaccination study on GBM patients [160]. Immune therapies applied to tumours maintained by proneural-like CSC should therefore also aim at the attenuation of $TGF\beta$ mediated immuno-suppression.

Thus, the identification of the CSC subtype in GBM based on the 24-genes signature will help to guide both $TGF\beta$ targeting and immune therapies by identifying GBM with strong $TGF\beta$-dependent immune paralysis. Conversely, the functions of $TGF\beta$ in GBM maintained by mesenchymal-like CSC are manifold and $TGF\beta$ may either inhibit or promote tumour growth. Thus extreme caution is called for when applying of $TGF\beta$ targeting therapies to mesenchymal GBM.
In conclusion the CSC subtype determines, if $TGF\beta$ signaling is active or not active and with that it determines the degree of immune infiltration and local immune paralysis. Thus, the identification of the CSC subtype driving a GBM will help to guide $TGF\beta$ targeting and immune therapy decisions.

Chapter 3

Lymphoma

Synopsis

In the first Section of this Chapter is an introduction to lymphoma biology. In the second Section human transformed germinal center B cells were used to identify patterns of gene expression changes in response to B cell specific stromal stimuli. These patterns indicate different pathway activaties in Diffuse Large B Cell Lymphomas and Burkitt Lymphomas. In the third Section I identify LEF1 target genes, allowing for the measurement of LEF1 activity in lymphoma. LEF1 activity in Burkitt Lymphomas is higher compared to DLBCL.

3.1 Biological introduction

Lymphomas are a heterogeneous group of cancers affecting cells of the immune system, called lymphocytes. Lymphomas are classified based on the cell they originate from: B cell lymphomas originate from B lymphocytes/B cells and T cell lymphomas originate from T lymphocytes/T cells.

95% of lymphomas are B cell lymphomas and develop during B cell differentiation [109]. The steps of B cell development can be associated with different structures of the B cell receptor (BCR). The BCR consists of two heavy-chain and two light-chain immunoglobulin (Ig) polypeptides, which are covalently connected by disulfide bridges.

Early B cell development takes place in the bone marrow, where the B cell precursor rearranges the DNA of Ig heavy-chain and light-chain genes and is equipped with a functional BCR. The BCR is responsible for detecting pathogens. Each B cell is equipped with a different BCR. Only when expressing a functional BCR does the B cell get signals to differentiate into a mature naive B cell and leave the bone marrow. If the mature naive B cell encounters an antigen that binds to the BCR, the B cell gets activated and enters into the germinal center reaction. Within the germinal centers (GC) the B cell undergoes clonal expansion and class-switch recombination. During these processes the BCR coding gene is further mutated and changed optimizing the fit to its antigen. The normal B cell depends on BCR expression, as well as on further signals from the micro-environment [109].
Most malignant growing B cell lymphomas use factors, necessary for normal B cell differentiation, to get survival and proliferation signals.

Lymphomas can be separated into two major categories: Hodgkin's lymphomas (HL) and non-Hodgkin's lymphomas (NHL). In 2008 the incidence rate for NHL was 5.1 new cases per $100,000$ persons. The Age-World-Standardized incidence rate (ASR(W)) is calculated by the WHO. The WHO collects a set of genetic and histopathological criteria for classifying tumours of hematopoietic and lymphoid tissues. Based on these criteria lymphomas are classified and treated. NHL comprises of many different types of B cell lymphoma, which can be distinguished in indolent (slow-growing) and aggressive (fast-growing) types [88].
Two aggressive forms of NHLs are Burkitt's Lymphoma (BL) and Diffuse Large B Cell Lymphoma (DLBCL) [66].

3.1.1 Burkitt's Lymphoma (BL)

The Burkitt Lymphoma (BL) was first described by Denis Burkitt in 1958 as the most prevalent childhood lymphoma located in the jaws of African children [24]. The BL accounts for 2% of all lymphomas [109], but in Europe and the United States sporadic BL is the most prominent childhood NHL accounting for 30% of all NHL cases [201]. There are three different types of BL: The endemic BL, the sporadic BL and the immunodeficiency associated BL. The endemic BL is associated with an infection of the Epstein-Barr-Virus (EBV) and is found in Equatorial Africa. Tumours with the same characteristics, but not associated with an EBV infection, can be found in many parts of the world and are called sporadic BL [21, 108]. Sporadic BLs account for 1-2% of all adult lymphomas in Western Europe and the United States.

The immunodeficiency BL is associated with an infection of the human immunodeficiency virus (HIV).

Immunophenotype characteristics of BL are high expression levels of $CD10$ and $BCL6$. There is no expression of $BCL2$ and $CD5$. The fraction of cells proliferating is at least 90%, assessed by immunohistochemistry using the proliferation marker protein Ki67 [43, 73]. Most BL harbor a deregulated c-Myc expression as a consequence of an IG-MYC translocation. In that case the proto-oncogene MYC is translocated into one of the immunoglobulin gene loci. The translocated allele of MYC is highly expressed whereas the normal allele is usually silent. High c-Myc activity can lead to a high proliferation rate and, in the absence of survival signals, to apoptosis [81]. Rare BL cases lacking the MYC rearrangement may up-regulate the expression of MYC by other mechanisms [116]. Recent gene expression studies have shown that typical BL cases have a characteristic molecular signature [81, 35]. This signature consists of 58 genes which can be used to calculate a continuous BL-index. Using this index, one can distinguish between molecular BL (mBL), non-molecular BL (non-mBL) and intermediate lymphoma [81].
BL is treated with regimens of chemotherapy including cyclophosphamide, vincristine, doxorubicin and high-dose methotrexate. A monoclonal antibody drug called rituximab may also be given, usually with chemotherapy. If the disease is not treated, BL is lethal within several months. Long-term remission can be achieved in approximately 50% of patients [149, 201].

3.1.2 Diffuse Large B Cell Lymphoma (DLBCL)

Compared to BL which is a quite defined group of lymphomas, Diffuse Large B Cell Lymphoma (DLBCL) is a heterogeneous disease. DLBCL accounts for 30-40% of all lymphomas and thus is the most frequent NHL. The morphological differences and the diversity in clinical presentation and outcome leads to the assumption that DLBCL comprises several disease entities [55, 109].
Different genetic aberrations are found in DLBCL. A translocation of the proto-oncogene $BCL6$ is found in 35% of DLBCL cases. Bcl6 is a transcription repressor and is essential for germinal center formation [41, 58]. In a subgroup of DLBCL cases the proto-oncogene MYC is translocated into one of the immunoglobulin gene loci [7, 111]. DLBCL show a high aberrant activity of survival signals. One of these signals is the activated NF-κB pathway which blocks apoptosis [51, 37]. Furthermore, DLBCL acquire survival properties by an aberrant activity of the BCR signaling pathway [38].

There are four major gene expression studies published investigating the heterogeneity of DLBCL.

Alizadeh and colleagues [2] substratified DLBCLs depending on their similarity to distinct differentiation stages of normal B cells into ABC-like and GCB-like DLBCLs. The activated B cell like or ABC-like DLBCLs are similar to activated post GC B cells, whereas the GC B cell like or GCB-like DLBCLs are similar to B cells during the germinal center reaction. ABC-like DLBCLs show a high aberrant activity of NF-κB and have a worse prognosis compared to GCB-like DLBCLs, which are dominated by abberant BCL6 [2, 168, 112].

Monti and colleagues [134] found three different subtypes of DLBCL: They name them by expression characteristics as "oxidative phosphorylation", "B cell receptor/proliferation" and "host inflammatory response".

Ci and colleagues [26] used a Bcl6 ChIP-on-Chip assay to obtain 1361 Bcl6 target genes. They associate the GCB-like DLBCLs [2] and the "B cell receptor/proliferation" lymphomas [134] with high Bcl6 activity. Bcl6 target genes were found to be repressed within these sub groups.

Bentink and colleagues [13] classified aNHL based on oncogene associated patterns, called gene modules. These gene modules were associated with the oncogenes: MYC, RAS, SRC, β-catenin and $E2F3$. They transfected quiescent primary human mammary epithelial cells (HMECs) singly with each of the five human oncogenes and used a semi-supervised machine learning methodology to generate these modules. They combined the modules to patterns of oncogene activity called Pathway Activation Patterns (PAPs). DLBCLs can be classified into four different pathway activation patterns (PAP1-PAP4). All DLBCLs defined as "intermediate" by the mBL index, which could not be allocated to one of the four major PAPs (PAP1-PAP4), were assigned to a fifth group, called "molecular individual lymphoma" (mind-L) [81, 13]. It is not yet clear which pathway activities define the mind-L group [13].

DLBCL tumours are treated with a combination of cyclophosphamide, hydroxydaunorubicin, vincristine and prednisolon (CHOP). Treatment was improved by adding Rituximab, a monoclonal antibody against the B cell specific marker $CD20$ (R-CHOP). However 30% of DLBCL patients do not completely respond to the therapy, are resistant to it, or relapse soon after treatment [92, 29].

3.2 Pathway activities in human transformed GC B cells and lymphomas

This project was done in cooperation with the group of D. Kube from the department of Haematology and Oncology, University Medical Center Göttingen. Experiments were done by A. Schrader. I performed the statistical analysis and evaluation of the microarray data. Parts of the work are published in Cell Communication and Signaling: "Global gene expression changes of in vitro stimulated human transformed germinal center B cells as surrogate for oncogenic pathway activation in individual aggressive B cell lymphomas" by A. Schrader, K. Meyer, F. von Bonin, M. Vockerodt, N. Walther, E. Hand, A. Ulrich, K. Matulewicz, D. Lenze, M. Hummel, A. Kieser, M. Engelke, L. Trmper, D. Kube [171].

3.2.1 Motivation

The current criteria to distinguish BL from DLBCL, like morphology, immunophenotype, and genetic abnormalities, do not allow for a reliably reproducible classification and the pathological mechanisms behind these criteria are poorly understood [181]. NHL cells proliferate actively and retain many of the immunophenotypic characteristics of B lymphocytes. Nevertheless, they are monoclonal tumour B cells, displaying characteristic nonrandom chromosomal abnormalities. Resulting from genetic aberrations the expression of tumour suppressor genes or protooncogenes can be deregulated, leading to a loss of cell growth and proliferation regulation. Cellular genes can be placed under different control elements and may switch off cellular growth regulation. In addition, specific combinations of signals for short or long term stimulation are provided to GC B cells through externally derived signals obtained from cells in the microenvironment [103, 110]. These include CD40L, BAFF and IL21. Furthermore, B cell specific activation can be mediated via antigenerecognition or via Toll like receptor activation by the bacterial component lipopolysaccharide. Theses stimuli are described in more detail in the following paragraph.

As already mentioned, a recent large-scale gene expression profiling of NHL tumour samples revealed a molecular definition for BL, by describing a Burkitt specific signature. This signature was used to model an index of "Burkitt-likeness" (mBL-index) and can be used to distinguish BL from DLBCL [81, 35].
We ask the question to which extent different pathways are responsible for the differences in gene expression that distinguish individual DLBCL and DLCBL from BL.

To get an insight into the cell signaling networks in individual human mature aNHL, human transformed GC B cells were stimulated with B cell specific stromal stimuli to identify patterns of gene expression changes.

3.2.2 Stromal signals affecting B cells

In peripheral secondary lymphoid organs B cells encounter foreign antigens. Antigen stimulated B cells develop to form germinal centers. The micro-environment within germinal centers not only provides antigen stimulation through the BCR, but additionally, B cells interact with T cells, macrophages, follicular dendritic and reticular cells. Signals influencing the development of a B cell can be discriminated into intrinsic signals, that do not depend on the surrounding micro-environment and paracrine signals, that are mediated by the micro-environment surrounding the B cell. Important paracrine signals are mediated by CD40, IL21 and BAFF. Central intrinsic signals include the engagement of the B cell receptor (BCR) and the Toll like receptors (TLRs). The TLRs are receptors of the innate immune response which are needed to recognize non-self molecules [109, 44, 98, 105, 173].

BCR: B cell receptors cluster upon antigen recognition on the surface of a B cell [198]. Here BCRs are activated by αIgM F$(ab)_2$ fragments. These induce the crosslinking and thereby activation of BCR. The BCR activation triggers a variety of signaling pathways including calcium signaling, PI3K/AKT, NF-κB and MAPK signals. These signal transduction pathways, initiated through the BCR, determine the fate of B cells within a context of BCR affinity to the antigen, expression levels of stimulatory or inhibitory coreceptors and the differentiation stage of B cells. Activation of the BCR complex initiates a signaling cascade that confers positive selection, proliferation, apoptosis or differentiation depending on defined combinations of costimulatory signals. These costimulatory signals can be intrinsic signals and paracrine signals. GC B cells undergo apoptosis, if not rescued through germinal center survival signals. (BCR signaling reviewed by Wang and Clark [193])

LPS: Bacterial lipopolysaccharide (LPS), which is a component of the cell wall of Gram-negative bacteria, is specifically recognized via Toll like receptor 4 (TLR4). TLR4 belongs to the family of pattern recognition receptors (PRRs). This recognition is part of the innate immune response [161, 78, 158, 25]. The activation of TLR4 triggers the activation of NF-κB and MAPK signaling [45, 161, 107].

It has already been shown that some lymphomas use the constant activation by a persisting microbial antigen for growth. For example, the chronic infection with Helicobacter pylori can lead to a chronic gastritis and subsequent to the development of MALT lymphomas [48, 203]. Furthermore, ABC DLBCLs were found to have highly recurrent oncogenic mutations affecting MYD88, an adaptor protein that mediates toll like receptor signaling [143].

CD40: CD40/TNFRSF5 belongs to the family of TNF receptors [119]. It is activated by using a specific ligand CD40L (called CD154). CD40L is expressed by CD4+ T helper cells. B cells are stimulated by T helper cells via CD40L [4, 114].
This co-stimulation helps to mediate the positive selection of the right B cell clone during immune response. CD40 mediates NF-κB and MAPK signals [52, 53, 15, 60]. It has been shown that CD40 activation can have positive effects on cell survival and cell proliferation. This includes examples were the CD40 signal rescues BL and DLBCL cells from apoptosis [173, 3, 79, 153].

IL21: The cytokine Interleukin-21 (IL21) is produced by activated CD4+ T cells and natural killer (NK) cells [36, 30, 71]. IL21 signals via the IL21R and mediates STAT1 and STAT3 activity. The IL21R receptor can trigger the activation of MAPK and PI3K/AKT signaling [202, 59]. IL21 is a potent B cell activator and it induces plasma cells to differentiate [36, 47]. The effect of IL21 stimulation on BLs and DLBCLs has not been described so far; but it has been shown that follicular lymphoma (FL) highly express IL21R and that the activation of these cells via IL21 induces apoptosis [1].

BAFF: The B cell activating factor of the TNF family (BAFF also called TNFSF13b or CD257) is essential for homeostasis of mature B cells. BAFF is expressed by T cells and macrophages. It can activate distinct TNF receptors which trigger non-canonical NF-κB and MAPK signaling. (BAFF signaling reviewed by Bossen and Schneider [17].)
An over-expression of BAFF can lead to severe B cell hyperplasia and autoimmune diseases [129, 100].
In lymphoma sections BAFF and BAFF receptors were detected [166] and seem to be associated with a bad prognosis for DLBCL patients [145]. Autocrine BAFF signaling has already been shown for chronic lymphoid leukemia, multiple myeloma, DLBCL and BL cells [77, 166, 74, 145, 144].

3.2.3 The effect of stromal signals on BL2 cells

Human transformed germinal center B cells (GC B cells) were used to identify patterns of gene expression changes in response to B cell specific stromal stimuli. To achieve *in vitro* perturbations, BL2 cells were stimulated in triplicates using CD40L, BAFF, IL21, αIgM $F(ab)_2$ fragments and lipopolysaccharide (LPS). The aim was to assess ligand specific signal transduction. At the same time the signal had to be strong enough to be monitored as gene expression change at the whole genome level. Probes of three independent biological experiments were hybridized to U133 plus 2.0 microarrays. Gene expression values were obtained by correcting for background and normalizing on probe level using the variance stabilization method by Huber and colleagues [80]. The normalized probe intensities were summarized into gene expression levels by using an additive model [84] fitted by the median polish procedure [187].

3.2.4 Gene expression patterns induced by stromal effects in BL2 indicate high pathway activation in DLBCL

Identifying global gene expression changes in response to stromal signals in BL2 cells

In order to identify pathway activities differing among aggressive NHL, we stimulated human transformed GC B cells with five stimuli: αIgM (BCR activation), LPS, CD40L, IL21 and BAFF. Using linear models implemented in the Bioconductor [64] package *limma* [176], I searched for genes showing the highest differential expression in response to each stimulus. The linear model was defined such that $E[y_j] = X\alpha_j$, where y_j is the expression data for gene j, X is the *design matrix* and α_j is the vector of coefficients for each gene j to be estimated .
The design matrix determines how the 8 coefficients are to be estimated. The model included the different stimulations and two of the three batches (pairs: $B1$,$B2$). The third batch $B3$ was implicitly modeled, such that $B3$ is defined by not being $B1$ or $B2$. The contrast β_j were computed:

$$\beta_j = C^T * \alpha_j,$$

where α_j is a vector containing the 8 estimated coefficients and C is the contrast matrix.

The following design matrix X and contrast matrix C were used:

$$X = \begin{array}{c} \\ \text{Crtl B1} \\ \text{CD40 B1} \\ \text{LPS B1} \\ \text{BAFF B1} \\ \text{BCR B1} \\ \text{IL21 B1} \\ \text{Crtl B2} \\ \text{CD40 B2} \\ \text{LPS B2} \\ \text{BAFF B2} \\ \text{BCR B2} \\ \text{IL21 B2} \\ \text{Crtl B3} \\ \text{CD40 B3} \\ \text{LPS B3} \\ \text{BAFF B3} \\ \text{BCR B3} \\ \text{IL21 B3} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{Crtl} & \text{CD40} & \text{LPS} & \text{BAFF} & \text{BCR} & \text{IL21} & \text{B1} & \text{B2} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

$$C = \begin{array}{c} \\ \text{Crtl} \\ \text{CD40} \\ \text{LPS} \\ \text{BAFF} \\ \text{BCR} \\ \text{IL21} \\ \text{B1} \\ \text{B2} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{CD40} & \text{LPS} & \text{BAFF} & \text{BCR} & \text{IL21} \end{array} \\ \begin{pmatrix} -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

The analysis resulted in five genelists (for each stimulation one) including significantly differentially expressed genes. A gene was included in a list if the corresponding coefficient β_j was significantly different from zero, thus if the p-value (adjusted after Benjamini Hochberg [12]) was smaller than 0.05.

Treatment with αIgM $F(ab)_2$ fragments led to differential expression of 6596 genes. CD40 activation changed the expression of 1194 genes. IL21 stimulation affected 902 genes. LPS stimulation lead to a significant expression change in 283 genes. BAFF stimulation affected the expression of 129 genes significantly. Expression values of the top 100 genes with the highest fold change between control samples and stimulations are shown in the heatmap in Figure 3.1.

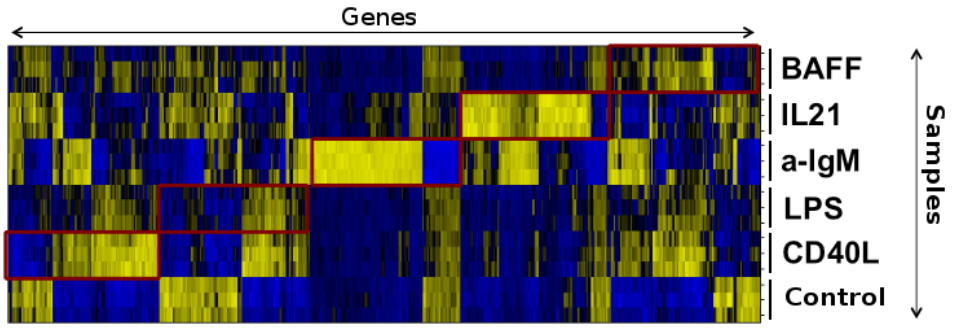


Figure 3.1: The top 100 genes showing the highest fold change per stimulation are shown in the heatmap. The gene expression value are color coded, where yellow is high and blue is low expression (Rows = BL2 Samples, Columns = Genes)

A gene ontology (GO) analysis of each genelists affected by the *in vitro* interventions was performed. Full results can be found in the Appendix (tables A2).
αIgM up-regulated genes are associated with MAP kinase, tyrosine/serine/threonine phosphatase activity and transmembrane transporter activity. We can summarize the corresponding biological processes that were affected as regulation of immune responses, MAP kinase activity, programmed cell death, regulation of metabolic processes or cell cycle and stress responses.
IL21 activated genes that are enriched in genesets associated with responses to viruses and cytokine production including type I interferon biosynthetic processes. IL21 up-regulated genes are enriched in genesets associated with regulation of programmed cell death, same as αIgM up-regulated genes are. The involvement of IL21 activated genes on cytokine signaling could explain its relation to IκB kinase/NF-κB cascade and NF-κB import into nucleus. Furthermore, IL21 affected genes are enriched in genesets associated with Toll-like receptor pathways, JakSTAT and chemokine signaling pathways.
Genes down-regulated by IL21 are enriched in genesets associated with nucleotidyltransferase activity, cytoskeletal protein or phospholipid binding and with cell shape, morphogenesis or chemotaxis.

Genes activated by BAFF are enriched in genesets associated with metabolic processes of amino acids and chromatin remodeling. Genes down-regulated by BAFF are enriched in genesets associated with lipoprotein metabolic process and protein amino acid acylation.

The genes up-regulated by CD40L are enriched in genesets associated with the major histocompatibility complex (MHC) class I receptor activity and hence antigen processing and presentation of peptide antigen, the regulation of membrane potential, small GTPase mediated signal transduction as well as metabolic processes. In contrast, genes down-regulated after CD40L intervention are enriched in genesets associated with phospholipase activity and negative regulation of transcription.

Lists of affected genes overlap significantly across stromal stimuli

To investigate whether the stimuli regulate the same genes, I analyzed the global responses of the stimuli in a pairwise manner using the method compareLists implemented in the Bioconductor package *OrderedList* [122].
The genelists for each stimulation were ranked according to the log fold change after treatment. Pairwise comparisons of the top and bottom ranks of the genelists regulated by IL21, CD40L, αIgM, BAFF and LPS were plotted (Figure 3.2). It was possible to observe high similarities between the genelists. The significant overlap can be observed by the difference between the blue lines, representing the number of overlapping genes at the respective position of the genelists, and the orange area giving the expected size of a random overlap. The genelists are additionally compared in reversed order represented by the green line. When the lists were compared pairwise and in not reversed order, they all showed a high similarity. Thus, all stimulations regulate the same genes in the same direction.

The strongest overlap can be seen between the genelists regulated by LPS, BAFF and CD40L, perhaps reflecting the NF-κB driven gene expression. But this could not be validated with the GO enrichment results, where only the genelist regulated by IL21 was associated to NF-κB signaling.

A strong overlap for the genelists regulated by IL21 and αIgM can be seen and is reflected in the results of the GO analyses. There IL21 and αIgM affected genes enriched for positive regulation of RNA metabolic processes or immune system processes and DNA-repair.

The shared functions of CD40L and αIgM affected genes were for example characterized by immune response, antigen processing and presentation or positive regulation of B cell activation, the BMP signaling pathway and phosphate metabolic processes. Thus, the stimulations of BL2 cells with αIgM, LPS, CD40, IL21 and BAFF all lead to changes in gene expression. Additionally, it was possible to show that the distinct stimuli influence common genes.

Next we addressed the question whether the stimulation data leads to a different stratification of patients, revealing that different pathway combinations are active in some patients and remain inactive in others. We used a dataset published by Hummel and colleagues [81] including 219 gene expression profiles from different lymphomas. The authors introduced an index of Burkitt-likeness for mature aggressive B cell lymphomas. This index reflects how Burkitt-like a sample is and allows classification into molecular BL (mostly sporadic BL), intermediate lymphoma and non-mBL (mostly diffuse large B cell lymphoma (DLBCL)).

The top 100 differentially expressed genes per stromal stimulation were transfered to the hgu133a chip (if possible) and plotted in the heatmap shown in Figure 3.3. The samples are clustered, using complete linkage. Most mBLs show compared to DLBCLs a low expression of genes stimulated by the stromal stimuli. Regarding the assignment of the patients to the ABC-like or the GCB-like subgroup, I could not observe a connection to the shown genesets. The DLBCLs show a continuous pattern but there is no further stratification possible using these genesets. Thus, it would be better to focus on genes that are not only regulated by one of the stromal stimuli, but that additionally stratify DLBCL samples.

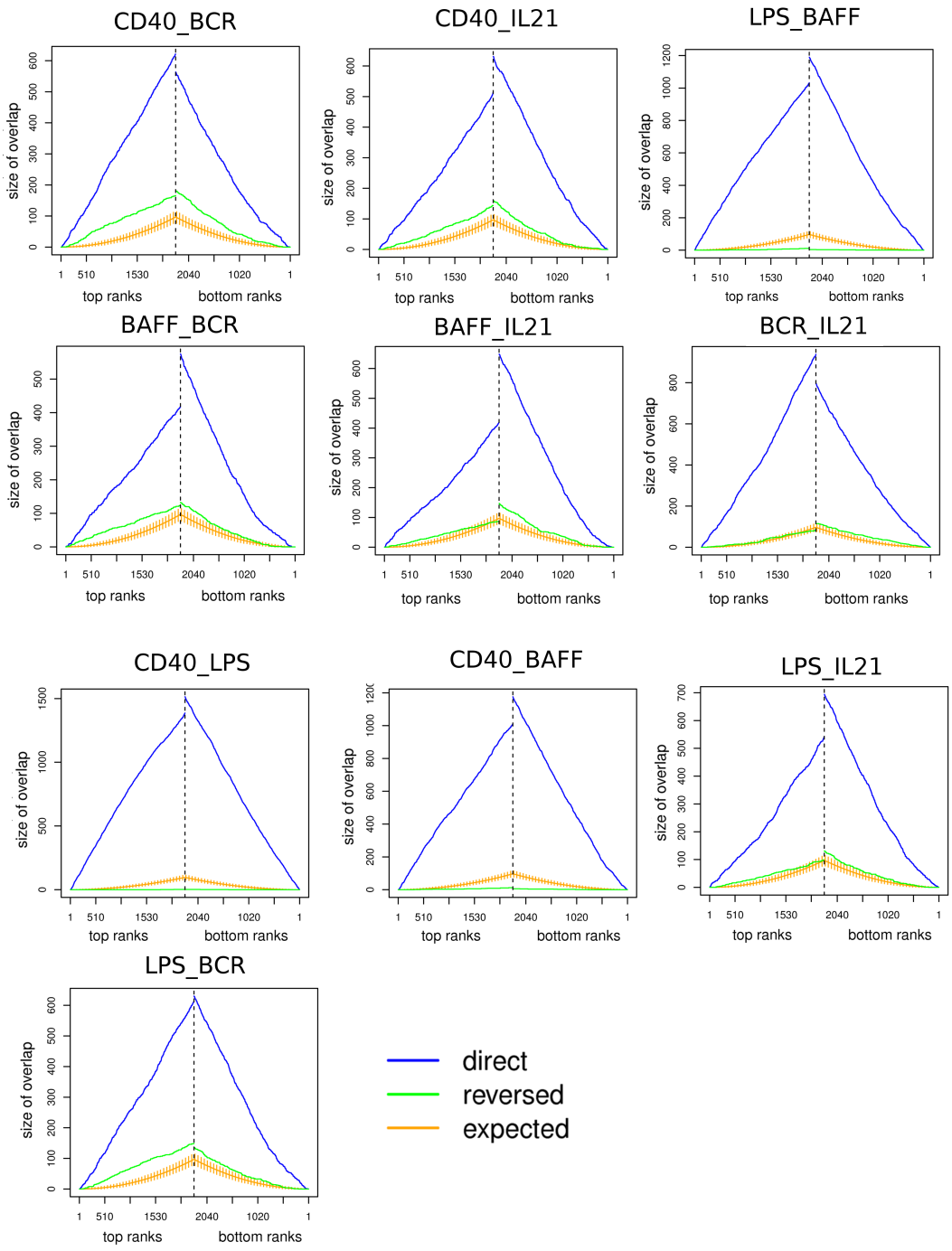


Figure 3.2: Pairwise Compare List results: Pairs indicated by the header of the single pictures. The blue lines represent the number of overlapping genes at the respective position of the genelists and the orange area gives the expected size of a random overlap. The genelists are additionally compared in reversed order represented by the green line.

