

## Supplementary information

Cell cycle arrest but not apoptosis is associated with a common transcriptional pathway  
in glucocorticoid treatment of childhood lymphoblastic leukemia

Tatsiana Aneichyk<sup>1</sup>, Stefan Schmidt<sup>2</sup>, Daniel Bindreither<sup>1,3</sup>, Armin Kroesbacher<sup>1,3</sup>, Nikola  
S Mueller<sup>4</sup>, Bernhard Meister<sup>5</sup>, Roman Crazzolara<sup>5</sup>, Georg Mann<sup>6</sup>, Renate Panzer-  
Gruemayer<sup>6</sup>, Reinhard Kofler<sup>1,3\*</sup>, Johannes Rainer<sup>1,3\*</sup>, Stephan Geley<sup>1\*§</sup>

Running title: Glucocorticoids in acute lymphoblastic leukemia

<sup>1</sup>Division of Molecular Pathophysiology, Biocenter, Medical University of Innsbruck,  
Innsbruck, 6020, Austria; <sup>2</sup>University Hospital for Internal Medicine V, Department  
Internal Medicine, Medical University of Innsbruck, Innsbruck, 6020, Austria; <sup>3</sup>Tyrolean  
Cancer Research Institute, Innsbruck, 6020, Austria; <sup>4</sup>Institute of Computational Biology,  
Helmholtz-Zentrum Muenchen, Muenchen, 85764, Germany

<sup>5</sup>Department of Pediatrics, Medical University of Innsbruck, Innsbruck, 6020, Austria

<sup>6</sup>Children's Cancer Research Institute, St. Anna Hospital Vienna, Medical University of  
Vienna, Vienna, 1010, Austria; \*These authors contributed equally to this work

<sup>§</sup>Correspondence: Stephan Geley, Division of Molecular Pathophysiology, Biocenter,  
Medical University of Innsbruck, Innrain 80/82, Innsbruck, Tyrol A-6020, Austria.

Phone: +43 (0) 512 9003 70370

Fax: +43 (0) 512 9003 73960

E-mail: Stephan.Geley@i-med.ac.at

## **Table of Contents**

1. Extended Material and Methods
2. Supplementary Figures and Legends to Supplementary Figures
3. List of supplementary Tables
4. Links to additional data available through the Internet
5. References

### **1. Extended Material and Methods**

#### **Data preprocessing**

Microarray raw data were preprocessed using frozen GCRMA – a combination of two preprocessing methods: GCRMA (1) and frozen RMA (2). Reference preprocessing parameters including those required for non-specific binding adjustment were generated from 156 high quality microarrays (3) (and unpublished). For more detailed description of frozen GCRMA and the underlying methodology see Bindreither (2014) (4). Preprocessing of the microarrays was done using Bioconductor (5) packages ‘GCRMA’ and ‘frma’.

#### **Patients ALL subtype classification and GC-response quantification**

Six molecular signatures comprised of probe sets detecting highly subtype specific sets of mRNAs, were generated based on a large publicly available leukemia patient data set, consisting of 693 ALL samples (6). Six hundred sixty six high quality arrays of this data set were utilized for signature generation. Thereby two state of the art machine-learning

algorithms, ‘Random Forest’ and ‘Nearest Shrunken Centroids’ were employed. The newly generated signatures were benchmarked to already published signatures using another microarray data set obtained from a different microarray platform (7). The data driven classification of patients was done by hierarchical cluster analysis based on their zero hour gene expression profiles. A detailed description of signature generation and patient classification can be found in Bindreither (2014) (4).

### Differential gene expression and GO analyses

For the identification of the common transcriptional response to GCs, a linear model including a factor for the time point was used and fitted to the M-values for each gene. The resulting coefficients and p-values for the 6 and 24 hour time point thus represent the average differential expression value (across all patients) and the significance thereof. For the subtype-specific response a linear model was defined that included a factor combining the patients’ subtype and samples’ time point resulting in per-subtype estimates of differential gene expression.

### Regression models

Two different regression models were used to evaluate the relationship between gene expression/regulation with response to GC\_treatment as measured by AUC after 72h. We will refer to them as “simple regression” and “extended regression”.

The simple regression model was defined as:

$$EM = \beta_0 + \beta_{AUC} AUC_{72} \quad (2)$$

where:

$EM$  - expression or regulation of the gene,

$AUC_{72}$  – area under the response curve representing the decrease in lymphoblast counts within the first 72h of GC therapy.

Interpretation of evaluated parameters:

$\beta_0$  - intercept, representing average expression/regulation of the gene when  $AUC = 0$  (no response to GC treatment).

