haematologica
Journal of the Ferrata Storti Foundation

Predictive biomarkers of sustained treatment-free remission in chronic myeloid leukemia: gene expression analyses from the ENESTfreedom and ENESTop studies

by Jerald P. Radich, Shalini Chaturvedi, Islam Sadek and Vanessa Obourn

Received: December 11, 2025.
Accepted: March 18, 2026.

Citation: Jerald P. Radich, Shalini Chaturvedi, Islam Sadek and Vanessa Obourn. Predictive biomarkers of sustained treatment-free remission in chronic myeloid leukemia: gene expression analyses from the ENESTfreedom and ENESTop studies.
Haematologica. 2026 Mar 26. doi: 10.3324/haematol.2025.300215 [Epub ahead of print]

Predictive biomarkers of sustained treatment-free remission in chronic myeloid leukemia: gene expression analyses from the ENESTfreedom and ENESTop studies

Jerald P. Radich[1], Shalini Chaturvedi[2,*], Islam Sadek[2], Vanessa Obourn[3, †]

[1]Fred Hutchinson Cancer Center, Seattle, WA, United States

[2]Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States

[3]Novartis, Cambridge, MA, United States

[*]Current affiliation: Bayer Pharmaceuticals, Whippany, NJ, United States

[†]Current affiliation: Bristol Myers Squibb, Cambridge MA, United States

Corresponding author: Jerald P. Radich, Translational Science and Therapeutics Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. Email: jradich@fredhutch.org

Running head: Biomarkers of TFR post-nilotinib treatment

Author contributions

J.P.R., S.C., I.S. and V.O. conceived the study. M.W. carried out analyses and validated data. All authors contributed to the drafting and review of the manuscript.

Conflict of interest

J.P.R. has served on ad-hoc advisory boards for Novartis. S.C. and V.O. are former Novartis employees and Novartis stockholders. I.S. is a Novartis employee and stockholder.

Trial registration

ClinicalTrials.gov identifier: NCT01784068; NCT01698905.

Data sharing statement

This study did not generate new unique reagents. Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author.

Acknowledgments

We thank the patients who participated in the trial and their families and caregivers and the staff members at each site who assisted with the trial. We also thank Matthew Wall, of Sage Bionetworks, Seattle, WA, United States, for help with analyses. Medical writing support was provided by Vanesa Martinez Lopez, Ph.D., of Novartis Ireland Ltd., which was funded by Novartis in accordance with Good Publication Practice (GPP 2022) guidelines (https://www.ismpp.org/gpp-2022).

Funding

This study was funded by Novartis.

Achieving treatment-free remission (TFR) is an increasingly important goal of therapy in chronic myeloid leukemia (CML). We previously showed that pre-treatment gene expression profiles (GEP) could predict deep versus poor responders to tyrosine kinase inhibitor (TKI) therapy, with evidence supporting this is attributable to differences in immune signaling. This study interrogated gene expression data from the ENESTfreedom and ENESTop studies to identify predictive biomarkers for sustained TFR in patients with CML treated with nilotinib. GEP of samples from patients who relapsed during nilotinib treatment versus those who relapsed during TFR, as well as those who relapsed shortly after treatment cessation versus those who achieved sustained TFR, were compared. Our analysis revealed a small set of differentially expressed genes between patients who relapsed on treatment and those who relapsed during TFR, but these genes were expressed at very low levels and a reliable predictive signature could not be established. In GEP analysis of patients who relapsed early in TFR vs those who sustained TFR, bulk RNA sequencing from samples collected just prior to therapy cessation showed no significant association between principal components and TFR outcome. Additionally, transcriptome panel analyses showed no significant differential gene expression between the two groups when GEP was assessed shortly after therapy cessation. Although this analysis included multiple assessments of RNA sequencing and target expression panel data from various time points throughout the treatment and TFR period, it did not identify RNA expression profiles that could reliably predict sustained TFR in patients with CML treated with nilotinib.

Introduction

The introduction of tyrosine kinase inhibitors (TKIs) for treating chronic myeloid leukemia in chronic phase (CML-CP) has been a pivotal milestone in modern medicine, transforming CML into a chronic disease and extending the overall survival of patients with CML to nearly equal that of the general population.(1, 2)

Treatment milestones in CML are largely guided by the quantitative assessment of *BCR::ABL1* transcripts in patient peripheral blood.(3) Failure to reach molecular burden milestones can prompt an investigation for resistance mutations and spark a TKI switch, while achieving significantly deep molecular responses (DMR) may prompt consideration of TKI discontinuation. Patients who have achieved a sustained response of at least MR^4 (*BCR::ABL1*^{IS} ≤0.01%) are considered to be in DMR and those maintaining DMR for a minimum of 2–3 years may be eligible to discontinue therapy and enter treatment-free remission (TFR).(3, 4)