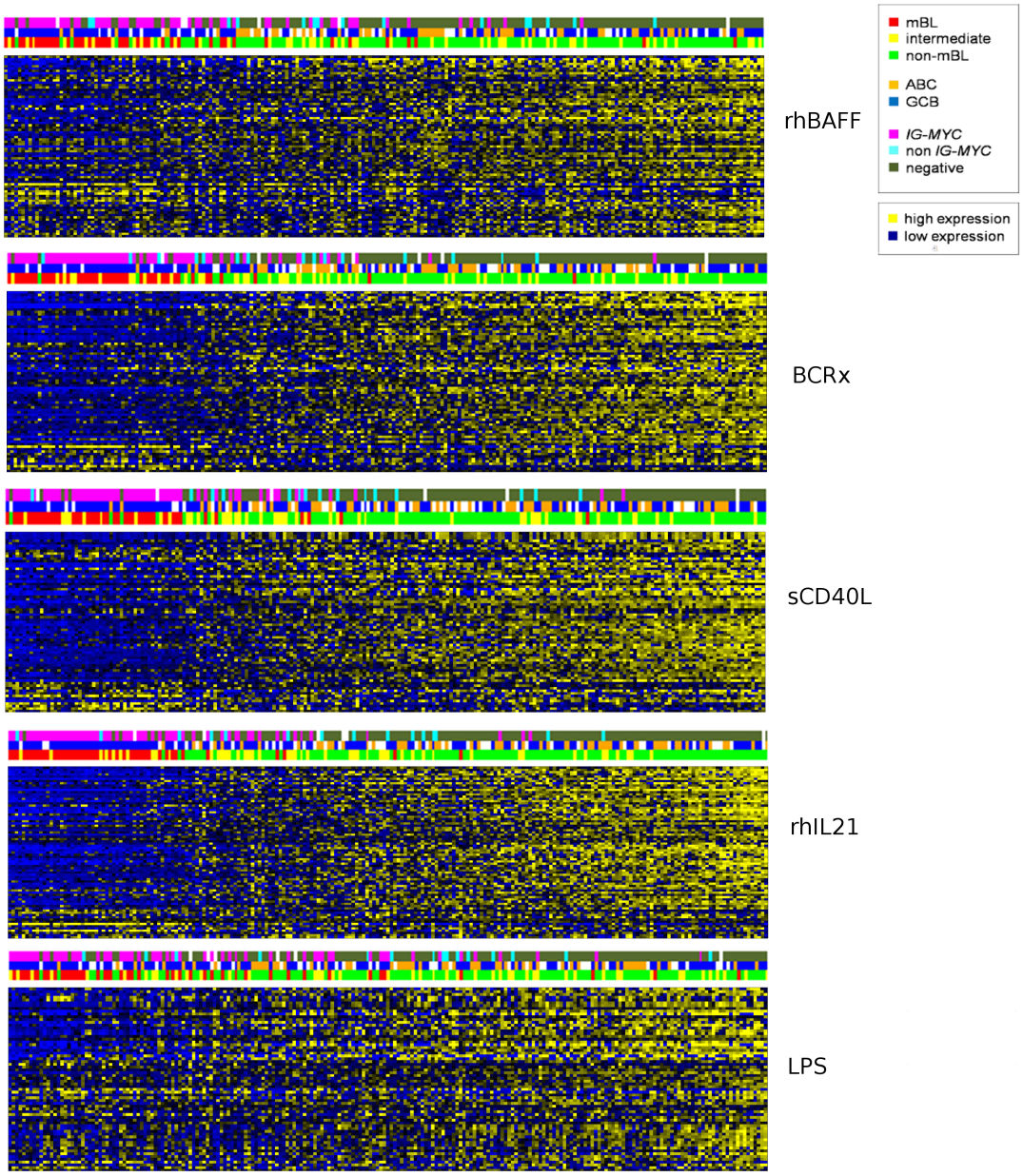


Figure 3.3: Heatmap of gene expression values of the top 100 regulated genes, after stimulation with the 5 distinct stimuli, in the 219 lymphoma samples published by Hummel and colleagues [81]. Samples in each heatmap are ordered differently, by the gene expression of the shown effected genes. Relative gene expression is encoded with yellow (high expression) and blue (low expression).

3.2.5 Identification of activation indices in DLBCL patients that are unique for a specific stromal stimulus

I used the guided clustering method [131] to identify genes that are specific for one of the five stromal stimuli and additionally stratify the DLBCLs in the dataset of Hummel and colleagues [81].
Guided clustering searches for clusters of genes that are highly correlated with a binary vector within the guided data and that show a high correlation structure within the patient data. I used this method five times, once for each stimulation (CD40L, BCR (αIgM), IL21, LPS and BAFF). As guiding data I used the stimulated BL2 cellline data, which was correlated to a binary vector that changed per guided clustering run. Within that binary vector, the samples belonging to one of the stimuli were labeled with 1, and the other stimulations, as well as the controls, were labeled with 0, making sure that index genes were only regulated by one of the five stimuli.
To find genesets that show differences within the DLBCL samples and not only between the DLBCL samples and the mBL samples, the patient data was reduced to the Myc negative DLBCL cases, thus, excluding all mBLs and all the samples that showed a Myc break of some kind.
For each stimulation I extracted up to three clusters depending on their direction and the amount of genes included, ensuring to get at least two clusters which were regulated in different directions. For example the IL21.1 cluster genes are down-regulated after IL21 stimulation, whereas the IL21.2 cluster genes are up-regulated upon stimulation.

But there is a trade-off between uniqueness (correlation on the guided data) and the correlation structure on the patient data. Hence there are some genes included that are not uniquely and/or not so strongly regulated by one of the stromal stimuli but show higher correlations on the patient data. The genelists of the different clusters can be found in the appendix (tables A3).

The retrieved gene clusters are shown in the heatmap plotted in Figure 3.5. All identified clusters are of different size and are only regulated by one of the specific stromal stimuli.

To visualize and to interpret the expression of the different genesets across patients, guided clustering [131] returns an index reflecting the activity of the pathway activated by the corresponding stromal stimulus. I recalculated the index on all 219 lymphoma samples, including mBL samples and all samples with Myc translocation previously excluded. To calculate the indices I used the method described in Section 1.1.

Thus, I received one index vector per gene cluster. These indices represent the activity of the corresponding pathway.

Next I compared the activity of the pathways, activated by the different stromal stimuli across patients by plotting a pairwise scatter plot of the indices (Figure 3.4). There are many indices highly correlated within the patients. This implies that the genes are similarly regulated within the patients and yield the same stratification. Looking at the scatter plot in Figure 3.4 there are two main groups of highly correlated indices. Group 1 includes the indices belonging to cluster IL21.1, BCR.1, BCR.2, CD40.1 and BAFF.1 (Text color green in Figure 3.4). The second group includes CD40.2, CD40.3, LPS.1, LPS.2 and BAFF.2 (text color blue in Figure 3.4). An exception is the index of IL21.2 (text color black in Figure 3.4) which is not correlated to any of the other indices. Thus, two different main effects within the patients driven by the five stimuli can be observed.

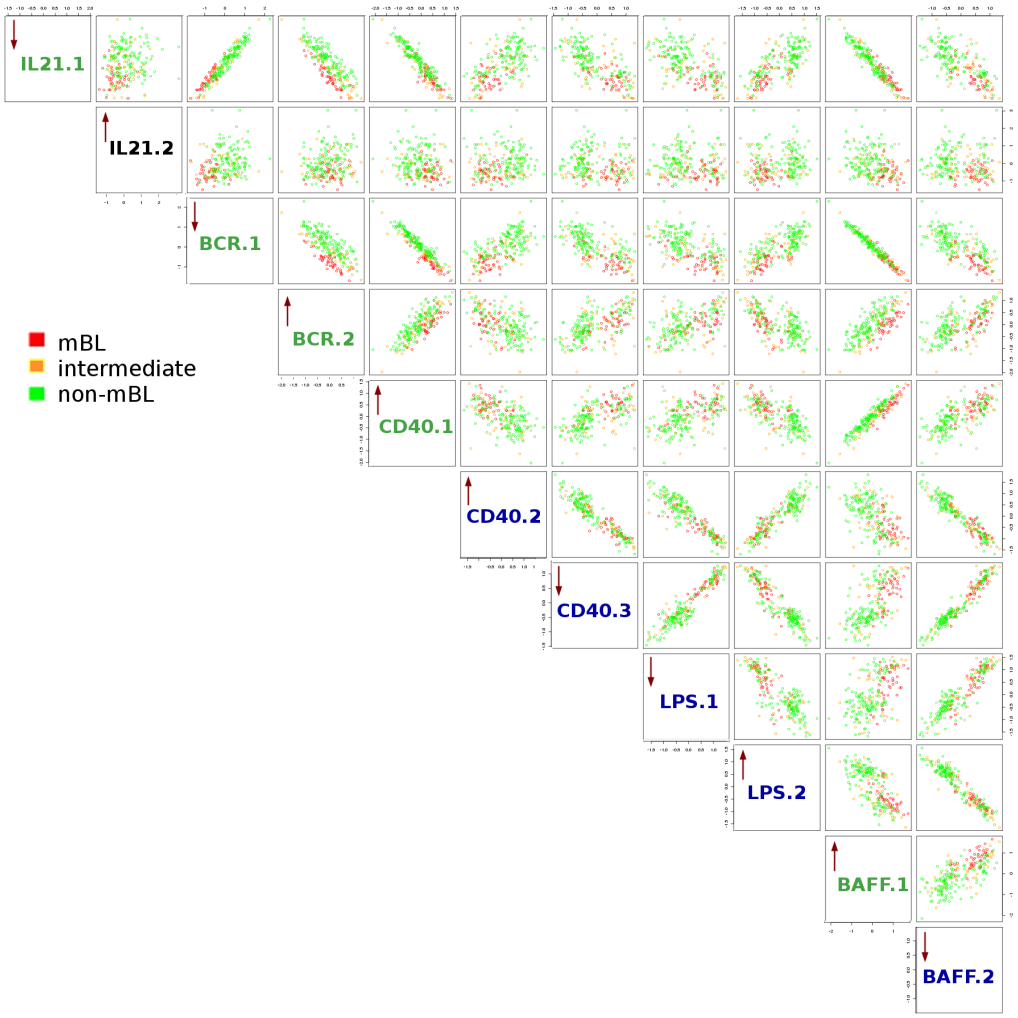


Figure 3.4: Pairwise scatter plot of indices calculated on unique gene clusters. Text color shows correlated groups of indices. The arrow within the first box in each row indicates if these cluster genes were up or down-regulated after stimulation with the respective stimuli.

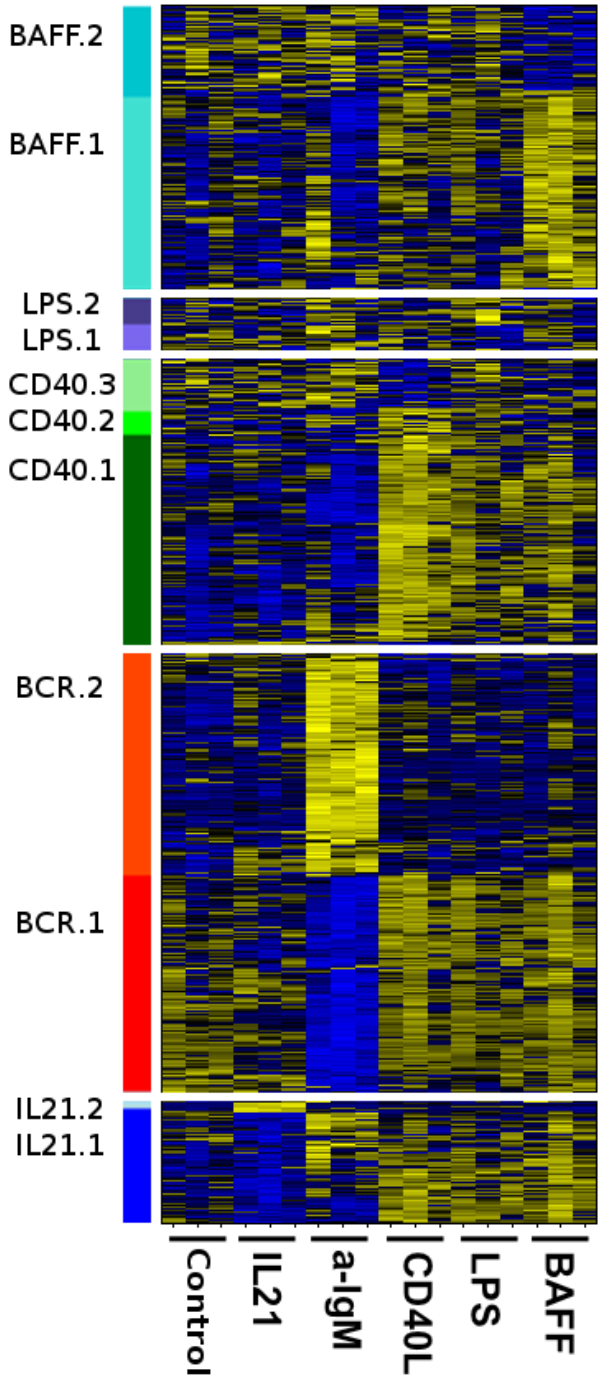


Figure 3.5: Gene expression of the different cluster genes found by guided clustering. The data shown was used as guiding data. Genes are ordered according to their belonging cluster which was found by guided clustering. Samples are ordered according to stimulation. Relative gene expression is encoded with yellow (high expression) and blue (low expression).

Indices hold independent prognostic information in DLBCLs

I asked the question, whether the indices hold prognostic information. I used multivariate linear Cox regression to model survival of DLBCL patients. For a subset of 104 DLBCL patients, the complete clinical data was available. Other prognostic factors, already known, were taken into account as confounding factors and were included in the Cox model. These confounding factors were the age of the patients (AGE), the stage of the disease (Ann Arbor stage, AAS) and the lymphoma definition as ABC- or GCB-like (GCB). For each index one Cox model was computed, including the corresponding index as continuous variable.
The Cox models were defined as follows:

$$h(t) = h_0(t) \times exp(b_1\ \mathrm{GCB} + b_2\ \mathrm{AAS} + b_3\ \mathrm{AGE} + b_4\ \mathrm{INDEX}),$$

where the hazard function $h(t)$ is determined by the baseline hazard h_0 and a set of 4 covariates: $\mathrm{GCB}, \mathrm{AAS}, \mathrm{AGE}, \mathrm{INDEX}$.

The aim of the analysis was to estimate the 4 coefficients b_i $(i = 1, \ldots, 4)$ for the 4 covariates $(\mathrm{GCB}, \mathrm{AAS}, \ldots, \mathrm{INDEX})$. The 4 covariates act as factors on the hazard $h(t)$ at any time point t. Thus the hazard of the event in any group was a constant multiple of the hazard in any other group. This proportionality implies that the quantities $exp(b_i)$ can be called hazard ratios [20]. I used the *coxph* function provided by the R-package *survival*, which fits the Cox proportional hazards regression model.

The fitted Cox model including the BAFF.2 index is returned by the *coxph* function in the following way:

```
   Call:
coxph(formula = surv.obj ~ as.factor(AAS) + GCB + AGE +  IndexBAFF2)

                   coef exp(coef) se(coef)      z   Pr(>|z|)
AASII           1.20827   3.34768  0.60507  1.997   0.0458 *
AASIII          1.49588   4.46325  0.58270  2.567   0.0103 *
AASIV           1.14474   3.14163  0.66990  1.709   0.0875 .
GCB            -0.29737   0.74277  0.36923 -0.805   0.4206
AGE             0.05533   1.05689  0.01421  3.894 9.88e-05 ***
IndexBAFF2     -3.58501   0.02774  1.69722 -2.112   0.0347 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The $coef$ column includes the estimated coefficients b_i $(i = 1, \ldots, 4)$. The $exp(coef)$ column, where $coef$ are the b_i $(i = 1, \ldots, 4)$, includes the estimated hazard ratios. We can see, that the Ann Arbor stages (AAS) and AGE have coefficients greater than zero, thus if the AAS is higher, or if the AGE increases, the event hazard increases and thus the length of survival decreases.

In contrast to that, the estimated coefficients for GCB and $IndexBAFF2$ are smaller than zero, thus, a GCB expression type or a low $IndexBAFF2$ value is associated with a better prognosis. The z column includes values that were calculated by dividing the estimated coefficients $(coef)$ by their corresponding standard error $se(coef)$. It is tested if these z values are significantly different from zero, yielding the p-values given in the column $PR(>|z|)$.

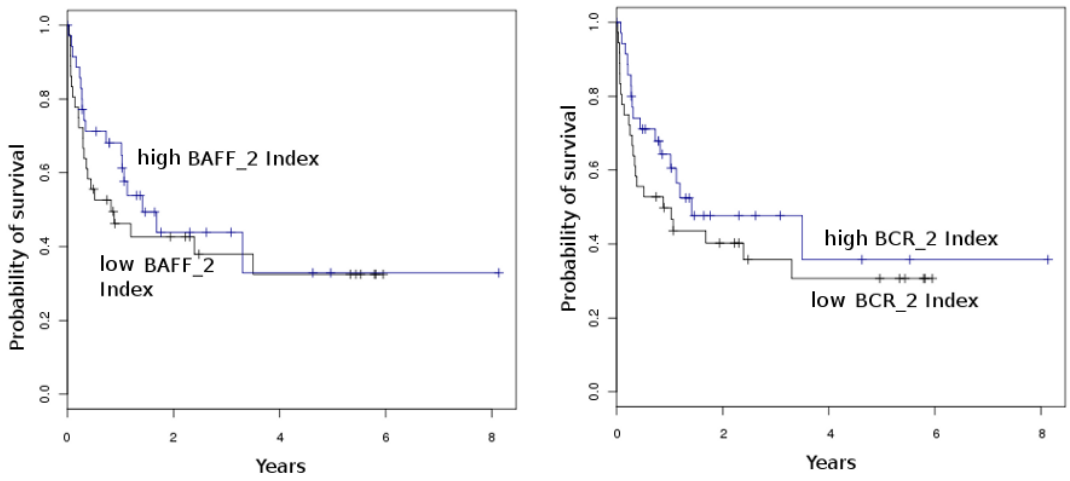
The fitted cox models using the function *coxph* including the BCR.2 index can be seen here:

```
   Call:
coxph(formula = surv.obj ~ as.factor(AAS) + GCB + AGE +  IndexBCR2)

                    coef  exp(coef) se(coef)     z  Pr(>|z|)
AASII            1.43033   4.18010  0.62227  2.299  0.02153 *
AASIII           1.59093   4.90830  0.58511  2.719  0.00655 **
AASIV            1.38878   4.00996  0.65252  2.128  0.03331 *
GCB             -0.26716   0.76555  0.37214 -0.718  0.47282
AGE              0.05504   1.05658  0.01377  3.998 6.39e-05 ***
IndexBCR2       -2.41835   0.08907  0.95973 -2.520  0.01174 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As expected, the influence of the covariate AAS, GCB and AGE on survival did not change. The estimated coefficient for $IndexBCR2$ is smaller than zero, thus a lower $IndexBCR2$ value is associated with a better prognosis. The two indices BAFF.2 and BCR.2 hold significant prognostic information, with p values $(PR(>|z|))$ smaller than 0.05.

In Figure 3.6 Kaplan-Meier curves for the two prognostic indices are shown. For the Kaplan-Meier curves the patients were grouped by index value into two groups. The first group comprised all patients with an index, smaller than the median index over all patients, and the second group, comprised all patients showing a higher index.



(a) (b)

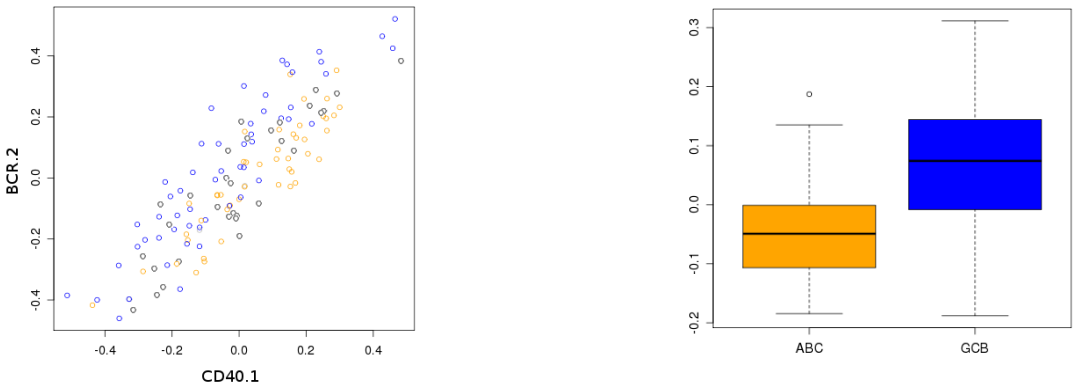
Figure 3.6: Kaplan-Meier estimates of survival using the (a) BAFF.2 Index and (b) BCR.2 Index. DLBCL patients are grouped by index value. (Sample-index smaller than the median or bigger than that).

Thus, the indices BAFF.2 and BCR.2 are of prognostic relevance for DLBCL patients.

GCB-like lymphomas are characterized by unbalanced BCR and CD40 signaling

Alizadeh and colleagues [2] sub-stratified DLBCLs depending on their similarity with normal cells in different differentiation stages of B cells into ABC-like and GCB-like DLBCLs. The ABC-like DLBCLs show a high aberrant activity of NF-κB and have a worse prognosis compared to GCB-like DLBCLs. I checked if it is possible to associate my indices with the two types of DLBCLs. The expression of the different target genes does not seem to characterize ABC-like or GCB-like lymphomas on their own. However, the difference of the two indices BCR.2 and CD40.1 (BCR.2-CD40.1) does (Figure 3.7).
To investigate whether the difference seen in the scatter plot (Figure 3.7 (a)) is a significant result, I combined the information given by the indices BCR.2 and CD40.1 by substracting the CD40.1 index from the BCR.2 index yielding a new combined index.

By analyzing the "joint" activity of these genesets in DLBCL patient samples, I found that there is a significant (t-test: p value $leq0.05$) difference between ABC-like DLBCL and GCB-like DLBCL. Thus, GCB-like lymphomas are associated with unbalanced BCR/CD40 signaling.



(a) (b)

Figure 3.7: (a) Scatterplot of BCR.2 index plotted against the CD40.1 index. The orange circles are ABC-like lyphomas the blue ones are GCB-like lymphomas and the grey ones were not classified as such (b) Boxplot of the combined index, calculated by substracting the CD40.1 index from the BCR.2 index ($BCR.2 - CD40.1$). A t-test was used to test if the difference between the two groups is significant yielding a p value of $1.385e-08$.

High c-Myc activity is associated with a low BCR.1 index

In Burkitt lymphoma (BL) aberrant c-Myc activity results from IG-MYC translocations [109]. However, MYC aberrations also occur in DLBCLs and have a negative prognostic impact [170]. Schrader and colleagues developed a "c-Myc index" which reflects the expression of c-Myc responsive genes and thereby the activity of c-Myc [170]. As expected, the "c-Myc index" is high in mBLs [170].

To see if we can link the c-Myc activity to one of our previously identified unique gene indices and thus to respective pathways, we compared the c-Myc index to our indices using 219 aNHL cases [81]. I found that the BCR.1 index is negatively correlated to the c-Myc index, with a Pearson correlation coefficient of -0.76 (Figure 3.8).
Lymphomas that exhibit high c-Myc activity are associated with a high expression of the genes included in the BCR.1 geneset, which are down-regulated upon BCR activation. The mBL cases show a high c-Myc index, as expected, and a low BCR.1 index.

However, this correlation can also be seen in the DLBCL cases. DLBCL cases that show a low c-Myc index have a high BCR.1 index and vice versa.

The results indicate a connection between c-Myc activity and the expression of the BCR.1 geneset. One possible explanation for the strong negative correlation could be that BCR.1 genes are regulated directly or indirectly through c-Myc. To get an idea of the function of these genes, a gene ontology (GO) analysis of the BCR.1 genelist was done.
The BCR.1 genes were enriched in genesets associated with cell cycle. The genes down regulated after BCR activation were enriched in genesets associated with "M phase of mitotic cell cycle" (GO:0000087), "cell cycle" (GO:0007049) , "M phase" (GO:0000279) and "cell cycle process" (GO:0022402). Thus, the differences of BCR.1 index in aNHL could explain the differences of growth behavior differences in aNHL.

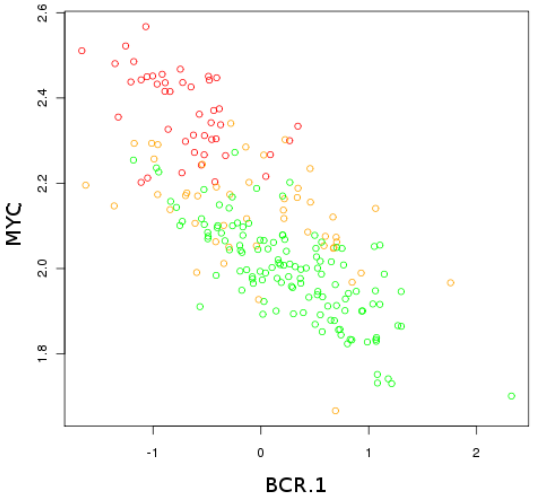


Figure 3.8: Scatterplot of the BCR.1 index plotted against the c-Myc index of 119 aNHL cases [81]. Circle colors represent molecular diagnosis: Red= mBL, green= non-mBL and yellow= intermediate

Discussion

We identified patterns of gene expression changes in response to LPS, IL21, BAFF, CD40L stimulation and αIgM crosslink and mapped them to expression profiles of lymphoma samples. It was possible to show that the distinct stimuli influence common genes and DLBCLs show a continuous pattern of gene module activity, whereas the mBLs show in general a lower pathway activation than DLBCLs. The identification of gene clusters regulated by the stromal stimuli offers new opportunities for the analysis of aberrant signaling in lymphomas.

We used guided clustering to identify genesets uniquely regulated by one of the five stimuli. From these regulated genesets we built indices on aNHL patients to summarize the corresponding pathway activity across patients. We revealed two different effects within the aNHL that hold prognostic information.

Furthermore, we found that the c-Myc index is negatively correlated to the newly identified BCR regulated gene cluster (BCR.1). One explanation for that could be that BCR.1 genes are regulated directly or indirectly via c-Myc. Furthermore, the expression of the BCR.1 geneset could be linked to differences in aNHL and to genesets mainly associated with the cell cycle. Further experiments are necessary to clarify the influence of BCR signalling on the growth of aNHLs.

