Energy metabolism is co-determined by genetic variants in chronic lymphocytic leukemia and influences drug sensitivity

by Junyan Lu, Martin Böttcher, Tatjana Walther, Dimitrios Mougiakakos, Thorsten Zenz, and Wolfgang Huber

Energy metabolism is co-determined by genetic variants in chronic lymphocytic leukemia and influences drug sensitivity

Junyan Lu[1], Martin Böttcher[2], Tatjana Walther[3], Dimitrios Mougiakakos[2], Thorsten Zenz[3,4] and Wolfgang Huber[1]

[1]European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

[2]Department of Internal Medicine 5, Hematology and Oncology, University of Erlangen-Nuremberg, Erlangen, Germany

[3]Molecular Therapy in Hematology and Oncology, National Center for Tumor Diseases and German Cancer Research Centre, Heidelberg, Germany

[4]Department of Medical Oncology and Hematology, University Hospital Zürich and University of Zürich, Zürich, Switzerland

JL and MB contributed equally to this work.

Running heads: Characterizing energy metabolism of CLL

Correspondence:

Wolfgang Huber, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. Phone: +49 6221 3878823; Email: wolfgang.huber@embl.de.

Thorsten Zenz, University Hospital Zurich, Department of Hematology, Rämistrasse 100, CH-8091 Zürich, Switzerland. Phone: 41.44.255.9469; Email: thorsten.zenz@usz.ch.

Dimitrios Mougiakakos, Department of Internal Medicine 5, Hematology and Oncology, University of Erlangen/Nuremberg, Ulmenweg 18, Erlangen, 90154 Germany; e-mail: dimitrios.mougiakakos@uk-erlangen.de.

Abstract

Chronic lymphocytic leukemia cells have an altered energy metabolism compared to normal B cells. While there is growing understanding of the molecular heterogeneity of the disease, the extent of metabolic heterogeneity and its relation to molecular heterogeneity has not been systematically studied. Here, we assessed 11 bioenergetic features, primarily reflecting cell's oxidative phosphorylation and glycolytic activity, in leukemic cells from 140 chronic lymphocytic leukemia patients using metabolic flux analysis. We surveyed these bioenergetic features for relationships with molecular profiles (including genetic aberrations, transcriptome and methylome profiles) of the tumors, their *ex vivo* responses to a panel of 63 compounds, and with clinical data.

We observed that leukemic cells with mutated immunoglobulin variable heavy-chain show significantly lower glycolytic activity than cells with unmutated immunoglobulin variable heavy-chain. Accordingly, several key glycolytic genes (PFKP, PGAM1 and PGK1) were found to be downregulated in samples harboring mutated immunoglobulin variable heavy-chain. In addition, 8q24 copy number gains, 8p12 deletions, 13q14 deletions and ATM mutations were identified as determinants of cellular respiration. The metabolic state of leukemic cells was associated with drug sensitivity, in particular, higher glycolytic activity was linked to increased resistance towards several drugs including rotenone, navitoclax, and

orlistat. In addition, we found glycolytic capacity and glycolytic reserve to be predictors of overall survival ($P < 0.05$) independent of established genetic predictors.

Taken together, our study shows that heterogeneity in the energy metabolism of chronic lymphocytic leukemia cells is influenced by genetic variants and may be exploitable for the choice of therapeutic strategies.

Introduction

Resistance to apoptosis rather than aberrant proliferation is regarded as the reason for CLL cell accumulation. However, active proliferation also contributes to CLL pathogenesis, as sizable clonal birth rates were observed in CLL[1,2]. This suggests a substantial bioenergetic demand of proliferating subsets of CLL cells in order to support cell growth and division. Deregulated energy metabolism is considered as one of the cancer hallmarks[3]. While molecular mechanisms promoting survival and proliferation of CLL cells have been extensively studied, fewer studies have addressed energy metabolism in CLL. Garcia-Manteiga et al. suggested oxidative phosphorylation as the primary source of energy[4]. This hypothesis is supported by subsequent findings that aerobic mitochondrial respiration results in high levels of oxidative stress of circulating CLL cells[5] and that targeting the respiratory machinery can be therapeutically exploited to achieve selective toxicity[6]. However, MacIntyre et al. reported increased concentrations of pyruvate and glutamate in serum samples from CLL patients as compared to healthy donors, which suggests active glycolysis[7].

It is well established that genetic heterogeneity contributes to the variable clinical outcomes of CLL. Based on the somatic mutation status in the variable regions of the immunoglobulin (Ig) heavy chain (IGHV) genes, CLL can be divided into two subgroups with distinct prognosis: CLL cells with unmutated IGHV genes (U-CLL) display higher BCR signaling activity and are more aggressive than CLL cells with mutated IGHV genes (M-CLL). Serum samples from U-CLL patients were found to contain higher levels of lactate, fumarate, and uridine than those from M-CLL patients[7], suggesting U-CLL cells might have higher rates of aerobic glycolysis. This finding is in line with the observation that normal B cells undergo a metabolic switch from oxidative phosphorylation towards glycolysis upon BCR stimulation[4]. However, considering the number of clinically relevant genetic alterations documented in CLL[8,9], the relationship between genetic heterogeneity and energy metabolism is largely

4

unexplored. Our previous work showed that many of the recurrent mutations influence drug sensitivities of CLL[10]. As metabolic reprogramming has been shown to affect drug responsiveness of various cancers[2,11,12], metabolism may serve as a promising target for overcoming drug resistance in CLL.

To gain a better understanding of the metabolic landscape of CLL tumor cells in relation to their genetic profile and to determine the role of metabolism in the response to drug treatments, we assessed the bioenergetic features of primary CLL samples (n = 140 patients) through extracellular flux assays investigating two major metabolic processes, namely aerobic glycolysis and oxidative phosphorylation. We performed an integrative analysis of these data with previously recorded *ex vivo* responses of the same samples to a panel of 63 drugs, somatic genome mutations, tumor transcriptomes, DNA methylomes, and clinical data[10]. We found multiple associations between the mutational status and bioenergetic features and found glycolysis activity of CLL cells to contribute to resistance towards compounds targeting mitochondria-related biological processes that include rotenone, orlistat, venetoclax, and navitoclax. In addition, glycolytic capacity and glycolytic reserve features were shown to provide additional information to known genomic markers, such as IGHV and TP53, for predicting overall survival.

Methods

Extracellular flux assays

Extracellular Flux analyses (illustrated in Supplementary Figure S1) were performed on 152 CLL samples and 9 B cell samples from healthy donors on a Seahorse XFe96 system as previously described[13]. The resulting data files (*.asyr) were converted to comma-separated values (CSV) files using the Wave Desktop software package (Agilent / Seahorse Bioscience) and imported into R for quality assessment and further analysis. The data for 140 of the 152 CLL samples passed quality control and were used for subsequent analyses. A detailed description of the workflow and criteria for quality control are described in the Supplementary Methods.

Integrative data analysis

Analyses were performed using R 3.4 and included univariate association tests, multivariate regression with and without lasso penalization, Cox regression, generalized linear models, principal component analysis and gene set enrichment analysis. For association tests between bioenergetic features and genetic variants (i.e., copy number variants and gene mutations), only those with five or more variant cases were included. Summary statistics of patients' demographic and clinical features are provided in Table S1. All p-values from association tests were adjusted for multiple testing by applying the Benjamini-Hochberg procedure to control false discovery rate (FDR). Further details are provided in Supplementary Methods.

Data availability

Our data and analysis are provided as a reader-reproducible pipeline supported by the R package seahorseCLL (https://github.com/lujunyan1118/seahorseCLL). A R Shiny based

online platform (http://mozi.embl.de/public/seahorseCLL) is also provided for querying and visualizing our dataset.

Study approval

The study was approved by the Ethics Committee Heidelberg (University of Heidelberg, Germany; S-206/2011; S-356/2013). Patients who donated tumor material provided written informed consent prior to study.

Results

CLL cells and B cells show distinct energy metabolic phenotypes

We first compared the energy metabolic profiles of the 140 CLL samples and 9 B cell samples from healthy donors. In a principal component analysis (PCA, Figure 1A), the CLL samples were clearly separated from the B cell samples, which indicates that CLL cells have a distinct metabolic phenotype. Nine of the 11 bioenergetic features showed altered levels between CLL cells and B cells (ANOVA test, multiple testing method of Benjamini and Hochberg for FDR = 5%; Table S2). In accordance with a previous report[6], mitochondrial respiration-related features, including basal respiration, maximal respiration, and ATP production were increased in CLL cells (Figure 1B).

With regard to aerobic glycolysis, the basal glycolysis activity did not show significant differences between CLL and B cells. However, CLL cells showed elevated glycolytic capacity and glycolytic reserve (Figure 1B). As those two features measure the maximum capability of cells for glycolysis and the flexibility of cells to respond to energetic demands, this observation suggests an increased adaptability of CLL cells to use glycolysis as an energy source when needed, although they do not primarily rely on it.

Molecular determinants of energy metabolism in CLL

Figure 1 shows a variability among the bioenergetic profiles of the CLL samples. We hypothesized that this variability may be related to the molecular heterogeneity of CLL[8,9]. Therefore, we tested the tumor-to-tumor variations of the bioenergetic features for possible correlations with 20 molecular features, including recurrent somatic mutations and copy number variations, IGHV status and methylation clusters (Figure 2A and Supplementary Figure S2).

The most prominent association identified was the IGHV status: IGHV mutated CLL (M-CLL) samples had lower glycolytic activity and glycolytic capacity than IGHV unmutated CLL (U-CLL) samples (Figure 2B). Patients with M-CLL and U-CLL have been observed to have distinct serum metabolite profiles, with U-CLL patients having higher lactate level in serum, which can be considered a sign of elevated glycolysis[7]. To our knowledge, our large sample size study provides the first direct proof that U-CLLs indeed have a higher glycolytic activity than M-CLLs. IGHV status is strongly associated with three subtypes of CLL defined by their global levels of CpG methylation[14]. Accordingly, we found that the high-programmed CLL (HP-CLL) subtype, which has higher global methylation level, had a lower glycolysis activity than the low-programmed CLL (LP-CLL) subtype (Figure 2C).

To further dissect the role of IGHV status in metabolic reprogramming, we analyzed transcriptome data that we had measured for 120 of these patient samples (of which 111 had annotation for IGHV status). We performed gene set enrichment analysis on the genes that were differentially expressed between M-CLL and U-CLL samples using the Hallmark gene sets from Molecular Signature Database (MSigDB)[15]. We found that genes down-regulated in M-CLL were enriched in the glycolysis pathway (Figure 3A). Thirty-four glycolysis-related

genes were down-regulated in M-CLL (Figure 3B), including several that encode key enzymes: PFKP (Phosphofructokinase, platelet), PGAM1 (Phosphoglycerate Mutase 1), and PGK1 (Phosphoglycerate kinase 1) (Figure 3C)[16–18]. This analysis suggests that IGHV status directly influences the expression of genes related to glycolysis resulting in the observed difference of glycolytic parameters between M-CLLs and U-CLLs. As IGHV status reflects the B-cell receptor (BCR) signaling activity[19], we queried two published datasets for the transcriptomic signatures of BCR stimulation in CLLs either by IgM[20] (GEO ID: GSE49695) or unmethylated bacterial DNA (CpG) (GEO ID: GSE30105). In both conditions, genes that were up-regulated after BCR stimulation were significantly enriched in glycolysis pathway (Supplementary Figure S3). Together these results indicate a causal link from BCR signaling to glycolysis activity in CLL; such a relation is also in line with previous evidence[21,22].

We also identified several other novel associations between bioenergetic features and genetic variants (Supplementary Figure S4). Gain of 8q24, deletion of 8p12, ATM mutation, EGR2 mutation and MED12 mutation were found to be associated with higher values of respiration-related features such as ATP production and maximal respiration, while tumors with chromothripsis showed lower OCR values.

Glycolytic activity contributes to drug resistance in CLL

Sensitivity to drugs is an informative cellular phenotype that reflects pathway dependencies of tumor cells[10] and therefore we asked how the 11 intrinsic bioenergetic features were related to the vulnerabilities of CLL cells towards a panel of 63 drugs applied *ex vivo*. This panel comprised clinically used drugs as well as small molecule probes of pathways important in leukemia. Using the Pearson correlation test, we identified 118 significant associations (FDR = 10%) between drug sensitivities and bioenergetic features (Figure 4A and Supplementary

Figure S5). Thirty-two drugs had at least one significant association with a bioenergetic feature. A significant association between a bioenergetic feature and an *ex vivo* drug response indicates that the sensitivity or resistance of CLL samples to the drug is affected by the intrinsic activity of the bioenergetic feature.

At an aggregate level, glycolysis-related features of the CLL cells were positively correlated with the viabilities of those cells after drug treatment, while respiration-related features were negatively correlated. This suggests that higher glycolysis activity of CLL cells reduces sensitivity to drugs, while higher respiration activity contributes to increased sensitivity ex vivo.

There were more specific patterns for drugs with different target profiles. CLL samples with higher respiration activity were more sensitive to kinase inhibitors, including the inhibitors of Bruton's tyrosine kinase (BTK), ibrutinib, and of spleen tyrosine kinase, tamatinib, both of which target the BCR pathway. In addition, two checkpoint kinase 1 (Chk1) inhibitors, AZD7762 and PF-477736, and the heat shock protein 90 (Hsp90) inhibitor AT13387 showed similar association patterns, which is in line with the report that they also target the BCR signaling cascade[10].