A substantial proportion of patients with CML-CP achieve TFR eligibility, with the cumulative incidence of DMR by 5 years reported as 66% for nilotinib and ranging between 42% and 68% for imatinib.(5) Among patients who choose to attempt TFR, approximately 50% experience CML relapse (defined as loss of major molecular response [MMR] or *BCR::ABL1*^{IS} ≤0.1%) within the first 6 to 12 months of discontinuing TKI treatment.(6, 7) Reliable prediction of patients who can successfully discontinue treatment holds significant value for those who are considering treatment cessation. Several clinical and biological features have been associated with successful TFR, including longer duration of TKI treatment, longer time in DMR, and a greater depth of molecular response while on treatment.(8, 9) The presence and persistence of certain additional genomic aberrations (AGA) within leukemic cells, such as those associated with clonal hematopoiesis of indeterminate potential (CHiP) or epigenetic regulation, have been associated with TFR success.(10, 11) Immune signatures have been described including levels of inflammatory cytokines, number and maturation signatures of natural killer (NK) cells and other immune cells, which show an association with TFR eligibility and success.(12-14) However, a reliable single biomarker or signature that can predict successful TKI discontinuation remains elusive.(9)

The ENESTfreedom and ENESTop studies investigated TFR in patients with CML-CP treated with nilotinib in first and second line, respectively.(7, 15) In ENESTfreedom, 52% of patients remained in MMR or better 48 weeks after ceasing frontline treatment with nilotinib; among patients who relapsed in the TFR phase, the majority did so within 24 weeks of stopping treatment.(7) Similar results were observed in ENESTop, with 58% of patients who attempted TFR remaining in MMR by 48 weeks. Again, most relapses during TFR occurred within 24 weeks of discontinuing nilotinib treatment.(15) On December 22, 2017, the Food and Drug Administration updated the nilotinib product label to include information on nilotinib discontinuation in patients with Philadelphia chromosome-positive CML-CP who have achieved a deep, sustained molecular response ($MR^{4.5}$, *BCR::ABL1*^{IS} <0.0032%), along with post-discontinuation

monitoring criteria and guidance for treatment re-initiation based on data from ENESTfreedom and ENESTop.(16)

We previously showed that pre-treatment gene expression profiles (GEP) could predict deep versus poor molecular response in patients with CML-CP receiving nilotinib therapy, and that most of the signal was driven by activated immune signaling involving T and natural killer (NK) cells in particular.(17, 18) We speculated that the immune system may also play a role in achieving TFR. Here, we present retrospective GEP analyses of RNA samples collected during the TFR studies ENESTfreedom and ENESTop, aiming to identify predictive gene expression markers of sustained TFR following nilotinib treatment while elucidating the biological mechanisms of sustained response after removal of TKI treatment.

Methods

Clinical Studies

ENESTfreedom (ClinicalTrials.gov identifier: NCT01784068) was a phase 2, single-arm study evaluating the potential for TFR in patients with CML-CP who had achieved $MR^{4.5}$ after at least 2 years of frontline treatment with nilotinib. Patients must have achieved a sustained DMR for at least 1 year of consolidation nilotinib therapy within the study before stopping treatment.(7) Patients re-initiated nilotinib treatment if they experienced loss of MMR while in the TFR phase.

ENESTop (ClinicalTrials.gov identifier: NCT01698905) was a phase 2, single-arm study assessing TFR in patients with CML-CP who achieved sustained DMR following a switch from imatinib to nilotinib. Patients in $MR^{4.5}$ after at least three years of TKI treatment (>4 weeks with imatinib, then ≥2 years with nilotinib) attempted TFR after one year of nilotinib consolidation therapy within the trial.(15) Nilotinib treatment was re-initiated in patients who lost MMR or had a confirmed loss of MR^4 in two consecutive assessments within 4 weeks.

For both studies, successful TFR was defined as no loss of MMR and no re-starting of nilotinib therapy following cessation of nilotinib. Patients were monitored for up to 528 weeks (approximately 10 years) after the last patient entered the TFR phase. Final trial data from ENESTfreedom and ENESTop have been published.(19, 20)

Whole blood samples collected from patients enrolled in the ENESTop and ENESTfreedom studies who consented for additional biomarker research were analyzed retrospectively with RNA sequencing (next generation sequencing [NGS]) or probe-based sequencing. Three distinct analyses, described below, examined the correlation of GEP at various collection timepoints with clinical outcomes to assess the

predictive capability of GEP on TFR success (**Figure 1**). Further details can be found in **Supplementary Methods**.

GEP of patients who relapsed during TFR vs relapsed on nilotinib therapy

Whole blood samples were collected at study screening from patients in ENESTfreedom and ENESTop, then retrospectively processed and analyzed at the Novartis NGDx labs via RNA sequencing (RNAseq). The data from the two studies were analyzed separately, as the patient populations and pre-treatment regimens used were different. A total of 15,036 genes in the ENESTfreedom study and 14,320 genes in the ENESTop study were used for downstream analysis and normalized after gene filtration. GEP from patients who relapsed on treatment (patients who entered the consolidation phase, but permanently discontinued nilotinib before the TFR phase) vs patients who relapsed on TFR (patients who entered the TFR phase and lost MMR before the clinical data cutoff date) were compared for differentially expressed (DE) genes between the responder groups. The analysis between groups to identify DE genes was based on negative binomial models and was conducted using the edgeR package(21) in R software.