3.3 Aberrant Lymphocyte enhancer binding factor 1 expression is characteristic for sporadic Burkitt lymphoma compared to Diffuse Large B cell Lymphomas

This project was done in cooperation with the group of Dieter Kube from the department of Haematology and Oncology, University Medical Center Göttingen. Most experiments were done by Antje Ulrich. I performed the statistical analysis and evaluation of the microarray data. Parts of the work are already published: "Aberrant Lymphocyte Enhancer-binding Factor 1 expression is characteristic for sporadic Burkitt lymphoma" by A. Ulrich, N. Walther, F. von Bonin, W. Klapper, K. Meyer, A. Schrader, M. Vockerodt, R. Spang, T. Pukrop, P. Murray,L. Trmper, D. Kube in The American Journal of Pathology [190].

BL is a highly malignant aNHL consisting of rapidly proliferating B cells [81, 102]. A characteristic feature of BL is a translocation of the MYC gene locus to one of the three IG gene loci [81, 102]. A recent large-scale gene expression profiling study introduced an index of Burkitt-likeness for mature aggressive B cell lymphomas [81]. Lymphocyte enhancer-binding factor 1 (LEF1) is one of 58 genes defining the Burkitt index.

LEF1 was first identified in T-lymphocytes. It is a member of the LEF/TCF family of transcription factors [185, 194]. LEF1-deficient mice exhibit defects in Wnt3a dependent pro-B cell proliferation and survival [164]. LEF1 is a central mediator of Wnt signaling. The Wnt pathway is important in cell development and differentiation (reviewed in [120, 28]). In immune cells the canonical Wnt signaling pathway regulates not only early steps of B cell development but also self renewal of hematopoietic stem cells [177].

The classical view is that upon binding of canonical Wnt ligands to corresponding receptors cytoplasmic β-catenin gets stabilized. It translocates into the nucleus and there it binds LEF/TCFs acting as a matrix for transcriptional coregulators of Wnt target genes. It is thought that LEF/TCF proteins do not independently activate target genes [65]. However, in BL and DLBCL no β-catenin was observed in the nucleus [67, 136]. LEF1 may not function in BLs in the classical way. This hypothesis is in line with growing evidence that LEF1 can also functionally interact with other proteins in the absence of β-catenin [34, 23].

In a number of hematopoietic neoplasia the activity of the Wnt pathway or its components, including LEF1, are dysregulated. This is supported by observations in B cell chronic or acute lymphatic leukemia (CLL/ALL) or in subsets of mantle cell lymphoma (MCL) and T cell leukemias as well as in certain celllines and mouse models [67, 150, 162, 200, 191, 167, 118, 46, 63, 139, 70, 126].

Nevertheless, detailed mechanisms are unclear. Studies on Wnt signaling in immune cells support the notion that the canonical Wnt signaling pathway regulates early B cell development [177] and not only self renewal of hematopoetic stem cells. LEF1-deficient mice exhibit defects in Wnt3a dependent pro-B cell proliferation and survival [164]. It is proposed that LEF1 expression is lost in peripheral B lineage cells. Nevertheless, a deregulation of LEF1 in mature B cells is supported by the observation of aberrant LEF1 expression in CLL.
Additionally, Qiang and colleagues found LEF1 expression in BL celllines [162, 67, 126, 70, 69].

Furthermore, in murine hematopoetic stem cells retrovirally transduced with a constitutive active LEF1 variant (fusion of β-catenin with LEF1) B-ALL and acute myeloid leukemia (AML) were observed [150]. LEF1-induced AMLs were characterized by immunoglobulin DH-JH rearrangements and expression of lymphoid and myeloid regulatory factors. The IG-gene rearrangements shown suggest that LEF1 induced AML was propagated by a leukemic stem cell with lymphoid characteristics. Thus LEF1 can be involved in defining leukocyte plasticity supporting the hypothesis that lymphomas of B cell origin may be characterized by reexpression of early markers in B cell differentiation providing B cell plasiticity during transformation [132].

We show that LEF1 expression is characteristic for BL in comparison to DLBCL. Aberrant LEF1 expression is involved in proliferation of BL cells. Using whole genome expression analysis new LEF1 target genes in lymphoma cells are identified. The joint expression of newly identified LEF1 target genes is a surrogate for the regulatory activity of LEF1 in individual lymphomas.

3.3.1 Data

127 B cell lymphoma specimens were obtained from the files of the Lymph Node Registry Kiel according to ethical approval. All lymphomas were classified according to the WHO classification using standard histological, immunohistochemical and molecular criteria [181].

In addition, 20 specimens obtained within the framework of the network project Molecular Mechanisms in malignant Lymphoma (MMML) were analysed. The latter cohort was diagnosed and classified based on molecular signatures [81]. All five lymphomas were analysed using tissue-micro-arrays (TMA) containing two cores of 1 mm or 0.6 mm (for the MMML specimen) for each case (ethical approval Ref. Nr. 1/1/05).

For gene expression analysis, RNA from BL2 cells was isolated after LEF1 specific RNA interference from three independent experiments. Fragmentation and hybridization of labeled anti sense RNA on Human Genome U133A 2.0 plus Arrays (Affymetrix) was performed according to manufacturers recommendations by the "Kompetenzzentrum für Fluoreszente Bioanalytik, Regensburg".
Gene expression values were obtained by first removing the background and normalizing on probe level using the variance stabilization method by Huber and colleagues [80]. The normalized probe intensities were summarized into gene expression levels by using an additive model [84] fitted by the median polish procedure [187].

3.3.2 LEF1 is aberrantly expressed in BL cells

In order to show aberrant expression of LEF1 in BL cells various experiments were conducted. A representative panel of B cell derived tumours, primary normal GC B cells and different celllines, were screened for LEF1 expression. Immunohistochemical (IHC) analysis revealed that strong LEF1 expression found in interfollicular areas of reactive lymph nodes, tonsils or spleen is restricted to CD3 positive T cells. A corresponding example of LEF1 staining in normal reactive tonsillar tissue sections can be seen in Figure 3.9 I and II. AB cells were devoid of detectable nuclear LEF1 expression.

The 127 lymphomas analysed by IHC were of B cell origin and 99 were classified as NHL. These samples included 10 BL and 31 DLBCL cases. In total, 22 tumours were found to show LEF1 expression. Due to insufficient staining (missing internal positive controls) two of the ten BL samples were not evaluable. Five of the remaining eight BL samples (63%) expressed nuclear LEF1 (Figure 3.9 III). Notably, LEF1 was found frequently in BL, which are known to be devoid of nuclear β-catenin [67, 136]. Only 1 of the 31 DLBCL (3%) was found to be positive for LEF1 expression (Figure 3.9 IV). All B-CLL ($n = 6$) and B-ALL ($n = 6$) showed nuclear LEF1 expression. No LEF1 expression was detected in mantle cell (MC), marginal zone or follicular lymphoma.

LEF1 expression was investigated in different lymphoma celllines. Immunoblot analysis revealed that 7/7 BL lines display increased LEF1 expression, while DLBCL and MCL celllines show weak or no LEF1 expression. qRT-PCR confirmed stronger expression of LEF1 in BLs compared to DLBCL cells. No mutations could be detected for LEF1 in BL or HL cells analysed. A 3‘-RACE was performed to detect variations in the usage of full length LEF1. Sequencing of cDNAs showed that in BL2 cells different 3‘-variations can be detected: full length LEF1 with and without exon 6 and corresponding variants with a truncation of exon 11.

Since in BL and DLBCL tissue no β-catenin was observed in the nucleus [67, 136], we tested whether celllines express β-catenin. Immunoblot analysis showed weak to no expression of β-catenin. Only DLBCL cells OCI-Ly3 were positive for β-catenin when compared to SW480 cells.
Taken together, LEF1 is aberrantly expressed in BL cells.

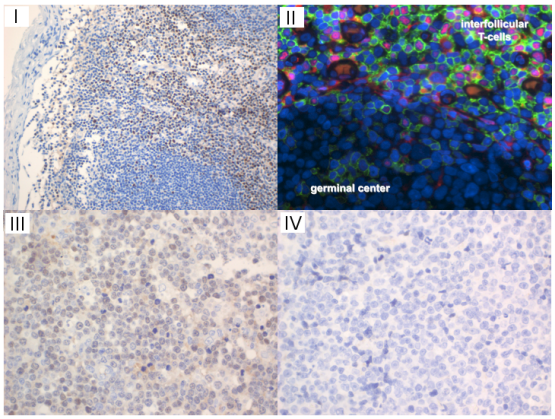


Figure 3.9: Representative examples of lymphoma precursor cells and lymphomas stained for LEF1. (I) Detection of LEF1 (brown) in the follical mantle of a germinal center in normal reactive tonsillar tissue sections, (II) CD3 positive T cells (green) express LEF1 (red) obtained from normal tonsil tissue (merged picture including DAPI for nuclear counterstain). (III) BL (IV) DLBCL

3.3.3 Identifying LEF1 target genes in BL cells

Next, transient siRNA-mediated knockdown of LEF1 in BL2, BL30, BL70 and Ramos cells was performed. Four independent siRNAs directed towards different exons were used in a pooled assay. LEF1 knockdown was confirmed by qRT-PCR and immunoblot (Figure 3.10 (A) and (B)).

To identify LEF1 target genes, RNA from BL2 cells was isolated after LEF1 specific RNA interference from three independent experiments and hybridized to U133 plus2.0 whole genome expression microarrays. I computed a list of differentially expressed genes between the expression profiles of perturbed and control celllines by using linear models implemented in the Bioconductor [64] package *limma* [176].
The linear model used was:

$$E[y_j] = X\alpha_j,$$

where y_j is the expression data for gene j, X is the *design matrix* and α_j is the vector of coefficients for each gene j to be estimated.
The following design matrix X and contrast matrix C were used:

$$X = \begin{array}{c} \\ \text{Scrb B1} \\ \text{LEF1 B1} \\ \text{Scrb B2} \\ \text{LEF1 B2} \\ \text{Scrb B3} \\ \text{LEF1 B3} \end{array} \begin{array}{c} \begin{array}{cccc} \text{Control} & \text{LEF1} & \text{B1} & \text{B2} \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \end{array} \qquad C = \begin{array}{c} \\ \text{Control} \\ \text{LEF1} \\ \text{B1} \\ \text{B2} \end{array} \begin{array}{c} \begin{array}{c} \text{Control-LEF1} \end{array} \\ \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} \end{array}$$

The design matrix determines how the 4 coefficients are to be estimated. The model took the treatment ($LEF1$ = siRNA-mediated knockdown of LEF1, controls)and the different batches $B1, B2$ and $B3$ into account. The third batch $B3$ was implicitly modeled, in a way that being batch $B3$ is determined by not being batch $B1$ or $B2$.
The contrasts β_j were computed as follows:

$$\beta_j = C^T * \alpha_j,$$

where α_j is a vector containing the 4 estimated coefficients and C is the contrast matrix. The result was a list of differentially expressed genes between control samples and samples with siRNA-mediated knockdown of LEF1. In the list 1453 genes were found to be significantly differentially expressed having a p value (adjusted after Benjamini and Hochberg [12]) smaller than 0.01. The top 100 genes were plotted in a heatmap shown in Figure 3.10.

These LEF1 targets genes are involved in gene regulation (HIST1H2BK, H2AFV, STAT2, ID1), cell signaling (BMPR2, CCDC88C, PRKAR1A, ITGB1), cell proliferation (CDC27, UBE2H, PSAT1, SATB1), regulation of metabolism (CTSH, ATP2B4, MAN1A1) and antigen presentation (HLADMB, FcRL3). Others are involved in B cell interaction with T cells (CD86), tumour progression (ARL6IP5, PSAT1, CTDSPL), and B cell differentiation (PBXIP1).

Genes down-regulated in cells with reduced LEF1 levels are LEF1 activated (positive regulated), whereas genes, that are up-regulated upon LEF1 knockdown are inhibited (negative regulated) by LEF1. Thus, for example, the expression of the genes for ARL6IP5, SATB1, STAT2, CCDC88C or CD86 is increased, whereas the expression of the genes CDC27, CTDSPL or PSAT1 is lower, when LEF1 expression is reduced by RNAi.

A set of LEF1 regulated genes was validated by qRT-PCR in BL2 and Ramos celllines, and showed a good agreement with the microarray data. In Ramos cells the effect was not obvious for $STAT2$ and $CTDSPL$. $CD86$ was not detectable in control siRNA transfected BL2 cells, but was expressed at low levels after LEF1 knockdown. Similarly, $SATB1$ is absent in Ramos cells, but appears after LEF1 silencing. Interestingly, expression of TCF1 or c-Myc was not affected.

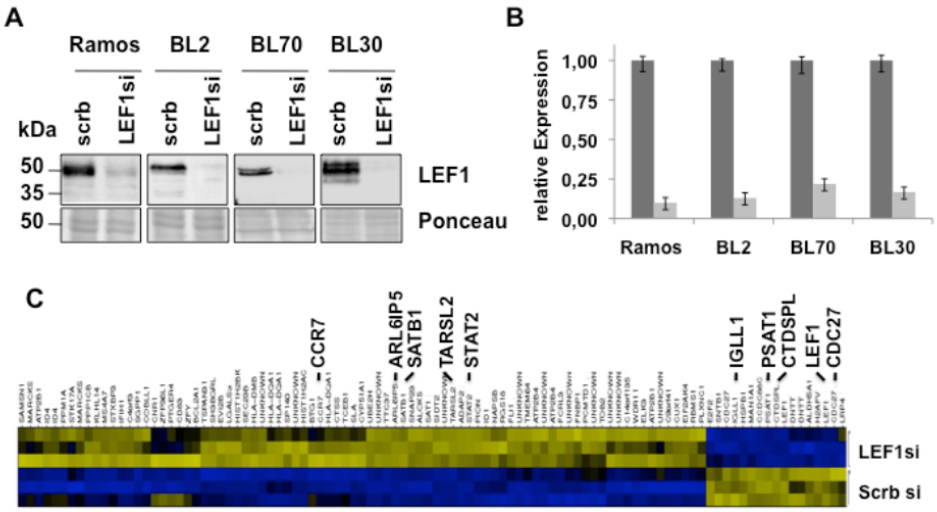
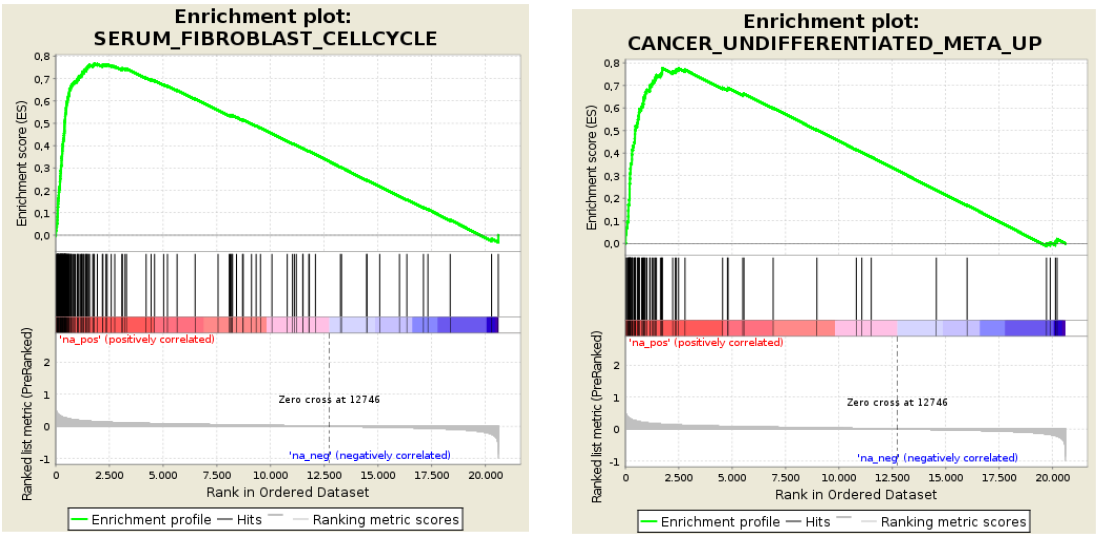


Figure 3.10: (A) Immunoblot analysis of LEF1 expression in Ramos, BL30, BL70 and BL2 lymphoma celllines after RNAi mediated knockdown of LEF1 (B) qRT-PCR analysis of the expression of LEF1 in lymphoma cells after RNAi mediated knockdown of LEF1. (C) Heatmap of LEF1 top 100 affected genes from RNAi mediated knockdown of LEF1 in BL2 cells. Each row in the heatmaps represents a microarray sample and each column represents a gene. Yellow and blue indicate high and low expression.

I ranked the list of differentially expressed genes by fold change and used geneset enrichment analysis (GSEA) [180] to detect functionally characterized genesets that are enriched at either end of the LEF1 target genelist.

Six genesets associated with cell cycle regulation were significantly enriched among LEF1 inducible genes, for one of the six genesets the enrichment plot is shown in Figure 3.11 (a). Most importantly, among LEF1 inducible genes a significant enrichment of genes that are up-regulated in tumour cells relative to matching normal tissue cells across many tumour entities [165] was observed.

Enrichment plots are shown in Figure 3.11. The results suggest that LEF1 target genes represent essential transcriptional features of neoplastic transformation.



(a) (b)

Figure 3.11: Example from geneset enrichment analysis of LEF1 affected genes. Enrichment plot from SERUM FIBROBLAST CELLCYCLE (a) and CANCER UNDIFFERENTIATED META UP (b) including the profile of the running ES Score and positions of geneset members on the rank ordered list.

3.3.4 LEF1 knockdown affects BL cell proliferation

Based on the enrichment of cell cycle regulators among LEF1 target genes, the effects of LEF1 silencing were assessed by cell cycle analysis, DNA synthesis experiments for proliferation and MTT assays for cell viability. DNA synthesis of cultured BL cells was monitored by 3H-thymidine.

Transient LEF1 knockdown in BL2 cells resulted in a marked inhibition of cell proliferation of about 40% when compared to control while cell viability was affected only by about 10%. Similar effects ($10 - 20\%$ reduction) were observed for Ramos, BL30 or BL70 cells in the absence of cell viability changes.
Additionally, the reduced expression of LEF1 in BL cells induces arrest at the G1 phase of the cell cycle. Increased numbers of cells in G1 phase were consistently observed in BL2, BL70, BL30 and Ramos cells, but were more pronounced in BL2 cells than in Ramos, BL70 and BL30 cells.

3.3.5 LEF1 target genes are affected by β-catenin and Wnt signaling

I evaluated the effect of canonical Wnt-signaling onto the newly identified LEF1 target genes. BL2 cells were transfected with a plasmid expressing constitutive β-catenin (β-cateninS32A) or an empty vector as control. Additionally, BL2 cells were stimulated by Wnt3a. Using linear models implemented in the Bioconductor [64] package *limma* [176], I searched for genes that are differentially expressed.
The linear model used was:

$$E[y_j] = X\alpha_j,$$

where y_j is the expression data for gene j, X is the *design matrix* and α_j is the vector of coefficients for each gene j to be estimated.
The following design matrix X and contrast matrix C were used:

$$X = \begin{array}{c} \\ \text{Scrb B1} \\ \text{Wnt3a B1} \\ \text{BetaCat B1} \\ \text{Scrb B2} \\ \text{Wnt3a B2} \\ \text{BetaCat B2} \\ \text{Scrb B3} \\ \text{Wnt3a B3} \\ \text{BetaCat B3} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Control} & \text{Wnt3a} & \text{BetaCat} & \text{B1} & \text{B2} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

$$C = \begin{array}{c} \\ \text{Control} \\ \text{Wnt3a} \\ \text{BetaCat} \\ \text{B1} \\ \text{B2} \end{array} \begin{array}{c} \begin{array}{cc} \text{Control-Wnt3a} & \text{Control-BetaCat} \end{array} \\ \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \end{array}$$

The design matrix determines how the 5 coefficients are to be estimated. The model took the treatments ($Wnt3a$ =stimulation with Wnt3a, $BetaCat$ = ectopic expression of β-catenin, controls) and the batches $B1, B2$ and $B3$ into account. The third batch $B3$ was implicitly modeled, in a way that being batch $B3$ is determined by not being batch $B1$ or $B2$.

The contrast β_j were computed as follows:

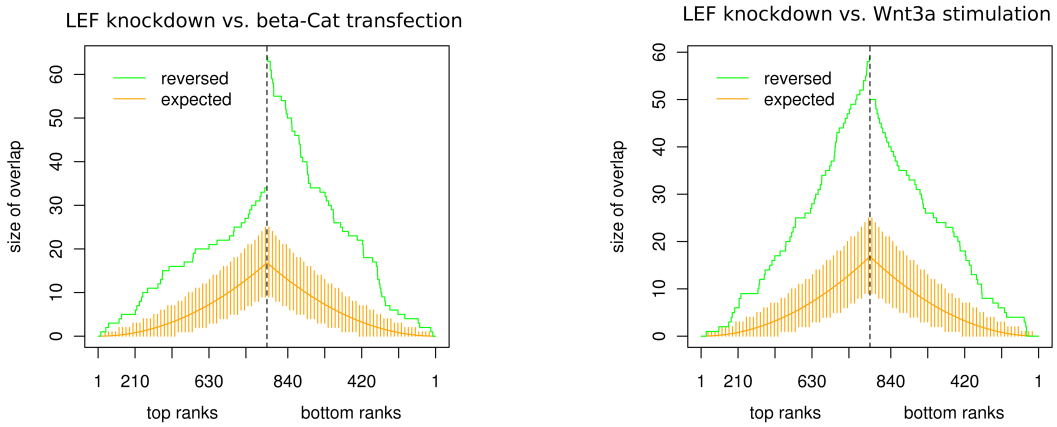
$$\beta_j = C^T * \alpha_j,$$

where α_j is a vector containing the estimated coefficients and C is the given contrast matrix. Two lists of differentially expressed genes were generated: One after stimulation with Wnt3a and one after ectopic expression of β-catenin.

I ranked the two lists by fold change between controls and interventions. Next I used the method *compareLists* implemented in the Bioconductor package *OrderedList* [122] to quantify the similarity between the LEF1 target genelist (sorted in the same way) and the two newly generated genelists.

Comparisons of the top and bottom ranks of the genelists regulated by LEF1, Wnt3a and β-catenin were plotted (Figure 3.12). Genes up-regulated by LEF1 knockdown showed a significant overlap with genes down-regulated by β-cateninS32A and vice versa (Figure 3.12 a). A similar relation was observed for the Wnt3a stimulation (Figure 3.12 b).

The significant overlap can be observed by the difference between the green lines, representing the number of overlapping genes at the respective position of the genelists and the orange area giving the expected size of a random overlap. Thus, on global level LEF1 target genes are also affected by Wnt3a and β-catenin signals in the respective direction.



(a) (b)

Figure 3.12: CompareLists Results: Comparison of the ranked genelist of the LEF1 knockdown experiment, (a) to the genelist affected by ectopic β-cateninS32A expression and (b) to the genelist resulting of a Wnt3a stimulation. The genelist after LEF1 knockdown was reversed to match the over-expression / stimulation contexts. The y-axis gives the number of genes included in the top n-genes (x-axis) of both genelists. The green line gives the overlap of the two genelists at each rank, where the order of one genelists was reversed. The orange line gives the expected size of a random overlap, and the vertical orange lines indicate 95% probability intervals of random overlaps.

3.3.6 LEF1 activity is higher in mBLs compared to DLBCLs

To further prove the functional relevance of LEF1 target genes identified *in vitro*, we investigated if the top 100 LEF1 target genes are also expressed preferentially in primary BL samples. Furthermore, we wanted to know if this LEF1 activity defines a group of genes of diagnostic relevance and can be used to discriminate subgroups of aggressive NHL. The expression of the top ranking LEF1 targets were combined to an index of LEF1 activity. From the top 100 LEF1 target genes, 68 genes were present on the Affymetrix HG-U133A gene chip and were used for the index. The gene expression values of theses 68 genes were condensed into an index of LEF1 target expression using the method getIndex explained in Section 1.1. This index quantitatively assesses to which degree an individual lymphoma expresses LEF1 target genes.

Lymphomas with a high LEF1 index showed more evidence of aberrant LEF1 activity than those with a low index. Genes activated by LEF1 were highest in mBL cases, low in non-mBLs and in between in those defined as intermediate lymphoma (Figure 3.13).

For those LEF1 target genes identified as suppressed by LEF1 an inverse correlation was observed (Figure 3.13). The expression of LEF1 target genes used to build the LEF1 index can be seen in Figure 3.14 showing the 219 lymphomas in a heatmap. Genes are split into activated or suppressed target genes of LEF1. Samples are sorted according to the LEF1 index. Thus, the combination of LEF1 target genes with large scale gene expression data from lymphoma patients enables us to measure the LEF1 activity in lymphoma subtypes. LEF1 activity is higher in BLs compared to DLBCLs.

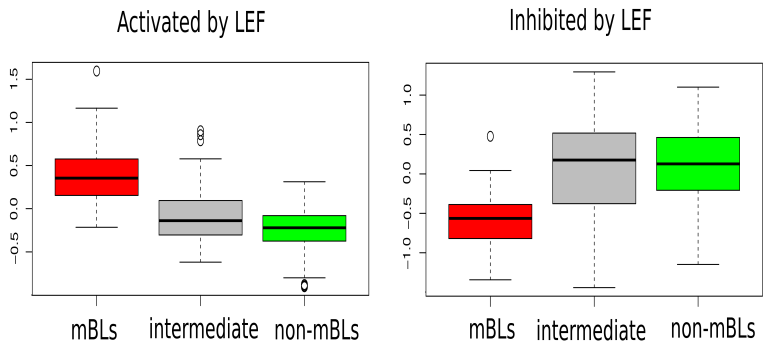


Figure 3.13: Boxplots of the LEF1 target gene expression index across mBL, intermediate and non-mBL lymphomas (Patient numbers: mBL=44, intermediates=48, non-mBLs=127). The index is once calculated for the target genes activated by LEF and once for the genes inhibited by LEF.

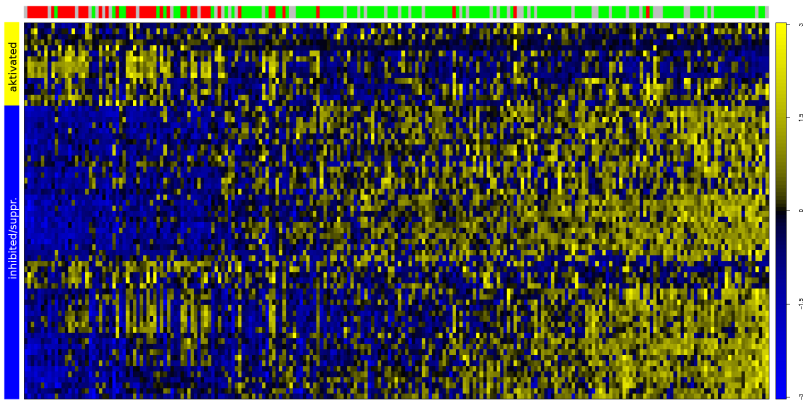


Figure 3.14: The heatmap shows the expression of LEF1 targets (rows) across 219 lymphoma samples (columns). Target genes are grouped into LEF1 activated (blue bar above) and LEF1 inhibited (yellow bar above) genes. Lymphoma samples are sorted by the LEF1 index. The color bar on top of the heatmap marks mBL in red, non-mBL in green and intermediate lymphoma in grey. Colors of the heatmap indicate strength of expression: Yellow high and blue low expression.