Viabilities after treatment of drugs targeting mitochondria-related biological processes (rotenone, venetoclax and navitoclax) were positively correlated with the glycolysis-related features (Figure 4A, Supplementary Figure S6) for most of the drug concentrations (Supplementary Figure S5); this finding is not merely due to confounding by IGHV status as shown by the multivariate test results (Supplementary Figure S7). Rotenone is a mitochondrial complex I inhibitor, which disrupts the electron transport chain and thus blocks cellular respiration. Therefore, the correlation between rotenone response and glycolysis activity can be explained by the fact that higher glycolysis activity or potential (with increased metabolic flexibility) can compensate for cytotoxic effects of respiration inhibition

by providing an alternative way of producing ATP. Venetoclax and navitoclax are BH3-mimetics that target the BCL2 protein and lead to mitochondrial damage and the inhibition of oxidative respiration[23]. Thus, lower reliance on oxidative respiration is a plausible explanation for the resistance to BH3-mimetics of CLL cells with high glycolysis activity. We also observed associations between glycolysis-related features and the responses to orlistat, an anti-obesity drug, which has also been identified as a pro-apoptotic agent in CLL by inhibiting lipoprotein lipase (LPL)[24], and KX2-391, an inhibitor of the proto-oncogene tyrosine-protein kinase Src (Supplementary Figure S6).

We previously showed that although drug response phenotypes of CLL cells were largely influenced by genetic variants, there was still substantial variance in the drug response phenotypes that were not explained by genetics. Thus, we asked whether the energy metabolism profile could add additional predictive information. For each drug, we built two multivariate linear regression models to predict its response profile: one included only the 20 genetic features shown in Supplementary Figure S2 as predictors, the other included these genetic features plus 11 bioenergetic features. As a measure of predictive strength, we compared the variance explained (R^2 value adjusted by numbers of predictors) between the two models. As shown in Figure 4B, for most drugs, including bioenergetic features in the model did not increase explanatory power (dots on diagonal line); moreover, responses to individual kinase inhibitors were well explained by the genetic features (blue dots in Figure 4B and Supplementary Figure S8). However, for five drugs, including venetoclax and rotenone, the variance explained increased by 10% or more upon inclusion of the bioenergetic features (red dots in Figure 4B). In addition, except for cephaeline, bioenergetic features were more significant than genetic features in the multivariate models (Figure 4C).

Association between clinical course and energy metabolism of CLL

11

The use of primary patient cells enabled us to investigate the associations between bioenergetic features with patient history or outcome in CLL. In our study cohort, 43 patients had received treatment before sample collection, in all cases with chemotherapeutic agents (Table S1), and none of them was undergoing treatment when samples were collected. Therefore, we first asked whether these completed treatments prior to sample collection affected the energy metabolism of primary tumor samples, as studies have shown chemotherapy or targeted therapy could drive clonal evolution leading to drug resistance or oxidative stress[25–27]. We found two bioenergetic features, namely glycolytic capacity and glycolytic reserve, associated with pretreatment status at a significance threshold of $P < 0.05$ (Table S3, Supplementary Figure S9). However, pretreatment status was also highly correlated with IGHV ($P = 0.0006$, Chi-square test). This reflects the fact that U-CLL patients more frequently receive treatment due to faster progression. Furthermore, glycolytic capacity and reserve are correlated with IGHV status based on our above analysis. Thus, to dissect confounding from more direct association, we included IGHV status as a blocking factor in a multivariate model. In this more careful analysis, no significant association between pretreatment status and bioenergetic features was detected ($P < 0.05$). In a second analysis to assess potential roles of pretreatment status on the biology of the tumor samples, we revisited our association tests between the bioenergetic features and (i) the genetic variants and (ii) the drug responses. Including pretreatment status as a blocking factor had negligible impact on directions, strengths and p-values of these associations (Supplementary Figure S10). Together, these results indicate that the treatments experienced by 43 of our patients led to no detectable differences between the metabolic phenotypes of their circulating CLL cell samples and those of the 97 other patients. Therefore, we proceeded with the subsequent analysis using the combined dataset of 140 samples.

Returning to clinical outcomes, we considered two endpoints: time to treatment (TTT) and overall survival (OS). Univariate Cox regression models indicated that glycolytic reserve,

maximal respiration, and spare respiratory capacity were associated with TTT, and glycolytic capacity and glycolytic reserve were associated with OS (P < 0.05, Supplementary Figure S11). Samples with higher values of these features were associated with worse clinical outcomes, i.e., shorter time to treatment and overall survival. In multivariate Cox models including age, trisomy 12, deletion of 11q22.3, deletion of 17p13, TP53 mutation and IGHV status as covariates, bioenergetic features were not picked up as predictive for TTT (Table S4–S6). However, glycolytic capacity and glycolytic reserve were the most significant predictors for OS also in the multivariate Cox models (Table 1), indicating those two glycolysis-related features provide additional OS-related information to established variables such as IGHV status, one of the most reliable prognostic markers in CLL. As shown in Figure 5 A-B, M-CLL patients with low glycolytic capacity or reserve showed best prognosis, U-CLL patients with high glycolytic capacity or reserve showed worst prognosis, while the other two groups lie in between.

We also investigated associations of each bioenergetic feature to clinical relevant phenotypes including CD38 expression, CD49d (IGTA4) expression and lymphocyte doubling time (LDT), which are considered as indicators for CLL progression[28–31]. Again, we considered IGHV status as a potential confounder (Table S7 and S8). Significant correlations existed of CD38 gene expression with glycolytic capacity, and with glycolytic reserve (5% FDR) (Figure 5 C-D). On top of the known fact that CD38 expression is highly associated with IGHV status[32], we found that it was positively correlated to glycolytic capacity or glycolytic reserve in either of the M-CLL or U-CLL disease subgroups (Supplementary Figure S12). This result suggests an IGHV status-independent link between CD38 activity and adaptability of CLL cells to glycolysis as an energy source.

The complex network of CLL energy metabolic predictors

13

While the analyses presented so far provide insights on pairwise associations between bioenergetic features and other tumor properties, we next aimed at a systems-level map of the network of gene mutations, DNA methylation, gene expression, *ex vivo* drug responses, and bioenergetic features. We used multivariate linear regression with lasso regularization to regressed each bioenergetic feature upon other available biological features and measured prediction performance using cross-validated R^2 (Figure 6).

We first assessed to what extent each omics data type alone, or the combination of all the datasets, explained each bioenergetic features. The gene expression data and the drug response data performed best in predicting bioenergetic features (Figure 6A). Combining all datasets slightly increased the prediction power for each metabolic feature, indicating that each set contains non-redundant information. Notably, the glycolysis-related features were better explained by the multi-omics data than the respiration-related features (Figure 6A and Supplementary Figure S13).

We visualized predictor profiles for individual bioenergetic features, focusing on the *ex vivo* drug responses, gene expressions, and genetic variants (Figure 6B and Supplementary Figure S13). In accordance with the above univariate analysis, the multivariate model identified IGHV status and response to mitochondria-targeting drugs like venetoclax and rotenone as important predictors for glycolysis-related features. In addition, SF3B1 mutation was identified as one of the top predictors for glycolytic capacity and reserve, its presence being associated with higher values. SF3B1 is an mRNA splicing factor that is frequently mutated in CLL and associated with more aggressive disease and worse survival, but its oncogenic mechanism is still elusive[33]. Another genomic aberration, deletion of 13q14, was selected as one of the top predictors for basal respiration and ATP production.

Several principal components (PC) from the gene expression datasets were also identified by the multivariate modeling. PC8 was the top predictor with positive coefficient for all

respiration related features. As the genes with high positive loadings on PC8 are enriched in E2F targets, this suggests that higher expression of E2F targets associates with higher respiratory activity in CLL cells. On the other hand, PC10 was the top predictor, with negative coefficient, for maximal respiration and spare respiratory capacity (Supplementary Figure S14). Based on enrichment analysis, genes with high negative loadings on PC10 are enriched in the mTOR pathway and therefore this also suggests higher mTOR pathway activity associates with high respiration capability. These findings are in line with previous reports that E2F transcription factors and mTOR pathway are key players in regulating mitochondrial activity[34,35].

PCs 2, 4, 6 and 11 were identified as predictors for several glycolysis-related features (Figure 6B and Supplementary Figure S13). Gene set enrichment analysis highlighted TNFα-NFκB signaling as the most enriched pathway for genes with high loadings on PC2, 4 and 6 (Supplementary Figure S14). This finding is consistent with previous reports that NFκB signaling pathway controls energy homeostasis in inflammatory and cancer cells[36]. As we also found NFκB activation signatures in the two published transcriptomic profiling datasets of BCR stimulation (Supplementary Figure S3), which is in line with previous reports that BCR stimulation activate NFκB, we suggest that NFκB activation may play a role in increased glycolysis after BCR activation[37,38].

Discussion

In this study, we identified molecular features that underlie the heterogeneity of energy metabolism in CLL and linked bioenergetic features with *ex vivo* drug responses and clinical course. We found that although CLL cells and B cells to have a similar basal glycolytic activity, CLL cells had a significantly higher glycolytic capacity and glycolytic reserve, which are both indicators for the cell's potential to switch to glycolysis as an energy source when necessary. Interestingly, we also found glycolytic capacity and reserve, but not basal

glycolysis, to be novel predictors for overall survival in our cohort – CLL patients with higher glycolytic capacity and reserve showed worse prognosis. In addition, higher glycolytic capacity and reserve were also found to be correlated with high expression of CD38 gene, a cell surface marker of B-cell activation and a negative prognostic marker in CLL. These observations can be viewed in context of a recent report of CLL cells' increased reliance on aerobic glycolysis to produce energy after a glycolytic switch induced by their contact with stromal cells[39]. Although we assayed circulating CLL cells for our study, the glycolytic capacity and reserve in the flux assay may actually measure the ability of CLL cells to adapt to glycolysis in a stimulated state, similar to the stimulation by stromal cells. Our findings thus imply that circulating CLL cells may have previously undergone such metabolic reprogramming and carry the metabolic repertoire that allows them to quickly switch to glycolysis when a suitable stimulation occurs, e.g., upon stromal contact, and the magnitude and efficiency of this switch can have further impact on the prognosis of CLL patients.

We showed that U-CLL has significantly higher glycolytic rates, which validates the previous hypothesis that U-CLL may have higher reliance on aerobic glycolysis due to higher BCR signaling pathway activity[4,7]. In addition, we illustrated that the glycolysis pathway is more active in U-CLLs than M-CLLs, accompanied by an up-regulation of key enzymes regulating cellular glycolysis. This indicates that the energy metabolism may be intrinsically different between M-CLLs and U-CLLs and that the BCR signaling pathway may have a direct impact on the metabolic reprogramming. As previous attempts to monitor circulating CLL cells *in vivo* by using fluorodeoxyglucose positron emission tomography (FDG-PET), which pinpoints anatomical locations with high rate of glycolysis, failed due to insufficient sensitivity[40], our results suggest that considering the difference between the M-CLL and U-CLL subtypes could increase the sensitivity of this diagnostic approach.

We found that the CLL patient samples with gain of 8q24 showed increased respiratory activity. The reason is likely the oncogenic activity of the extra copy of the MYC proto-oncogene. Previous studies have shown that MYC substantially contributes to mitochondrial biogenesis, and the over-expression of MYC leads to increased respiratory capability in several cell line models, which is in line with our observation[41].

In our study, we also highlighted the possibility of exploiting heterogeneity of energy metabolism to improve individualized patient care. We show that higher glycolytic flexibility can contribute to the resistance of CLL samples to the treatment by drugs that affect mitochondria, such as rotenone, venetoclax, and navitoclax. We postulate that the cytotoxic effects of those drugs may partially result from restricting the energy supply by blocking cellular respiration and thus, cells with higher glycolytic potential can counteract their effect due to higher metabolic flexibility.

We are aware of certain limitations of the current study. Firstly, while most of the proliferative activity of CLL cells appears in lymph node and bone marrow, we only used circulating CLL cells in the study due to the easier availability of patient material, which was instrumental in enabling the study size. In addition, although we observed many biologically meaningful associations, they are generally weak, as indicated by relatively small effect sizes or correlation coefficients. While there is the possibility that biological variables not measured by us contribute to the heterogeneity in energy metabolism, a likely explanation comprises biological noise (since we are using patient samples instead of cell lines) and technical noise of the Seahorse extracellular flux measurements, and the other assays used. Indeed, our study is to our knowledge the first that uses such a dynamic assay to systematically interrogate energy metabolism at such large scale.

All in all, our in-depth characterization of the energy metabolism and integrative analyses provide valuable insights on mechanisms underlying the metabolic regulation of CLL cells

and reveal the possibilities of guiding clinical diagnosis and individualized patient care based on metabolic profiles. Our large-scale energy metabolism dataset complements the current traditional omics datasets, such as RNA sequencing, DNA sequencing, and methylation profiling and provide another layer for a better understanding of CLL biology.

References

1.   Messmer BT, Messmer D, Allen SL, et al. In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia B cells. J Clin Invest. 2005;115(3):755–764.

2.   van Gent R, Kater AP, Otto SA, et al. In vivo Dynamics of Stable Chronic Lymphocytic Leukemia Inversely Correlate with Somatic Hypermutation Levels and Suggest No Major Leukemic Turnover in Bone Marrow. Cancer Res. 2008;68(24):10137–10144.

3.   Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–674.

4.   Garcia-Manteiga JM, Mari S, Godejohann M, et al. Metabolomics of B to plasma cell differentiation. J Proteome Res. 2011;10(9):4165–4176.

5.   Moran EC, Kamiguti AS, Cawley JC, Pettitt AR. Cytoprotective antioxidant activity of serum albumin and autocrine catalase in chronic lymphocytic leukaemia. Br J Haematol. 2002;116(2):316–328.