GEP of patients who relapsed during TFR vs those with sustained TFR following nilotinib therapy

Whole blood samples were collected at week 48 of nilotinib consolidation from patients in ENESTfreedom and ENESTop and retrospectively analyzed via RNAseq. Patients who remained in TFR at week 144 were defined as ‘good responders’ and those who relapsed by week 24 of the TFR phase were defined as ‘poor responders’; good and poor responder categories were defined from the *BCR::ABL1* ratio data (MMR/DMR, yes or no).

The remaining clinical data was filtered for predictive analysis to include only variables that were reported at study baseline. The clinical features were compared against responder status using either a chi-square (for categorical variables) or Wilcoxon rank-sum test (for continuous variables); clinical variables which were significantly associated with responder status are presented in **Supplementary Table S1**. Inference of relative cell type abundance in each sample was performed using the MCPcounter algorithm applied to the $\log_2$(CPM+1) gene expression data. MCPcounter assigns a score for 10 different cell types that is proportional to the inferred amount of that cell type in a given sample.

Penalized logistic regression models were constructed from gene expression, clinical variables, inferred immune cell type compositions, and combinations thereof using the glmnet library in the R programming language.

Transcriptome panel analyses of patients who relapsed during TFR vs those with sustained TFR following nilotinib therapy

Whole blood samples were collected at week 12 post-TKI cessation from patients in ENESTfreedom and ENESTop and retrospectively processed and analysed for GEP using the HTG transcriptome panel (HTG Molecular Diagnostics, Inc., Tucson, AZ). (22, 23) Those who remained in TFR at week 144 were defined as 'good responders' and those who relapsed by week 36 of the TFR phase were defined as 'poor responders'.

The HTG transcriptome assay contains a total of 19,616 probes, including four positive process control probes, 100 negative process control probes, 22 genomic DNA probes, and 92 External RNA Controls Consortium (ERCC) probes; further information about this platform is presented in **Supplementary Methods**. Samples were processed as a single batch, with a total of three HTG EdgeSeq (HTG Molecular Diagnostics, Inc., Tucson, AZ) processing runs, resulting in 12 sequencing runs.

Ethics

ENESTfreedom and ENESTop studies were designed and conducted in accordance with the ethical principles of the Declaration of Helsinki, the International Conference on Harmonization (ICH) Harmonized Tripartite Guidelines for Good Clinical Practice and local laws and regulations. Written informed consent was provided by all patients before any study procedures took place. The study protocols and their amendments were reviewed and approved by an independent ethics committee or institutional review board for each study center. Analyses were carried out only for those archival RNA samples for which patient consent was available.

Results

GEP of patients who relapsed during TFR vs relapsed on nilotinib therapy

Drivers of relapse in the absence of TKI therapy may be different from the factors driving resistance on TKI therapy. This analysis compared GEP of patients who relapsed on nilotinib treatment (before attempting TFR) and patients who relapsed after treatment cessation at any time before Week 144 of TFR. GEP was assessed in 20 samples from patients who relapsed on treatment (n=10 each from ENESTfreedom and ENESTop) and 136 samples from patients who relapsed during TFR (n=84 from ENESTfreedom and n=52 from ENESTop). Nine samples from the ENESTfreedom study and seven samples from the ENESTop study were excluded from the analysis due to quality control (QC) failure. Ten samples from ENESTfreedom and five samples from ENESTop included in the analysis had lower than optimal library complexity.

Overall, 74 genes from samples in the ENESTfreedom study were identified as differentially expressed (DE) between the two patient groups using a threshold of adjusted p-value <0.05, whereas 406 genes were identified as DE with a threshold of p-value <0.2 (**Supplementary Table S2**). For samples from the ENESTop study, no DE genes were identified at an adjusted p-value of <0.05 and 9 genes were identified as DE using a threshold of adjusted p-value <0.2 (**Supplementary Table S3**). The log-abundance ratio (e.g., $\log_2$ fold-change), the log-average concentration/abundance (e.g., $\log_2$ CPM), p-values, and false discovery rate (FDR)-adjusted p-values are provided for each individual gene. The 9 DE genes identified in the ENESTop study data did not overlap with DE genes in the ENESTfreedom study, and most of the genes identified as DE across both studies were expressed at low levels.