3.3.7 Discussion

The distinction of subgroups of mature aggressive B cell lymphomas and the understanding of the molecular events leading to different subgroups is an important prerequisite to further improve personalized treatment and to detoxify current chemotherapy strategies. LEF1 could be one new candidate for treatment in BL. In contrast to DLBCLs, nuclear LEF1 expression is detectable in the majority (63%) of analysed BL. Since normal GC B cells are negative for LEF1 the expression of LEF1 in BLs is aberrant. By combining LEF1 target genes identified in BL celllines with large scale gene expression data from lymphoma patients, it was possible to establish a LEF1 expression index that reflects LEF1 activity. These observations further support the importance of molecular stratifications of lymphomas.
LEF1 expression is involved in cell proliferation and cell cycle control of BL celllines. The joint expression of LEF1 target genes as a surrogate for the regulatory activity of LEF1 in primary BLs supports its role in cell proliferation and cell cycle. GSEA revealed an enrichment of cell cycle regulators affected by LEF1 knockdown in BL celllines. Thus, it is tempting to speculate that LEF1 is involved in a higher proliferative rate observed in BLs compared to DLBCLs. Thus LEF1 contributes to the biology of BLs but not DLBCLs [81, 101, 35].
Whether LEF1 acts as a cofactor for c-Myc affecting cell proliferation and cell cycle progression is not yet clear. Under our experimental conditions, LEF1 affects c-Myc expression only at low levels.

In addition, the LEF1 index can now be used to compare LEF1 activity in ALL or CLL patients. The aberrant expression of LEF1 in ALL or CLL is stronger compared to BLs and most probably depends on β-catenin [126, 97, 146, 178]. Interestingly the modern therapeutic concepts to treat BLs are closer to that of ALLs as that for DLBCLs [95]. It should be discussed if the application of LEF1 specific inhibitors should be integrated into respective protocols. However, so far small molecule inhibitors often lack precise specificity and those published so far still need approval [62, 125].
The comparison of LEF1 target genes with those from β-cateninS32A/Wnt3a intervened BL cells showed a significant target gene overlap (Figure 3.12). This suggests that canonical Wnt signals further enhance pre-existing transcriptional activities of LEF1 in BL cells.

Nevertheless, in BL and DLBCL tissue no β-catenin was observed in the nucleus suggesting β-catenin independent LEF1 activity [67, 136].

This is supported to some extend by immunoblot analysis that showed weak to no expression of β-catenin in BL and most DLBCL celllines (data not shown). Whether this weak β-catenin expression could be involved in aberrant LEF1 activity needs to be further evaluated.
Furthermore, it would be interesting to analyse other potential binding partners of LEF1 and the effect of Epstein-Barr virus in endemic BLs onto LEF1 and Wnt signaling [136, 34, 23, 89].
In conclusion, the findings demonstrate aberrant LEF1 expression in sporadic BL. Thus, the current study has potential clinical implications for patients with aggressive lymphoma by describing new LEF1 target genes, which now should be investigated further and evaluated as potentially therapeutic targets.

Appendix

Table A1: Sample-Name information

Overview of tumor samples used for microarray experiments. Tumors developed upon transplantation of p53-deficient postnatal neural stem/progenitor cells over-expressing different gene combinations.

Initial combinatorial perturbation experiments			
Gene combination	Array sample number	Transplantation	Transplanted cell number
HRAS + MYC	15	Secondary	50
HRAS + MYC	16	Secondary	50
HRAS + MYC	17	Primary	300000
HRAS + MYC	18	Primary	300000
HRAS + MYC	14	Primary	300000
HRAS + MYC	4	Primary	300000
HRAS + MYC	3	Primary	300000
MYC + Ezh2	1	Primary	300000
MYC + Ezh2	5	Secondary	50
MYC + Ezh2	8	Primary	300000
MYC + Ezh2	19	Primary	300000
MYC + Ezh2	20	Secondary	50
MYC + Bmi1	6	Secondary	5000
MYC + Bmi1	21	Secondary	5000
MYC + Bmi1	7	Secondary	5000
HRAS + Bmi1	22	Primary	300000
HRAS + Bmi1	9	Primary	300000
HRAS	23	Primary	300000
HRAS	2	Secondary	50
HRAS	10	Primary	300000
HRAS	24	Secondary	50
MYC	11	Primary	300000
MYC	12	Primary	300000
MYC	25	Primary	300000
MYC	13	Secondary	50
MYC	26	Tertiary	50
Consecutive perturbation experiments			
Gene combination	Array sample number	Transplantation	Transplanted cell number
HRAS (Glioma 2)	49	Primary	500000
Additional MYC (daughter 2a)	52	Secondary	5000
Additional MYC (daughter 2b)	53	Secondary	5000
MYC (PNET 1)	48	Primary	500000
Additional HRAS (daughter 1a)	50	Secondary	5000
Additional HRAS (daughter 1b)	51	Secondary	5000

Figure A1: $TGF\beta$ induced SMAD2 phosphorylation

SMAD2 phosphorylation after incubation with $TGF\beta$ 1 and 2 (10 ng/ml) for 0, 10, 30, and 60 min was analyzed by immunoblotting. Beta-actin was included as loading control. (Figure taken from publication by Beier et al [9].)

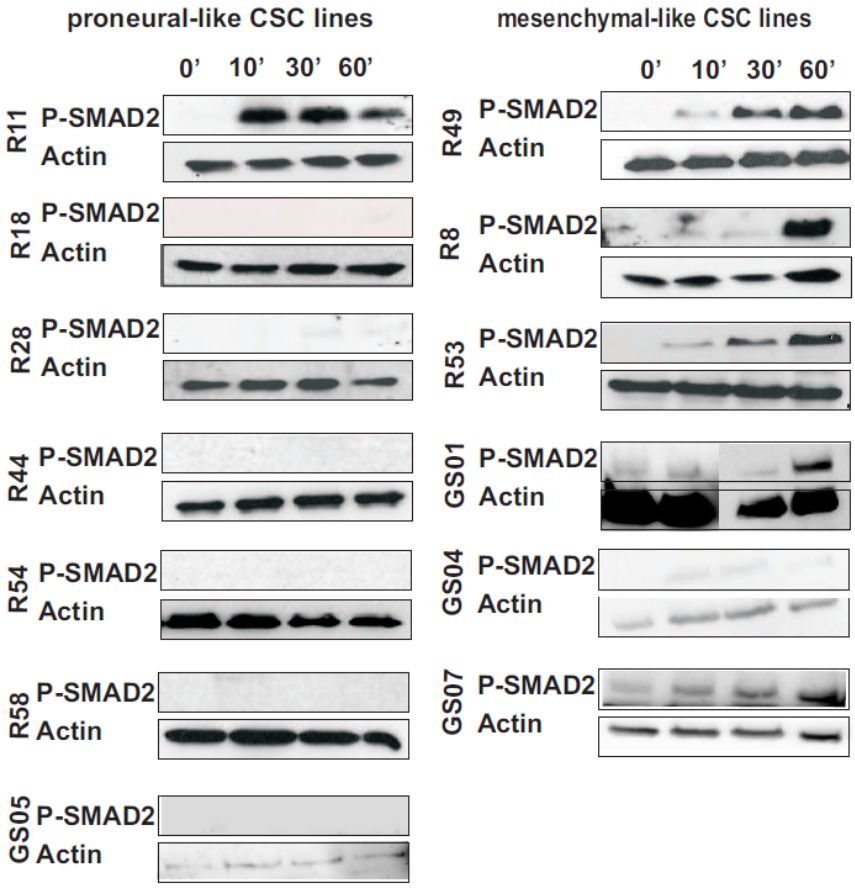


Figure A2: Responsiveness to TGFβ in proneural- and mesenchymal-like CSC lines

(a,c) *TGFβ* 1 and 2 expression in GBM (n=19). The amount of TGFβ expressing cells was scored as described in the Material and Methods section. There was no significant difference of TGFβ expression between proneural and mesenchymal GBM. (b) Relative TGFβR2 mRNA expression normalized to 18S RNA is given in CSC lines with and without SMAD phosphorylation. (d) Clonogenicity of the CSC line R11 after treatment with TGF-beta in the medium supplemented or not with LIF. The number of clones is indicated relative to the untreated control. (Figures taken from publication by Beier et al [9].)

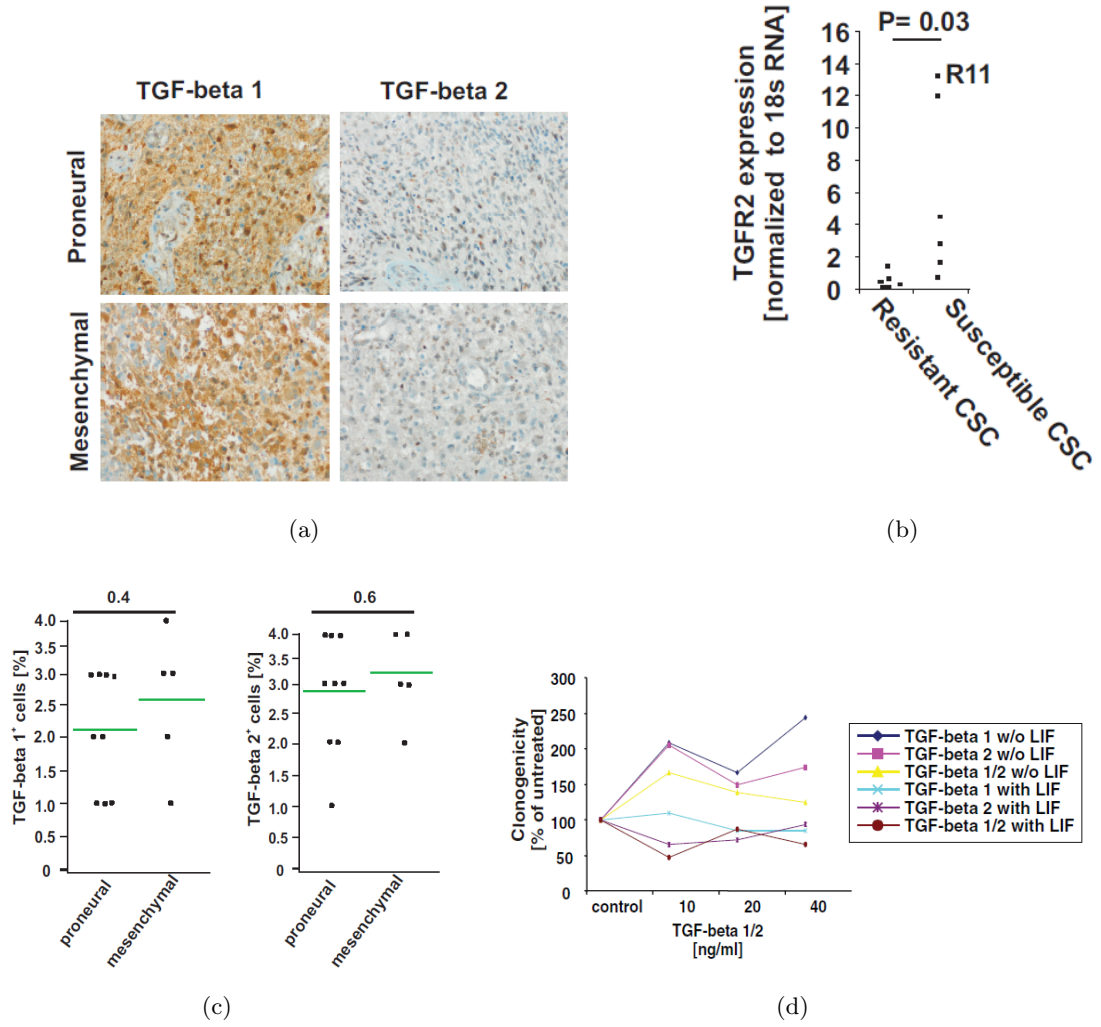


Figure 3.15

Tables A2: Results GO-Analyses

In order to identify pathway activities differing among aggressive NHL, human transformed GC B cells were stimulated with five distinct stimuli: αIgM (BCR activation), LPS, CD40L, IL21 and BAFF.
Using linear models implemented in the Bioconductor [64] package limma [176] I searched for genes showing the highest differential expression in response to each stimulus and a significant regulation (adjusted p value ≤ 0.05). The experimental batches were considered in the linear model.
A gene ontology (GO) analysis of each gene lists affected by the different *in vitro* interventions was performed using the different GO gene lists of molecular functions, biological processes and cellular components. Results are shown in the tables on the following pages.

BCR_Up								
Molecular Function			Biological Process			Cellular Component		
GO ID	GO Term	p-value	GO ID	GO Term	p-value	GO ID	GO Term	p-value
GO:0003824	catalytic activity	0,0000	GO:0006974	response to DNA damage stimulus	0,0000	GO:0005622	intracellular	0,0000
GO:0016799	hydrolase activity, hydrolyzing N-glycosyl compounds	0,0000	GO:0006259	DNA metabolic process	0,0000	GO:0044424	intracellular part	0,0000
GO:0016772	transferase activity, transferring phosphorus-containing groups	0,0000	GO:0008152	metabolic process	0,0000	GO:0043231	intracellular membrane-bounded organelle	0,0000
GO:0016740	transferase activity	0,0000	GO:0034984	cellular response to DNA damage stimulus	0,0000	GO:0043227	membrane-bounded organelle	0,0000
GO:0003677	DNA binding	0,0000	GO:0044238	primary metabolic process	0,0000	GO:0044464	cell part	0,0000
GO:0003676	nucleic acid binding	0,0000	GO:0006281	DNA repair	0,0000	GO:0005623	cell	0,0000
GO:0003684	damaged DNA binding	0,0000	GO:0044237	cellular metabolic process	0,0000	GO:0043229	intracellular organelle	0,0000
GO:0019104	DNA N-glycosylase activity	0,0000	GO:0043170	macromolecule metabolic process	0,0000	GO:0043226	organelle	0,0000
GO:0016798	hydrolase activity, acting on glycosyl bonds	0,0001	GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0,0000	GO:0005634	nucleus	0,0000
GO:0000287	magnesium ion binding	0,0001	GO:0044260	cellular macromolecule metabolic process	0,0000	GO:0005737	cytoplasm	0,0000
GO:0043566	structure-specific DNA binding	0,0002	GO:0043283	biopolymer metabolic process	0,0000	GO:0005654	nucleoplasm	0,0000
GO:0016783	sulfurtransferase activity	0,0002	GO:0034960	cellular biopolymer metabolic process	0,0000	GO:0044428	nuclear part	0,0000
GO:0046914	transition metal ion binding	0,0002	GO:0033554	cellular response to stress	0,0000	GO:0070013	intracellular organelle lumen	0,0000
GO:0016301	kinase activity	0,0003	GO:0007049	cell cycle	0,0000	GO:0031981	nuclear lumen	0,0000
GO:0005488	binding	0,0003	GO:0006261	DNA-dependent DNA replication	0,0000	GO:0043233	organelle lumen	0,0000
GO:0016773	phosphotransferase activity, alcohol group as acceptor	0,0004	GO:0022403	cell cycle phase	0,0000	GO:0031974	membrane-enclosed lumen	0,0000
GO:0005515	protein binding	0,0004	GO:0051301	cell division	0,0000	GO:0005815	microtubule organizing center	0,0000
GO:0016788	hydrolase activity, acting on ester bonds	0,0005	GO:0051716	cellular response to stimulus	0,0000	GO:0005813	centrosome	0,0000
GO:0008270	zinc ion binding	0,0005	GO:0006260	DNA replication	0,0000	GO:0044446	intracellular organelle part	0,0000
GO:0016779	nucleotidyltransferase activity	0,0006	GO:0022402	cell cycle process	0,0000	GO:0044422	organelle part	0,0000
GO:0003690	double-stranded DNA binding	0,0008	GO:0034961	cellular biopolymer biosynthetic process	0,0000	GO:0005819	spindle	0,0000
GO:0030145	manganese ion binding	0,0008	GO:0043284	biopolymer biosynthetic process	0,0000	GO:0015630	microtubule cytoskeleton	0,0000
GO:0004536	deoxyribonuclease activity	0,0008	GO:0000279	M phase	0,0000	GO:0000775	chromosome, centromeric region	0,0000
GO:0005072	transforming growth factor beta receptor, cytoplasmic mediator activity	0,0010	GO:0009059	macromolecule biosynthetic process	0,0000	GO:0044444	cytoplasmic part	0,0000
GO:0005543	phospholipid binding	0,0010	GO:0034645	cellular macromolecule biosynthetic process	0,0000	GO:0044451	nucleoplasm part	0,0000
GO:0003697	single-stranded DNA binding	0,0020	GO:0044249	cellular biosynthetic process	0,0000	GO:0005694	chromosome	0,0000
GO:0005071	transmembrane receptor protein serine/threonine kinase signaling protein activity	0,0020	GO:0009058	biosynthetic process	0,0000	GO:0044427	chromosomal part	0,0000
GO:0005524	ATP binding	0,0020	GO:0000278	mitotic cell cycle	0,0000	GO:0000777	condensed chromosome kinetochore	0,0000
GO:0030554	adenyl nucleotide binding	0,0030	GO:0010468	regulation of gene expression	0,0000	GO:0000776	kinetochore	0,0001
GO:0004520	endodeoxyribonuclease activity	0,0040	GO:0060255	regulation of macromolecule metabolic process	0,0000	GO:0000779	condensed chromosome, centromeric region	0,0001
GO:0003712	transcription cofactor activity	0,0040	GO:0000087	M phase of mitotic cell cycle	0,0000	GO:0005876	spindle microtubule	0,0001
GO:0032559	adenyl ribonucleotide binding	0,0040	GO:0006996	organelle organization	0,0001	GO:0005828	kinetochore microtubule	0,0002
GO:0001727	lipid kinase activity	0,0050	GO:0000280	nuclear division	0,0001	GO:0044450	microtubule organizing center part	0,0002
GO:0004518	nuclease activity	0,0050	GO:0007067	mitosis	0,0001	GO:0005657	replication fork	0,0004
GO:0017076	purine nucleotide binding	0,0060	GO:0009057	macromolecule catabolic process	0,0001	GO:0016604	nuclear body	0,0008
GO:0000217	DNA secondary structure binding	0,0060	GO:0048285	organelle fission	0,0001	GO:0043228	non-membrane-bounded organelle	0,0010

BCR_Down

Molecular Function			Biological Process			Cellular Component		
GO ID	GO Term	p-value	GO ID	GO Term	p-value	GO ID	GO Term	p-value
GO:0005515	protein binding	0,0000	GO:0016070	RNA metabolic process	0,0000	GO:0005622	intracellular	0,0000
GO:0003676	nucleic acid binding	0,0000	GO:0010467	gene expression	0,0000	GO:0044424	intracellular part	0,0000
GO:0030528	transcription regulator activity	0,0000	GO:0006915	apoptosis	0,0000	GO:0043231	intracellular membrane-bounded organelle	0,0000
GO:0008134	transcription factor binding	0,0000	GO:0012501	programmed cell death	0,0000	GO:0043227	membrane-bounded organelle	0,0000
GO:0016564	transcription repressor activity	0,0000	GO:0034960	cellular biopolymer metabolic process	0,0000	GO:0044464	cell part	0,0000
GO:0003712	transcription cofactor activity	0,0000	GO:0008219	cell death	0,0000	GO:0005623	cell	0,0000
GO:0005525	GTP binding	0,0000	GO:0016265	death	0,0000	GO:0043229	intracellular organelle	0,0000
GO:0032395	MHC class II receptor activity	0,0000	GO:0044260	cellular macromolecule metabolic process	0,0000	GO:0043226	organelle	0,0000
GO:0016563	transcription activator activity	0,0000	GO:0043283	biopolymer metabolic process	0,0000	GO:0005634	nucleus	0,0000
GO:0032561	guanyl ribonucleotide binding	0,0001	GO:0034645	cellular macromolecule biosynthetic process	0,0000	GO:0042611	MHC protein complex	0,0000
GO:0019001	guanyl nucleotide binding	0,0001	GO:0010468	regulation of gene expression	0,0000	GO:0042470	melanosome	0,0000
GO:0003743	translation initiation factor activity	0,0001	GO:0009059	macromolecule biosynthetic process	0,0000	GO:0048770	pigment granule	0,0000
GO:0015175	neutral amino acid transmembrane transporter activity	0,0001	GO:0043170	macromolecule metabolic process	0,0000	GO:0005737	cytoplasm	0,0000
GO:0003723	RNA binding	0,0001	GO:0010556	regulation of macromolecule biosynthetic process	0,0000	GO:0042613	MHC class II protein complex	0,0000
GO:0000166	nucleotide binding	0,0001	GO:0044249	cellular biosynthetic process	0,0000	GO:0012505	endomembrane system	0,0000
GO:0008135	translation factor activity, nucleic acid binding	0,0001	GO:0009058	biosynthetic process	0,0000	GO:0005783	endoplasmic reticulum	0,0000
GO:0003714	transcription corepressor activity	0,0002	GO:0060255	regulation of macromolecule metabolic process	0,0000	GO:0031090	organelle membrane	0,0000
GO:0005488	binding	0,0002	GO:0019882	antigen processing and presentation	0,0000	GO:0044431	Golgi apparatus part	0,0001
GO:0032553	ribonucleotide binding	0,0003	GO:0006351	transcription, DNA-dependent	0,0000	GO:0005635	nuclear envelope	0,0001
GO:0032555	purine ribonucleotide binding	0,0003	GO:0051252	regulation of RNA metabolic process	0,0000	GO:0005793	ER-Golgi intermediate compartment	0,0001
GO:0003924	GTPase activity	0,0003	GO:0006355	regulation of transcription, DNA-dependent	0,0000	GO:0000139	Golgi membrane	0,0001
GO:0003700	transcription factor activity	0,0004	GO:0032774	RNA biosynthetic process	0,0000	GO:0044444	cytoplasmic part	0,0001
GO:0003677	DNA binding	0,0005	GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0,0000	GO:0005788	endoplasmic reticulum lumen	0,0002
GO:0045182	translation regulator activity	0,0006	GO:0034961	cellular biopolymer biosynthetic process	0,0000	GO:0005773	vacuole	0,0002
GO:0017076	purine nucleotide binding	0,0007	GO:0043284	biopolymer biosynthetic process	0,0000	GO:0005774	vacuolar membrane	0,0003
GO:0019955	cytokine binding	0,0010	GO:0031326	regulation of cellular biosynthetic process	0,0000	GO:0005768	endosome	0,0003
GO:0003756	protein disulfide isomerase activity	0,0010	GO:0009889	regulation of biosynthetic process	0,0000	GO:0044432	endoplasmic reticulum part	0,0004
GO:0016864	intramolecular oxidoreductase activity, transposing S-S bonds	0,0010	GO:0045449	regulation of transcription	0,0000	GO:0042175	nuclear envelope-endoplasmic reticulum network	0,0004
GO:0004911	interleukin-2 receptor activity	0,0010	GO:0006350	transcription	0,0000	GO:0044428	nuclear part	0,0004
GO:0019976	interleukin-2 binding	0,0010	GO:0031323	regulation of cellular metabolic process	0,0000	GO:0005765	lysosomal membrane	0,0004
GO:0005114	type II transforming growth factor beta receptor binding	0,0020	GO:0019222	regulation of metabolic process	0,0000	GO:0005789	endoplasmic reticulum membrane	0,0005
GO:0016791	phosphatase activity	0,0020	GO:0044238	primary metabolic process	0,0000	GO:0005770	late endosome	0,0005
GO:0016862	intramolecular oxidoreductase activity, interconverting keto- and enol-groups	0,0020	GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	0,0000	GO:0044437	vacuolar part	0,0006
GO:0017017	MAP kinase tyrosine/serine/threonine phosphatase activity	0,0020	GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0,0000	GO:0000323	lytic vacuole	0,0006
GO:0033549	MAP kinase phosphatase activity	0,0020	GO:0050794	regulation of cellular process	0,0000	GO:0005764	lysosome	0,0006
GO:0046983	protein dimerization activity	0,0020	GO:0006916	anti-apoptosis	0,0000	GO:0042612	MHC class I protein complex	0,0008

BAFF_Up

Molecular Function

GO ID	GO Term	p-value
GO:0016706	oxidoreductase activity, acting on paired donors	0,0009
GO:0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	0,005
GO:0051213	dioxygenase activity	0,005
GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	0,005
GO:0016564	transcription repressor activity	0,006

Biological Process

GO ID	GO Term	p-value
GO:0010629	negative regulation of gene expression	0,00003
GO:0008634	negative regulation of survival gene product expression	0,00007
GO:0048523	negative regulation of cellular process	0,0002
GO:0010605	negative regulation of macromolecule metabolic process	0,0002
GO:0009892	negative regulation of metabolic process	0,0003
GO:0048519	negative regulation of biological process	0,0003
GO:0045884	regulation of survival gene product expression	0,0004
GO:0007405	neuroblast proliferation	0,0007
GO:0050768	negative regulation of neurogenesis	0,0007
GO:0010721	negative regulation of cell development	0,0007
GO:0030262	apoptotic nuclear changes	0,0007
GO:0045665	negative regulation of neuron differentiation	0,0009
GO:0043433	negative regulation of transcription factor activity	0,001
GO:0051607	defense response to virus	0,001
GO:0043065	positive regulation of apoptosis	0,001
GO:0043068	positive regulation of programmed cell death	0,001
GO:0043392	negative regulation of DNA binding	0,001
GO:0051100	negative regulation of binding	0,002
GO:0050794	regulation of cellular process	0,002
GO:0006997	nucleus organization	0,002
GO:0002252	immune effector process	0,002
GO:0016481	negative regulation of transcription	0,003
GO:0050789	regulation of biological process	0,003
GO:0030855	epithelial cell differentiation	0,004
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	0,004
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0,004
GO:0010558	negative regulation of macromolecule biosynthetic process	0,005
GO:0006800	oxygen and reactive oxygen species metabolic process	0,005
GO:0031327	negative regulation of cellular biosynthetic process	0,006
GO:0009890	negative regulation of biosynthetic process	0,006
GO:0051707	response to other organism	0,006
GO:0051090	regulation of transcription factor activity	0,006
GO:0051093	negative regulation of developmental process	0,006
GO:0006917	induction of apoptosis	0,006
GO:0012502	induction of programmed cell death	0,007
GO:0051094	positive regulation of developmental process	0,007

BAFF_Down

Molecular Function

GO ID	GO Term	p-value
GO:0032393	MHC class I receptor activity	0,0000
GO:0032395	MHC class II receptor activity	0,0000
GO:0004888	transmembrane receptor activity	0,0020

Biological Process

GO ID	GO Term	p-value
GO:0019882	antigen processing and presentation	0,0000
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	0,0000
GO:0048002	antigen processing and presentation of peptide antigen	0,0000
GO:0006955	immune response	0,0000
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	0,0000
GO:0002376	immune system process	0,0000
GO:0050896	response to stimulus	0,0030

Cellular Component

GO ID	GO Term	p-value
GO:0042611	MHC protein complex	0,0000
GO:0042612	MHC class I protein complex	0,0000
GO:0042613	MHC class II protein complex	0,0000
GO:0044459	plasma membrane part	0,0006
GO:0032991	macromolecular complex	0,0006
GO:0043234	protein complex	0,0010

CD40_Up

Molecular Function

GO ID	GO Term	p-value
GO:0015269	calcium-activated potassium channel activity	0,001
GO:0005515	protein binding	0,002
GO:0022839	ion gated channel activity	0,002
GO:0003873	6-phosphofructo-2-kinase activity	0,002
GO:0017089	glycolipid transporter activity	0,002
GO:0051861	glycolipid binding	0,002
GO:0005227	calcium activated cation channel activity	0,003
GO:0005516	calmodulin binding	0,004
GO:0005521	lamin binding	0,004
GO:0030371	translation repressor activity	0,005
GO:0030234	enzyme regulator activity	0,006
GO:0008047	enzyme activator activity	0,006
GO:0005543	phospholipid binding	0,007
GO:0004331	fructose-2,6-bisphosphate 2-phosphatase activity	0,007
GO:0008443	phosphofructokinase activity	0,007
GO:0008092	cytoskeletal protein binding	0,009
GO:0019864	IgG binding	0,01