6.   Jitschin R, Hofmann AD, Bruns H, et al. Mitochondrial metabolism contributes to oxidative stress and reveals therapeutic targets in chronic lymphocytic leukemia. Blood. 2014;123(17):2663–2672.

7.   MacIntyre DA, Jiménez B, Lewintre EJ, et al. Serum metabolome analysis by 1H-NMR reveals differences between chronic lymphocytic leukaemia molecular subgroups. Leukemia. 2010;24(4):788–797.

8.   Zenz T, Mertens D, Küppers R, Döhner H, Stilgenbauer S. From pathogenesis to treatment of chronic lymphocytic leukaemia. Nat Rev Cancer. 2010;10(1):37–50.

9.   Fabbri G, Dalla-Favera R. The molecular pathogenesis of chronic lymphocytic leukaemia. Nat Rev. Cancer 2016;16(3):145–162.

10.   Dietrich S, Oleś M, Lu J, et al. Drug-perturbation-based stratification of blood cancer. J Clin Invest. 2018;128(1):427-445.

11.   Wu S-H, Bi J-F, Cloughesy T, Cavenee WK, Mischel PS. Emerging function of mTORC2 as a

core regulator in glioblastoma: metabolic reprogramming and drug resistance. Cancer Biol Med. 2014;11(4):255–263.

12.   Long Y, Tsai W-B, Wangpaichitr M, et al. Arginine Deiminase Resistance in Melanoma Cells Is Associated with Metabolic Reprogramming, Glucose Dependence, and Glutamine Addiction. Mol Cancer Ther. 2013;12(11):2581–2590.

13.   Böttcher M, Renner K, Berger R, et al. D-2-hydroxyglutarate interferes with HIF-1α stability skewing T-cell metabolism towards oxidative phosphorylation and impairing Th17 polarization. Oncoimmunology. 2018;e1445454.

14.   Oakes CC, Seifert M, Assenov Y, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat Genet. 2016;48(3):253–264.

15.   Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417–425.

16.   Semenza GL, Roth PH, Fang HM, Wang GL. Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. J Biol Chem. 1994;269(38):23757–23763.

17.   Hitosugi T, Zhou L, Elf S, et al. Phosphoglycerate Mutase 1 Coordinates Glycolysis and Biosynthesis to Promote Tumor Growth. Cancer Cell. 2012;22(5):585–600.

18.   Dunaway GA, Kasten TP, Sebo T, Trapp R. Analysis of the phosphofructokinase subunits and isoenzymes in human tissues. Biochem J. 1988;251(3):677–683.

19.   Stevenson FK, Krysov S, Davies AJ, Steele AJ, Packham G. B-cell receptor signaling in chronic lymphocytic leukemia. Blood. 2011;118(16):4313–4320.

20.   Tavolaro S, Colombo T, Chiaretti S, et al. Increased chronic lymphocytic leukemia proliferation upon IgM stimulation is sustained by the upregulation of miR-132 and miR-212. Genes Chromosom Cancer. 2015;54(4):222–234.

21.   Doughty CA, Bleiman BF, Wagner DJ, et al. Antigen receptor-mediated changes in glucose metabolism in B lymphocytes: Role of phosphatidylinositol 3-kinase signaling in the glycolytic

control of growth. Blood. 2006;107(11):4458–4465.

22.   Qorraj M, Bruns H, Böttcher M, et al. The PD-1/PD-L1 axis contributes to immune metabolic dysfunctions of monocytes in chronic lymphocytic leukemia. Leukemia. 2017;31(2):470–478.

23.   Lagadinou ED, Sach A, Callahan K, et al. BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells. Cell Stem Cell. 2013;12(3):329–341.

24.   Pallasch CP, Schwamb J, Königs S, et al. Targeting lipid metabolism by the lipoprotein lipase inhibitor orlistat results in apoptosis of B-cell chronic lymphocytic leukemia cells. Leukemia. 2008;22(3):585–592.

25.   Ahn IE, Underbayev C, Albitar A, et al. Clonal evolution leading to ibrutinib resistance in chronic lymphocytic leukemia. Blood. 2017;129(11):1469–1479.

26.   Udensi UK, Tchounwou PB. Dual effect of oxidative stress on leukemia cancer induction and treatment. J Exp Clin Cancer Res. 2014;33:106.

27.   Malcikova J, Stano-Kozubik K, Tichy B, et al. Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia. Leukemia. 2015;29(4):877–885.

28.   Bulian P, Shanafelt TD, Fegan C, et al. CD49d is the strongest flow cytometry-based predictor of overall survival in chronic lymphocytic leukemia. J Clin Oncol. 2014;32(9):897–904.

29.   Ibrahim S, Keating M, Do KA, et al. CD38 expression as an important prognostic factor in B-cell chronic lymphocytic leukemia. Blood. 2001;98(1):181–186.

30.   Molica S, Alberti A. Prognostic value of the lymphocyte doubling time in chronic lymphocytic leukemia. Cancer. 1987;60(11):2712–2716.

31.   Viñolas N, Reverter JC, Urbano-Ispizua A, Montserrat E, Rozman C. Lymphocyte doubling time in chronic lymphocytic leukemia: an update of its prognostic significance. Blood Cells. 1987;12(2):457–470.

32.   Cruse JM, Lewis RE, Webb RN, Sanders CM, Suggs JL. Zap-70 and CD38 as predictors of IgVH mutation in CLL. Exp Mol Pathol. 2007;83(3):459–461.

33. Cazzola M, Rossi M, Malcovati L, Associazione Italiana per la Ricerca sul Cancro Gruppo Italiano Malattie Mieloproliferative. Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. Blood. 2013;121(2):260–269.

34. Benevolenskaya E V, Frolov M V. Emerging links between E2F control and mitochondrial function. Cancer Res. 2015;75(4):619–623.

35. Morita M, Gravel S-P, Chénard V, et al. mTORC1 controls mitochondrial activity and biogenesis through 4E-BP-dependent translational regulation. Cell Metab. 2013;18(5):698–711.

36. Tornatore L, Thotakura AK, Bennett J, Moretti M, Franzoso G. The nuclear factor kappa B signaling pathway: integrating metabolism with inflammation. Trends Cell Biol. 2012;22(11):557–566.

37. Saba NS, Liu D, Herman SEM, et al. Pathogenic role of B-cell receptor signaling and canonical NF-κB activation in mantle cell lymphoma. Blood. 2016;128(1):82–92.

38. Weil R, Israël A. T-cell-receptor- and B-cell-receptor-mediated activation of NF-κB in lymphocytes. Curr Opin Immunol. 2004;16(3):374–381.

39. Jitschin R, Braun M, Qorraj M, et al. Stromal cell-mediated glycolytic switch in CLL cells involves Notch-c-Myc signaling. Blood. 2015;125(22):3432–3436.

40. Karam M, Novak L, Cyriac J, Ali A, Nazeer T, Nugent F. Role of fluorine-18 fluoro-deoxyglucose positron emission tomography scan in the evaluation and follow-up of patients with low-grade lymphomas. Cancer. 2006;107(1):175–183.

41. Morrish F, Hockenbery D. MYC and mitochondrial biogenesis. Cold Spring Harb Perspect Med. 2014;4(5):a014225–a014225.

Table 1. Results of multivariate Cox regression model for overall survival (n=119, events =18) by including either glycolytic reserve or glycolytic capacity as a predictor

Multivariate cox model including glycolytic reserve

Factor	p value	Hazard Ratio	lower 95% CI	upper 95% CI
glycolytic reserve	0.033	1.10	1.00	1.20
U-CLL	0.095	3.00	0.83	11.00
treatment	0.206	2.50	0.61	9.90
trisomy12	0.265	2.40	0.52	11.00
age	0.413	1.20	0.79	1.80
TP53 mutations	0.504	1.60	0.42	5.90
11q22.3 deletions	0.629	0.71	0.17	2.90
17p13 deletions	0.790	0.80	0.16	4.00

Multivariate cox model including glycolytic capacity

Factor	p value	Hazard Ratio	lower 95% CI	upper 95% CI
glycolytic capacity	0.046	1.10	1.00	1.10
U-CLL	0.101	2.90	0.81	10.00
treatment	0.178	2.60	0.65	10.00
trisomy12	0.312	2.20	0.48	9.70
TP53 mutation	0.469	1.70	0.42	6.50
11q22.3 deletions	0.494	0.61	0.15	2.50
age	0.546	1.10	0.76	1.70
17p13 deletions	0.644	0.68	0.13	3.60

CI: Confidence interval

Figure legends

Figure 1: Difference of energy metabolism between CLL cells and normal B cells. (A) Scatterplot of the top two principal components of the 11 tested bioenergetic features. Each dot represents a CLL patient sample (blue) or a healthy-donor derived B cell (red). (B) Beeswarm plots showing differences of six of the bioenergetic features between B cell samples (n=9) and CLL samples (n=140).

Figure 2: Associations between genetic variants and bioenergetic features. (A) The distribution of P-values of the associations between each genetic variant and each energy metabolic feature (ANOVA test). Associations that did not pass a threshold corresponding to 5% FDR (method of Benjamini and Hochberg) are colored in gray. The associations with higher bioenergetic values in mutated cases are colored by red while the associations with lower bioenergetic values in mutated cases (or high-programmed subtype) are colored by blue. (B, C) Exemplary associations, visualized in beeswarm plots. (B) Glycolysis and IGHV status. (C) Glycolysis and DNA methylation cluster.

Figure 3: Genes from the glycolysis pathway are down-regulated in M-CLL samples. (A) Hallmark gene sets that are significantly enriched (method of Benjamini and Hochberg for FDR = 10%) among genes differentially expressed between M-CLL and U-CLL. (B) The heatmap shows the z-score of the expression values of glycolysis pathway genes that are differentially expressed between M-CLL and U-CLL samples. (C) Beeswarm plots for the expression values of three key genes in the glycolysis pathway, PFKP (Phosphofructokinase, platelet), PGAM1 (Phosphoglycerate Mutase 1), and PGK1 (Phosphoglycerate kinase 1).

Figure 4: Correlation test results between drug response phenotypes and bioenergetic features. A) The y-axis shows the negative logarithm of the Pearson correlation test p-values. Only drugs with at least one significant association with bioenergetic features are shown (method of Benjamini and Hochberg for FDR = 10%). Viabilities across different drug concentrations were aggregated using Tukey's median polish method. Correlations with glycolysis-related features are colored by warm colors and correlations with respiration-related features are colored by cold colors. The dotted line indicates the P-value threshold given by the method of Benjamini and Hochberg for FDR=10%. B) Comparison of explained variance of drug responses between the multivariate model including only genetic features and the model including genetic and bioenergetic features. C) Predictors with significant p values (<0.05) in multivariate models for the drugs colored by red in figure 4B. A red bar indicates a positive association with drug responses (higher drug sensitivity is associated with presence of the mutation or higher value of the bioenergetic feature), and a blue bar indicates negative association.

Figure 5. Associations between bioenergetic features and clinical course. A, B) Kaplan-Meier plots for overall survival stratified by IGHV status and glycolytic capacity (A) or glycolytic reserve (B). The cutoff to define high and low bioenergetic groups was estimated by maximally selected rank test. The cutoff value and number of samples in each group are shown inside the parentheses in the figure legends. C, D) Scatter plots for associations of CD38 expression with glycolytic capacity and glycolytic reserve.

Figure 6: Multivariate regression models for energy metabolism features. (A) Explanatory power (cross-validated R^2) of datasets of different data types for the prediction of the energy metabolic features. The error bars represent standard deviations of R^2 over 100 repeated cross-validations. The numbers in parentheses after dataset names indicate the number of variables in the dataset. (B) Visualization of fitted adaptive L1 (lasso) regularization multivariate models using drug responses, gene mutations, IGHV status, pretreatment status and the top 20 principal components of the gene expression (RNASeq) data. The numbers in parentheses indicate the number of samples used for the regression. The Z-scores of the energy metabolic features (i.e., centered by mean and scaled by standard deviation) are shown in the scatter plot at the bottom. The heatmap in the middle shows the predictor values. The continuous variables (drug responses and gene expression PCs) are shown centered and scaled using the red-white-blue color representation, the binary variables (genetic variants, IGHV status) in black and white (black: mutation present). The average model coefficients over 100 repeated cross-validations are shown by horizontal bars on the left. Only the features that were selected at least 80 times out of 100 repeats are shown.

Figure 1

Figure 2

A — Seahore measurements (y-axis): spare respiratory capacity, proton leak, OCR, maximal respiration, glycolytic reserve, glycolytic capacity, glycolysis, ECAR/OCR, ECAR, basal respiration, ATP production. x-axis: $-\log_{10}(p\ value)$. 5% FDR line shown. Labels: MED12, del8p12, gain8q24, EGR2, Chromothripsis, ATM, Methylation_Cluster, IGHV status.

B — glycolysis ~ IGHV status ($p = 3.2 \times 10^{-5}$). ECAR (pMol/min). unmutated (n=48), mutated (n=62).

C — glycolysis ~ Methylation_Cluster ($p = 3.5 \times 10^{-5}$). ECAR (pMol/min). LP (n=44), IP (n=27), HP (n=46).