To help with the interpretation of DE analysis results and test whether the identified individual DE genes were over-represented within specific gene sets, gene-set enrichment analysis (GSEA) was performed. For ENESTfreedom, from a total of 327 input DE genes, 237 genes were expressed at higher levels in patients who relapsed on treatment and 90 genes were expressed at higher levels in patients who relapsed during TFR. For ENESTop, a total of 182 input DE genes were identified (see **Supplementary Methods** for further detail), of which 60 genes were expressed at higher rates in patients who relapsed on treatment and 122 genes expressed at higher rates in patients who relapsed in TFR. Subsequent pathway analysis from both DE input gene sets identified 14 pathways using all 327 input DE genes from ENESTfreedom, while for ENESTop, pathway analysis identified 11 pathways using all 182 input DE genes (**Supplementary Table S4 and S5**). Despite the large number of gene sets identified, and additional enrichment analysis using more refined gene sets, no over-represented pathways were detected. (data not shown).

Plots of the logged intensity ratio (M) versus the mean logged intensities (A) (MA plots), heatmaps, and violin plots were used to aid visualization of DE genes among patient groups (**Figure 2** and **Supplementary Figure S1**). The visualization plots did not show any strong expression patterns for most of the DE identified genes.

GEP of patients who relapsed during TFR vs those with sustained TFR following nilotinib therapy

Prior to treatment discontinuation, any differences in GEP between patients with CML who relapsed before Week 24 versus those who remained in TFR are unlikely to originate from the disease itself, but rather from other elements of the microenvironment, including differences in immune system function. We then interrogated patient samples collected during the nilotinib consolidation phase to explore whether GEP assessed prior to therapy discontinuation could predict success in TFR. Samples were collected from 164 patients enrolled in ENESTfreedom (87 good responders and 77 poor responders) and 106 patients enrolled in ENESTop (61 good responders and 45 poor responders) at week 48 of consolidation therapy, just prior to TKI discontinuation.

Kernel density plots of all samples (i.e., expression profiles from individual patients) were inspected to confirm that no sample had an outlying distribution (**Supplementary Figure S2**). The density of undetected transcripts (i.e., expression =0) was greater than 40% for nearly all samples from both ENESTfreedom and ENESTop. However, mean expression of genes on the Y-chromosome was bimodally distributed and the labels provided for sex agreed with the expectation of higher expression in males (**Supplementary Figure S3A and B**). Principal component analysis (PCA) indicated that the data possessed underlying structure and did not cluster into batches requiring correction (**Supplementary Figure S3C and D**).

The variance explained by the first principal component (22% for ENESTfreedom and 24% for ENESTop) fell within the typically observed range and did not raise concerns of unidentified batch effects. The first 10 principal components (PC1 to PC10) were compared against clinical features: significant associations were observed between PC6 and PC10 with sex and PC6 with age in ENESTfreedom samples, whereas significant associations were observed between PC6, PC7 and PC10 with sex, PC5 and PC8 with age, and PC2 with time between first achieving $MR^{4.5}$ and beginning TFR in ENESTop samples (**Figure 3A and B**). However, these associations did not warrant batch correction to the expression data

The gene expression data was used to create two additional types of predictors for each patient sample: inferred immune cell-type infiltration and biological pathway enrichments. The principal components were compared against the immune cell type infiltration scores, revealing strong associations between immune cells and principal components PC1 and PC4 in samples from the ENESTfreedom study and component PC1 in samples from the ENESTop study (**Figure 3C and D**). Comparison of the first 10 principal components against responder status using a Wilcoxon rank-sum test revealed that no components were significantly associated with response (**Figure 3E and F**).

Prior to constructing regression models, the FDR of prediction was assessed by generating a univariate p-value histogram (**Supplementary Figure S4**). Differential expression analysis of the responder groups was performed and the $\log_2$(fold-change) values were used as coefficients for GSEA. Several pathways and biological processes related to ribosomal gene expression, cell cycle, and immune response to bacteria were correlated to good responders. However, there was no unifying biological explanation for the mechanism differentiating good and poor outcomes after stopping nilotinib treatment.

The results of all bootstrapped prediction analysis are listed in **Table 1 and 2** and shown in **Figure 4A and B**. Neither LASSO nor Ridge regression generated models with robust performance. Specifically, the area under the curve (AUC) of all models overlapped with the AUC distribution of random models. These results suggest that the gene expression data did not contain sufficient information to stratify the good and poor responder groups.

Transcriptome panel analyses of patients who relapsed during TFR vs those with sustained TFR following nilotinib therapy

It was possible that GEP would find no differences in TFR outcome groups before discontinuation, but after discontinuation there may be evidence of immune activation as the CML clones begin to grow in the absence of TKI suppression. A final analysis was conducted on ENESTfreedom and ENESTop samples collected just after TKI cessation (at Week 12 after stopping nilotinib) using a probe-based RNA expression platform from HTG technologies to identify biomarkers that could stratify patients who had relapsed by Week 36 of TFR (poor responders) or remained in TFR at Week 144 (good responders). Overall, 274 RNA samples were processed (ENESTfreedom: 81 from patients who relapsed and 84 from patients who remained in TFR; ENESTop: 50 from patients who relapsed and 59 from patients who remained in TFR). The overall QC failure rate between clinical studies differed, with 128 ENESTfreedom samples (78%) and 69 ENESTop samples (63%) passing all post-sequencing QC metrics (**Supplementary Table S6**).