Cellular Component

GO ID	GO Term	p-value
GO:0044464	cell part	0,003
GO:0005623	cell	0,003

Biological Process

GO ID	GO Term	p-value
GO:0002376	immune system process	0,00001
GO:0030183	B cell differentiation	0,00002
GO:0002521	leukocyte differentiation	0,00003
GO:0042113	B cell activation	0,00003
GO:0007165	signal transduction	0,00005
GO:0030098	lymphocyte differentiation	0,00007
GO:0006955	immune response	0,00007
GO:0030888	regulation of B cell proliferation	0,00008
GO:0046649	lymphocyte activation	0,0001
GO:0007154	cell communication	0,0001
GO:0050794	regulation of cellular process	0,0001
GO:0051251	positive regulation of lymphocyte activation	0,0002
GO:0045321	leukocyte activation	0,0002
GO:0008634	negative regulation of survival gene product expression	0,0002
GO:0050789	regulation of biological process	0,0003
GO:0030097	hemopoiesis	0,0003
GO:0048523	negative regulation of cellular process	0,0004
GO:0002696	positive regulation of leukocyte activation	0,0004
GO:0050864	regulation of B cell activation	0,0005
GO:0010629	negative regulation of gene expression	0,0005
GO:0050867	positive regulation of cell activation	0,0005
GO:0050871	positive regulation of B cell activation	0,0005
GO:0048519	negative regulation of biological process	0,0005
GO:0030890	positive regulation of B cell proliferation	0,0005
GO:0042100	B cell proliferation	0,0006
GO:0048534	hemopoietic or lymphoid organ development	0,0007
GO:0001775	cell activation	0,0008
GO:0045628	regulation of T-helper 2 cell differentiation	0,001
GO:0002520	immune system development	0,001
GO:0065007	biological regulation	0,002
GO:0009607	response to biotic stimulus	0,002
GO:0009892	negative regulation of metabolic process	0,002
GO:0002252	immune effector process	0,002
GO:0002902	regulation of B cell apoptosis	0,002
GO:0045064	T-helper 2 cell differentiation	0,002
GO:0055091	phospholipid homeostasis	0,002

CD40_Down

Molecular Function			Biological Process			Cellular Component		
GO ID	GO Term	p-value	GO ID	GO Term	p-value	GO ID	GO Term	p-value
GO:0032395	MHC class II receptor activity	0,0000	GO:0019882	antigen processing and presentation	0,0000	GO:0042611	MHC protein complex	0,0000
GO:0032393	MHC class I receptor activity	0,0000	GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	0,0000	GO:0042613	MHC class II protein complex	0,0000
GO:0005515	protein binding	0,0000	GO:0048002	antigen processing and presentation of peptide antigen	0,0000	GO:0005737	cytoplasm	0,0000
GO:0004718	Janus kinase activity	0,0001	GO:0006955	immune response	0,0000	GO:0042612	MHC class I protein complex	0,0000
GO:0046978	TAP1 binding	0,0008	GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	0,0000	GO:0042824	MHC class I peptide loading complex	0,0000
GO:0015197	peptide transporter activity	0,0020	GO:0002376	immune system process	0,0000	GO:0042825	TAP complex	0,0002
GO:0015433	peptide antigen-transporting ATPase activity	0,0020	GO:0002250	adaptive immune response	0,0001	GO:0044464	cell part	0,0005
GO:0015440	peptide-transporting ATPase activity	0,0020	GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	0,0003	GO:0005623	cell	0,0005
GO:0046977	TAP binding	0,0020	GO:0002483	antigen processing and presentation of endogenous peptide antigen	0,0004	GO:0005788	endoplasmic reticulum lumen	0,0006
GO:0046979	TAP2 binding	0,0020	GO:0019885	antigen processing and presentation of endogenous peptide antigen via MHC class I	0,0004	GO:0044424	intracellular part	0,0007
GO:0004716	receptor signaling protein tyrosine kinase activity	0,0040	GO:0006694	steroid biosynthetic process	0,0007	GO:0005622	intracellular	0,0010
GO:0030695	GTPase regulator activity	0,0040	GO:0016126	sterol biosynthetic process	0,0007	GO:0005886	plasma membrane	0,0020
GO:0042288	MHC class I protein binding	0,0050	GO:0019883	antigen processing and presentation of endogenous antigen	0,0010	GO:0042405	nuclear inclusion body	0,0020
GO:0005085	guanyl-nucleotide exchange factor activity	0,0070	GO:0009966	regulation of signal transduction	0,0010	GO:0005901	caveola	0,0020
GO:0005131	growth hormone receptor binding	0,0070	GO:0006695	cholesterol biosynthetic process	0,0010	GO:0043229	intracellular organelle	0,0040
GO:0015923	mannosidase activity	0,0080	GO:0016125	sterol metabolic process	0,0020	GO:0043226	organelle	0,0040
			GO:0007242	intracellular signaling cascade	0,0020	GO:0001725	stress fiber	0,0060
			GO:0051056	regulation of small GTPase mediated signal transduction	0,0020	GO:0032432	actin filament bundle	0,0060
			GO:0051249	regulation of lymphocyte activation	0,0020	GO:0000323	lytic vacuole	0,0070
			GO:0051168	nuclear export	0,0020	GO:0005764	lysosome	0,0070
			GO:0007264	small GTPase mediated signal transduction	0,0030	GO:0044459	plasma membrane part	0,0070
			GO:0010627	regulation of protein kinase cascade	0,0030	GO:0005783	endoplasmic reticulum	0,0070
			GO:0008203	cholesterol metabolic process	0,0030	GO:0005773	vacuole	0,0070
			GO:0002286	T cell activation during immune response	0,0040	GO:0012505	endomembrane system	0,0080
			GO:0046649	lymphocyte activation	0,0040	GO:0043231	intracellular membrane-bounded organelle	0,0080
			GO:0006916	anti-apoptosis	0,0050	GO:0043227	membrane-bounded organelle	0,0090
			GO:0007042	lysosomal lumen acidification	0,0050	GO:0031967	organelle envelope	0,0090
			GO:0031119	tRNA pseudouridine synthesis	0,0050	GO:0005643	nuclear pore	0,0090
			GO:0042110	T cell activation	0,0050	GO:0031975	envelope	0,0090
			GO:0002694	regulation of leukocyte activation	0,0050			
			GO:0006400	tRNA modification	0,0050			
			GO:0016311	dephosphorylation	0,0060			
			GO:0006664	glycolipid metabolic process	0,0060			
			GO:0002443	leukocyte mediated immunity	0,0070			
			GO:0010646	regulation of cell communication	0,0070			
			GO:0050865	regulation of cell activation	0,0070			

IL21_Up

Molecular Function

GO ID	GO Term	p-value
GO:0003714	transcription corepressor activity	0,0000
GO:0016564	transcription repressor activity	0,0000
GO:0003684	damaged DNA binding	0,0007
GO:0003696	satellite DNA binding	0,0020
GO:0043028	caspase regulator activity	0,0040
GO:0003712	transcription cofactor activity	0,0050
GO:0032451	demethylase activity	0,0050
GO:0001530	lipopolysaccharide binding	0,0090
GO:0005095	GTPase inhibitor activity	0,0090
GO:0005515	protein binding	0,0090
GO:0030165	PDZ domain binding	0,0100

Cellular Component

GO ID	GO Term	p-value
GO:0044464	cell part	0,0020
GO:0005623	cell	0,0020
GO:0005634	nucleus	0,0080

Biological Process

GO ID	GO Term	p-value
GO:0009968	negative regulation of signal transduction	0,0000
GO:0010648	negative regulation of cell communication	0,0000
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	0,0001
GO:0048523	negative regulation of cellular process	0,0001
GO:0048519	negative regulation of biological process	0,0002
GO:0006974	response to DNA damage stimulus	0,0002
GO:0010646	regulation of cell communication	0,0005
GO:0009966	regulation of signal transduction	0,0005
GO:0010629	negative regulation of gene expression	0,0006
GO:0006281	DNA repair	0,0006
GO:0034984	cellular response to DNA damage stimulus	0,0006
GO:0016481	negative regulation of transcription	0,0009
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide	0,0010
GO:0033554	cellular response to stress	0,0010
GO:0044249	cellular biosynthetic process	0,0010
GO:0045892	negative regulation of transcription	0,0020
GO:0044260	cellular macromolecule metabolic process	0,0020
GO:0010558	negative regulation of macromolecule biosynthetic process	0,0020
GO:0006950	response to stress	0,0020
GO:0034960	cellular biopolymer metabolic process	0,0020
GO:0051253	negative regulation of RNA metabolic process	0,0020
GO:0043170	macromolecule metabolic process	0,0020
GO:0051716	cellular response to stimulus	0,0020
GO:0031327	negative regulation of cellular biosynthetic process	0,0020
GO:0043283	biopolymer metabolic process	0,0020
GO:0009890	negative regulation of biosynthetic process	0,0020
GO:0009058	biosynthetic process	0,0020
GO:0050794	regulation of cellular process	0,0020
GO:0044237	cellular metabolic process	0,0030
GO:0042110	T cell activation	0,0030
GO:0008152	metabolic process	0,0030
GO:0006366	transcription from RNA polymerase II promoter	0,0030
GO:0006357	regulation of transcription from RNA polymerase II promoter	0,0030
GO:0006350	transcription	0,0040
GO:0044238	primary metabolic process	0,0040
GO:0010605	negative regulation of macromolecule metabolic process	0,0040

IL21_Down

Molecular Function

GO ID	GO Term	p-value
GO:0030528	transcription regulator activity	0,0000
GO:0003700	transcription factor activity	0,0000
GO:0003950	NAD+ ADP-ribosyltransferase activity	0,0001
GO:0003676	nucleic acid binding	0,0001
GO:0016763	transferase activity, transferring pentosyl groups	0,0002
GO:0005487	nucleocytoplasmic transporter activity	0,0003
GO:0008134	transcription factor binding	0,0004
GO:0016564	transcription repressor activity	0,0004
GO:0003677	DNA binding	0,0004
GO:0017017	MAP kinase tyrosine/serine/threonine phosphatase activity	0,0007
GO:0033549	MAP kinase phosphatase activity	0,0007
GO:0019955	cytokine binding	0,0007
GO:0005062	hematopoietin/interferon-class cytokine receptor signal transducer activity	0,0020
GO:0032393	MHC class I receptor activity	0,0020
GO:0046914	transition metal ion binding	0,0030
GO:0005515	protein binding	0,0030
GO:0008270	zinc ion binding	0,0040
GO:0005488	binding	0,0040
GO:0003712	transcription cofactor activity	0,0050
GO:0016757	transferase activity, transferring glycosyl groups	0,0060
GO:0016563	transcription activator activity	0,0100

Cellular Component

GO ID	GO Term	p-value
GO:0005622	intracellular	0,0000
GO:0044424	intracellular part	0,0000
GO:0005634	nucleus	0,0000
GO:0043231	intracellular membrane-bounded organelle	0,0000
GO:0043227	membrane-bounded organelle	0,0000
GO:0033256	I-kappaB/NF-kappaB complex	0,0000
GO:0043229	intracellular organelle	0,0001
GO:0043226	organelle	0,0001
GO:0033257	Bcl3/NF-kappaB2 complex	0,0003
GO:0009897	external side of plasma membrane	0,0003
GO:0031965	nuclear membrane	0,0030
GO:0042612	MHC class I protein complex	0,0050
GO:0044464	cell part	0,0060
GO:0005737	cytoplasm	0,0060
GO:0005623	cell	0,0060

Biological Process

GO ID	GO Term	p-value
GO:0002376	immune system process	0,0000
GO:0006955	immune response	0,0000
GO:0009615	response to virus	0,0000
GO:0006915	apoptosis	0,0000
GO:0051704	multi-organism process	0,0000
GO:0012501	programmed cell death	0,0000
GO:0008219	cell death	0,0000
GO:0001775	cell activation	0,0000
GO:0016265	death	0,0000
GO:0046649	lymphocyte activation	0,0000
GO:0045321	leukocyte activation	0,0000
GO:0007249	I-kappaB kinase/NF-kappaB cascade	0,0000
GO:0009607	response to biotic stimulus	0,0000
GO:0050794	regulation of cellular process	0,0000
GO:0050789	regulation of biological process	0,0000
GO:0007243	protein kinase cascade	0,0000
GO:0051707	response to other organism	0,0000
GO:0001773	myeloid dendritic cell activation	0,0000
GO:0010556	regulation of macromolecule biosynthetic process	0,0000
GO:0048519	negative regulation of biological process	0,0000
GO:0042110	T cell activation	0,0000
GO:0060255	regulation of macromolecule metabolic process	0,0000
GO:0048534	hemopoietic or lymphoid organ development	0,0000
GO:0010468	regulation of gene expression	0,0000
GO:0002263	cell activation during immune response	0,0000
GO:0002366	leukocyte activation during immune response	0,0000
GO:0008632	apoptotic program	0,0000
GO:0002520	immune system development	0,0000
GO:0065007	biological regulation	0,0000
GO:0031326	regulation of cellular biosynthetic process	0,0000
GO:0048523	negative regulation of cellular process	0,0000
GO:0045449	regulation of transcription	0,0000
GO:0009889	regulation of biosynthetic process	0,0000
GO:0046632	alpha-beta T cell differentiation	0,0000
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0,0000
GO:0042981	regulation of apoptosis	0,0000

IL21_Down

Molecular Function

GO ID	GO Term	p-value
GO:0030528	transcription regulator activity	0,0000
GO:0003700	transcription factor activity	0,0000
GO:0003950	NAD+ ADP-ribosyltransferase activity	0,0001
GO:0003676	nucleic acid binding	0,0001
GO:0016763	transferase activity, transferring pentosyl groups	0,0002
GO:0005487	nucleocytoplasmic transporter activity	0,0003
GO:0008134	transcription factor binding	0,0004
GO:0016564	transcription repressor activity	0,0004
GO:0003677	DNA binding	0,0004
GO:0017017	MAP kinase tyrosine/serine/threonine phosphatase activity	0,0007
GO:0033549	MAP kinase phosphatase activity	0,0007
GO:0019955	cytokine binding	0,0007
GO:0005062	hematopoietin/interferon-class (D200-domain) cytokine receptor signal transducer activity	0,0020
GO:0032393	MHC class I receptor activity	0,0020
GO:0046914	transition metal ion binding	0,0030
GO:0005515	protein binding	0,0030
GO:0008270	zinc ion binding	0,0040
GO:0005488	binding	0,0040
GO:0003712	transcription cofactor activity	0,0050
GO:0016757	transferase activity, transferring glycosyl groups	0,0060
GO:0016563	transcription activator activity	0,0100

Biological Process

GO ID	GO Term	p-value
GO:0002376	immune system process	0,0000
GO:0006955	immune response	0,0000
GO:0009615	response to virus	0,0000
GO:0006915	apoptosis	0,0000
GO:0051704	multi-organism process	0,0000
GO:0012501	programmed cell death	0,0000
GO:0008219	cell death	0,0000
GO:0001775	cell activation	0,0000
GO:0016265	death	0,0000
GO:0046649	lymphocyte activation	0,0000
GO:0045321	leukocyte activation	0,0000
GO:0007249	I-kappaB kinase/NF-kappaB cascade	0,0000
GO:0009607	response to biotic stimulus	0,0000
GO:0050794	regulation of cellular process	0,0000
GO:0050789	regulation of biological process	0,0000
GO:0007243	protein kinase cascade	0,0000
GO:0051707	response to other organism	0,0000
GO:0001773	myeloid dendritic cell activation	0,0000
GO:0010556	regulation of macromolecule biosynthetic process	0,0000
GO:0048519	negative regulation of biological process	0,0000
GO:0042110	T cell activation	0,0000
GO:0060255	regulation of macromolecule metabolic process	0,0000
GO:0048534	hemopoietic or lymphoid organ development	0,0000
GO:0010468	regulation of gene expression	0,0000
GO:0002263	cell activation during immune response	0,0000
GO:0002366	leukocyte activation during immune response	0,0000
GO:0008632	apoptotic program	0,0000
GO:0002520	immune system development	0,0000
GO:0065007	biological regulation	0,0000
GO:0031326	regulation of cellular biosynthetic process	0,0000
GO:0048523	negative regulation of cellular process	0,0000
GO:0045449	regulation of transcription	0,0000
GO:0009889	regulation of biosynthetic process	0,0000
GO:0046632	alpha-beta T cell differentiation	0,0000
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nu	0,0000
GO:0042981	regulation of apoptosis	0,0000

Cellular Component

GO ID	GO Term	p-value
GO:0005622	intracellular	0,0000
GO:0044424	intracellular part	0,0000
GO:0005634	nucleus	0,0000
GO:0043231	intracellular membrane-bounded organelle	0,0000
GO:0043227	membrane-bounded organelle	0,0000
GO:0033256	I-kappaB/NF-kappaB complex	0,0000
GO:0043229	intracellular organelle	0,0001
GO:0043226	organelle	0,0001
GO:0033257	Bcl3/NF-kappaB2 complex	0,0003
GO:0009897	external side of plasma membrane	0,0003
GO:0031965	nuclear membrane	0,0030
GO:0042612	MHC class I protein complex	0,0050
GO:0044464	cell part	0,0060
GO:0005737	cytoplasm	0,0060
GO:0005623	cell	0,0060

LPS_Up

Molecular Function

GO ID	GO Term	p-value
GO:0003873	6-phosphofructo-2-kinase activity	0,0003
GO:0005521	lamin binding	0,0004
GO:0016564	transcription repressor activity	0,0009
GO:0004331	fructose-2,6-bisphosphate 2-phosphatase activity	0,0009
GO:0008443	phosphofructokinase activity	0,0009
GO:0003714	transcription corepressor activity	0,0010
GO:0016303	1-phosphatidylinositol-3-kinase activity	0,0020
GO:0019203	carbohydrate phosphatase activity	0,0020
GO:0035004	phosphoinositide 3-kinase activity	0,0020
GO:0019200	carbohydrate kinase activity	0,0070

Cellular Component

GO ID	GO Term	p-value
GO:0044464	cell part	0,0060
GO:0005623	cell	0,0060

Biological Process

GO ID	GO Term	p-value
GO:0010629	negative regulation of gene expression	0,0001
GO:0048523	negative regulation of cellular process	0,0006
GO:0010605	negative regulation of macromolecule metabolic process	0,0006
GO:0008634	negative regulation of survival gene product expression	0,0009
GO:0009892	negative regulation of metabolic process	0,001
GO:0050896	response to stimulus	0,001
GO:0048519	negative regulation of biological process	0,001
GO:0001938	positive regulation of endothelial cell proliferation	0,002
GO:0009607	response to biotic stimulus	0,004
GO:0007165	signal transduction	0,004
GO:0002285	lymphocyte activation during immune response	0,005
GO:0010558	negative regulation of macromolecule biosynthetic process	0,005
GO:0043534	blood vessel endothelial cell migration	0,005
GO:0009615	response to virus	0,005
GO:0002263	cell activation during immune response	0,006
GO:0002366	leukocyte activation during immune response	0,006
GO:0045884	regulation of survival gene product expression	0,006
GO:0031327	negative regulation of cellular biosynthetic process	0,006
GO:0009890	negative regulation of biosynthetic process	0,006
GO:0002376	immune system process	0,006
GO:0016481	negative regulation of transcription	0,008
GO:0007154	cell communication	0,01
GO:0006955	immune response	0,01

LPS_Down

Molecular Function

GO ID	GO Term	p-value
GO:0032393	MHC class I receptor activity	0,0000
GO:0032395	MHC class II receptor activity	0,0000
GO:0004888	transmembrane receptor activity	0,0001
GO:0004872	receptor activity	0,0010
GO:0004871	signal transducer activity	0,0010
GO:0060089	molecular transducer activity	0,0010
GO:0016866	intramolecular transferase activity	0,0060

Cellular Component

GO ID	GO Term	p-value
GO:0042611	MHC protein complex	0,0000
GO:0042612	MHC class I protein complex	0,0000
GO:0042613	MHC class II protein complex	0,0000
GO:0005886	plasma membrane	0,0002
GO:0016020	membrane	0,0030
GO:0044459	plasma membrane part	0,0030

Biological Process

GO ID	GO Term	p-value
GO:0019882	antigen processing and presentation	0,0000
GO:0048002	antigen processing and presentation of peptide antigen	0,0000
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	0,0000
GO:0006955	immune response	0,0000
GO:0002376	immune system process	0,0000
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	0,0000
GO:0050896	response to stimulus	0,0008
GO:0042517	positive regulation of tyrosine phosphorylation of Stat3 protein	0,0010
GO:0002286	T cell activation during immune response	0,0010
GO:0006400	tRNA modification	0,0020
GO:0002285	lymphocyte activation during immune response	0,0020
GO:0042516	regulation of tyrosine phosphorylation of Stat3 protein	0,0020
GO:0042503	tyrosine phosphorylation of Stat3 protein	0,0020
GO:0002263	cell activation during immune response	0,0030
GO:0002366	leukocyte activation during immune response	0,0030
GO:0042531	positive regulation of tyrosine phosphorylation of STAT protein	0,0030
GO:0042110	T cell activation	0,0030
GO:0010740	positive regulation of protein kinase cascade	0,0040
GO:0046427	positive regulation of JAK-STAT cascade	0,0040
GO:0050776	regulation of immune response	0,0050
GO:0042509	regulation of tyrosine phosphorylation of STAT protein	0,0050
GO:0050863	regulation of T cell activation	0,0060
GO:0006694	steroid biosynthetic process	0,0070
GO:0007260	tyrosine phosphorylation of STAT protein	0,0070
GO:0008203	cholesterol metabolic process	0,0070
GO:0010627	regulation of protein kinase cascade	0,0080
GO:0002682	regulation of immune system process	0,0080
GO:0046425	regulation of JAK-STAT cascade	0,0080
GO:0006695	cholesterol biosynthetic process	0,0080
GO:0045580	regulation of T cell differentiation	0,0080
GO:0002250	adaptive immune response	0,0100
GO:0016125	sterol metabolic process	0,0100

Tables A3: Unique gene clusters

The method guided clustering [131] was used to identify genes that are specific for one of the five stimuli and additionally stratify the DLBCLs in the dataset of Hummel and colleagues [81]. As guiding data I used the stimulated BL2 cell line data, which was correlated to a binary vector that changed per guided clustering run. Within that binary vector the samples belonging to one of the stimuli were labeled with 1, and the other stimulations as well as the controls were labeled with 0, making sure that genes were only regulated by one of the five stimulations. To find gene sets that showed differences within the DLBCL samples and not only between the DLBCL samples and the mBL samples, the patient data was reduced to the Myc negative DLBCL cases. Thus, excluding all mBLs and in addition all the samples that showed a Myc break of some kind. For each stimulation I extracted up to three clusters depending on their direction and the amount of genes included, making sure to get at least two clusters which are regulated in different directions. The genes included in the different clusters can be found on the following pages.