$\beta_{AUC}$  - effect of the  $\log_2$ -fold change in AUC on average gene expression/regulation.

The model was fitted to the data for each gene and p-values for each coefficient were adjusted for multiple hypothesis testing using the method from Benjamini and Hochberg (8).

The extended regression model for gene expression/regulation was defined as:

$$EM = \beta_0 + \beta_{T_6}T_6 + \beta_{T_{24}}T_{24} + \beta_{T_{ALL}}S_{T_{ALL}} + \beta_{ETV/RX}S_{ETV/RX} + \beta_{HYPER}S_{Hyper} + \beta_G G_m + \beta_{AUC}AUC_{72} \quad (1)$$

where

$EM$  - expression or regulation of the gene,

$T_6, T_{24}$  - factorial variable, indicating if the observation was made at 6-8 or 24h. The model for gene regulation has only a term with  $T_{24}$ , as  $T_6$  is used as a base level of regulation.

$S_{T_{ALL}}, S_{ETV/RX}, S_{Hyper}$  - factorial variable specifying the patient's subtype (T-ALL, ETV6/RUNX1, hyperdiploid or "other preB"),

$G_m$  - factorial variable indicating whether the observation was made in a male patient,

$AUC_{72}$  – area under the response curve quantifying the decrease in lymphoblast counts recorded in the first 72h of GC therapy.

Interpretation of evaluated parameters:

$\beta_0$  - intercept, representing the average expression/regulation of the gene at time point 0 (time point 6-8h for regulation) in female patients of the “other preB-ALL” subtype with AUC = 0 (no response to treatment).

$\beta_{T_6}$ ,  $\beta_{T_{24}}$  - change of the gene’s average expression/regulation after 6-8 (only for expression) and 24h of GC treatment, respectively.

$\beta_{TALL}$ ,  $\beta_{ETV/RX}$ ,  $\beta_{hyper}$  – difference of the gene’s average expression/regulation in TALL, ETV6/RUNX1 or hyperdiploid subtype to the base level (i.e. expression in “other preB-ALL” subtype).

$\beta_G$  - difference of average expression/regulation levels in male compared to female patients.

$\beta_{AUC}$  - effect of log2-fold change in AUC on average gene expression/regulation.

The above model was fitted for each gene and the resulting p-values for the coefficients were adjusted for multiple hypothesis testing using the method from Benjamini and Hochberg (8).

### Feature selection: elastic-net regression

For elastic-net regression we combined expression and regulation estimates of pre-selected probe sets in order to identify a possible combination of genes that may act together in order to facilitate GC-response in patients. First, the gene expression was adjusted for subtype, gender and time point by fitting a regression model:

$$E = \beta_0 + \beta_{T_6}T_6 + \beta_{T_{24}}T_{24} + \beta_{TALL}S_{TALL} + \quad (3)$$

$$+\beta_{ETV/RX}S_{ETV/RX} + \beta_{HYPER}S_{Hyper} + \beta_G G_m,$$

and then adjusting the observed expression values using fitted values:

$$E^{adj} = E^{obs} - E^{fit} \quad (4)$$

Pre-selection of the probe sets was based on median absolute deviation (MAD) and absolute value for estimated coefficient  $\beta_{AUC}$ . A probe set was considered for the elastic net analysis, if its MAD was higher than 50% of all probe sets and its  $|\beta_{AUC}|$  was within the top 5%.

Elastic net regression was applied to the combined data set of adjusted average expression values and average regulation values of approximately 1000 probe sets each using the R package glmnet (9) with the elastic net mixing parameter  $\alpha$  set to 0.1.

The regularization parameter  $\lambda$  was estimated using a 10x10-fold cross validation selecting the average value within 1 standard error from the minimum of the cross-validation error curve in the direction of increased regularization.

To ensure a robust analysis, the above approach was performed with the following bootstrapping: in each of in total 100 iterations, 90% of all patients were randomly selected and the above described regression models, feature pre-selection and elastic net regression was applied on the data of this subset. Probe sets with non-zero coefficients in more than 50 iterations were reported as significant probe sets.