PCA and Differential Expression Analysis (DEA) were conducted with the HTG transcriptome analysis pipeline for all samples that had passed QC. Samples were split into responder and non-responder groups with data from both studies pooled, as well as into responder and non-responder groups within each clinical study. PCA plots for overall responders versus non-responders showed no separation between pooled responder and non-responder groups (**Figure 5A**). Furthermore, no separation between groups was observed when responders versus non-responder groups within each study were compared to each other (**Supplementary Figure S5**).

Volcano plots were generated to further analyze potential clustering of responder groups on the PCA plots. No differentially expressed genes were observed when good responders were compared to poor responders in the pooled study data (**Figure 5B**). When good responder and poor responder groups within each study were compared directly, a small number of differentially expressed genes were observed in some group pairings (**Supplementary Figure S6**). However, due to the overall similarities in the groups (**Supplementary Figure S5**), the number of differentially expressed genes in any group comparison was very low. Additionally, the volcano plots generated did not exhibit a typical volcano plot shape, indicating that fold changes observed may be in genes with very low levels of expression. As an example, the *ORM1* gene (coding for an acute phase reactant) was significantly differentially expressed in some individual group comparisons. However, fold change was small for all comparisons and further investigation did not show notable differences in expression between non-responders and responders ($p=0.0924$).

Discussion

Prior exploratory analyses from the phase 3 trial ENESTnd (Evaluating Nilotinib Efficacy and Safety in Clinical Trials-Newly Diagnosed Patients, NCT00471497) identified a gene expression signature to predict good and poor response to nilotinib in patients with CML.(18) Based on these results, we explored whether gene expression could also be used to predict sustained TFR. We used samples from ENESTfreedom and ENESTop clinical studies to conduct three individual analyses to explore the potential of GEP to predict TFR outcome in patients who received nilotinib therapy. To our knowledge, this is the largest assessment of bulk RNAseq in patients attempting TFR after nilotinib treatment. We assessed samples from study screening, week 48 consolidation of nilotinib treatment and week 12 of TFR for differential expression between good and poor responders to treatment or in TFR, as defined by each sub-group analysis. The systematic and comprehensive nature of this analysis, with data collected at multiple time points from patients at different stages of treatment, was expected to provide answers on the feasibility of identifying biomarkers of TFR success using bulk sequencing. Overall, our analyses did not reveal gene expression biomarkers that could predict TFR outcomes in patients with CML-CP treated with nilotinib in first or second line. Our results are in agreement with those from other studies and suggest that this type of approach may not be optimal to identify biomarkers of TFR success in patients with CML.

In GEP analyses of samples from patients on nilotinib treatment prior to TFR attempt, PCA revealed good signal strength and no batch effects in the RNAseq dataset, which indicated that there were no systematic issues with the quality of the RNAseq or the reliability of the data outputs from the sequencing. However, principal component features did not associate at a rate better than random with patient ability to maintain TFR, indicating that even a strong RNAseq data set in this study population does not harbor an expression signature that is associated with sustained TFR. RNAseq of screening samples revealed a small set of DE genes between patients who relapsed on treatment and those who relapsed during TFR, as well as between patients who relapsed during TFR and those who successfully discontinued treatment with nilotinib. However, DE genes identified from both studies were expressed at very low levels and the reliability of these signals is uncertain. No over-represented biological pathways were identified that could differentiate good versus poor responders; consequently, no clear signature for predicting sustained TFR could be identified from either of these data sets. A possible explanation for this is that in bulk analyses, the most highly expressed genes from the most numerically abundant cell clusters are likely disproportionately represented, whereas genes expressed by type and activation state of cells that are critical for sustained TFR (such as rare immune cell subpopulations) may be present only at low frequencies in peripheral blood and not quantifiable in the respective sample.(18) Additionally, the high proportion of undetected transcripts and lack of signal-stratifying response may result from the fact that samples were obtained while patients were receiving therapy and mostly in deep response, therefore harboring very low numbers of cancer cells and/or LSCs. Given that clinically useful biomarkers need

consistent detectability, adequate abundance, and reproducible effect sizes across different patient cohorts and platforms, such low-level expression in DE genes is unlikely to have clinical significance. Our data support the fact that a predictive signature for TFR cannot be identified from GEP analysis using bulk RNAseq of peripheral blood samples from patients with CML.