BCR.1			BCR.2			IL21.1		IL21.2
XPO1	PEX14	RAD1	MAPKSP1	RBMX	RAB7A	SEC24B	RIOK2	STAT1
PMS1	TOMM70A	GCDH	ATF1	DUSP11	THAP1	RSRC2	QTRTD1	IRF1
RNASEN	TSN	CPOX	CCDC59	ATP6V1H	RNMT	DYNC1LI1	DDX50	IFIT3
LBR	SFRS4	CLPX	RSL24D1	MAP1LC3B	SAT1	USP8	NMD3	CXCL10
CETN3	HSPH1	MORC2	BZW1	IDI1	CMTM6	RABGGTB	SUCLA2	APOL6
DCK	FOSL1	RFC5	BNIP2	NTAN1	WDR26	FAM178A	DENR	OASL
GPN3	KIAA0406	RAD51C	ETF1	DCTN5	PYROXD1	NOC3L	TCERG1	IFI44L
SLBP	DSN1	RCN2	TAF9	CENPN	PIKFYVE	MTERFD1	DNAJA1	IFIT1
RHOT1	PRPF4B	PIGF	POMP	HNRNPC	NIPA2	ANKRD27	SEPHS2	RSAD2
CAND1	C13orf34	ACTR6	GNAI3	MARS	RHOA	ZDHHC6	DNAJC7	MX1
RACGAP1	BAG2	CDC73	ARF6	YARS	DAZAP2	METAP1	CDC42EP3	IFI44
KIF20A	ZNHIT3	MDM1	YWHAZ	EDEM1	TBC1D15	C2orf49	REV1	IFITM1
AURKA	PDHB	KIAA0528	GABPB1	DDX18	C12orf35	TRA2B	TIMM9	IFITM3
HMMR	CCDC56	PARG	CEBPG	BUD31	WIPF1	SFRS1	XPOT	NA
NDC80	CUTC	THAP11	MAPK6	RCHY1	C10orf26	KRR1	HIRA	
CDC20	C14orf104	C12orf52	RBM22	PRKD3	SKP1	RBM25	TXNL1	
CENPA	COX11	PTCD3	C6orf62	SWAP70	KIAA0232	NUP50	ABHD3	
PLK1	CDK8	CASP6	ERBB2IP	LY75	CHIC2	ZNHIT6	HSPA8	
NEK2	C8orf41	CTR9	UBE2B	CHD1	ARL8B	GNL2	C3orf64	
CCNA2	AURKAIP1	MRPL46	CSNK1A1	EIF1	PPP2R2A	CCT6A	PNN	
KIF18A	IMP3	GPSM2	NFE2L2	IMPA1	RNF111	POLR3E	CSTF3	
RMI1	MINA	NDUFC1	PLEKHB2	CHMP2B	AURKB	RPL23AP7	EXOSC4	
PBK	PUM2	UBE2G1	MTMR6	OSBP	TOP2B	APIP	NUDC	
PRC1	DHX29	PRPSAP1	TRIP12	RAB5A	BCL2L13	PSMC6	DICER1	
BUB1B	COASY	NCAPD3	TM9SF3	LZTS1	ATP6V1E1	ZC3H15	KIAA1012	
PLK4	THAP7	DPF2	DDX5	NCRNA00081	RAB1A	SART3	C11orf58	
CDCA8	MRPS34	HEATR3	YIPF4	RAB21	CD164	CHORDC1	KIF23	
NCAPH	CCDC51	SHCBP1	CEP57	UGCG	PLEKHF2	C14orf104	FBXL4	
TMEM48	MRPL17	C4orf41	MIS12	GABARAPL2	HECA	WDR3	HAUS3	
OIP5	COIL	SKP2	ZNF330	FBXO28	ITSN2	FUBP3	YEATS2	
CEP55	ATMIN	MTX2	LARP7	HSD17B12	PPP2R3C	TAF2	KIAA0753	
KIF14	SMARCAL1	MUDENG	NUP88	YTHDF3	RNF11	ZMYM4	HNRNPD	
ESPL1	COQ9	C15orf44	WAPAL	ECD	ARF4	ZBTB11	FASTKD3	
POLA2	COBRA1	GINS1	VPS26A	ABCF1	SLC3A2	PRPF38B	ATF2	
FEN1	MED20	ZWILCH	UBE2D3	C15orf24	SGPL1	BTAF1	PUS1	
BRCA1	CCDC99	MRPL18	RBM15	TSPYL1	MORF4L1	TNFAIP8	JRKL	
TUBG1	SIP1	C5orf22	PITPNB	ZEB2	PIP4K2A	KIAA0406	HSPA4	
MRPL16	FANCG	TTC33	TARS	CD63	CHSY1	TM9SF3	RPAP3	
TACC3	MCM2	PSME4	MAK16	BACH1	FAF2	PPWD1	TBK1	
SAC3D1	CRIPT	PPM1B	MTHFD2	PRNP	GOLT1B	UBXN2B	EIF4E	
ASPM	GAPVD1	PIAS4	SNX5	HEXB	CYP20A1	LUC7L3	UTP11L	
WDHD1	CNP	WBP4	PDXDC1	PTTG1IP	GOT1	GTF2H1	SUV39H2	
BIRC5	RAB11A	ACAP2	DHX15	CEBPB	ADO	PCNP	TMEM165	
CCNB1	MRPS31	GRSF1	DENR	CTSB	PSEN1	ANAPC10	HNRNPU	
ASF1B	MRPL34	GADD45GIP1	ZNF410	M6PR	VAMP3	RECQL	YME1L1	
ORC1L	PHB	MSH2	DERL2	LAMP1	AGA	KLHL2	ETFA	
KIF2C	POP4	DTYMK	HNRNPK	IFI30	CNIH4	TIMM17A	NUP54	
CDCA3	AGGF1	FRAT2	C5orf28	RIN2	HNMT	CEP170	LIN37	
FOXM1	KIF22	STAMBP	DIMT1L	TFEC	ARL4C	RRS1		
RFC3	POP7	C17orf75	CDK7	GPX4	C3AR1	LARP4		
ECT2	DDX23	FANCL	RAP2C	TLR1	SIRPA	DNAJB14		
DSCC1	ZBED5	SFRS2B	SSFA2	TNFRSF14	EMP3	SERBP1		
C12orf48	BBS10	HMBS	JMJD1C	GPR137B	MKLN1	RBM28		
WRAP53	ZNF638	PAAF1	EIF1B	LY96	ITM2B	SSB		
PARP2	SRBD1	NAA40	ZC3H15	IFNGR1	CCPG1	MPHOSPH10		
TRIP13	GTF2E1	NDUFS3	SAR1A	SRGN	CDC42	C1orf109		
KIF11	MEN1	DUT	BCAS2	CRIM1	KIAA0430	NOL7		
AASDHPPT	LSM2	STRA13	ETFDH	CLEC2B	TNPO1	TCEB1		
FARSA	SRPK1	HADH	MMADHC	CD97	TXNDC9	MAGOH		
CTCF	TTRAP	SEPHS1	MRPS30	GPR65	EIF2S2	RCHY1		
RNASEH2A	ZNF184	ABHD10	MRPL39	CYFIP1	WARS	C9orf82		
NEIL3	MICB	SLC4A1AP	GMFB	IFNGR2	CCL4	FBXW2		
STIL	TUBD1	STRADA	TSG101	ATP6V0E1	NINJ1	ENOPH1		
TARDBP	NFYB	CBX1	CSNK1D	OSTM1	LGALS8	RARS		
ARMC1	KIF18B	MRPS27	GDI2	ACSL1	RRAGD	TAF1A		
MRPL35	PRDM10	WRB	LASP1	WSB2	ACP5	VEZF1		
WDR67	HMGN4	DERA	CSGALNACT2	SGK1	C6orf145	CEBPZ		
NUP37	RAD54B	SAR1B	ZBTB1	CSTB	IL32	AGPAT5		
MRPL12	DTWD1	MAP3K4	SLC2A3P1	UPP1	MARCH7	EIF3J		
RAD54L	BRD8	KIAA1279	ZNF217	LRP10	KLF10	ZCCHC8		
C16orf53	TRMT61B	PPCS	CAPZA1	SERINC1	SMURF2	CNIH		
ALG6	RNF34	C5orf15	HMGCR	RAB22A	SFT2D2	UNC50		
TROAP	CDC27	RIOK2	ATP2C1	AP2B1	FAM18B	SGMS1		
CDC7	EXOC1	CKAP2	HSPA5	C1orf54	UBL3	MLX		
RFC4	SACM1L	NARS2	PPP6C	DENND4A	MAFF	AHCTF1		
UNG	VPS33B	CEP76	TOR1A	USP12	TMEM30A	CEP57		
PPAT	SUCLA2	CACYBP	JOSD1	STK17B	TMBIM6	PIBF1		
FASTKD1	MRS2	TMEM97	PTBP1	GHITM	MON1B	PSMD12		
KIF15	C6orf211	POLE2	YBX1	EIF2S1	DOCK2	SDHAF1		
ANP32A	HMGB3	USP13	EPRS	TOR1B	NR3C1	C12orf44		
SRRD	GEMIN6	RTF1	SDCBP	HSPA13	GCH1	MFSD1		
LRRC47	MRPS16	MRPL49	RNF4	CD86	SAP30BP	NPTN		
PREB	MCM10	CSE1L	CAB39	FRG1	MATR3	SOAT1		
ZC3H14	AP1AR	PLA2G12A	YAF2	PSMD12	GCC2	SLC35A1		
TTK	C4orf27	ELP4	TMEM123	TMEM66	PLXNC1	MTO1		
EFTUD1	FASTKD3	UBTF	TMEM222	ATF4	HEXA	GSPT1		
OSBPL11	CLCN3	GTF3C1	TFAM	POLR2D		UTP6		
MYCBP	PRMT5	BCOR	CDK17	SHMT2		ANKRD40		
DYNLL1	TDP1		USP16	SFPQ		BANP		
NIF3L1	LARS2		GARS	USP15		TMEM33		
MRFAP1L1	BARD1		CNIH	OSBPL8		TMED5		
NARG2	MSH3		RAN	LPXN		SLC30A5		
SPAST	MPHOSPH6		C11orf57	MAP2K3		SNW1		
ZW10	MTIF2		LYRM1	ZHX2		C14orf135		
ELF2	ATR		SARS	RHOG		POP1		
ADH5	PPP2R5E		CLIC1	CYTSA		ZDHHC13		
ORC4L	PSMC6		MAPRE1	H3F3B		KIAA1704		
ORC2L	ZNF107		SLC25A32	PDXK		RSC1A1		
CLINT1	EIF2B4		ADIPOR2	RAB8B		ATP11B		
DLAT	C9orf40		E2F6	NFYA		CUL4A		

BAFF.1			BAFF.2		LPS.1	LPS.2	CD40.1			CD40.2	CD40.3
KPNB1	PDCD5	BPTF	EPHB4	C1orf183	LAMA1	GRIN2B	SFRS2	TFCP2	MGAT2	CCDC134	PICK1
SLTM	ACOT7	NKTR	AGRN	SOX14	VNN3	POU2AF1	MRPL46	PSME3	ATMIN	AMN	SOX15
REV1	MCM10	SART3	VCX2	SLC14A2	ELAVL3	NRD1	MUS81	MAP2K4	UMPS	DOK4	HP
EIF3A	DDX52	MED24	TBKBP1	BBOX1	TSHZ2	FAF1	TOE1	EFR3A	MRPL17	HABP4	TRIM17
LSM14A	NCAPD3	FAM98A	DNAJC28	CNTNAP2	MEGF8	TRBV10-2	FARSA	EPS15	ELAC2	CST2	CYP2A13
CKAP5	LIG1	CCDC99	GPR110	FBXL12	FOXO4	PDGFB	EIF4A3	TCTN3	ZNF706	TUT1	SLC9A3R2
CENPE	RAD51AP1	MRPS15	ALDOAP2	NFASC	GRK1	SLC34A1	MCM3	EIF4ENIF1	PRKRA	SYNPO	SDK2
ORC6L	MRFAP1L1	ANKRD17	SLC18A3	PMS2L4	ADAMTSL2	NR2E3	NOC3L	ANKRD40	CDC5L	CRYBB2	APOC4
KIF18B	TPR	UBE2K	MLXIPL	CCR3	TMEM179B	CRTAC1	ANKRD27	HPS6	TUBG1	OTOR	DCLK2
ATAD2	PRPF3	CSTF3	DLG4	TAS2R9	DOHH	GABRD	DNAJA2	SART3	SAMM50	ANP32C	CELA2B
TOP2A	WHSC2	LSM6	OSGIN1	RGS12	PSD	PDIA2	GPN2	ANAPC10	FBXL15	OR1F2P	CNPY4
ASPM	HAT1	POLR2E	ENPP3	CPLX2	SNAPC2	ZNF205	C16orf80	MRS2	ERCC3	SOX14	LTB4R2
WHSC1	DNAJC13	EED	ALPI	C14orf79	NIN	POU4F3	VRK1	KPNA3	SAE1	KCNG1	GPATCH2
SMC2	ATXN2	MTERFD1	SOX15	C19orf42	FOXE1	ADORA1	NDUFV2	APIP	CDC25A	HUNK	KCNQ1
MSH2	MTDH	TAF5	CADM4	GRIN2A	ZNF646	KCNJ5	COX4NB	BCL10	UCHL3	KCNQ1DN	SULT4A1
CACYBP	TCERG1	SS18L1	STBD1	C6orf54	C1orf183	IGF2AS	SHQ1	CTSH	ACD	MAP1A	IGF2AS
WBP4	RAD50	LRDD	SCNN1A	UTP14C	GJA8	ABO	MCAT	TBK1	PSMB2	TACR2	IGFBP5
ZC3H13	YY1AP1	NCAPG2	PFKFB4	DUSP26	SIX5	TRGV5	PRPF4	RRP1B	NAT10	NGB	STMN3
SFRS1	DLX2	SPC25	MATN4	SLC9A3R2	PTPRS	STMN3	LCMT1	NUFIP1	C21orf59	APOBEC3F	PYY2
NASP	DUOX2	NCOA6	PCYT1B	AGER	TEX11	PKLR	RFC2	GRPEL1	RNF115	GRTP1	SLC9A3
ACYP1	NDUFS1	PSPC1	LRAT	GNG7	FOXC2	WFDC8	RAD51C	CHMP6	ELOVL1	SPINT3	ADRA2C
SDCCAG1	CEBPZ	LIN37	OPRD1	MUC6	SLC2A4	NRXN2	FEN1	YIF1A	NECAP2	KCNF1	NRXN2
LUC7L3	NDUFB8	HAUS6	NBR2		C2orf72	SPTB	TRIP13	NSMCE4A	NBN	BRD7P3	AKAP4
KIF20B	EXOSC7	EXOSC10	NPTXR		NFKBIL2	OPRL1	TRA2B	UBE3B	GCNT4	NYX	FAM179B
SUCLA2	HSPE1	IMMT	KRT16		EPO	CEMP1	SFRS1	PPCS	SEMA3B	TM6SF2	PLCH2
SPAG5	C14orf104	NAA16	TNXB		GPR32	SLC9A3R2	HNRNPR	ZBTB11	CLCN1	SLC12A4	ANKRD2
MKI67	ATR	USP48	TUBB4Q		PRDM8	TLX2	PEF1	NPL	GFM1	NFKBIL2	UCP3
PLK4	PNN	SPEN	SH2D4A		PLEKHA4	SYT5	C2orf47	NOTCH2	PNLIP	MATK	ALOX12B
NCAPH	PHIP	PIAS4	TGFB2		CYP11B2	NFIX	DLD	LAMP2	RCN2	GHSR	HIST1H2AK
PRDX3	ZFC3H1	ABHD3	WIF1		BARX1	MMP16	SFRS3	RDX	NPAS1		MYCN
ADK	NUP43	CCDC91	MYH15		KRT3	SNCA	DDX19A	TMEM165	DNAJA1		SCN2B
PAICS	NCOR1	PSMD1	HOXD10		CEBPE	MAP1LC3C	CIAPIN1	PIK3CB	SF3B2		ZNF205
EIF2B3	TRIM24	RECQL4	GDF2		CDKN2B	APCS	CYC1	HLA-J	MRPL39		EPB41L1
NUDC	PSMC6	MTHFD1	CCL27		ZCCHC4	PSPN	PEX14	HLA-F	DDA1		SLC34A1
KIN	POLR2B	PPP2R5E	MAPT			MTHFR	GCDH	HLA-B	WBP11		NHEJ1
C12orf48	NCL	SEC24A	CRYBB2			LRIT1	EXOSC4	HLA-G	CMAS		HDAC11
SSB	RECQL	SAR1B	TPPP3				PUF60	HLA-C	CCT6A		FKBP6
SMC3	SYNCRIP	CEP63	FGF20				REEP4	HCP5	TARDBP		PCYT1B
KIF14	MCCC2	UCHL5	PCDHB17				SF4	HLA-A	IMP3		SNCG
POLD1	HSPA4	RANGAP1	ADAMTS7				UBTF	BTN2A2	RIOK2		OR3A2
SCRIB	PSMD14	WDHD1	CABP2				ABCF3	PLEKHO1	MED22		PTCRA
REXO4	C1orf109	DNA2	PRB3				RRP9	CSF2RB	FTSJ3		ROM1
MSH3	SLC4A1AP	C11orf48	UBTD1				STX4	HLA-DPB1	ZFR		NPHS1
MPHOSPH10	MED4	C10orf2	HSPB2				CTCF	ZNF672	PWP1		SIRT5
FASTKD3	WDR61	CACNA1S	PDGFB				SMARCAL1	RFX5	MINA		CATSPERB
ARL6IP4	AHCTF1	SAP30L	NR2E3				NOC4L	NUBP1	RARS		TCL6
SAFB2	MAGOH	CTNNBL1	CRTAC1				WDR18	TRIP4	HNRNPA0		HGFAC
CHERP	BRD7	DHX30	C14orf115				PDCD5	RAB27A	PRLR		CCNJL
SRRM1	NSMCE4A	C14orf166	GPR37L1				MAT2A	BID	HSPA4		HOXC13
ZNF146	PHF3	DHFR	LCN1				AIMP2	ARIH1	SIX3		ADCY10
SUPT16H	PPP1R12A	ACD	NDST1				MTIF2	CD59	ADRA1D		ACAP2
KIAA0406	MAVS	GPN3	TRPV6				HEATR3	SLC31A2	WDR41		HTR5A
UPF2	ST13	PDHB	HOXA3				DDRGK1	FLVCR2	ALCAM		TBL2
LUC7L2	SNX1	PIBF1	TP53TG5				THOC1	CTSC	PALB2		SLC38A3
BMS1	CBX1	E4F1	TSKS				MTHFS	LGALS9	SYNGR2		ATP6V1B1
CHAF1A	HNRNPH3		KRT19P2				TBL3	RAC2	PSMC4		PAX3
TCOF1	RNASEN		RNLS				PPP3CA	HNRNPF	RAP2B		RBP3
SAFB	PRR11		XPNPEP3				RPS6KA1	PSMC6	TAF1B		C8orf30A
TOPBP1	HNRNPD		DDX25				TYMS	SLURP1	NUDT21		PBOV1
SSRP1	PRPF19		HSD3B2				RASSF2	TIMM17A	YARS2		GNG7
SFRS12	DDX49		ACCN2				MRPL12	MINPP1	DCLRE1B		NPPA
BAZ1A	ESPL1		RGS7				HSPA14	UBAC1	CINP		CLCN7
DAZAP1	EDF1		TACR2				BRMS1	C1orf35	IFRD2		PDGFRA
RPAP3	CSF3		PSG3				COPS7B	DCC	SERBP1		TASP1
HSPH1	NIPBL		OR7C1				PIK3CD	MTMR4	TTC4		EXOSC1
RSF1	ASCC3		BICC1				ACOX3	RPS6KB1	C16orf88		WNT8B
TOP1	NUP133		SYN2				IMPDH1	UBE3C	PSMA7		ADORA1
PPWD1	TECR		CAMKV				SARS2	TMEM70	MCCC2		DHX29
KTN1	CCNE1		AMBN				SPIB	KIAA0947	TESC		TLX2
CENPF	KIAA0101		EPB41L1				SH3BP5	LARS2	LMNB2		DUSP9
IARS2	AK2		AP3D1				DOCK10	C12orf43	THAP11		KCNJ5
BRAP	EXOSC9		SPAG9				ARHGAP17	STAM2	FNBP1		
ACTR6	CHCHD3		OR11A1				RAB7L1	ISOC1	EIF4H		
FABP5	RPIA		PKLR				EIF2B5	METAP1	CD160		
POLR2F	KIAA0562		SHB				MICALL1	API5	TRAPPC3		
ITGB1BP1	STK39		ADCY10				POP4	SRPK1	FASTKD3		
GNB1	LRRC40		TGFBRAP1				TTC35	MFN2	CTR9		
NUP107	HNRNPH1		TRIM49				UAP1	SEC23IP	HCK		
TCP1	C14orf156		PCDH1				ICT1	CAPRIN1	C3orf37		
CCT6A	PPIH		ATP2B3				TACO1	SEC24B	PHF16		
PRPF38B	RIOK2		RAX				MRPL34	ARFGEF1	TRIM44		
CPSF6	RCBTB1		CDHR5				C8orf33	CLPX	KCNJ1		
AHSA1	PRPF4B		WNT8B				UCK2	UBE2L3	NOL11		
ERCC3	SMS		RHO				TOMM70A	PDHX	CPSF4		
YY1	ACIN1		KRT15				AGFG1	C14orf104	C7orf23		
CHD8	BRCA1		CACNA1E				LSM12	BRD7	RSRC2		
PPIG	EBNA1BP2		LMAN1L				G3BP1	METTL1	WDR1		
DUT	POLR3K		GUCA1B				ANXA7	C9orf114	WDR3		
CHORDC1	GRPEL1		DNAH9				EXOC1	SLC35B1			
DDX23	CISD1		NGFR				SLC25A46	SNX4			
RUVBL1	SNRNP25		GJA3				TRPM3	DCTD			
EXO1	MELK		MYL10				DDX50	PHB			
EIF5B	TACC3		NR0B1				EIF4E	FAM35A			
PABPN1	SNRPF		GRAP2				ZC3H14	CNPY3			
NDUFS8	PFAS		IFT140				HEATR2	ABCB11			
TAF2	USP24		LOC645961				GTF3C2	KEAP1			
BUB1B	CMAS		PTPRU				TIMM23	DDX42			
DNM1L	ROCK1		KRT83				POLR3E	CCDC51			
FIP1L1	WDR18		S100A14				DHX30	BRIX1			
SMARCA4	IK		NLGN3				DPH2	RRS1			

Bibliography

[1] N. Akamatsu, Y. Yamada, H. Hasegawa, K. Makabe, R. Asano, I. Kumagai, K. Murata, Y. Imaizumi, K. Tsukasaki, K. Tsuruda, K. Sugahara, S. Atogami, K. Yanagihara, and S. Kamihira. High il-21 receptor expression and apoptosis induction by il-21 in follicular lymphoma. *Cancer Lett*, 256(2):196–206, Oct 2007.

[2] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt. Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769):503–511, Feb 2000.

[3] S. An and K. A. Knox. Ligation of cd40 rescues ramos-burkitt lymphoma b cells from calcium ionophore- and antigen receptor-triggered apoptosis by inhibiting activation of the cysteine protease cpp32/yama and cleavage of its substrate parp. *FEBS Lett*, 386(2-3):115–122, May 1996.

[4] R. J. Armitage, W. C. Fanslow, L. Strockbine, T. A. Sato, K. N. Clifford, B. M. Macduff, D. M. Anderson, S. D. Gimpel, T. Davis-Smith, and C. R. Maliszewski. Molecular and biological characterization of a murine ligand for cd40. *Nature*, 357(6373):80–82, May 1992.

[5] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet*, 25(1):25–29, May 2000.

[6] J. Baehring and J. Piepmeier. Pathology and classification of tumors of the nervous system. 2006.

[7] B. W. Baron, G. Nucifora, N. McCabe, R. Espinosa, M. M. L. Beau, and T. W. McKeithan. Identification of the gene associated with the recurring chromosomal translocations t(3;14)(q27;q32) and t(3;22)(q27;q11) in b-cell lymphomas. *Proc Natl Acad Sci U S A*, 90(11):5262–5266, Jun 1993.

[8] A. Behdad and A. Perry. Central nervous system primitive neuroectodermal tumors: a clinicopathologic and genetic study of 33 cases. *Brain Pathol*, 20(2):441–450, Mar 2010.

[9] C. P. Beier, P. Kumar, K. Meyer, P. Leukel, V. Bruttel, I. Aschenbrenner, M. J. Riemenschneider, A. Fragoulis, P. Rümmele, K. Lamszus, J. B. Schulz, J. Weis, U. Bogdahn, J. Wischhusen, P. Hau, R. Spang, and D. Beier. The cancer stem cell subtype determines immune infiltration of glioblastoma. *Stem Cells Dev*, Jun 2012.

[10] D. Beier, P. Hau, M. Proescholdt, A. Lohmeier, J. Wischhusen, P. J. Oefner, L. Aigner, A. Brawanski, U. Bogdahn, and C. P. Beier. Cd133(+) and cd133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles. *Cancer Res*, 67(9):4010–4015, May 2007.

[11] D. Beier, S. Röhrl, D. R. Pillai, S. Schwarz, L. A. Kunz-Schughart, P. Leukel, M. Proescholdt, A. Brawanski, U. Bogdahn, A. Trampe-Kieslich, B. Giebel, J. Wischhusen, G. Reifenberger, P. Hau, and C. P. Beier. Temozolomide preferentially depletes cancer stem cells in glioblastoma. *Cancer Res*, 68(14):5706–5715, Jul 2008.

[12] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.

[13] S. Bentink, S. Wessendorf, C. Schwaenen, M. Rosolowski, W. Klapper, A. Rosenwald, G. Ott, A. H. Banham, H. Berger, A. C. Feller, M.-L. Hansmann, D. Hasenclever, M. Hummel, D. Lenze, P. Möller, B. Stuerzenhofecker, M. Loeffler, L. Truemper, H. Stein, R. Siebert, R. Spang, and M. M. in Malignant Lymphomas Network Project of the. Pathway activation patterns in diffuse large b-cell lymphomas. *Leukemia*, 22(9):1746–1754, Sep 2008.

[14] M. Bi, C. Naczki, M. Koritzinsky, D. Fels, J. Blais, N. Hu, H. Harding, I. Novoa, M. Varia, J. Raleigh, D. Scheuner, R. J. Kaufman, J. Bell, D. Ron, B. G. Wouters, and C. Koumenis. Er stress-regulated translation increases tolerance to extreme hypoxia and promotes tumor growth. *EMBO J*, 24(19):3470–3481, Oct 2005.

[15] G. A. Bishop. The many faces of cd40: multiple roles in normal immunity and disease. *Semin Immunol*, 21(5):255–256, Oct 2009.

[16] R. Bjerkvig, B. B. Tysnes, K. S. Aboody, J. Najbauer, and A. J. A. Terzis. Opinion: the origin of the cancer stem cell: current controversies and new insights. *Nat Rev Cancer*, 5(11):899–904, Nov 2005.

[17] C. Bossen and P. Schneider. Baff, april and their receptors: structure, function and signaling. *Semin Immunol*, 18(5):263–275, Oct 2006.

[18] T. Boveri. Eine fr die erste orientierung geeignete darstellung dieser und anderer chromosomenprobleme findet sich in meiner schrift: Ergebnisse ber die konstitution der chromatischen substanz des zellkerns. *Zellenstudien. VI.*, 1907.

[19] T. Boveri. Concerning the origin of malignant tumours by theodor boveri. translated and annotated by henry harris. *J Cell Sci*, 121 Suppl 1:1–84, Jan 2008.

[20] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman. Survival analysis part ii: multivariate data analysis–an introduction to concepts and methods. *Br J Cancer*, 89(3):431–436, Aug 2003.

[21] G. Brady, G. J. Macarthur, and P. J. Farrell. Epstein-barr virus and burkitt lymphoma. *Postgrad Med J*, 84(993):372–377, Jul 2008.

[22] C. Brennan, H. Momota, D. Hambardzumyan, T. Ozawa, A. Tandon, A. Pedraza, and E. Holland. Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations. *PLoS One*, 4(11):e7752, 2009.

[23] L. Bruhn, A. Munnerlyn, and R. Grosschedl. Aly, a context-dependent coactivator of lef-1 and aml-1, is required for tcralpha enhancer function. *Genes Dev*, 11(5):640–653, Mar 1997.

[24] D. Burkitt. A sarcoma involving the jaws in african children. *Br J Surg*, 46(197):218–223, Nov 1958.

[25] D. Chiron, I. Bekeredjian-Ding, C. Pellat-Deceunynck, R. Bataille, and G. Jego. Toll-like receptors: lessons to learn from normal and malignant human b cells. *Blood*, 112(6):2205–2213, Sep 2008.

[26] W. Ci, J. M. Polo, L. Cerchietti, R. Shaknovich, L. Wang, S. N. Yang, K. Ye, P. Farinha, D. E. Horsman, R. D. Gascoyne, O. Elemento, and A. Melnick. The bcl6 transcriptional program features repression of multiple oncogenes in primary b cells and is deregulated in dlbcl. *Blood*, 113(22):5536–5548, May 2009.

[27] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman. Survival analysis part i: basic concepts and first analyses. *Br J Cancer*, 89(2):232–238, Jul 2003.

[28] H. Clevers. Wnt/beta-catenin signaling in development and disease. *Cell*, 127(3):469–480, Nov 2006.

[29] B. Coiffier. Current strategies for the treatment of diffuse large b cell lymphoma. *Curr Opin Hematol*, 12(4):259–265, Jul 2005.

[30] J. M. Coquet, K. Kyparissoudis, D. G. Pellicci, G. Besra, S. P. Berzins, M. J. Smyth, and D. I. Godfrey. Il-21 is produced by nkt cells and modulates nkt cell activation and cytokine production. *J Immunol*, 178(5):2827–2834, Mar 2007.

[31] V. Coskun, H. Wu, B. Blanchi, S. Tsao, K. Kim, J. Zhao, J. C. Biancotti, L. Hutnick, R. C. Krueger, G. Fan, J. de Vellis, and Y. E. Sun. Cd133+ neural stem cells in the ependyma of mammalian postnatal forebrain. *Proc Natl Acad Sci U S A*, 105(3):1026–1031, Jan 2008.

[32] D. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

[33] H. Cushing and P. Bailey. Classification of tumors of the glioma group on a histogenetic basis with a correlated study of prognosis. *JE Lippincott Comp., Philadelphia, London & Montreal*, 1926.

[34] D. L. Daniels and W. I. Weis. Beta-catenin directly displaces groucho/tle repressors from tcf/lef in wnt-mediated transcription activation. *Nat Struct Mol Biol*, 12(4):364–371, Apr 2005.

[35] S. S. Dave, K. Fu, G. W. Wright, L. T. Lam, P. Kluin, E.-J. Boerma, T. C. Greiner, D. D. Weisenburger, A. Rosenwald, G. Ott, H.-K. Müller-Hermelink, R. D. Gascoyne, J. Delabie, L. M. Rimsza, R. M. Braziel, T. M. Grogan, E. Campo, E. S. Jaffe, B. J. Dave, W. Sanger, M. Bast, J. M. Vose, J. O. Armitage, J. M. Connors, E. B. Smeland, S. Kvaloy, H. Holte, R. I. Fisher, T. P. Miller, E. Montserrat, W. H. Wilson, M. Bahl, H. Zhao, L. Yang, J. Powell, R. Simon, W. C. Chan, L. M. Staudt, and L. M. P. Project. Molecular diagnosis of burkitt's lymphoma. *N Engl J Med*, 354(23):2431–2442, Jun 2006.

[36] I. D. Davis, K. Skak, M. J. Smyth, P. E. G. Kristjansen, D. M. Miller, and P. V. Sivakumar. Interleukin-21 signaling: functions in cancer and autoimmunity. *Clin Cancer Res*, 13(23):6926–6932, Dec 2007.

[37] R. E. Davis, K. D. Brown, U. Siebenlist, and L. M. Staudt. Constitutive nuclear factor kappab activity is required for survival of activated b cell-like diffuse large b cell lymphoma cells. *J Exp Med*, 194(12):1861–1874, Dec 2001.