## **2. Supplementary Figures and Legends to Supplementary Figures**

### **Supplementary Figure S1: Examples of response curves**

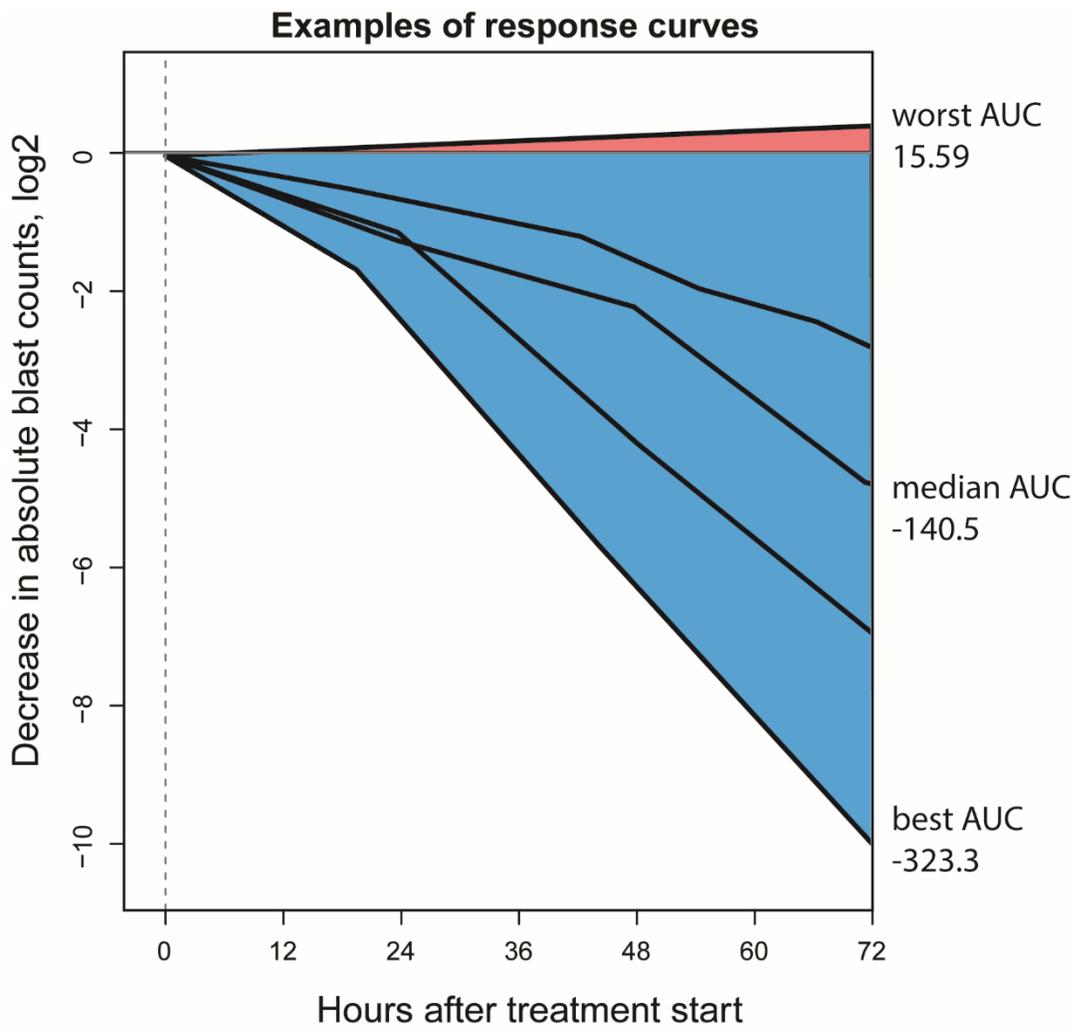
The graph depicts 5 response curves of individual children (see Supplementary Table S1). On the y-axis, the decrease in absolute blast relative to the blast counts at initiation of treatment counts is plotted against different time point during the first 72h of systemic GC monotherapy. The area under the curve (AUC) was calculated. One patient showed an increase in blast during the first 72h, hence a positive AUC (indicated as red area), all others (like the 4 depicted in the graph) showed decreases in blast counts, hence negative AUC values (green areas).