Figure 3

A Gene enrichment analysis

Hallmark gene sets:
- GLYCOLYSIS (33) — Down
- PANCREAS_BETA_CELLS (5) — Down
- TNFA_SIGNALING_VIA_NFKB (54) — Up
- P53_PATHWAY (42) — Up
- E2F_TARGETS (35) — Up
- MITOTIC_SPINDLE (46) — Up

x-axis: $-\log_{10}(p)$

Direction:
- Down (red)
- Up (blue)

B HALLMARK_GLYCOLYSIS

IGHV:
- M (red)
- U (grey)

Genes: SDC3, CHST6, EGLN3, AK4, KDELR3, CLDN3, PFKP, CHPF, TPST1, BIK, ME2, CHST2, TALDO1, ENO1, CASP6, NANP, RPE, PGAM1, SOD1, TXN, TPI1, TSTA3, PMM2, GFPT1, GNPDA1, PPP2CB, SDHC, MDH1, PGK1, PGLS, AK3, MDH2, HDLBP

C normalized expression

PFKP — p = 7.1×10^{-8} — M-CLL (n=58), U-CLL (n=53)

PGAM1 — p = 6.3×10^{-5} — M-CLL (n=58), U-CLL (n=53)

PGK1 — p = 1.1×10^{-2} — M-CLL (n=58), U-CLL (n=53)

Figure 4

Figure 5

Figure 6

Energy metabolism is co-determined by genetic variants in chronic lymphocytic leukemia and influences drug sensitivity

(Supplementary Materials)

1 Supplementary Methods

1.1 Experimental details for extracellular flux assays

Seahorse XFe96 culture plates (Agilent/Seahorse Bioscience) were prepared by coating each well with Corning™ Cell-Tak Cell and Tissue Adhesive (BD, 354241) according to the manufacturer's recommendations. Additionally, a Seahorse XFe96 cartridge (Agilent, Seahorse Bioscience) was loaded with 200 µl Calibrant solution (Agilent, Seahorse Bioscience) per well and incubated overnight in a CO_2-free atmosphere. The next day, healthy donor-derived, magnetic-bead isolated B cells or CLL peripheral blood mononuclear cells (PBMCs) were thawed from cryo-frozen aliquots, washed in assay specific medium according to the manufacturer's recommendations, and viable cells were automatically counted on a Muse®Cell Analyzer (Merck Millipore). Cells were seeded at a density of 2.4×10^5 cells in 175 µl medium per well. Samples were run with 2-7 technical replicates depending on the material's availability. The ports of the Seahorse cartridge were loaded with 25 µl of each 80 mM glucose, 9 µM oligomycin and 1M 2DG for the glycolysis stress test (GST) and 20 µl of 10 µM oligomycin, 22 µl of 15 µM FCCP and 25 µl of 30 µM antimycin A/rotenone for the mitochondrial stress test (MST). After sensor calibration, assays were run as prescribed in the manufacturer's manual to record ECAR (extracellular acidification rate) and OCR (oxygen consumption rate) over time.

1.2 Quality control criteria for extracellular flux assay measurements

First, we filtered out failed measurements in MST and GST by examining the changes of OCR and ECAR values after compound application, as follows. In MST, there are four stages, and each stage is separated by the application of tool compounds. Based on the biology of mitochondrial respiration, the following criteria were defined (see Supplementary Figure S1 for illustration): the OCR values of stage 2 (after oligomycin and before FCCP injection) should be lower than the OCR values of stage 1 (before oligomycin injection); the OCR values of stage 3 (after FCCP and before rotenone & antimycin A injection) should be higher or equal to the OCR values from stage 1; the OCR values of stage 4 (after rotenone & antimycin A injection) should be lower than the OCR values from stage 1. Similarly, in GST, there are also four stages: the ECAR values of stage 2 (after glucose and before oligomycin injection) should be higher than the ECAR values of stage 1 (before glucose injection); the ECAR values of stage 3 (after oligomycin and before 2-DG injection) should be higher or equal to the ECAR values in stage 2; the ECAR values of stage 4 (after 2-DG injection) should be lower than those in stage 2. Measurements that did not meet these criteria were considered invalid and set aside.

Next, outlier samples were identified based on the modified Z-score of the OCR and ECAR values; the modified Z-score (Z_i) of a measurement point was defined as $Z_i = 0.6745 \times (x_i - \tilde{x})/\text{MAD}$, with $\tilde{x}$ denoting the median of the values from a certain measurement point across all samples and MAD the median absolute deviation. If a certain sample contained more than 40% measurements with modified Z-score higher than 3.5, this sample was considered as an outlier sample and excluded from the subsequent analysis.

Due to the large number of samples, MST and GST were performed over periods of 18 and 16 days, respectively. Therefore, each day was defined as a batch, and batch effects were estimated and adjusted. Specifically, to test for associations between bioenergetic features and categorical variables, i.e., the genomic variants and cell types (B cell VS CLL cell), ANOVA test by including batch as a blocking factor was used. For associations with continuous variables (i.e., drug responses and gene expressions) the batch effect in bioenergetic features was firstly adjusted by using the comBat function in the sva package [1] and then the Pearson correlation test was used.

After the quality control process described above, totally 12 out of 152 samples that did not pass quality control were excluded from subsequent analysis.

1.3 Summarizing bioenergetic features

Based on the Seahorse assay (illustrated in Supplementary Figure S1), five mitochondrial respiration-related bioenergetic features (basal respiration, ATP production, proton leak, maximal respiration and spare respiratory capacity) were calculated from the oxygen consumption rate (OCR) time course during a mitochondrial stress test (MST), and three glycolysis-related features (glycolysis, glycolytic capacity and glycolytic reserve) were calculated from the extracellular acidification rate (ECAR) during a glycolysis stress test (GST). The stress tests employ metabolically interfering compounds as described in the Supplementary Methods. In addition, the baseline OCR and ECAR values and the ECAR/OCR values were also defined as bioenergetic features.

1.4 Multi-omics profiling and ex-vivo drug sensitivity assay

Multi-omics profiling, including whole-exome sequencing, targeted sequencing, DNA methylation profiling and RNA sequencing, were previously performed on the same set of patient samples; in addition, the sensitivities of these samples to a panel of 63 small molecule compounds at 5 concentrations each were characterized [2]. Clinical outcomes of those sample were also recorded. Those data are available in the R data package BloodCancerMultiOmics2017, from the Bioconductor project (`http://bioconductor.org`).

1.5 Gene enrichment analysis

For the $n = 120$ patient samples for which we had both bioenergetic data and RNASeq data, the RNAseq data were used for identifying expression signatures of IGHV mutation status and for defining biological meanings of gene expression principal components selected by multivariate regression models. To characterize expression signatures of IGHV status, differentially expressed genes (FDR = 10%, method of Benjamini and Hochberg) were firstly identified by using DESeq2 [3] and then raked by their test statistics. As for defining the biological meanings of gene expression principal component, genes were ranked by their loadings on each principal components. Gene set enrichment analysis was then performed on the ranked lists using the Parametric Analysis of Gene Set Enrichment (PAGE) method [4] with the KEGG and H gene set selections from the MSigDB database (`http://software.broadinstitute.org/gsea/msigdb`).

1.6 Penalized multivariate regression

We performed multivariate regression to explain bioenergetic features by a large feature data space. We used a Gaussian linear model with L1-penalty (i.e., lasso regression) as implemented

in the R package glmnet version 2.0 with mixing parameter $\alpha = 1$ [5]. Before analysis the expression data were normalized and transformed using the *varianceStabilizingTransformation* function from DESeq2, and both expression and methylation data were filtered to include only the top 5000 most variable features each. Genetic mutations were only included in the model if present in at least 5 samples. Features with more than 20% missing values were excluded. Remaining missing values were imputed by the mean for methylation data and by the most common mutation status for genetic data.

As predictors in the lasso model the genetic mutations and IGHV status (coded as 0-1), demographics (age, sex) and the top 20 principal components of gene expression and methylation data were used. All features were scaled to unit variance and mean zero before using lasso to achieve fair treatment of all predictors by the penalty constraint. To compare explanatory power of different datasets a separate model was fit including only predictors of one omic type at a time as well as a joint model including all predictors. Using 3-fold cross-validation, the optimal penalty parameter λ was chosen to minimize the cross-validated R^2 of the model using the function *cv.glmnet*. The cross-validation process was repeated 100 times for each model to reduce the model variance, and then the average coefficient and feature selection frequency over 100 repeats were calculated. As a measure of explained variance, the reduction in cross-validated mean squared error relative to the null model was calculated and then averaged over 100 repeats. For single features, i.e. IGHV the R^2 from a standard linear model was used as corresponding quantity.