[38] R. E. Davis, V. N. Ngo, G. Lenz, P. Tolar, R. M. Young, P. B. Romesser, H. Kohlhammer, L. Lamy, H. Zhao, Y. Yang, W. Xu, A. L. Shaffer, G. Wright, W. Xiao, J. Powell, J.-K. Jiang, C. J. Thomas, A. Rosenwald, G. Ott, H. K. Muller-Hermelink, R. D. Gascoyne, J. M. Connors, N. A. Johnson, L. M. Rimsza, E. Campo, E. S. Jaffe, W. H. Wilson, J. Delabie, E. B. Smeland, R. I. Fisher, R. M. Braziel, R. R. Tubbs, J. R. Cook, D. D. Weisenburger, W. C. Chan, S. K. Pierce, and L. M. Staudt. Chronic active b-cell-receptor signalling in diffuse large b-cell lymphoma. *Nature*, 463(7277):88–92, Jan 2010.

[39] D. Defays. An efficient algorithm for a complete link method. *The Computer Journal*, 20(4):364–366, 1977.

[40] C. Denoyelle, G. Abou-Rjaily, V. Bezrookove, M. Verhaegen, T. M. Johnson, D. R. Fullen, J. N. Pointer, S. B. Gruber, L. D. Su, M. A. Nikiforov, R. J. Kaufman, B. C. Bastian, and M. S. Soengas. Anti-oncogenic role of the endoplasmic reticulum differentially activated by mutations in the mapk pathway. *Nat Cell Biol*, 8(10):1053–1063, Oct 2006.

[41] A. L. Dent, A. L. Shaffer, X. Yu, D. Allman, and L. M. Staudt. Control of inflammation, cytokine expression, and germinal center formation by bcl-6. *Science*, 276(5312):589–592, Apr 1997.

[42] T. Denysenko, L. Gennero, M. A. Roos, A. Melcarne, C. Juenemann, G. Faccani, I. Morra, G. Cavallo, S. Reguzzi, G. Pescarmona, and A. Ponzetto. Glioblastoma cancer stem cells: heterogeneity, microenvironment and related therapeutic strategies. *Cell Biochem Funct*, 28(5):343–351, Jul 2010.

[43] A. Dogan, E. Bagdi, P. Munson, and P. G. Isaacson. Cd10 and bcl-6 expression in paraffin sections of normal lymphoid tissue and b-cell lymphomas. *Am J Surg Pathol*, 24(6):846–852, Jun 2000.

[44] O. D'Orlando, G. Gri, G. Cattaruzzi, S. Merluzzi, E. Betto, V. Gattei, and C. Pucillo. Outside inside signalling in cd40-mediated b cell activation. *J Biol Regul Homeost Agents*, 21(3-4):49–62, 2007.

[45] S. L. Doyle and L. A. J. O'Neill. Toll-like receptors: from the discovery of nfkappab to new insights into transcriptional regulations in innate immunity. *Biochem Pharmacol*, 72(9):1102–1113, Oct 2006.

[46] G. Dsen, E. Tenstad, M. K. Nygren, H. Stubberud, S. Funderud, and E. Rian. Wnt expression and canonical wnt signaling in human bone marrow b lymphopoiesis. *BMC Immunol*, 7:13, 2006.

[47] R. Ettinger, S. Kuchen, and P. E. Lipsky. The role of il-21 in regulating b-cell function in health and disease. *Immunol Rev*, 223:60–86, Jun 2008.

[48] P. Farinha and R. D. Gascoyne. Molecular pathogenesis of mucosa-associated lymphoid tissue lymphoma. *J Clin Oncol*, 23(26):6370–6378, Sep 2005.

[49] E. R. Fearon and B. Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, 61(5):759–767, Jun 1990.

[50] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *Int J Cancer*, 127(12):2893–2917, Dec 2010.

[51] F. Feuerhake, J. L. Kutok, S. Monti, W. Chen, A. S. LaCasce, G. Cattoretti, P. Kurtin, G. S. Pinkus, L. de Leval, N. L. Harris, K. J. Savage, D. Neuberg, T. M. Habermann, R. Dalla-Favera, T. R. Golub, J. C. Aster, and M. A. Shipp. Nfkappab activity, function, and target-gene signatures in primary mediastinal large b-cell lymphoma and diffuse large b-cell lymphoma subtypes. *Blood*, 106(4):1392–1399, Aug 2005.

[52] T. M. Foy, J. D. Laman, J. A. Ledbetter, A. Aruffo, E. Claassen, and R. J. Noelle. gp39-cd40 interactions are essential for germinal center formation and the development of b cell memory. *J Exp Med*, 180(1):157–163, Jul 1994.

[53] T. M. Foy, D. M. Shepherd, F. H. Durie, A. Aruffo, J. A. Ledbetter, and R. J. Noelle. In vivo cd40-gp39 interactions are essential for thymus-dependent humoral immunity. ii. prolonged suppression of the humoral immune response by an antibody to the ligand for cd40, gp39. *J Exp Med*, 178(5):1567–1575, Nov 1993.

[54] K. A. Frazer, L. Pachter, A. Poliakov, E. M. Rubin, and I. Dubchak. Vista: computational tools for comparative genomics. *Nucleic Acids Res*, 32(Web Server issue):W273–W279, Jul 2004.

[55] J. W. Friedberg and R. I. Fisher. Diffuse large b-cell lymphoma. *Hematol Oncol Clin North Am*, 22(5):941–52, ix, Oct 2008.

[56] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer Series in Statistics, 2001.

[57] M. A. Friese, J. Wischhusen, W. Wick, M. Weiler, G. Eisele, A. Steinle, and M. Weller. Rna interference targeting transforming growth factor-beta enhances nkg2d-mediated antiglioma immune response, inhibits glioma cell migration and invasiveness, and abrogates tumorigenicity in vivo. *Cancer Res*, 64(20):7596–7603, Oct 2004.

[58] T. Fukuda, T. Yoshida, S. Okada, M. Hatano, T. Miki, K. Ishibashi, S. Okabe, H. Koseki, S. Hirosawa, M. Taniguchi, N. Miyasaka, and T. Tokuhisa. Disruption of the bcl6 gene results in an impaired germinal center formation. *J Exp Med*, 186(3):439–448, Aug 1997.

[59] C. F. Fuqua, R. Akomeah, J. O. Price, and S. E. Adunyah. Involvement of erk-1/2 in il-21-induced cytokine production in leukemia cells and human monocytes. *Cytokine*, 44(1):101–107, Oct 2008.

[60] E. Gallagher, T. Enzler, A. Matsuzawa, A. Anzelon-Mills, D. Otero, R. Holzer, E. Janssen, M. Gao, and M. Karin. Kinase mekk1 is required for cd40-dependent activation of the kinases jnk and p38, germinal center formation, b cell proliferation and antibody production. *Nat Immunol*, 8(1):57–63, Jan 2007.

[61] R. Galli, E. Binda, U. Orfanelli, B. Cipelletti, A. Gritti, S. D. Vitis, R. Fiocco, C. Foroni, F. Dimeco, and A. Vescovi. Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res*, 64(19):7011–7021, Oct 2004.

[62] R. K. Gandhirajan, P. A. Staib, K. Minke, I. Gehrke, G. Plickert, A. Schlösser, E. K. Schmitt, M. Hallek, and K.-A. Kreuzer. Small molecule inhibitors of wnt/beta-catenin/lef-1 signaling induces apoptosis in chronic lymphocytic leukemia cells in vitro and in vivo. *Neoplasia*, 12(4):326–335, Apr 2010.

[63] P. Gelebart, M. Anand, H. Armanious, A. C. Peters, J. D. Bard, H. M. Amin, and R. Lai. Constitutive activation of the wnt canonical pathway in mantle cell lymphoma. *Blood*, 112(13):5171–5179, Dec 2008.

[64] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, 5(10):R80, 2004.

[65] K. Giese, J. Cox, and R. Grosschedl. The hmg domain of lymphoid enhancer factor 1 bends dna and facilitates assembly of functional nucleoprotein structures. *Cell*, 69(1):185–195, Apr 1992.

[66] D. J. Good and R. D. Gascoyne. Classification of non-hodgkin's lymphoma. *Hematol Oncol Clin North Am*, 22(5):781–805, vii, Oct 2008.

[67] R. W. J. Groen, M. E. C. M. Oud, E. J. M. Schilder-Tol, M. B. Overdijk, D. ten Berge, R. Nusse, M. Spaargaren, and S. T. Pals. Illegitimate wnt pathway activation by beta-catenin mutation or autocrine stimulation in t-cell malignancies. *Cancer Res*, 68(17):6969–6977, Sep 2008.

[68] H. S. Guenther, N. O. Schmidt, H. S. Phillips, D. Kemming, S. Kharbanda, R. Soriano, Z. Modrusan, H. Meissner, M. Westphal, and K. Lamszus. Glioblastoma-derived stem cell-enriched cultures form distinct subgroups according to molecular and phenotypic criteria. *Oncogene*, 27(20):2897–2909, May 2008.

[69] A. Gutierrez, R. C. Tschumper, X. Wu, T. D. Shanafelt, J. Eckel-Passow, P. M. Huddleston, S. L. Slager, N. E. Kay, and D. F. Jelinek. Lef-1 is a prosurvival factor in chronic lymphocytic leukemia and is expressed in the preleukemic state of monoclonal b-cell lymphocytosis. *Blood*, 116(16):2975–2983, Oct 2010.

[70] N. C. Gutirrez, E. M. Ocio, J. de Las Rivas, P. Maiso, M. Delgado, E. Fermin, M. J. Arcos, M. L. Snchez, J. M. Hernndez, and J. F. S. Miguel. Gene expression profiling of b lymphocytes and plasma cells from waldenström's macroglobulinemia: comparison with expression patterns of the same cell counterparts from chronic lymphocytic leukemia, multiple myeloma and normal individuals. *Leukemia*, 21(3):541–549, Mar 2007.

[71] M. Harada, K. Magara-Koyanagi, H. Watarai, Y. Nagata, Y. Ishii, S. Kojo, S. Horiguchi, Y. Okamoto, T. Nakayama, N. Suzuki, W.-C. Yeh, S. Akira, H. Kitamura, O. Ohara, K. ichiro Seino, and M. Taniguchi. Il-21-induced bepsilon cell apoptosis mediated by natural killer t cells suppresses ige responses. *J Exp Med*, 203(13):2929–2937, Dec 2006.

[72] H. P. Harding, Y. Zhang, H. Zeng, I. Novoa, P. D. Lu, M. Calfon, N. Sadri, C. Yun, B. Popko, R. Paules, D. F. Stojdl, J. C. Bell, T. Hettmann, J. M. Leiden, and D. Ron. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell*, 11(3):619–633, Mar 2003.

[73] R. P. Hasserjian, G. Ott, K. S. J. Elenitoba-Johnson, O. Balague-Ponz, D. de Jong, and L. de Leval. Commentary on the who classification of tumors of lymphoid tissues (2008): "gray zone" lymphomas overlapping with burkitt lymphoma or classical hodgkin lymphoma. *J Hematop*, Jun 2009.

[74] B. He, A. Chadburn, E. Jou, E. J. Schattner, D. M. Knowles, and A. Cerutti. Lymphoma b cells evade apoptosis through the tnf family members baff/blys and april. *J Immunol*, 172(5):3268–3279, Mar 2004.

[75] S. J. M. Healy, A. M. Gorman, P. Mousavi-Shafaei, S. Gupta, and A. Samali. Targeting the endoplasmic reticulum-stress response as an anticancer strategy. *Eur J Pharmacol*, 625(1-3):234–246, Dec 2009.

[76] F. Hertwig, K. Meyer, S. Braun, S. Ek, R. Spang, C. V. Pfenninger, I. Artner, G. Prost, X. Chen, J. A. Biegel, A. R. Judkins, E. Englund, and U. A. Nuber. Definition of genetic events directing the development of distinct types of brain tumors from postnatal neural stem/progenitor cells. *Cancer Res*, 72(13):3381–3392, Jul 2012.

[77] J. M. Hildebrand, Z. Luo, M. K. Manske, T. Price-Troska, S. C. Ziesmer, W. Lin, B. S. Hostager, S. L. Slager, T. E. Witzig, S. M. Ansell, J. R. Cerhan, G. A. Bishop, and A. J. Novak. A baff-r mutation associated with non-hodgkin lymphoma alters traf recruitment and reveals new insights into baff-r signaling. *J Exp Med*, 207(12):2569–2579, Nov 2010.

[78] K. Hoshino, O. Takeuchi, T. Kawai, H. Sanjo, T. Ogawa, Y. Takeda, K. Takeda, and S. Akira. Cutting edge: Toll-like receptor 4 (tlr4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for tlr4 as the lps gene product. *J Immunol*, 162(7):3749–3752, Apr 1999.

[79] K. K. Hristov, K. A. Knox, and V. I. Mitev. Regulation of tyrosine phosphorylation during the cd40-mediated rescue of ramos-bl b cells from bcr-triggered apoptosis. *Int J Mol Med*, 16(5):937–941, Nov 2005.

[80] W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, and M. Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18 Suppl 1:S96–104, 2002.

[81] M. Hummel, S. Bentink, H. Berger, W. Klapper, S. Wessendorf, T. F. E. Barth, H.-W. Bernd, S. B. Cogliatti, J. Dierlamm, A. C. Feller, M.-L. Hansmann, E. Haralambieva, L. Harder, D. Hasenclever, M. Kühn, D. Lenze, P. Lichter, J. I. Martin-Subero, P. Möller, H.-K. Müller-Hermelink, G. Ott, R. M. Parwaresch, C. Pott, A. Rosenwald, M. Rosolowski, C. Schwaenen, B. Stürzenhofecker, M. Szczepanowski, H. Trautmann, H.-H. Wacker, R. Spang, M. Loeffler, L. Trümper, H. Stein, R. Siebert, and M. M. in Malignant Lymphomas Network Project of the Deutsche Krebshilfe. A biologic definition of burkitt's lymphoma from transcriptional and genomic profiling. *N Engl J Med*, 354(23):2419–2430, Jun 2006.

[82] J. T. Huse, H. S. Phillips, and C. W. Brennan. Molecular subclassification of diffuse gliomas: seeing order in the chaos. *Glia*, 59(8):1190–1199, Aug 2011.

[83] T. N. Ignatova, V. G. Kukekov, E. D. Laywell, O. N. Suslov, F. D. Vrionis, and D. A. Steindler. Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. *Glia*, 39(3):193–206, Sep 2002.

[84] R. A. Irizarry, B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed. Summaries of affymetrix genechip probe level data. *Nucleic Acids Res*, 31(4):e15, Feb 2003.

[85] R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264, Apr 2003.

[86] M. S. Isakoff, C. G. Sansam, P. Tamayo, A. Subramanian, J. A. Evans, C. M. Fillmore, X. Wang, J. A. Biegel, S. L. Pomeroy, J. P. Mesirov, and C. W. M. Roberts. Inactivation of the snf5 tumor suppressor stimulates cell cycle progression and cooperates with p53 loss in oncogenic transformation. *Proc Natl Acad Sci U S A*, 102(49):17745–17750, Dec 2005.

[87] H. Iwasaki, S. ichi Mizuno, Y. Arinobu, H. Ozawa, Y. Mori, H. Shigematsu, K. Takatsu, D. G. Tenen, and K. Akashi. The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. *Genes Dev*, 20(21):3010–3021, Nov 2006.

[88] E. S. Jaffe, N. L. Harris, J. Diebold, and H. K. Müller-Hermelink. World health organization classification of lymphomas: a work in progress. *Ann Oncol*, 9 Suppl 5:S25–S30, 1998.

[89] K. L. Jang, J. Shackelford, S. Y. Seo, and J. S. Pagano. Up-regulation of beta-catenin by a viral oncogene correlates with inhibition of the seven in absentia homolog 1 in b lymphoma cells. *Proc Natl Acad Sci U S A*, 102(51):18431–18436, Dec 2005.

[90] J. Jeuken, C. van den Broecke, S. Gijsen, S. Boots-Sprenger, and P. Wesseling. Ras/raf pathway activation in gliomas: the result of copy number gains rather than activating mutations. *Acta Neuropathol*, 114(2):121–133, Aug 2007.

[91] L. Johnson, K. Mercer, D. Greenbaum, R. T. Bronson, D. Crowley, D. A. Tuveson, and T. Jacks. Somatic activation of the k-ras oncogene causes early onset lung cancer in mice. *Nature*, 410(6832):1111–1116, Apr 2001.

[92] B. Kahl. Chemotherapy combinations with monoclonal antibodies in non-hodgkin's lymphoma. *Semin Hematol*, 45(2):90–94, Apr 2008.

[93] M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, Jan 2000.

[94] E. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[95] Y. L. Kasamon and L. J. Swinnen. Treatment advances in adult burkitt lymphoma and leukemia. *Curr Opin Oncol*, 16(5):429–435, Sep 2004.

[96] P. A. Kenny, G. Y. Lee, C. A. Myers, R. M. Neve, J. R. Semeiks, P. T. Spellman, K. Lorenz, E. H. Lee, M. H. Barcellos-Hoff, O. W. Petersen, J. W. Gray, and M. J. Bissell. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol Oncol*, 1(1):84–96, Jun 2007.

[97] N. I. Khan, K. F. Bradstock, and L. J. Bendall. Activation of wnt/beta-catenin pathway mediates growth and survival in b-cell progenitor acute lymphoblastic leukaemia. *Br J Haematol*, 138(3):338–348, Aug 2007.

[98] W. N. Khan. B cell receptor and baff receptor signaling regulation of b cell homeostasis. *J Immunol*, 183(6):3561–3567, Sep 2009.

[99] J. B. Kim, V. Sebastiano, G. Wu, M. J. Arazo-Bravo, P. Sasse, L. Gentile, K. Ko, D. Ruau, M. Ehrich, D. van den Boom, J. Meyer, K. Hübner, C. Bernemann, C. Ortmeier, M. Zenke, B. K. Fleischmann, H. Zaehres, and H. R. Schöler. Oct4-induced pluripotency in adult neural stem cells. *Cell*, 136(3):411–419, Feb 2009.

[100] S. J. Kim, S. J. Lee, I. Y. Choi, Y. Park, C. W. Choi, I. S. Kim, W. Yu, H. S. Hwang, and B. S. Kim. Serum baff predicts prognosis better than april in diffuse large b-cell lymphoma patients treated with rituximab plus chop chemotherapy. *Eur J Haematol*, 81(3):177–184, Sep 2008.

[101] W. Klapper, H. Stoecklein, S. Zeynalova, G. Ott, F. Kosari, A. Rosenwald, M. Loeffler, L. Trümper, M. Pfreundschuh, R. Siebert, and G. H.-G. N.-H. L. S. Group. Structural aberrations affecting the myc locus indicate a poor prognosis independent of clinical risk factors in diffuse large b-cell lymphomas treated within randomized trials of the german high-grade non-hodgkin's lymphoma study group (dshnhl). *Leukemia*, 22(12):2226–2229, Dec 2008.

[102] E. Klein and G. Klein. Burkitt lymphoma. *Semin Cancer Biol*, 19(6):345–346, Dec 2009.

[103] U. Klein and R. Dalla-Favera. Germinal centres: role in b-cell physiology and malignancy. *Nat Rev Immunol*, 8(1):22–33, Jan 2008.

[104] D. Kleinbaum and M. Klein. Survival analysis, a self-learning text. *Survival analysis*, pages 55–96, 2012.

[105] D. Konforte, N. Simard, and C. J. Paige. Il-21: an executor of b cell fate. *J Immunol*, 182(4):1781–1787, Feb 2009.

[106] C. Koumenis, C. Naczki, M. Koritzinsky, S. Rastani, A. Diehl, N. Sonenberg, A. Koromilas, and B. G. Wouters. Regulation of protein synthesis by hypoxia via activation of the endoplasmic reticulum kinase perk and phosphorylation of the translation initiation factor eif2alpha. *Mol Cell Biol*, 22(21):7405–7416, Nov 2002.

[107] D. Krappmann, E. Wegener, Y. Sunami, M. Esen, A. Thiel, B. Mordmuller, and C. Scheidereit. The ikappab kinase complex and nf-kappab act as master regulators of lipopolysaccharide-induced gene expression and control subordinate activation of ap-1. *Mol Cell Biol*, 24(14):6488–6500, Jul 2004.

[108] R. Küppers. B cells under influence: transformation of b cells by epstein-barr virus. *Nat Rev Immunol*, 3(10):801–812, Oct 2003.

[109] R. Küppers. Mechanisms of b-cell lymphoma pathogenesis. *Nat Rev Cancer*, 5(4):251–262, Apr 2005.

[110] R. Küppers and R. Dalla-Favera. Mechanisms of chromosomal translocations in b cell lymphomas. *Oncogene*, 20(40):5580–5594, Sep 2001.

[111] M. Ladanyi, K. Offit, S. C. Jhanwar, D. A. Filippa, and R. S. Chaganti. Myc rearrangement and translocations involving band 8q24 in diffuse large cell lymphomas. *Blood*, 77(5):1057–1063, Mar 1991.

[112] L. T. Lam, G. Wright, R. E. Davis, G. Lenz, P. Farinha, L. Dang, J. W. Chan, A. Rosenwald, R. D. Gascoyne, and L. M. Staudt. Cooperative signaling through the signal transducer and activator of transcription 3 and nuclear factor-kappab pathways in subtypes of diffuse large b-cell lymphoma. *Blood*, 111(7):3701–3713, Apr 2008.

[113] D. Latchman. *Gene regulation: a eukaryotic perspective*. Taylor & Francis Group, 2005.

[114] S. Lederman, M. J. Yellin, A. Krichevsky, J. Belko, J. J. Lee, and L. Chess. Identification of a novel surface protein on activated cd4+ t cells that induces contact-dependent b cell differentiation (help). *J Exp Med*, 175(4):1091–1101, Apr 1992.

[115] J. Lee, S. Kotliarova, Y. Kotliarov, A. Li, Q. Su, N. M. Donin, S. Pastorino, B. W. Purow, N. Christopher, W. Zhang, J. K. Park, and H. A. Fine. Tumor stem cells derived from glioblastomas cultured in bfgf and egf more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell*, 9(5):391–403, May 2006.

[116] E. Leucci, M. Cocco, A. Onnis, G. D. Falco, P. van Cleef, C. Bellan, A. van Rijk, J. Nyagol, B. Byakika, S. Lazzi, P. Tosi, H. van Krieken, and L. Leoncini. Myc translocation-negative classical burkitt lymphoma cases: an alternative pathogenetic mechanism involving mirna deregulation. *J Pathol*, 216(4):440–450, Dec 2008.

[117] A. Li, J. Walling, S. Ahn, Y. Kotliarov, Q. Su, M. Quezado, J. C. Oberholtzer, J. Park, J. C. Zenklusen, and H. A. Fine. Unsupervised analysis of transcriptomic profiles reveals six glioma subtypes. *Cancer Res*, 69(5):2091–2099, Mar 2009.

[118] H. Liang, Q. Chen, A. H. Coles, S. J. Anderson, G. Pihan, A. Bradley, R. Gerstein, R. Jurecic, and S. N. Jones. Wnt5a inhibits b cell proliferation and functions as a tumor suppressor in hematopoietic tissue. *Cancer Cell*, 4(5):349–360, Nov 2003.

[119] R. M. Locksley, N. Killeen, and M. J. Lenardo. The tnf and tnf receptor superfamilies: integrating mammalian biology. *Cell*, 104(4):487–501, Feb 2001.

[120] C. Y. Logan and R. Nusse. The wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol*, 20:781–810, 2004.

[121] C. Lottaz, D. Beier, K. Meyer, P. Kumar, A. Hermann, J. Schwarz, M. Junker, P. J. Oefner, U. Bogdahn, J. Wischhusen, R. Spang, A. Storch, and C. P. Beier. Transcriptional profiles of cd133+ and cd133- glioblastoma-derived cancer stem cell lines suggest different cells of origin. *Cancer Res*, 70(5):2030–2040, Mar 2010.

[122] C. Lottaz, X. Yang, S. Scheid, and R. Spang. Orderedlist–a bioconductor package for detecting similarity in ordered gene lists. *Bioinformatics*, 22(18):2315–2316, Sep 2006.

[123] D. N. Louis. Molecular pathology of malignant gliomas. *Annu Rev Pathol*, 1:97–117, 2006.

[124] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues. The 2007 who classification of tumours of the central nervous system. *Acta Neuropathol*, 114(2):97–109, Aug 2007.

[125] D. Lu, J. X. Liu, T. Endo, H. Zhou, S. Yao, K. Willert, I. G. H. Schmidt-Wolf, T. J. Kipps, and D. A. Carson. Ethacrynic acid exhibits selective toxicity to chronic lymphocytic leukemia cells by inhibition of the wnt/beta-catenin pathway. *PLoS One*, 4(12):e8294, 2009.

[126] D. Lu, Y. Zhao, R. Tawatao, H. B. Cottam, M. Sen, L. M. Leoni, T. J. Kipps, M. Corr, and D. A. Carson. Activation of the wnt signaling pathway in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*, 101(9):3118–3123, Mar 2004.

[127] H.-I. Ma, C.-L. Kao, Y.-Y. Lee, G.-Y. Chiou, L.-K. Tai, K.-H. Lu, C.-S. Huang, Y.-W. Chen, S.-H. Chiou, I.-C. Cheng, and T.-T. Wong. Differential expression profiling between atypical teratoid/rhabdoid and medulloblastoma tumor in vitro and in vivo using microarray analysis. *Childs Nerv Syst*, 26(3):293–303, Mar 2010.

[128] M. Mabry, T. Nakagawa, B. D. Nelkin, E. McDowell, M. Gesell, J. C. Eggleston, R. A. Casero, and S. B. Baylin. v-ha-ras oncogene insertion: a model for tumor progression of human small cell lung cancer. *Proc Natl Acad Sci U S A*, 85(17):6523–6527, Sep 1988.

[129] F. Mackay, F. Sierro, S. T. Grey, and T. P. Gordon. The baff/april system: an important player in systemic rheumatic diseases. *Curr Dir Autoimmun*, 8:243–265, 2005.

[130] M. Maisel, A. Herr, J. Milosevic, A. Hermann, H.-J. Habisch, S. Schwarz, M. Kirsch, G. Antoniadis, R. Brenner, S. Hallmeyer-Elgner, H. Lerche, J. Schwarz, and A. Storch. Transcription profiling of adult and fetal human neuroprogenitors identifies divergent paths to maintain the neuroprogenitor cell state. *Stem Cells*, 25(5):1231–1240, May 2007.

[131] M. Maneck, A. Schrader, D. Kube, and R. Spang. Genomic data integration using guided clustering. *Bioinformatics*, 27(16):2231–2238, Aug 2011.

[132] S. Mathas. The pathogenesis of classical hodgkin's lymphoma: a model for b-cell plasticity. *Hematol Oncol Clin North Am*, 21(5):787–804, Oct 2007.

[133] G. McLachlan, K. Do, and C. Ambroise. *Analyzing microarray gene expression data*, volume 422. John Wiley & Sons, 2004.

[134] S. Monti, K. J. Savage, J. L. Kutok, F. Feuerhake, P. Kurtin, M. Mihm, B. Wu, L. Pasqualucci, D. Neuberg, R. C. T. Aguiar, P. D. Cin, C. Ladd, G. S. Pinkus, G. Salles, N. L. Harris, R. Dalla-Favera, T. M. Habermann, J. C. Aster, T. R. Golub, and M. A. Shipp. Molecular profiling of diffuse large b-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood*, 105(5):1851–1861, Mar 2005.

[135] S. Monti, P. Tamayo, J. Mesirov, and T. Golub. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning*, 52(1):91–118, 2003.

[136] J. A. Morrison, M. L. Gulley, R. Pathmanathan, and N. Raab-Traub. Differential signaling pathways are activated in the epstein-barr virus-associated malignancies nasopharyngeal carcinoma and hodgkin lymphoma. *Cancer Res*, 64(15):5251–5260, Aug 2004.