In certain cases, a panel-based approach to expression profiling offers advantages over bulk RNAseq, including targeted gene lists that can enable deeper coverage of fewer genes and potential for grouping pathway genes or elimination of overlapping expressors, both of which have the potential to reduce the "noise" of bulk RNAseq when exploring expression patterns that may be difficult to detect. The HTG transcriptome panel has been reported to perform better in samples that are old or of poorer quality, and offers a simplified analysis pipeline and automated outputs, which are well suited to comparative group analysis.(24, 25) Following the lack of detectable predictive signature using bulk RNAseq, these characteristics provided the rationale for using a panel-based approach to investigate the hypothesis that cessation of TKI treatment may allow relapse-driving genes to increase their expression, thus becoming detectable soon after TKI pressure has been removed. However, similar to previous results, this method failed to discover differential gene expression between groups, both when interrogated as overall responder vs non-responder groups and as individual patient groups (responder and non-responder patients in ENESTfreedom and ENESTop, four groups in total). It could be speculated that assessing GEP in patients who relapsed on nilotinib treatment and those who sustained TFR at week 144 could potentially provide some differences, given that these two groups of patients are the most dissimilar; however, GEP analysis has consistently failed to predict TFR.

Bone marrow samples may provide more comprehensive information than peripheral blood samples, given the role of the bone marrow immune microenvironment in protecting leukemic stem cells and thus enhancing resistance to CML therapy.(26) However, acquiring sequential bone marrow samples from patients in a clinical trial only for exploratory research purposes poses significant ethical and operational hurdles, and a predictive TFR signature that is present only in bone marrow samples would face challenges for adoption into CML patient care.

It has been speculated that leukemic stem cells (LSCs) that persist despite long-term TKI therapy can be controlled or eradicated by an efficient immune response;(13, 27, 28) in this regard, LSCs may play a role in disease re-emergence despite no detectable disease at treatment cessation, and a higher-functioning immune system could suppress emergent disease driven by LSCs to sustain TFR. Although the differences in gene and pathway expression identified in this study were not statistically significant, expression signals were consistently associated with genes implicated in immune regulation, aligning with the emerging consensus on the key role of the immune response in maintaining low disease burden in CML. Different immune cell populations such as γδ+ T cells, CD4+ regulatory T cells, CD8+ PD-1+ cells, and NK cells have been reported to be differentially regulated in patients who successfully discontinue treatment compared with those that relapse during TFR.(27-34) While the work presented here did not

support GEP as a robust method for predicting TFR in CML patients, our data does support the existing body of work demonstrating the role of immune-related genes and immune profiles in successful TFR, which used alternative methods such as flow cytometry. Considering that T cells appear to play a significant role in sustained remission, continued characterization of T cells isolated from peripheral blood may provide insights that ultimately help identifying strong predictors of TFR success. Furthermore, continued advancement of more sensitive detection methodologies or multimodal biomarker sets enabling more advanced statistical approaches may prove more effective to identify predictors of TFR.

We did observe a relatively high QC failure rate of the samples subjected to transcriptomic analysis, which is expected from archival samples due to RNA degradation during processing, storage and extraction and is a known limitation of this study. (35) However, the number of samples passing QC (>200) was sufficient to support the statistical analysis described herein and larger than that used in previous studies, (11-14) supporting the strength of our conclusions.

Our comprehensive GEP analyses of patients undergoing nilotinib therapy for CML-CP have revealed significant insights into the challenges of identifying predictive biomarkers for TFR. The observed lack of differential gene expression and the high similarity between responder and non-responder groups underscore the complexity of the biological mechanisms underlying sustained remission. Although bulk RNAseq analyses have been useful in detecting predictors of response to treatment,(17, 18, 36) this technique does not appear to be suitable to identify statistically significant predictors of success in TFR. Our data support the fact that a predictive signature for TFR is unlikely to be identified from GEP analysis of peripheral blood samples using bulk RNAseq in patients with CML. New approaches regarding methodology for analysis and sample types, among other factors, will be needed in future attempts to identify predictors of TFR success in patients with CML. Among these, techniques that offer promise include spatial transcriptomics methods that can measure gene expression systematically across tissue space, (37) single-cell RNAseq methods (such as CITE-seq) to combine multiplexed protein marker detection with single-cell transcriptome profiling, (38) and complementing RNAseq immunophenotyping findings with functional immune assays. Individual techniques such as scRNAseq are highly applicable to biomarker identification but have significant limitations for broad implementation in a clinical setting. Likely, a multimodal approach, combining sensitive molecular detection methods with clinical, patient and disease characteristics, will have the highest chance of success in identifying predictors of TFR success in patients with CML.

References

1. Bower H, Björkholm M, Dickman PW, Höglund M, Lambert PC, Andersson TM. Life expectancy of patients with chronic myeloid leukemia approaches the life expectancy of the general population. J Clin Oncol. 2016;34(24):2851-2857.

2. Hughes TP, Ross DM. Moving treatment-free remission into mainstream clinical practice in CML. Blood. 2016;128(1):17-23.

3. Cross NCP, Ernst T, Branford S, et al. European LeukemiaNet laboratory recommendations for the diagnosis and management of chronic myeloid leukemia. Leukemia. 2023;37(11):2150-2167.

4. Atallah E, Sweet K. Treatment-free remission: the new goal in CML therapy. Curr Hematol Malig Rep. 2021;16(5):433-439.

5. Hochhaus A, Baccarani M, Silver RT, et al. European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. Leukemia. 2020;34(4):966-984.