2 Supplementary Figures
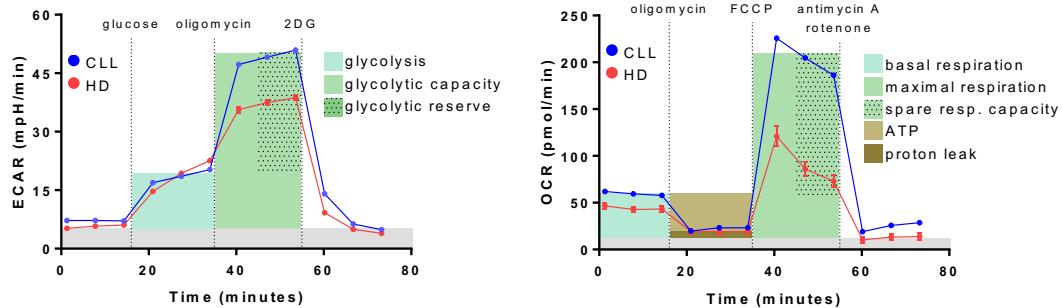
Figure S1: Representative scheme of the glycolytic stress test (left panel) and the mitochondrial stress test (right panel) depicting the extracellular acidification rate (ECAR) and the oxygen consumption rate (OCR), respectively. The calculation of the different metabolic parameters after sequential injection of metabolically active compounds is illustrated by colored boxes as indicated. The grey box symbolizes non-glycolytic acidification as well as non-mitochondrial respiration as background.
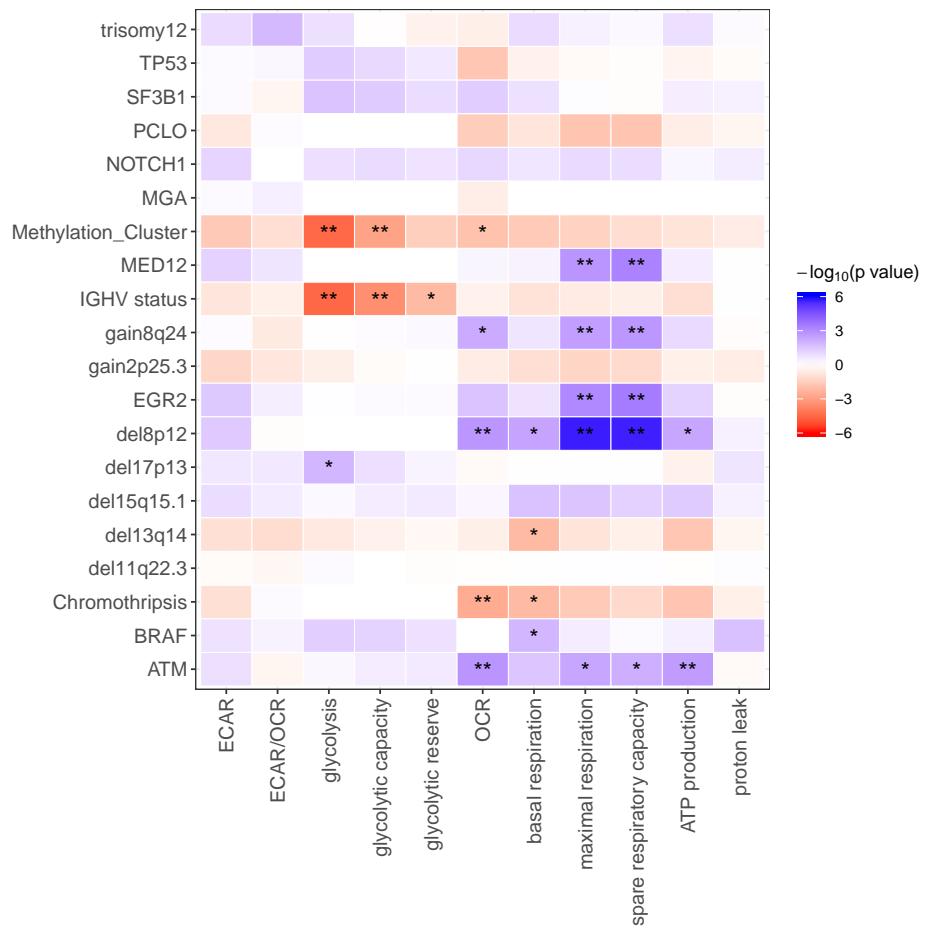
Figure S2: **A heatmap plot showing the p values for all tested associations between bioenergetic features and genetic variants.** A positive p value (colored by blue) indicates higher bioenergetic value in mutated cases (or highly methylated group for methylation cluster). A negative p value (colored by red) indicates lower bioenergetic value in mutated cases. ** indicates the association passed 5% FDR control and * indicates the association passed 10% FDR control, using the method of Benjamin and Hochberg on the set of 220 (20 × 11) raw p-values.
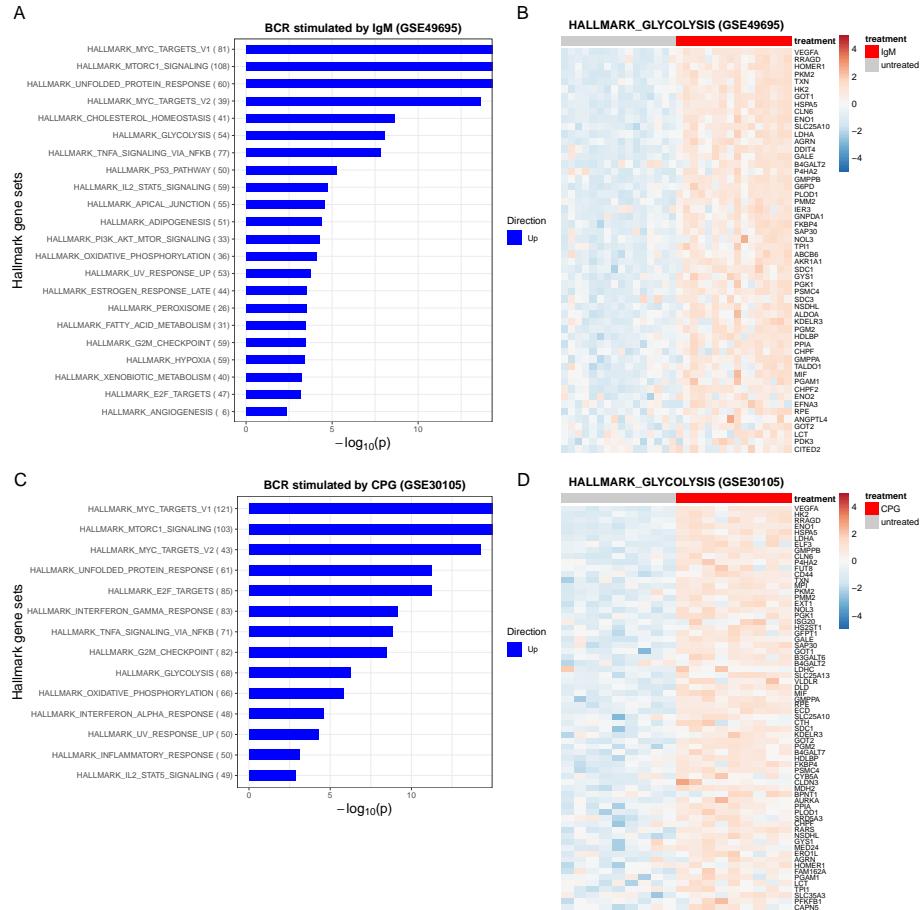
Figure S3: **Gene expression signatures of B-cell receptor stimulation queried from two public datasets** (A, C) *Hallmark* gene sets that are significantly enriched (method of Benjamini and Hochberg for FDR = 1%) among genes differentially expressed after BCR stimulation by IgM (GEO accession ID: GSE49695) or by CPG (GEO accession ID: GSE30105). (B,D) The heatmaps show the *z*-score of the expression values of glycolysis pathway genes that are differentially expressed after BCR stimulation by IgM or CPG.
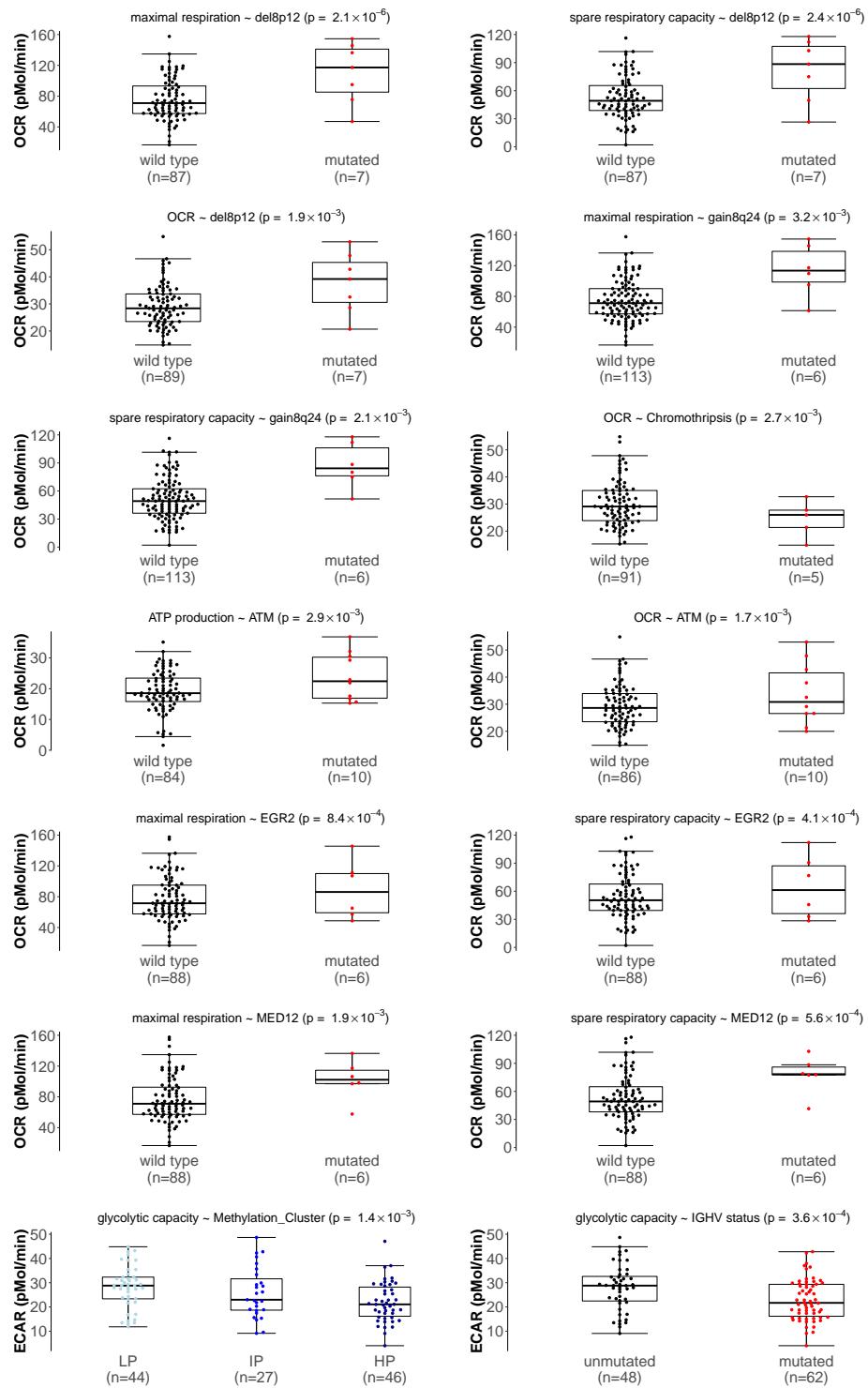
Figure S4: Beeswarm plots for all significant associations (method of Benjamini and Hochberg for FDR = 5%) between genetic variants and energy bioenergetic features