[137] H. Motohashi and M. Yamamoto. Nrf2-keap1 defines a physiologically important stress response mechanism. *Trends Mol Med*, 10(11):549–557, Nov 2004.

[138] H. Muaddi, M. Majumder, P. Peidis, A. I. Papadakis, M. Holcik, D. Scheuner, R. J. Kaufman, M. Hatzoglou, and A. E. Koromilas. Phosphorylation of eif2alpha at serine 51 is an important determinant of cell survival and adaptation to glucose deficiency. *Mol Biol Cell*, 21(18):3220–3231, Sep 2010.

[139] C. G. Mullighan, S. Goorha, I. Radtke, C. B. Miller, E. Coustan-Smith, J. D. Dalton, K. Girtman, S. Mathew, J. Ma, S. B. Pounds, X. Su, C.-H. Pui, M. V. Relling, W. E. Evans, S. A. Shurtleff, and J. R. Downing. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*, 446(7137):758–764, Apr 2007.

[140] A. Murat, E. Migliavacca, T. Gorlia, W. L. Lambiv, T. Shay, M.-F. Hamou, N. de Tribolet, L. Regli, W. Wick, M. C. M. Kouwenhoven, J. A. Hainfellner, F. L. Heppner, P.-Y. Dietrich, Y. Zimmer, J. G. Cairncross, R.-C. Janzer, E. Domany, M. Delorenzi, R. Stupp, and M. E. Hegi. Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. *J Clin Oncol*, 26(18):3015–3024, Jun 2008.

[141] M. Nakagawa, N. Takizawa, M. Narita, T. Ichisaka, and S. Yamanaka. Promotion of direct reprogramming by transformation-deficient myc. *Proc Natl Acad Sci U S A*, 107(32):14152–14157, Aug 2010.

[142] C. G. A. R. Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–1068, Oct 2008.

[143] V. N. Ngo, R. M. Young, R. Schmitz, S. Jhavar, W. Xiao, K.-H. Lim, H. Kohlhammer, W. Xu, Y. Yang, H. Zhao, A. L. Shaffer, P. Romesser, G. Wright, J. Powell, A. Rosenwald, H. K. Muller-Hermelink, G. Ott, R. D. Gascoyne, J. M. Connors, L. M. Rimsza, E. Campo, E. S. Jaffe, J. Delabie, E. B. Smeland, R. I. Fisher, R. M. Braziel, R. R. Tubbs, J. R. Cook, D. D. Weisenburger, W. C. Chan, and L. M. Staudt. Oncogenically active myd88 mutations in human lymphoma. *Nature*, 470(7332):115–119, Feb 2011.

[144] A. J. Novak, J. R. Darce, B. K. Arendt, B. Harder, K. Henderson, W. Kindsvogel, J. A. Gross, P. R. Greipp, and D. F. Jelinek. Expression of bcma, taci, and baff-r in multiple myeloma: a mechanism for growth and survival. *Blood*, 103(2):689–694, Jan 2004.

[145] A. J. Novak, D. M. Grote, M. Stenson, S. C. Ziesmer, T. E. Witzig, T. M. Habermann, B. Harder, K. M. Ristow, R. J. Bram, D. F. Jelinek, J. A. Gross, and S. M. Ansell. Expression of blys and its receptors in b-cell non-hodgkin lymphoma: correlation with disease activity and patient outcome. *Blood*, 104(8):2247–2253, Oct 2004.

[146] M. K. Nygren, G. Dsen, M. E. Hystad, H. Stubberud, S. Funderud, and E. Rian. Wnt3a activates canonical wnt signalling in acute lymphoblastic leukaemia (all) cells and inhibits the proliferation of b-all cell lines. *Br J Haematol*, 136(3):400–413, Feb 2007.

[147] R. Pallini, L. Ricci-Vitiani, G. L. Banna, M. Signore, D. Lombardi, M. Todaro, G. Stassi, M. Martini, G. Maira, L. M. Larocca, and R. D. Maria. Cancer stem cell analysis and clinical outcome in patients with glioblastoma multiforme. *Clin Cancer Res*, 14(24):8205–8212, Dec 2008.

[148] D. W. Parsons, S. Jones, X. Zhang, J. C.-H. Lin, R. J. Leary, P. Angenendt, P. Mankoo, H. Carter, I.-M. Siu, G. L. Gallia, A. Olivi, R. McLendon, B. A. Rasheed, S. Keir, T. Nikolskaya, Y. Nikolsky, D. A. Busam, H. Tekleab, L. A. Diaz, J. Hartigan, D. R. Smith, R. L. Strausberg, S. K. N. Marie, S. M. O. Shinjo, H. Yan, G. J. Riggins, D. D. Bigner, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V. E. Velculescu, and K. W. Kinzler. An integrated genomic analysis of human glioblastoma multiforme. *Science*, 321(5897):1807–1812, Sep 2008.

[149] A. S. Perkins and J. W. Friedberg. Burkitt lymphoma in adults. *Hematology Am Soc Hematol Educ Program*, pages 341–348, 2008.

[150] K. Petropoulos, N. Arseni, C. Schessl, C. R. Stadler, V. P. S. Rawat, A. J. Deshpande, B. Heilmeier, W. Hiddemann, L. Quintanilla-Martinez, S. K. Bohlander, M. Feuring-Buske, and C. Buske. A novel role for lef-1, a central transcription mediator of wnt signaling, in leukemogenesis. *J Exp Med*, 205(3):515–522, Mar 2008.

[151] S. Peuelas, J. Anido, R. M. Prieto-Snchez, G. Folch, I. Barba, I. Cuartas, D. Garca-Dorado, M. A. Poca, J. Sahuquillo, J. Baselga, and J. Seoane. Tgf-beta increases glioma-initiating cell self-renewal through the induction of lif in human glioblastoma. *Cancer Cell*, 15(4):315–327, Apr 2009.

[152] C. V. Pfenninger, T. Roschupkina, F. Hertwig, D. Kottwitz, E. Englund, J. Bengzon, S. E. Jacobsen, and U. A. Nuber. Cd133 is not present on neurogenic astrocytes in the adult subventricular zone, but on embryonic neural stem cells, ependymal cells, and glioblastoma cells. *Cancer Res*, 67(12):5727–5736, Jun 2007.

[153] L. V. Pham, A. T. Tamayo, L. C. Yoshimura, Y.-C. Lin-Lee, and R. J. Ford. Constitutive nf-kappab and nfat activation in aggressive b-cell lymphomas synergistically activates the cd154 gene and maintains lymphoma cell survival. *Blood*, 106(12):3940–3947, Dec 2005.

[154] H. S. Phillips, S. Kharbanda, R. Chen, W. F. Forrest, R. H. Soriano, T. D. Wu, A. Misra, J. M. Nigro, H. Colman, L. Soroceanu, P. M. Williams, Z. Modrusan, B. G. Feuerstein, and K. Aldape. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell*, 9(3):157–173, Mar 2006.

[155] S. G. M. Piccirillo, B. A. Reynolds, N. Zanetti, G. Lamorte, E. Binda, G. Broggi, H. Brem, A. Olivi, F. Dimeco, and A. L. Vescovi. Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature*, 444(7120):761–765, Dec 2006.

[156] K. H. Plate and W. Risau. Angiogenesis in malignant gliomas. *Glia*, 15(3):339–347, Nov 1995.

[157] S. M. Pollard, K. Yoshikawa, I. D. Clarke, D. Danovi, S. Stricker, R. Russell, J. Bayani, R. Head, M. Lee, M. Bernstein, J. A. Squire, A. Smith, and P. Dirks. Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell*, 4(6):568–580, Jun 2009.

[158] A. Poltorak, X. He, I. Smirnova, M. Y. Liu, C. V. Huffel, X. Du, D. Birdwell, E. Alejos, M. Silva, C. Galanos, M. Freudenberg, P. Ricciardi-Castagnoli, B. Layton, and B. Beutler. Defective lps signaling in c3h/hej and c57bl/10sccr mice: mutations in tlr4 gene. *Science*, 282(5396):2085–2088, Dec 1998.

[159] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. H. Kim, L. C. Goumnerova, P. M. Black, C. Lau, J. C. Allen, D. Zagzag, J. M. Olson, T. Curran, C. Wetmore, J. A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J. P. Mesirov, E. S. Lander, and T. R. Golub. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, 415(6870):436–442, Jan 2002.

[160] R. M. Prins, H. Soto, V. Konkankit, S. K. Odesa, A. Eskin, W. H. Yong, S. F. Nelson, and L. M. Liau. Gene expression profile correlates with t-cell infiltration and relative survival in glioblastoma patients vaccinated with dendritic cell immunotherapy. *Clin Cancer Res*, 17(6):1603–1615, Mar 2011.

[161] E. M. Plsson-McDermott and L. A. J. O'Neill. Signal transduction by the lipopolysaccharide receptor, toll-like receptor-4. *Immunology*, 113(2):153–162, Oct 2004.

[162] Y.-W. Qiang, Y. Endo, J. S. Rubin, and S. Rudikoff. Wnt signaling in b-cell neoplasia. *Oncogene*, 22(10):1536–1545, Mar 2003.

[163] R. Reece. *Analysis of genes and genomes*. John Wiley & Sons, 2004.

[164] T. Reya, M. O'Riordan, R. Okamura, E. Devaney, K. Willert, R. Nusse, and R. Grosschedl. Wnt signaling regulates b lymphocyte proliferation through a lef-1 dependent mechanism. *Immunity*, 13(1):15–24, Jul 2000.

[165] D. R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, and A. M. Chinnaiyan. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A*, 101(25):9309–9314, Jun 2004.

[166] S. J. Rodig, A. Shahsafaei, B. Li, C. R. Mackay, and D. M. Dorfman. Baff-r, the major b cell-activating factor receptor, is expressed on most mature b cells and b-cell lymphoproliferative disorders. *Hum Pathol*, 36(10):1113–1119, Oct 2005.

[167] J. Romn-Gmez, L. Cordeu, X. Agirre, A. Jimnez-Velasco, E. S. Jos-Eneriz, L. Garate, M. J. Calasanz, A. Heiniger, A. Torres, and F. Prosper. Epigenetic regulation of wnt-signaling pathway in acute lymphoblastic leukemia. *Blood*, 109(8):3462–3469, Apr 2007.

[168] A. Rosenwald, G. Wright, W. C. Chan, J. M. Connors, E. Campo, R. I. Fisher, R. D. Gascoyne, H. K. Muller-Hermelink, E. B. Smeland, J. M. Giltnane, E. M. Hurt, H. Zhao, L. Averett, L. Yang, W. H. Wilson, E. S. Jaffe, R. Simon, R. D. Klausner, J. Powell, P. L. Duffey, D. L. Longo, T. C. Greiner, D. D. Weisenburger, W. G. Sanger, B. J. Dave, J. C. Lynch, J. Vose, J. O. Armitage, E. Montserrat, A. Lpez-Guillermo, T. M. Grogan, T. P. Miller, M. LeBlanc, G. Ott, S. Kvaloy, J. Delabie, H. Holte, P. Krajci, T. Stokke, L. M. Staudt, and L. M. P. Project. The use of molecular profiling to predict survival after chemotherapy for diffuse large-b-cell lymphoma. *N Engl J Med*, 346(25):1937–1947, Jun 2002.

[169] D. M. Schewe and J. A. Aguirre-Ghiso. Inhibition of eif2alpha dephosphorylation maximizes bortezomib efficiency and eliminates quiescent multiple myeloma cells surviving proteasome inhibitor therapy. *Cancer Res*, 69(4):1545–1552, Feb 2009.

[170] A. Schrader, S. Bentink, R. Spang, D. Lenze, M. Hummel, M. Kuo, J. R. Arrand, P. G. Murray, L. Trümper, D. Kube, and M. Vockerodt. High myc activity is an independent negative prognostic factor for diffuse large b cell lymphomas. *Int J Cancer*, 131(4):E348–E361, Aug 2012.

[171] A. Schrader, K. Meyer, F. von Bonin, M. Vockerodt, N. Walther, E. Hand, A. Ulrich, D. L. Kamila Matulewicz, M. Hummel, A. Kieser, M. Engelke, L. Trmper, and D. Kube. Global gene expression changes of in vitro stimulated human transformed germinal centre b cells as surrogate for oncogenic pathway activation in individual aggressive b cell lymphomas. *Cell Communication and Signaling*, 2012. MEINS.

[172] J. A. Schwartzbaum, K. Huang, S. Lawler, B. Ding, J. Yu, and E. A. Chiocca. Allergy and inflammatory transcriptome is predominantly negatively correlated with cd133 expression in glioblastoma. *Neuro Oncol*, 12(4):320–327, Apr 2010.

[173] A. L. Shaffer, R. M. Young, and L. M. Staudt. Pathogenesis of human b cell lymphomas. *Annu Rev Immunol*, 30:565–610, 2012.

[174] S. K. Singh, C. Hawkins, I. D. Clarke, J. A. Squire, J. Bayani, T. Hide, R. M. Henkelman, M. D. Cusimano, and P. B. Dirks. Identification of human brain tumour initiating cells. *Nature*, 432(7015):396–401, Nov 2004.

[175] G. Smyth. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, 3:Article3, 2004.

[176] G. Smyth. Limma: linear models for microarray data. *Bioinformatics and computational biology solutions using R and Bioconductor*, pages 397–420, 2005.

[177] F. J. T. Staal and H. C. Clevers. Wnt signalling and haematopoiesis: a wnt-wnt situation. *Nat Rev Immunol*, 5(1):21–30, Jan 2005.

[178] F. J. T. Staal and T. C. Luis. Wnt signaling in hematopoiesis: crucial factors for self-renewal, proliferation, and cell fate decisions. *J Cell Biochem*, 109(5):844–849, Apr 2010.

[179] C. D. Stiles and D. H. Rowitch. Glioma stem cells: a midterm exam. *Neuron*, 58(6):832–846, Jun 2008.

[180] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–15550, Oct 2005.

[181] S. Swerdlow, E. Campo, N. Harris, E. Jaffe, S. Pileri, H. Stein, et al. Who classification of tumours of haematopoietic and lymphoid tissues (international agency for research on cancer, lyon, france). 2008.

[182] R. Team et al. R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna Austria*, (01/19), 2010.

[183] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A*, 99(10):6567–6572, May 2002.

[184] J. Tonn and M. Westphal. *Neuro-oncology of CNS tumors*. Springer, 2006.

[185] A. Travis, A. Amsterdam, C. Belanger, and R. Grosschedl. Lef-1, a gene encoding a lymphoid-specific protein with an hmg domain, regulates t-cell receptor alpha enhancer function [corrected]. *Genes Dev*, 5(5):880–894, May 1991.

[186] R. J. T. C. Trevor J. Hastie, Balasubramanian Narasimhan. Pam: prediction analysis of microarraysusers guide and manual.

[187] J. Tukey. Exploratory data analysis. *Reading, MA*, 1977.

[188] M. Uhl, S. Aulwurm, J. Wischhusen, M. Weiler, J. Y. Ma, R. Almirez, R. Mangadu, Y.-W. Liu, M. Platten, U. Herrlinger, A. Murphy, D. H. Wong, W. Wick, L. S. Higgins, and M. Weller. Sd-208, a novel transforming growth factor beta receptor i kinase inhibitor, inhibits growth and invasiveness and enhances immunogenicity of murine and human glioma cells in vitro and in vivo. *Cancer Res*, 64(21):7954–7961, Nov 2004.

[189] L. Uhrbom, C. Dai, J. C. Celestino, M. K. Rosenblum, G. N. Fuller, and E. C. Holland. Ink4a-arf loss cooperates with kras activation in astrocytes and neural progenitors to generate glioblastomas of various morphologies depending on activated akt. *Cancer Res*, 62(19):5551–5558, Oct 2002.

[190] A. Ulrich, N. Walthe, F. von Bonin, W. Klapper, K. Meyer, A. Schrader, M. Vockerodt, R. Spang, T. Pukrop, P. Murray, L. Trmper, and D. Kube. Aberrant lymphocyte enhancer-binding factor 1 expression is characteristic for sporadic burkitt lymphoma. *The American Journal of Pathology*, 2012. MEINS.

[191] A. Valencia, J. Romn-Gmez, J. Cervera, E. Such, E. Barragn, P. Bolufer, F. Moscard, G. F. Sanz, and M. A. Sanz. Wnt signaling pathway is epigenetically regulated by methylation of wnt antagonists in acute myeloid leukemia. *Leukemia*, 23(9):1658–1666, Sep 2009.

[192] R. G. W. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov, G. Alexe, M. Lawrence, M. O'Kelly, P. Tamayo, B. A. Weir, S. Gabriel, W. Winckler, S. Gupta, L. Jakkula, H. S. Feiler, J. G. Hodgson, C. D. James, J. N. Sarkaria, C. Brennan, A. Kahn, P. T. Spellman, R. K. Wilson, T. P. Speed, J. W. Gray, M. Meyerson, G. Getz, C. M. Perou, D. N. Hayes, and C. G. A. R. Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in pdgfra, idh1, egfr, and nf1. *Cancer Cell*, 17(1):98–110, Jan 2010.

[193] L. D. Wang and M. R. Clark. B-cell antigen-receptor signalling in lymphocyte development. *Immunology*, 110(4):411–420, Dec 2003.

[194] M. L. Waterman, W. H. Fischer, and K. A. Jones. A thymus-specific member of the hmg protein family regulates the human t cell receptor c alpha enhancer. *Genes Dev*, 5(4):656–669, Apr 1991.

[195] R. L. Wiseman and W. E. Balch. A new pharmacology–drugging stressed folding pathways. *Trends Mol Med*, 11(8):347–350, Aug 2005.

[196] D. Y. Wu, D. C. Tkachuck, R. S. Roberson, and W. H. Schubach. The human snf5/ini1 protein facilitates the function of the growth arrest and dna damage-inducible protein (gadd34) and modulates gadd34-bound protein phosphatase-1 activity. *J Biol Chem*, 277(31):27706–27715, Aug 2002.

[197] S. Yamagishi, Y. Koyama, T. Katayama, M. Taniguchi, J. Hitomi, M. Kato, M. Aoki, Y. Itoyama, S. Kato, and M. Tohyama. An in vitro model for lewy body-like hyaline inclusion/astrocytic hyaline inclusion: induction by er stress with an als-linked sod1 mutation. *PLoS One*, 2(10):e1030, 2007.

[198] J. Yang and M. Reth. The dissociation activation model of b cell antigen receptor triggering. *FEBS Lett*, 584(24):4872–4877, Dec 2010.

[199] X. Yang, S. Bentink, S. Scheid, and R. Spang. Similarities of ordered gene lists. *J Bioinform Comput Biol*, 4(3):693–708, Jun 2006.

[200] J. Ying, H. Li, J. Yu, K. M. Ng, F. F. Poon, S. C. C. Wong, A. T. C. Chan, J. J. Y. Sung, and Q. Tao. Wnt5a exhibits tumor-suppressive activity through antagonizing the wnt/beta-catenin signaling, and is frequently methylated in colorectal cancer. *Clin Cancer Res*, 14(1):55–61, Jan 2008.

[201] J. T. Yustein and C. V. Dang. Biology and treatment of burkitt's lymphoma. *Curr Opin Hematol*, 14(4):375–381, Jul 2007.

[202] R. Zeng, R. Spolski, E. Casas, W. Zhu, D. E. Levy, and W. J. Leonard. The molecular basis of il-21-mediated proliferation. *Blood*, 109(10):4135–4142, May 2007.

[203] E. Zucca, F. Bertoni, E. Roggero, G. Bosshard, G. Cazzaniga, E. Pedrinis, A. Biondi, and F. Cavalli. Molecular analysis of the progression from helicobacter pylori-associated chronic gastritis to mucosa-associated lymphoid-tissue lymphoma of the stomach. *N Engl J Med*, 338(12):804–810, Mar 1998.

List of Figures

Abbreviations

ABC	Activated B Cell
ALL	Actue Lymphatic Leukemia
AML	Acute Myeloid Leukemia
aNSC	adult Neural Stem Cells
ASR	Age-World-Standardized incidence Rate
AT/RT	Atypical Teratoid/Rhabdoid Tumor
BCR	B-Cell Receptor
BL	Burkitt's Lymphoma
CLL	Chronic Lymphatic Leukemia
CNC	Cap 'n' Collar
CNS	Central Nervous System
CSC	Cancer Stem Cells
DLBCL	Diffuse Large B Cell Lymphoma
EBV	Epstein-Barr-Virus
ER	Endoplasmic Reticulum
ES	Enrichment Score
ESC	Embryonic Stem Cell
FACS	Fluorecense Activated Cells Sorting
FDR	False discovery rate
FL	Follicular Lymphoma
fNSC	fetal Neural Stem Cells
GBM	Glioblastoma Multiform
GC	Germinal Center
GCB	Germinal Center B Cell
GO	Gene Ontology
GSEA	Gene Set Enrichment Analyses
HL	Hodgkin's Lymphoma
Ig	Immunoglobulin
IHC	Immunohistochemical

KEGG ..... Kyoto Encyclopedia of Genes and Genomes
KM ..... Kaplan-Meier
LEF1 ..... Lymphocyte Enhancer-binding Factor 1
LPS ..... Lipopolysaccharide
mBL ..... molecular Burkitt lymphoma
MCL ..... Mantle Cell Lymphoma
MHC ..... Major Histocompatibility Complex
MMML ..... Molecular Mechanisms in Malignant Lymphoma
MSC ..... Mesenchymal Stem Cells
MSigDB ..... Molecular Signature Database
NES ..... Normalized Enrichment Score
NHL ..... Non-Hodgkin's Lymphoma
NK cells ..... Natural Killer cells
PAM ..... Prediction Analysis of Microarrays
PBMC ..... Peripheral Blood Mononuclear Cells
PNET ..... Primitive Neuroectodermal Tumor
PRRs ..... Pattern Recognition Receptors
RNA ..... Ribonucleic Acid
SVZ ..... Sub-Ventricular Zone
TIC ..... Tumor Initiating Cells
TLR ..... Toll Like Receptor
TMA ..... Tissue-Micro-Arrays
UPR ..... Unfolded Protein Response
WHO ..... World Health Organization

Curriculum Vitae

Name: Katharina Meyer
Date of birth: 23/4/1983
Place of birth: Bonn (Germany)

Academic degree

Diploma biomathematics

Date:	October 2008
Supervisors:	Prof. Dr. Cieslik (Ernst Moritz Arndt University, Greifswald), Dr. Geue (Friedrich Löffler Institute)
Thesis:	Clusteranalyse der RFLP-Analyse-Daten von STEC-026 Stämmen

Education

- since February 2009: PhD student of the Institute of Functional Genomics, Department of Statistical Bioinformatics, University of Regensburg.
- October 2008: Diploma in Biomathematics from Ernst Moritz Arndt University, Greifswald.
- 2006: Semester abroad at the Massey University in Palmerston North, New Zealand.
- 2003-2008: Study of Biomathematics at the Ernst Moritz Arndt University, Greifswald.
- 2003: Abitur, Erich-Klausener-Gymnasium, Adenau

Publications

[1] C. Beier, P. Kumar, **K. Meyer**, P. Leukel, V. Bruttel, I. Aschenbrenner, M. Riemenschneider, A. Fragoulis, P. Rümmele, K. Lamszus, J. Schulz, J. Weis, U. Bogdahn, J. Wischhusen, P. Hau, R. Spang, and D. Beier. The cancer stem cell subtype determines immune infiltration of glioblastoma. *Stem Cells Development*, Jun 2012.

[2] L. Geue, S. Klare, C. Schnick, B. Mintel, **K. Meyer**, and F. Conraths. Analysis of the clonal relationship of serotype o26: H11 enterohemorrhagic escherichia coli isolates from cattle. *Applied and environmental microbiology*, 75(21):6947–6953, 2009.

[3] F. Hertwig, **K. Meyer**, S. Braun, S. Ek, R. Spang, C. Pfenninger, I. Artner, G. Prost, X. Chen, J. A. Biegel, A. Judkins, E. Englund, and U. Nuber. Definition of genetic events directing the development of distinct types of brain tumors from postnatal neural stem/progenitor cells. *Cancer Research*, 72(13):3381–3392, Jul 2012.

[4] C. Lottaz, D. Beier, **K. Meyer**, P. Kumar, A. Hermann, J. Schwarz, M. Junker, P. J. Oefner, U. Bogdahn, J. Wischhusen, R. Spang, A. Storch, and C. P Beier. Transcriptional profiles of cd133+ and cd133- glioblastoma-derived cancer stem cell lines suggest different cells of origin. *Cancer Research*, 70(5):2030–2040, Mar 2010.

[5] Richter, M. Schlesner, S. Hoffmann, M. Kreuz, E. Leich, B. Burkhardt, D. Lenz, M. Szczepanowski, O. Ammerpohl, S. H. Bernhart, A.Claviez, D.Hasenclever, V. Hovestadt, N. Hornig, J. Korbel, D. Kube, D. Langerberger, C. Lawerenz, J. Lisfeld, **K. Meyer**, S. Picelli, J.Pischimarov, B. Radlwimmer, T. Rausch, M. Rohde, M. Rosolowski, M. Schilhabel, R. Scholtysik, R. Spang, H. Trautmann, R. Wagener, T. Zenz, A. Borkhardt, H. Drexler, P. Müller, R. MacLeod, C. Pott, S. Schreiber, L. Trümper, M. Loeffler, P. Lichter, R. Eils, R. Küppers, M. Hummel, W. Klapper, P. Rosenstiel, A. Rosenwald, B. Brors, and R.Siebert. Recurrent mutation of the id3 gene in burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nature Genetics*, 2012.

[6] A. Schrader, **K. Meyer**, F. von Bonin, M. Vockerodt, N. Walther, E. Hand, A. Ulrich, K. Matulewicz, D. Lenze, M. Hummel, A. Kieser, M. Engelke, L. Trümper, and D. Kube. Global gene expression changes of in vitro stimulated human transformed germinal center b cells as surrogate for oncogenic pathway activation in individual lymphomas. *Cell Communication and Signaling*, 2012.

Acknowledgment

This work was carried out in the Department of Statistical Bioinformatics of the Institute of Functional Genomics at the University of Regensburg. I thank all past and present colleagues for the good working atmosphere and fruitful scientific and non-scientific discussions.

I am very grateful to my supervisors Rainer Spang, Wolfram Gronwald and Anja Bosserhoff for their constant advice. Especially, I thank Rainer Spang for giving me the opportunity to carry out my thesis in an interdisziplinary field, which allowed me to collaborate with experts from many different disciplines. The thesis would be impossible without collaboration, especially I want to thank Alexandra Schrader, Elisabeth Hand, Dieter Kube, Dagmar und Christoph Beier and Ulrike Nuber.

While working on this thesis I met many people who made contributions either direct or indirect. In particular I greatfully thank Daniela, Franzi, Julia, Eva and Christa. I also wish to thank my Kicker-group: Mohammad, Toni, Martin and both Christians for cheering me up after lunch.

Special thanks go to Daniela Schlüter (and family), Anna Hölting and Maike Heinrichs who read drafts of this thesis and greatly improved it by their comments.

Additionally I would like to thank my “Greifswalder”-family for encouraging me throughout the time it took to finish the presented work.
And last, but not least, I am more than grateful to my whole family, especially to my mum and my twin sister, you’re the best.

Eidesstattliche Erklärung

1. Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus anderen Quellen direkt oder indirekt bernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet.

2. Weitere Personen waren an der inhaltlich-materiellen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe eines Promotionsberaters oder anderer Personen in Anspruch genommen. Niemand hat von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

3. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Regensburg, den 04.02.2013 Katharina Meyer