### **Supplementary Figure S2: Volcano plot 24h**

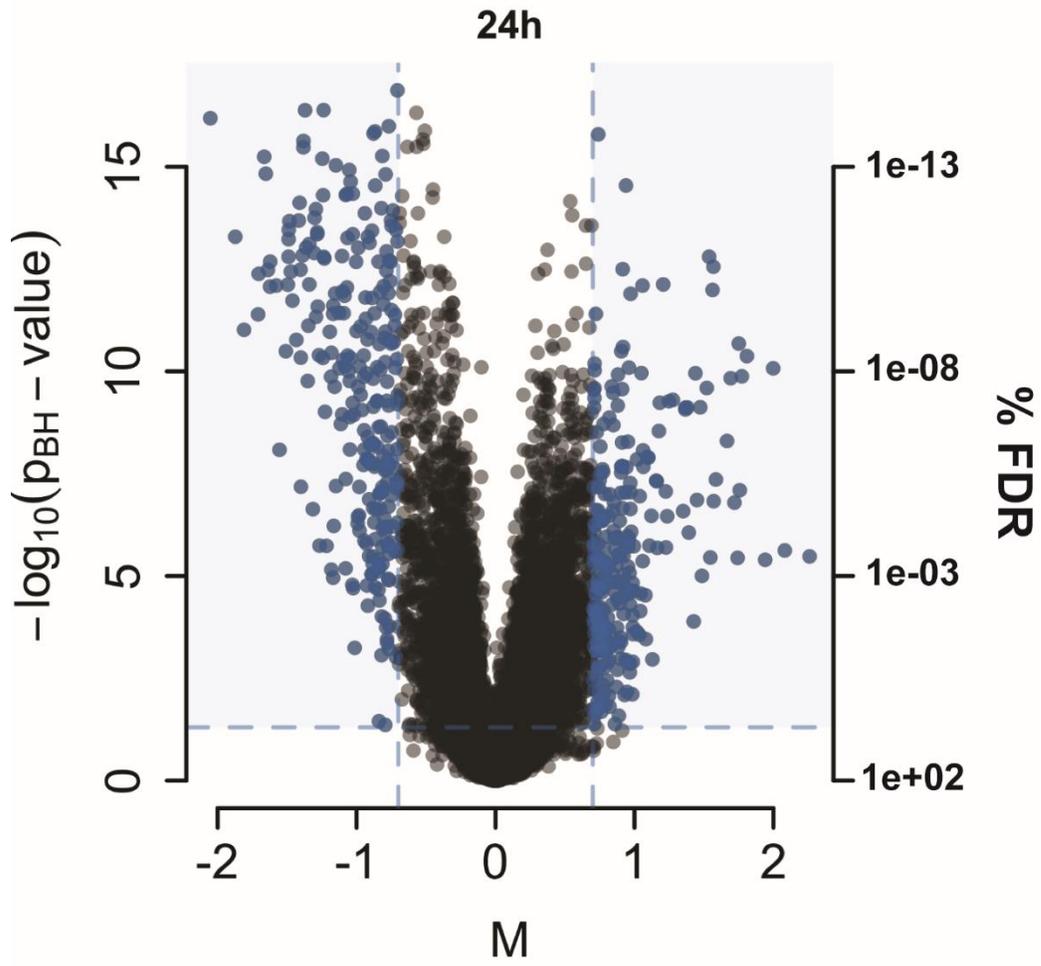
A “Volcano plot” showing the mean M values of all 46 patients GC-treated for 24h (M) and the Benjamini-Hochberg adjusted p values (left scale in log10) or percent false discovery rate (right scale in log2). The corresponding graph for 6-8h is depicted in Figure 2B in the main text.