7
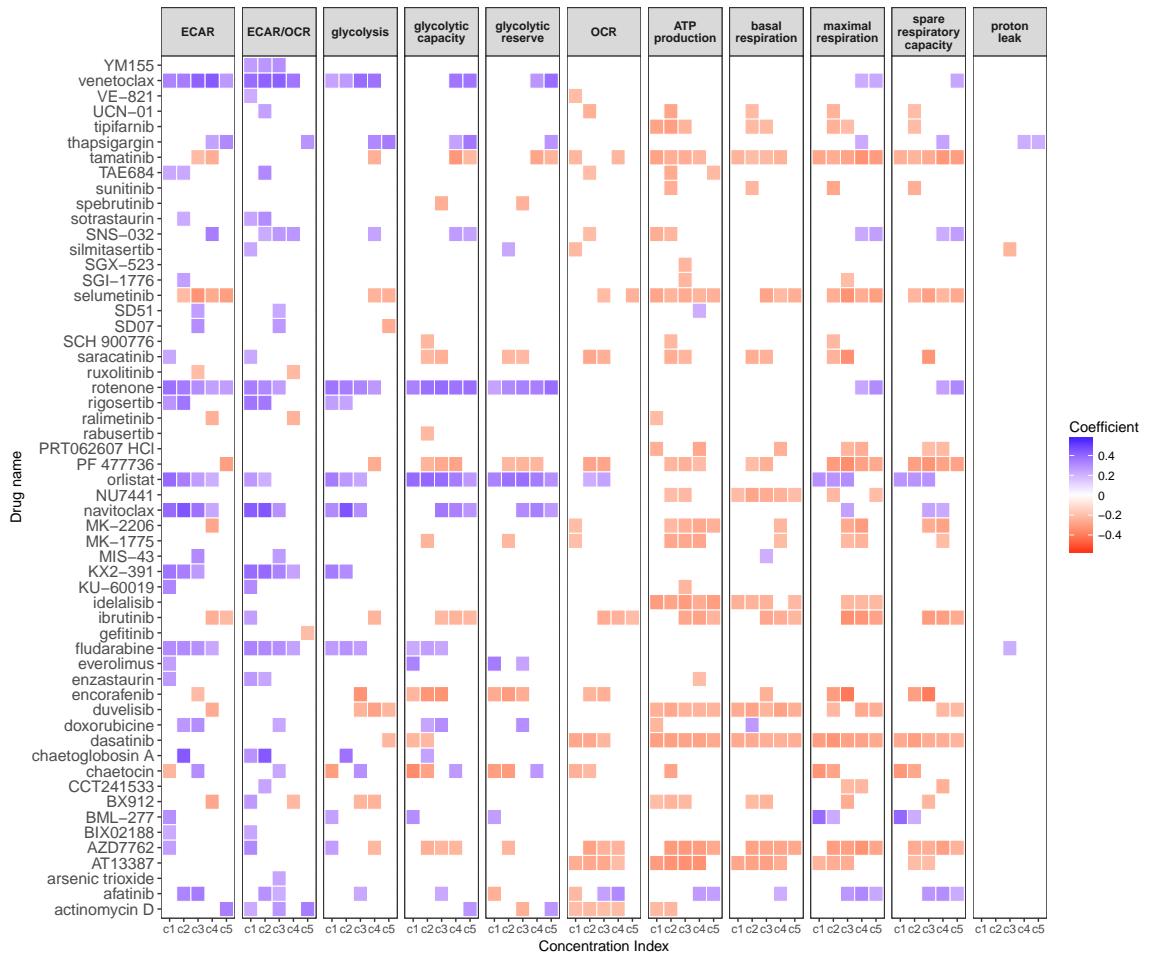
Figure S5: Correlations between bioenergetic features and drug response for all drug concentrations. Only the correlations passed 10% FDR threshold are shown. c1 indicates the highest concentration and c5 indicates the lowest concentration.
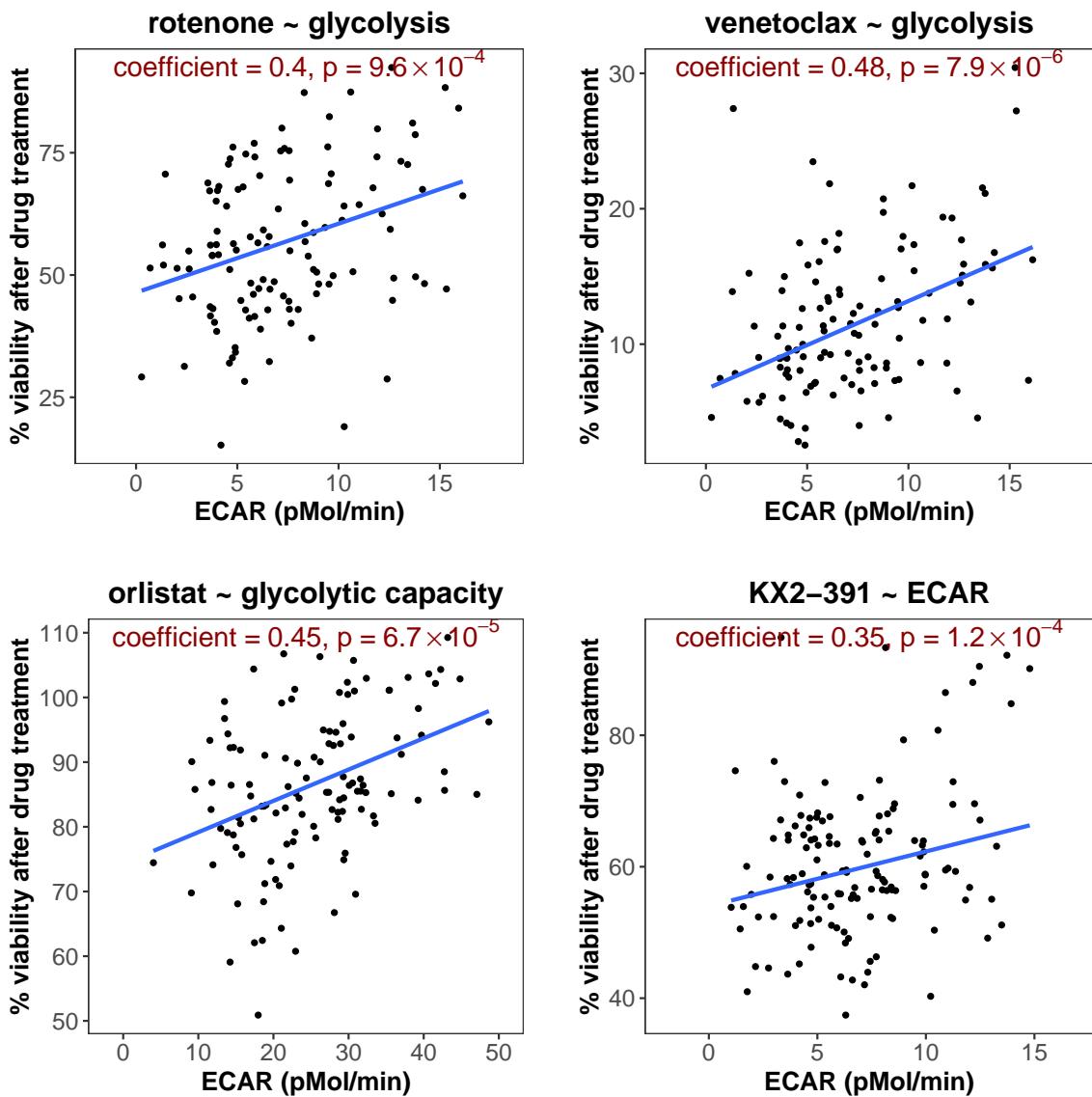
Figure S6: Exemplary scatter plots of the significant correlations between glycolysis-related features and drugs that target mitochondria. The x-axis shows the values of measured bioenergetic features, the y-axis shows the viabilities of the CLL samples after drug treatment.
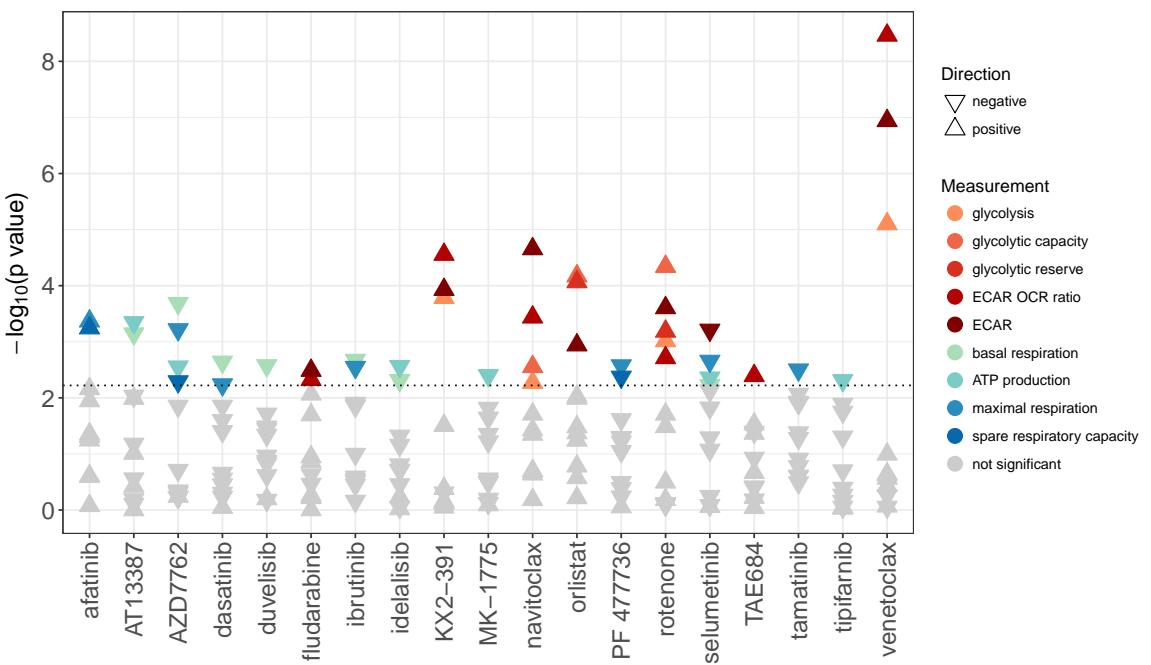
Figure S7: Correlation test results between drug response phenotype and bioenergetic features. Similar to main text Figure 4A, but IGHV status was included as a cofactor in a multivariate model in order to exclude the confounding effect of IGHV stauts. The dashed line indicate 5% FDR. The associations between glycolysis-related features and drugs such as venetoclax, navitoclax, rotenone, orlistat and KX2-391 are still significant in the multi-variate model, indicating the increased resistance to those drugs in CLL samples with higher glycolysis activity is not confounded by IGHV status.
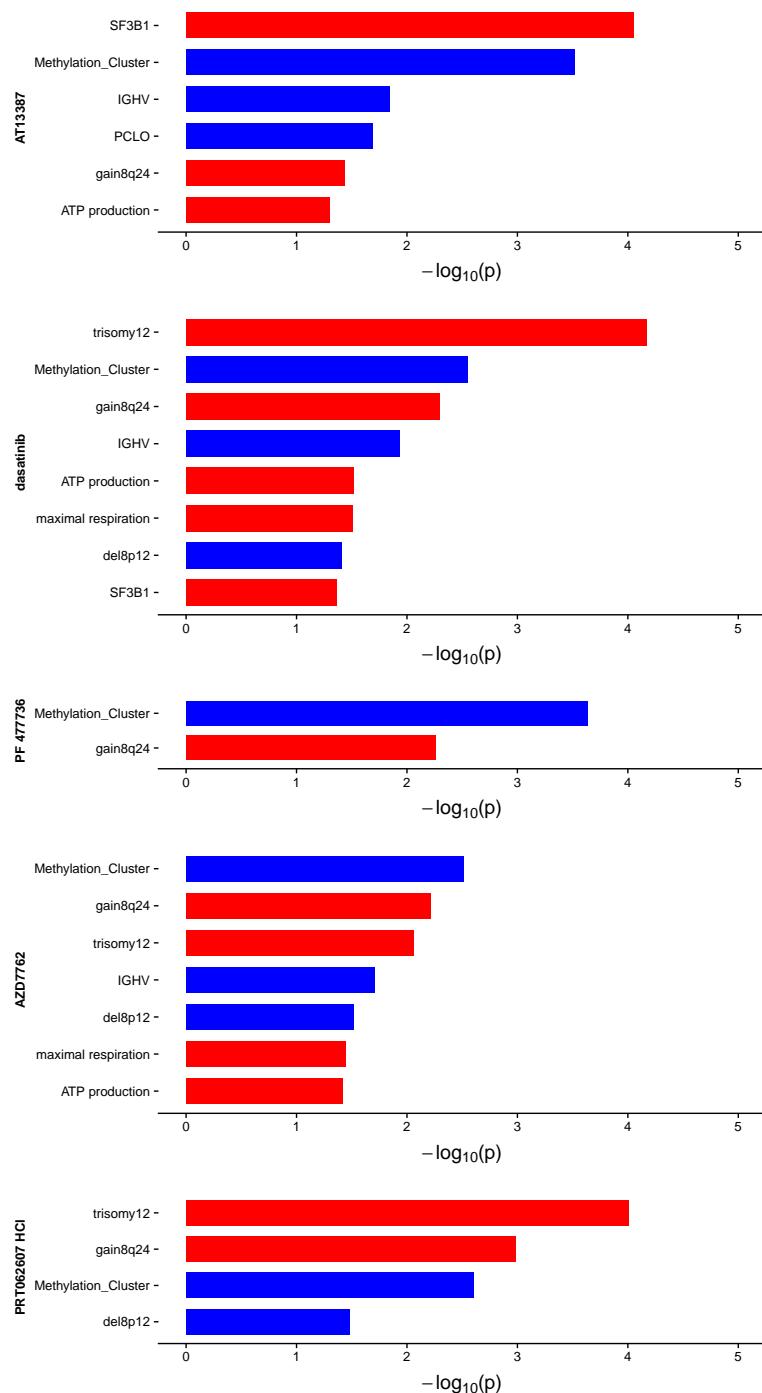
Figure S8: Drugs whose variance in patients responses can be well explained by genetic variants alone (drugs colored by blue in main text Figure 4B). The bar plots show the $log_{10}(p$ value) of the corresponding feature. Red indicates a positive association with drug responses (higher drug sensitivity associates to mutation or higher bioenergetic feature value) while blue indicates a negative association with drug responses (higher drug resistance associates to mutation or higher bioenergetic feature value). Only the features with p values less than 0.05 in the multivariate linear regression models are shown.
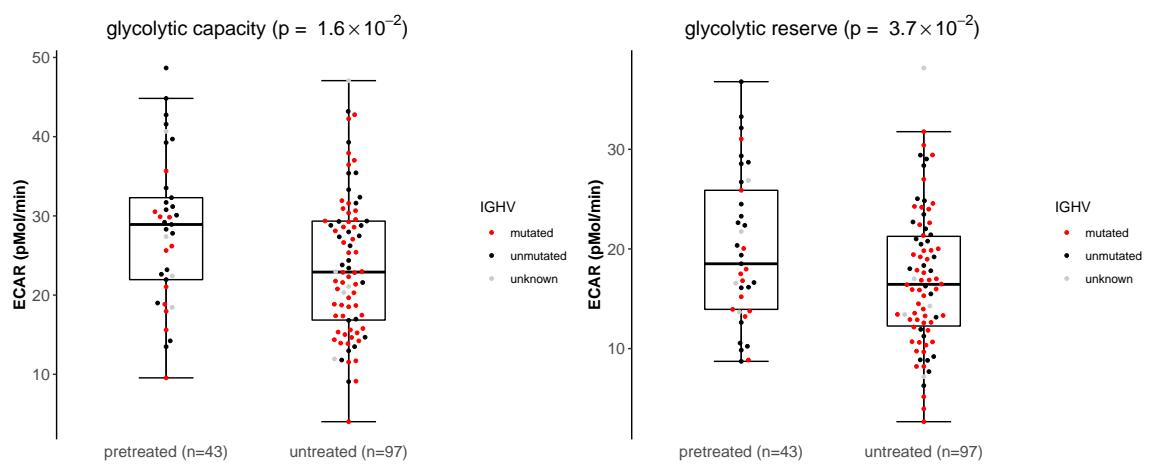
Figure S9: Beeswarm plots for associations of pretreatment status with glycolytic capacity (A) and glycolytic reserve (B).
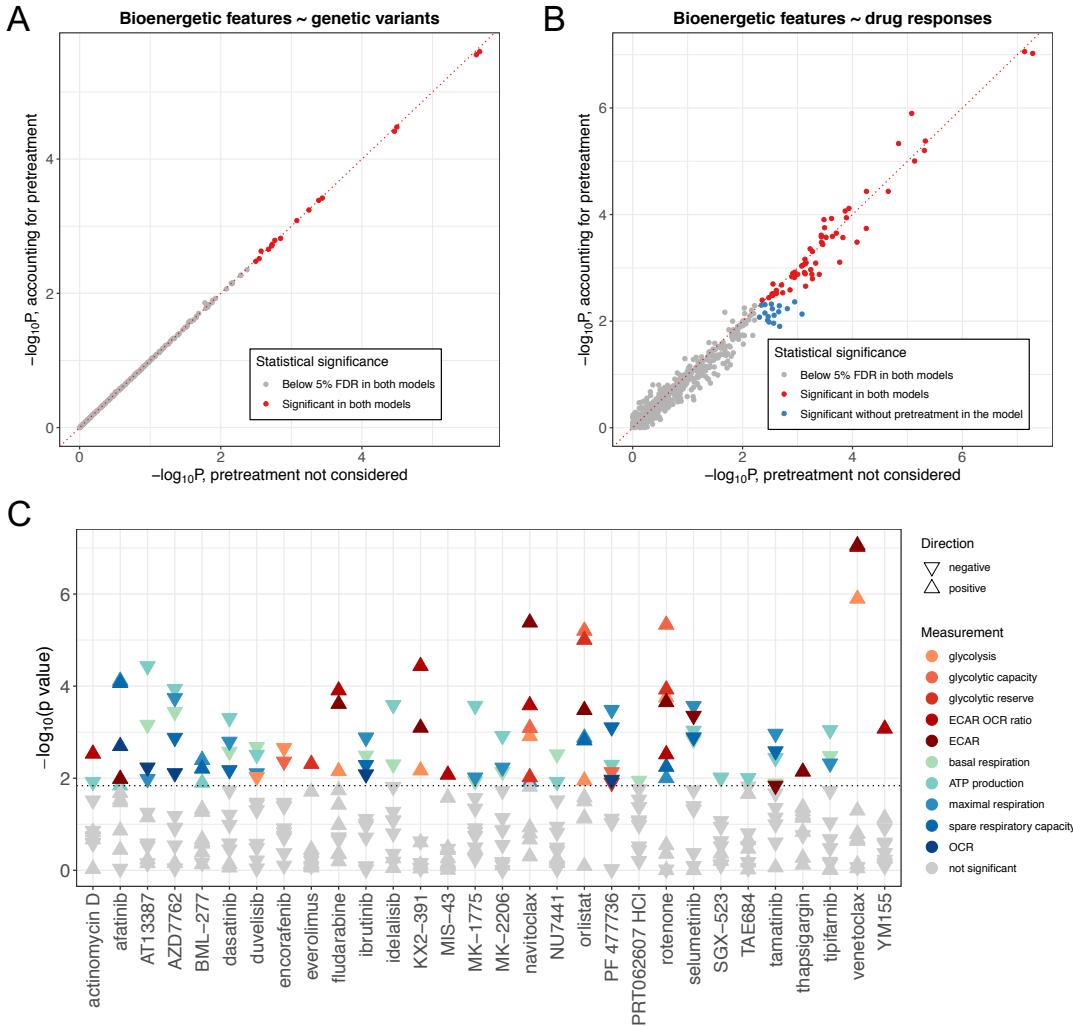