6. Ross DM, Masszi T, Gómez Casares MT, et al. Durable treatment-free remission in patients with chronic myeloid leukemia in chronic phase following frontline nilotinib: 96-week update of the ENESTfreedom study. J Cancer Res Clin Oncol. 2018;144(5):945-954.

7. Hochhaus A, Masszi T, Giles FJ, et al. Treatment-free remission following frontline nilotinib in patients with chronic myeloid leukemia in chronic phase: results from the ENESTfreedom study. Leukemia. 2017;31(7):1525-1531.

8. Fujioka Y, Sugiyama D, Matsumura I, et al. Regulatory T Cell as a biomarker of treatment-free remission in patients with chronic myeloid leukemia. Cancers (Basel). 2021;13(23):5904.

9. Söderlund S, Persson I, Ilander M, et al. Plasma proteomics of biomarkers for inflammation or cancer cannot predict relapse in chronic myeloid leukaemia patients stopping tyrosine kinase inhibitor therapy. Leuk Res. 2020;90:106310.

10. Branford S, Wadham C, Shanmuganathan N, et al. Age-related clonal hematopoiesis mutations detected at the tme of stopping tyrosine kinase inhibitor therapy predict the achievement of treatment-free remission for patients with CML. Blood. 2023;142(Supplement 1):447.

11. Adnan Awad S, Brück O, Shanmuganathan N, et al. Epigenetic modifier gene mutations in chronic myeloid leukemia (CML) at diagnosis are associated with risk of relapse upon treatment discontinuation. Blood Cancer J. 2022;12(4):69.

12. Sanchez MB, Vasconcelos Cordoba B, Pavlovsky C, et al. In-depth characterization of NK cell markers from CML patients who discontinued tyrosine kinase inhibitor therapy. Front Immunol. 2023;14:1241600.

13. Huuhtanen J, Adnan-Awad S, Theodoropoulos J, et al. Single-cell analysis of immune recognition in chronic myeloid leukemia patients following tyrosine kinase inhibitor discontinuation. Leukemia. 2024;38(1):109-125.

14. Kok CH, Saunders VA, Shanmuganathan N, et al. Increased inflammatory cytokines in plasma are associated with sustained treatment-free remission in chronic myeloid leukaemia. Blood. 2024;144(Supplement 1):993.

15. Mahon FX, Boquimpani C, Kim DW, et al. Treatment-free remission after second-line nilotinib treatment in patients with chronic myeloid leukemia in chronic phase: results from a single-group, phase 2, open-label study. Ann Intern Med. 2018;168(7):461-470.

16. Pulte ED, Wroblewski T, Bloomquist E, et al. U.S. Food and drug administration benefit-risk assessment of nilotinib treatment discontinuation in patients with chronic phase chronic myeloid leukemia in a sustained molecular remission. Oncologist. 2019;24(5):e188-e195.

17. Radich JP, Larson RA, Kantarjian HM, et al. Gene expression signature predicts deep molecular response (DMR) in chronic myeloid leukemia (CML): an exploratory biomarker analysis from ENESTnd. Blood. 2019;134(Supplement_1):665.

18. Radich JP, Wall M, Branford S, et al. Molecular response in newly diagnosed chronic-phase chronic myeloid leukemia: prediction modeling and pathway analysis. Haematologica. 2023;108(6):1567-1578.

19. Hughes TP, Clementino NCD, Fominykh M, et al. Long-term treatment-free remission in patients with chronic myeloid leukemia after second-line nilotinib: ENESTop 5-year update. Leukemia. 2021;35(6):1631-1642.

20. Radich JP, Hochhaus A, Masszi T, et al. Treatment-free remission following frontline nilotinib in patients with chronic phase chronic myeloid leukemia: 5-year update of the ENESTfreedom trial. Leukemia. 2021;35(5):1344-1355.

21. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.

22. Ran D, Moharil J, Lu J, et al. Platform comparison of HTG EdgeSeq and RNA-Seq for gene expression profiling of tumor tissue specimens. J Clin Oncol. 2020;38(15_suppl):3566-3566.

23. Jaramillo MC, Reinholz G, Rounseville M, LaRoche D, Kankipati V. HTG transcriptome panel (HTP): An accurate and robust tool for transcriptome-wide gene expression profiling. J Clin Oncol. 2022;40(16_suppl):e15063.

24. O'Rourke D, Sanchez-Garcia JF, Rolfe PA, et al. Abstract 2016: Comparison of HTG-edge targeted RNA sequencing platform with whole transcriptome RNA sequencing for clinical biomarker studies. Cancer Res. 2020;80(16_Supplement):2016.

25. Qi Z, Wang L, Desai K, et al. Reliable gene expression profiling from small and hematoxylin and eosin-stained clinical formalin-fixed, paraffin-embedded specimens using the HTG EdgeSeq Platform. J Mol Diagn. 2019;21(5):796-807.

26. Patterson SD, Copland M. The bone marrow immune microenvironment in CML: treatment responses, treatment-free remission, and therapeutic vulnerabilities. Curr Hematol Malig Rep. 2023;18(2):19-32.