### **Supplementary Figure S3: Hierarchical clustering of GC response genes with GBSs**

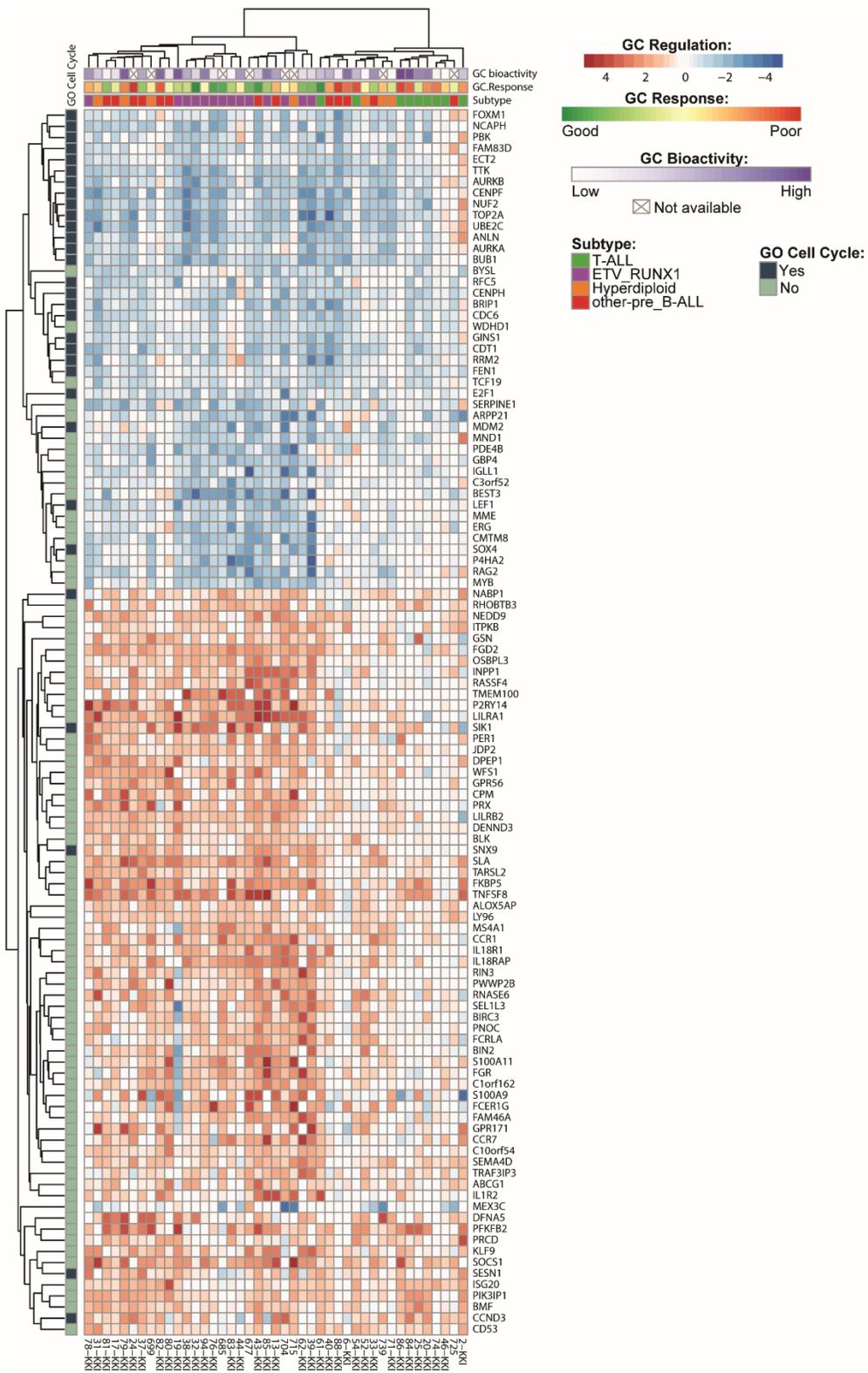
A “Heat map” depicting the 110 genes with GBS regulated after 24h systemic GC monotherapy. A corresponding heat-map for 6-8h is shown in Figure 3 in the main text, the values for the individual genes are summarized in Supplementary Table S2A.



Supplementary  
Figure S1



SUPPLEMENTARY  
FIGURE S2



SUPPLEMENTARY FIGURE S3

### **3. List of Supplementary Tables**

All 6 Supplementary Tables are supplied as separate searchable and sortable xlsx-files with their legends presented at the top or the end of each table.

Supplementary Table 1: Patient's characteristics

Supplementary Table 2A: 2307 "GC-regulated" probe sets in vivo

Table 2B: M-values of the 2307 "GC-regulated" probe sets of all 46 patients

Supplementary Table 3: Gene ontology analysis

Supplementary Table 4: Extended linear regression (to identify individual "effectors")

Supplementary Table 5: Ordinary least squares regression (to identify individual "modulators")

Supplementary Table 6: 1283 probe sets selected by the "Elastic Net" approach

### **4. Links to additional data available through the Internet**

Raw and preprocessed expression data has been deposited at the Gene Expression Omnibus (Accession number: GSE73578); during review process the data is accessible using the link:

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=sfktcoyghzavrup&acc=GSE73578>

8.

## 5. References

- (1) Wu Z, Irizarry RA, Gentleman RC, Martinez Murillo F, Spencer F. A model based background adjustment for oligonucleotide expression arrays. Internet 2004;<http://www.bepress.com/jhubiostat/paper1/>.
- (2) McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). *Biostatistics* 2010 Apr;11:242-53.
- (3) Schmidt S, Rainer J, Riml S, Ploner C, Jesacher S, Achmüller C, et al. Identification of glucocorticoid response genes in children with acute lymphoblastic leukemia. *Blood* 2006 Mar 1;107:2061-9.
- (4) Bindreither D. Advanced bioinformatic analysis of mRNA expression profiles with special focus on childhood acute lymphoblastic leukemia cells and their response to glucocorticoids 2014. PhD Thesis, Innsbruck Medical University, Innsbruck, Austria.
- (5) Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
- (6) Haferlach T, Kohlmann A, Wieczorek L, Basso G, Kronnie GT, Bene MC, et al. Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. *J Clin Oncol* 2010 May 20;28:2529-37.
- (7) Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK, et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* 2003 Nov 15;102:2951-9.
- (8) Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Statist Soc* 2015;57:289-300.
- (9) Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33:1-22.