Figure S10: **Impact of pretreatment status on the associations of bioenergetic features to genetic variants and drug responses.** (A,B) Comparison of p values between the model without considering pretreatment status and the model considering pretreatment status. The scatter plot indicates that the test results from the two models are highly concordant. (C) Correlation test results between drug response phenotype and bioenergetic features. Similar to main text Figure 4A, but pretreatment status was included as a cofactor in a multivariate model in order to exclude the confounding effect of pretreatment stauts. The dashed line indicate 5% FDR. The two plots are highly similar, suggesting a very minimal impact of pretreatment status on the association between bioenergetic features and drug responses.
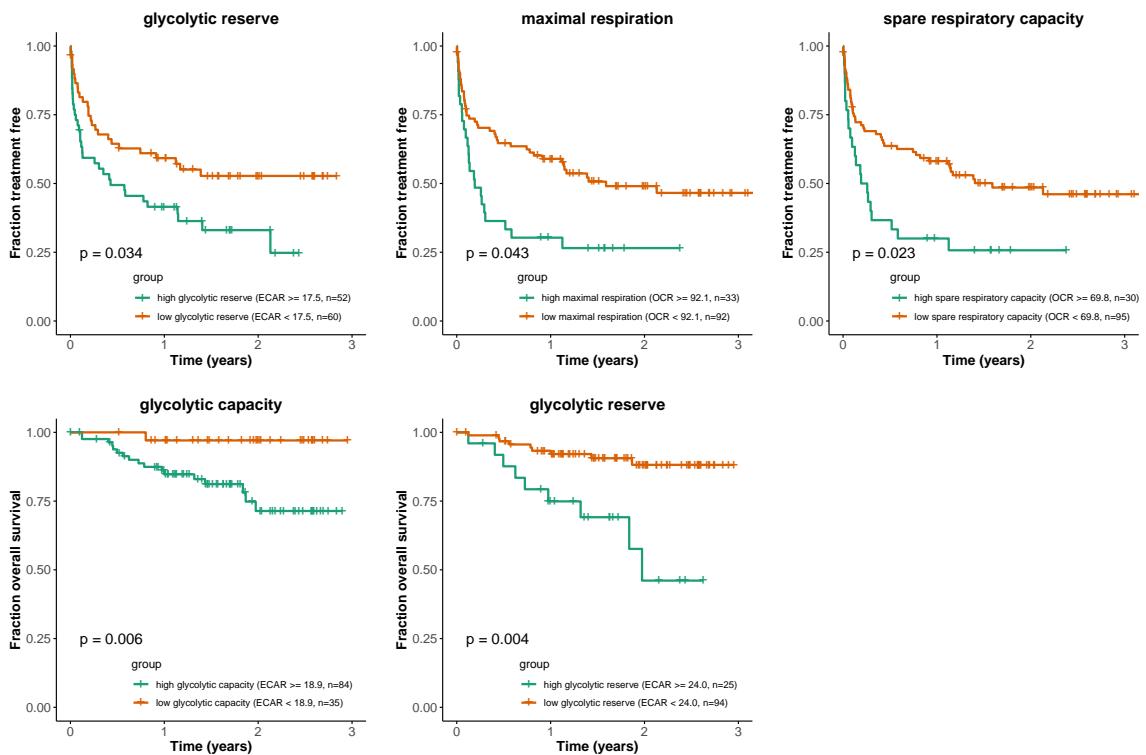
Figure S11: Kaplan-Meier plots for overall survival (OS) and time to treatment (TTT) stratified by bioenergetic features. P-values were calculated from univariate Cox regression with bioenergetic features treated as continuous variables. To visualize the effect using Kaplan-Meier plots, the patients were categorized into two groups, with high and low energy metabolic activity, using maximally selected rank statistics as cut-offs. The cutoff value and number of samples in each group are shown inside the parentheses in the figure legends.
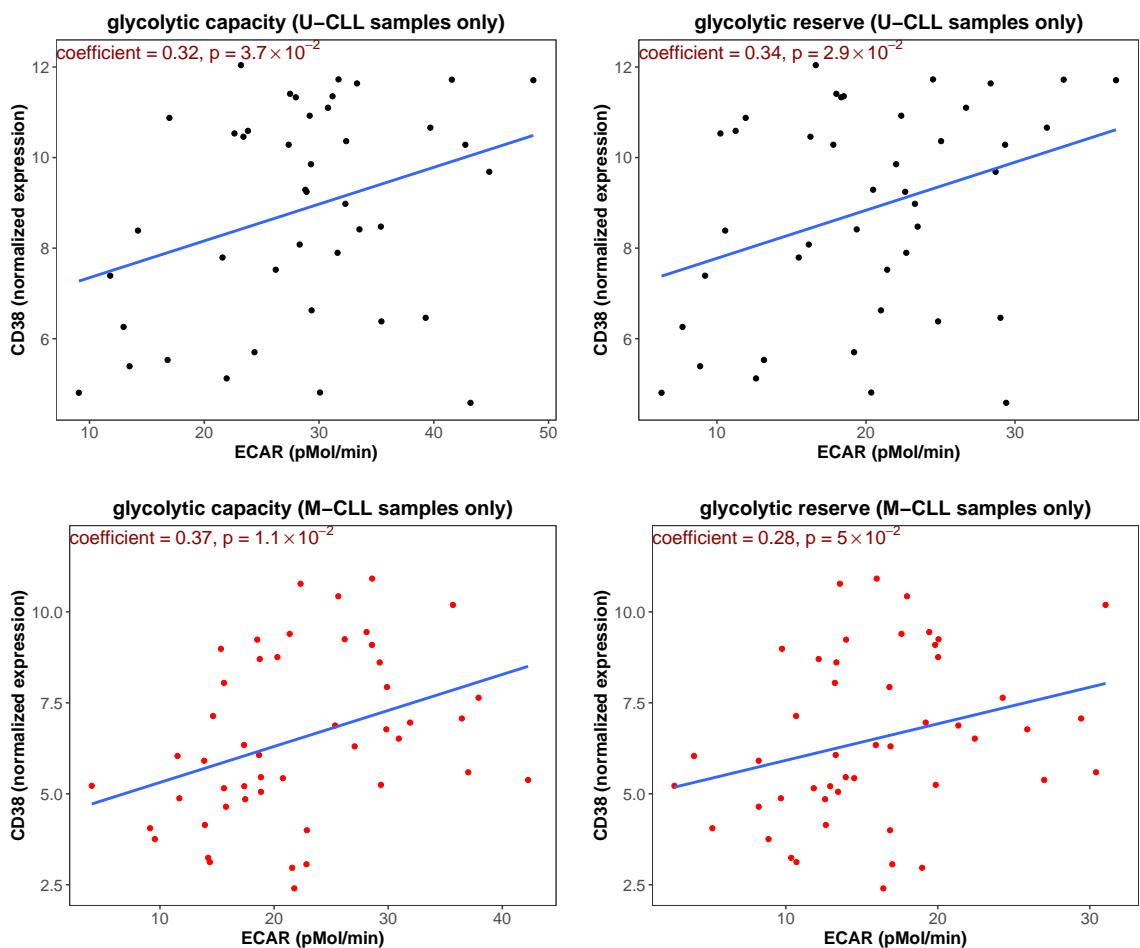
Figure S12: Scatter plots for associations of CD38 expression with glycolytic capacity and glycolytic reserve in U-CLL samples and M-CLL samples separately.
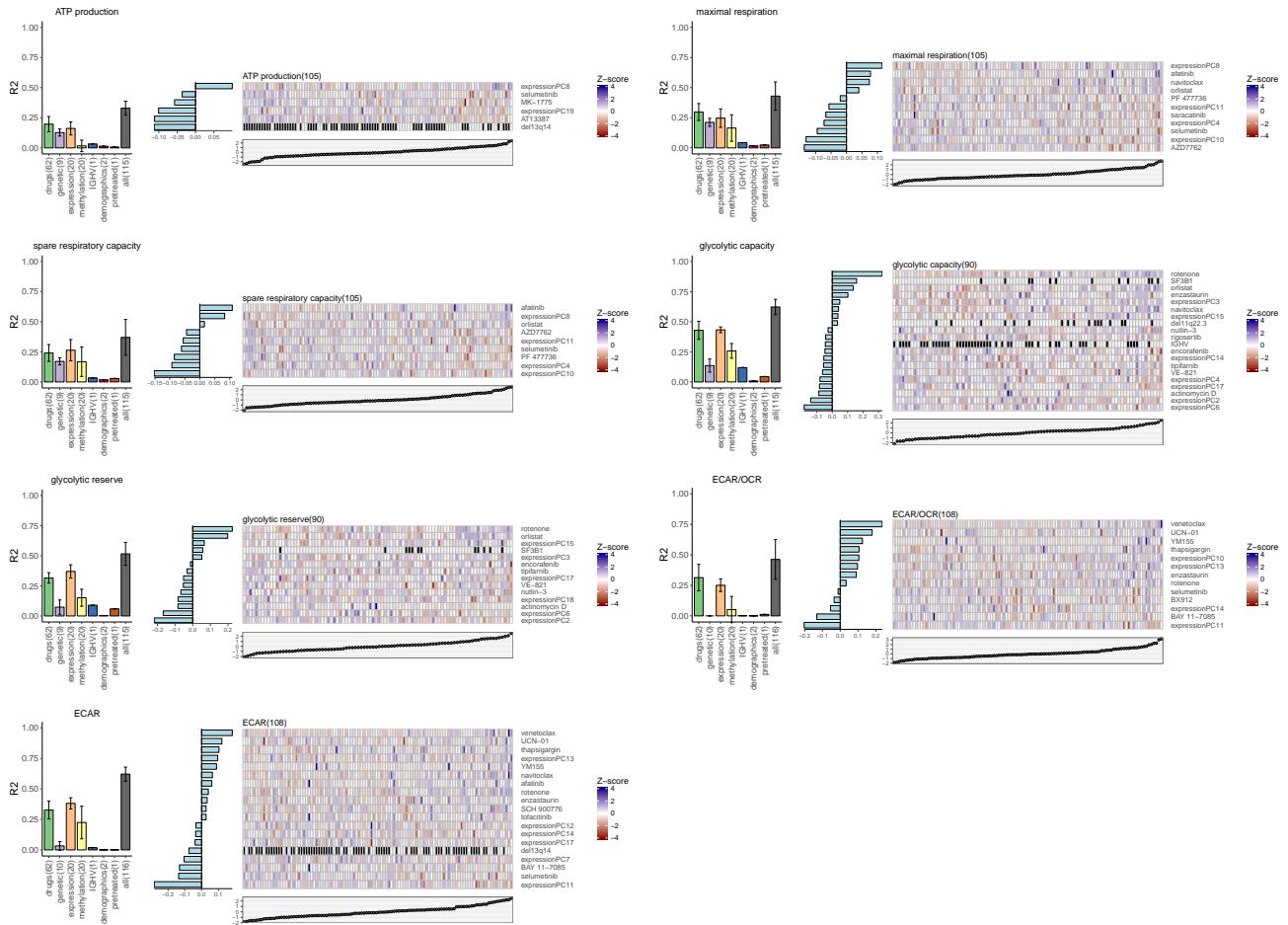
Figure S13: Barplots for explanatory power (R2) of features from different data types for the prediction of energy metabolism and visualization of fitted adaptive LASSO regularization multivariate models, for bioenergetic features not shown in main text Figure 6.
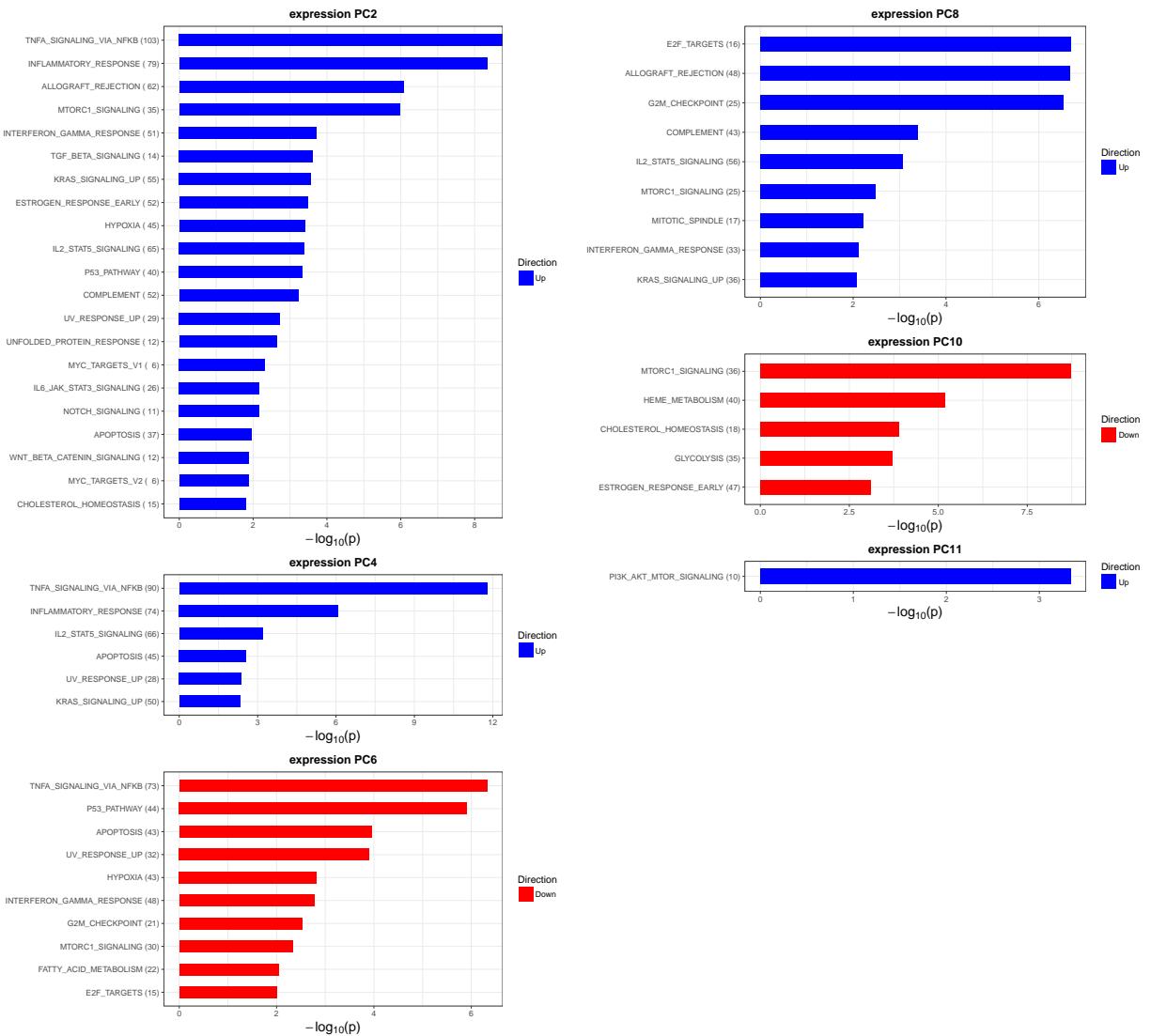
Figure S14: Gene set enrichment analysis on the genes correlated with principal components that are frequently picked by multivariate models as predictors for energy metabolism phenotype.

3 Supplementary Tables

Table S1: Background information of patients included in the study. (n.d. - no data available)

No.	Patient ID	Sex	IGHV	Age	Methylation_Cluster	Pretreated	Type of treatment
1	H017	m	U	56	LP	no	
2	H015	f	U	62	LP	no	
3	H023	f	U	70	LP	yes	Chemoimmunotherapy
4	H033	f	M	62	HP	no	
5	H035	f	M	79	IP	yes	Chemoimmunotherapy
6	H036	f	M	75	HP	no	
7	H040	f	M	83	IP	no	
8	H042	f	U	71	LP	yes	Chemoimmunotherapy
9	H046	m	M	88	HP	no	
10	H014	f	U	86	LP	yes	Chemoimmunotherapy
11	H028	f	M	72	HP	no	
12	H062	m	M	53	n.d.	no	
13	H065	f	U	77	LP	yes	Chemoimmunotherapy
14	H010	f	U	72	LP	no	
15	H027	m	U	57	LP	no	
16	H069	f	U	76	LP	yes	Chemoimmunotherapy
17	H063	f	M	49	IP	no	
18	H082	m	M	82	IP	no	
19	H072	m	U	57	IP	no	
20	H056	m	M	83	HP	no	
21	H021	m	M	49	HP	no	
22	H011	f	M	72	HP	no	
23	H078	m	U	68	LP	yes	Chemoimmunotherapy
24	H012	f	U	61	LP	yes	Chemoimmunotherapy
25	H016	m	M	55	IP	no	
26	H057	m	M	66	HP	no	
27	H045	m	U	90	LP	yes	Chemoimmunotherapy
28	H013	m	U	77	LP	yes	Chemoimmunotherapy
29	H094	m	M	45	HP	no	
30	H060	m	U	75	HP	no	
31	H039	f	M	54	HP	no	
32	H090	f	M	70	IP	yes	Chemoimmunotherapy
33	H095	f	U	52	LP	no	
34	H029	f	M	75	IP	yes	Chemoimmunotherapy
35	H020	m	M	64	HP	no	
36	H019	f	U	70	IP	yes	Chemoimmunotherapy
37	H041	m	M	75	HP	no	
38	H100	m	M	74	HP	no	
39	H032	m	U	67	LP	yes	Chemoimmunotherapy
40	H101	f	M	72	HP	no	
41	H102	f	U	78	LP	no	
42	H044	m	U	59	IP	yes	Chemoimmunotherapy
43	H083	m	n.d.	69	HP	no	
44	H104	m	U	79	LP	no	
45	H058	f	M	74	IP	no	
46	H077	f	U	69	LP	no	
47	H031	f	M	62	IP	no	
48	H005	m	M	75	IP	yes	Chemoimmunotherapy
49	H105	m	M	49	HP	no	
50	H081	f	M	64	HP	no	
51	H106	m	M	70	HP	no	
52	H054	f	M	49	HP	no	
53	H089	f	M	54	HP	no	

54	H108	m	M	57	HP	no	
55	H047	m	U	68	LP	yes	Chemoimmunotherapy
56	H064	m	n.d.	71	LP	yes	Chemoimmunotherapy
57	H113	m	M	69	IP	no	
58	H066	m	U	47	LP	yes	Chemoimmunotherapy
59	H111	m	U	54	LP	yes	Chemoimmunotherapy
60	H043	f	U	44	LP	yes	Chemoimmunotherapy
61	H088	f	M	59	HP	no	
62	H107	m	U	43	LP	no	
63	H051	f	U	78	IP	yes	Chemoimmunotherapy
64	H118	m	M	49	IP	yes	Chemoimmunotherapy
65	H093	f	U	76	LP	no	
66	H030	m	U	52	LP	no	
67	H053	f	M	83	IP	no	
68	H059	m	M	54	HP	no	
69	H096	f	n.d.	61	HP	no	
70	H080	m	U	81	LP	yes	Chemoimmunotherapy
71	H133	m	n.d.	68	HP	no	
72	H073	m	M	64	IP	yes	Chemoimmunotherapy
73	H135	f	M	76	HP	yes	Chemoimmunotherapy
74	H079	m	U	47	LP	no	
75	H136	m	U	65	LP	yes	Chemoimmunotherapy
76	H103	m	M	70	IP	no	
77	H148	f	U	33	LP	yes	Chemoimmunotherapy
78	H164	f	U	73	LP	no	
79	H165	f	U	57	LP	no	
80	H166	f	U	62	LP	no	
81	H099	f	M	54	HP	no	
82	H115	m	M	72	IP	no	
83	H037	m	M	71	IP	no	
84	H067	f	M	77	HP	no	
85	H169	f	M	41	HP	no	
86	H170	f	M	74	HP	yes	Chemoimmunotherapy
87	H184	m	M	74	HP	no	
88	H084	m	M	87	LP	no	
89	H186	f	M	72	IP	no	
90	H070	m	n.d.	71	HP	no	
91	H171	m	U	73	LP	yes	Chemoimmunotherapy
92	H173	f	M	73	IP	yes	Chemoimmunotherapy
93	H228	m	U	64	IP	no	
94	H229	f	M	74	n.d.	yes	Chemoimmunotherapy
95	H191	m	n.d.	38	IP	yes	Chemoimmunotherapy
96	H230	m	U	70	IP	yes	Chemoimmunotherapy
97	H055	m	M	64	HP	no	
98	H168	m	n.d.	57	IP	yes	Chemoimmunotherapy
99	H192	f	M	72	IP	no	
100	H185	f	M	86	HP	no	
101	H187	m	U	59	LP	no	
102	H109	m	U	85	LP	no	
103	H049	m	M	58	IP	no	
104	H163	m	M	65	HP	no	
105	H231	m	U	46	LP	no	
106	H234	m	U	68	LP	no	
107	H235	m	M	73	HP	no	
108	H236	m	M	67	HP	no	
109	H225	f	M	46	HP	no	
110	H038	m	M	73	HP	no	
111	H237	f	M	73	IP	no	

112	H238	m	U	74	LP	no	
113	H240	m	M	82	IP	no	
114	H233	m	U	54	LP	no	
115	H242	m	U	48	LP	no	
116	H243	m	U	79	LP	no	
117	H167	f	U	64	LP	no	
118	H247	f	M	46	HP	no	
119	H194	m	M	75	HP	no	
120	H248	f	M	62	HP	no	
121	H050	f	M	62	n.d.	no	
122	H246	m	U	74	LP	no	
123	H249	m	U	83	LP	no	
124	H137	m	M	53	HP	no	
125	H110	m	M	66	HP	no	
126	H252	m	U	69	LP	no	
127	H255	m	U	67	LP	yes	Chemoimmunotherapy
128	H239	f	U	69	LP	no	
129	H256	f	n.d.	62	HP	yes	Chemoimmunotherapy
130	H257	f	U	65	LP	no	
131	H258	m	M	64	HP	no	
132	H259	m	U	60	LP	yes	Chemoimmunotherapy
133	H264	m	M	77	HP	yes	Chemoimmunotherapy
134	H265	m	U	59	LP	yes	Chemoimmunotherapy
135	H260	m	U	62	LP	yes	Chemoimmunotherapy
136	H268	m	n.d.	82	HP	no	
137	H266	m	M	74	HP	yes	Chemoimmunotherapy
138	H270	f	M	66	IP	no	
139	H271	m	M	65	HP	no	
140	H272	m	U	55	LP	yes	Chemoimmunotherapy

Table S2: ANOVA test results (adjusted for batch effect) of bioenergetic features between CLL cells and normal B cells

Seahorse measurement	p	Difference of mean	adjusted p
ATP production	0.000	15.333	0.000
basal respiration	0.000	16.308	0.000
ECAR	0.013	-2.890	0.016
ECAR/OCR	0.000	-0.622	0.000
glycolysis	0.972	-0.044	0.972
glycolytic capacity	0.000	13.896	0.000
glycolytic reserve	0.000	13.940	0.000
maximal respiration	0.000	58.044	0.000
OCR	0.000	17.166	0.000
proton leak	0.629	0.975	0.692
spare respiratory capacity	0.000	41.736	0.000

Table S3: Association test results of bioenergetic features related to pretreatment status

Seahorse mearuement	p value	adjusted p	p value (IGHV blocked)	adjusted p (IGHV blocked)
ATP production	0.755	0.831	0.867	0.934
basal respiration	0.641	0.821	0.934	0.934
ECAR	0.672	0.821	0.426	0.670
ECAR/OCR	0.245	0.385	0.271	0.597
glycolysis	0.065	0.185	0.229	0.597
glycolytic capacity	0.016	0.174	0.083	0.597
glycolytic reserve	0.037	0.185	0.114	0.597
maximal respiration	0.101	0.185	0.489	0.672
OCR	0.098	0.185	0.224	0.597
proton leak	0.851	0.851	0.925	0.934
spare respiratory capacity	0.085	0.185	0.409	0.670

Table S4: Multivariate Cox regression model for time to treatment with glycolytic reserve as a covariate

factor	p value	hazard ratio	lower 95% CI	upper 95% CI
age	0.0397	0.77	0.61	0.99
trisomy12	0.594	1.3	0.53	3.1
11q22.3 deletions	0.622	1.2	0.55	2.7
17p13 deletions	0.556	1.3	0.55	3
TP53 mutations	0.0125	2.6	1.2	5.6
U-CLL	0.108	1.8	0.88	3.6
glycolytic reserve	0.095	1	0.99	1.1

Table S5: Multivariate Cox regression model for time to treatment with maximal respiration as a covariate

factor	p value	hazard ratio	lower 95% CI	upper 95% CI
age	0.0169	0.77	0.62	0.95
trisomy12	0.336	1.5	0.66	3.3
11q22.3 deletions	0.18	1.6	0.79	3.4
17p13 deletions	0.581	1.3	0.54	3
TP53 mutations	0.00532	3	1.4	6.4
U-CLL	0.0354	2	1	3.8
maximal respiration	0.0743	1	1	1

Table S6: Multivariate Cox regression model for time to treatment with spare respiratory capacity as a covariate

factor	p value	hazard ratio	lower 95% CI	upper 95% CI
age	0.0191	0.77	0.62	0.96
trisomy12	0.328	1.5	0.67	3.4
11q22.3 deletions	0.187	1.6	0.79	3.4
17p13 deletions	0.572	1.3	0.54	3
TP53 mutations	0.00743	2.9	1.3	6.1
U-CLL	0.0332	2	1.1	3.8
spare respiratory capacity	0.0672	1	1	1

Table S7: Correlation tests between each Seahorse measurements and lymphocyte doubling time

Seahorse mearuement	p value	adjusted p	p value (IGHV blocked)	adjusted p (IGHV blocked)
ATP production	0.170	0.442	0.519	0.815
basal respiration	0.666	0.814	0.786	0.959
ECAR	0.838	0.862	0.435	0.815
ECAR/OCR	0.862	0.862	0.161	0.815
glycolysis	0.032	0.199	0.778	0.959
glycolytic capacity	0.036	0.199	0.488	0.815
glycolytic reserve	0.099	0.364	0.452	0.815
maximal respiration	0.505	0.729	0.959	0.959
OCR	0.201	0.442	0.143	0.815
proton leak	0.330	0.605	0.269	0.815
spare respiratory capacity	0.530	0.729	0.882	0.959

Table S8: Associations of bioenergetic features with CD38 and IGTA4(CD49d) expression

Measurement	Gene	p value	p value (IGHV blocked)	adjusted p value	adjusted p value (IGHV blocked)
glycolytic capacity	CD38	0.000	0.001	0.001	0.021
glycolytic reserve	CD38	0.000	0.003	0.003	0.033
glycolysis	CD38	0.001	0.019	0.004	0.083
maximal respiration	CD38	0.002	0.032	0.009	0.083
glycolysis	ITGA4	0.003	0.028	0.013	0.083
ECAR	CD38	0.005	0.009	0.015	0.067
spare respiratory capacity	CD38	0.005	0.072	0.015	0.121
basal respiration	ITGA4	0.005	0.028	0.015	0.083
basal respiration	CD38	0.006	0.027	0.015	0.083
OCR	CD38	0.009	0.070	0.021	0.121
ATP production	CD38	0.013	0.070	0.026	0.121

References

[1] Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics (Oxford, England). 2012 mar;28(6):882–3. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/22257669http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3307112`.

[2] Dietrich S, Oleś M, Lu J, Sellner L, Anders S, Velten B, et al. Drug-perturbation-based stratification of blood cancer. The Journal of clinical investigation. 2017 dec;128(1). Available from: `https://www.jci.org/articles/view/93801http://www.ncbi.nlm.nih.gov/pubmed/29227286`.

[3] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014 dec;15(12):550. Available from: `http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8`.

[4] Kim SY, Volsky DJ. PAGE: parametric analysis of gene set enrichment. BMC bioinformatics. 2005 jun;6:144. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/15941488http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1183189`.

[5] Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software. 2010 feb;33(1):1–22. Available from: `http://www.jstatsoft.org/v33/i01/`.