27. Ilander M, Olsson-Strömberg U, Schlums H, et al. Increased proportion of mature NK cells is associated with successful imatinib discontinuation in chronic myeloid leukemia. Leukemia. 2017;31(5):1108-1116.

28. Rea D, Henry G, Khaznadar Z, et al. Natural killer-cell counts are associated with molecular relapse-free survival after imatinib discontinuation in chronic myeloid leukemia: the IMMUNOSTIM study. Haematologica. 2017;102(8):1368-1377.

29. Imagawa J, Tanaka H, Okada M, et al. Discontinuation of dasatinib in patients with chronic myeloid leukaemia who have maintained deep molecular response for longer than 1 year (DADI trial): a multicentre phase 2 trial. Lancet Haematol. 2015;2(12):e528-535.

30. Nievergall E, Reynolds J, Kok CH, et al. TGF-α and IL-6 plasma levels selectively identify CML patients who fail to achieve an early molecular response or progress in the first year of therapy. Leukemia. 2016;30(6):1263-1272.

31. Kimura S, Imagawa J, Murai K, et al. Treatment-free remission after first-line dasatinib discontinuation in patients with chronic myeloid leukaemia (first-line DADI trial): a single-arm, multicentre, phase 2 trial. Lancet Haematol. 2020;7(3):e218-e225.

32. Kwaśnik P, Zaleska J, Link-Lenczowska D, et al. High level of CD8(+)PD-1(+) cells in patients with chronic myeloid leukemia who experienced loss of MMR after imatinib discontinuation. Cells. 2024;13(8):723.

33. Irani YD, Hughes A, Clarson J, et al. Successful treatment-free remission in chronic myeloid leukaemia and its association with reduced immune suppressors and increased natural killer cells. Br J Haematol. 2020;191(3):433-441.

34. Shah NP, García-Gutiérrez V, Jiménez-Velasco A, et al. Dasatinib discontinuation in patients with chronic-phase chronic myeloid leukemia and stable deep molecular response: the DASFREE study. Leuk Lymphoma. 2020;61(3):650-659.

35. Zou C, Ji C, Zhu Y, et al. Effects of freezing and rewarming methods on RNA quality of blood samples. Biopreserv Biobank. 2023;21(2):176-183.

36. McWeeney SK, Pemberton LC, Loriaux MM, et al. A gene expression signature of CD34+ cells to predict major cytogenetic response in chronic-phase chronic myeloid leukemia patients treated with imatinib. Blood. 2010;115(2):315-325.

37. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596(7871):211-220.

38. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017;14(9):865-868.

Table 1. Performance of TFR predictors in ENESTfreedom.

Model features	Number of features	LASSO AUC	Random LASSO AUC	Ridge AUC	Random Ridge AUC
Gene expression	12,355	0.55 ± 0.05	0.58 ± 0.07	0.55 ± 0.05	0.54 ± 0.06
Clinical	14	0.55 ± 0.05	0.55 ± 0.05	0.57 ± 0.06	0.57 ± 0.06
Gene expression + clinical	12,369	0.54 ± 0.05	0.55 ± 0.05	0.55 ± 0.05	0.55 ± 0.04
Pathways	1,482	0.55 ± 0.05	0.55 ± 0.05	0.54 ± 0.05	0.54 ± 0.04
Immune cell type infiltration	10	0.56 ± 0.06	0.53 ± 0.04	0.55 ± 0.05	0.55 ± 0.05

AUC values are reported as the mean ± standard deviation. AUC, area under the receiver operating characteristic curve; TFR, treatment-free remission.

Table 2. Performance of TFR predictors in ENESTop.

Model features	Number of features	LASSO AUC	Random LASSO AUC	Ridge AUC	Random Ridge AUC
Gene expression	12,328	0.55 ± 0.06	0.56 ± 0.06	0.57 ± 0.07	0.55 ± 0.05
Clinical	15	0.55 ± 0.07	0.57 ± 0.06	0.56 ± 0.06	0.56 ± 0.06
Gene expression + clinical	12,343	0.56 ± 0.07	0.57 ± 0.06	0.56 ± 0.06	0.55 ± 0.06
Pathways	3,271	0.56 ± 0.06	0.56 ± 0.07	0.55 ± 0.06	0.54 ± 0.05
Immune cell type infiltration	10	0.56 ± 0.07	0.56 ± 0.07	0.59 ± 0.08	0.57 ± 0.07

AUC values are reported as the mean ± standard deviation. AUC, area under the receiver operating characteristic curve; TFR, treatment-free remission.

Figure legends

Figure 1. **Sample collection timings for all analyses.** All samples were collected from patients enrolled in the ENESTfreedom and ENESTop studies. Samples were collected at screening to compare GEP from patients who relapsed on treatment (patients who entered the consolidation phase, but permanently discontinued nilotinib before the TFR phase) with patients who relapsed on TFR (patients who entered the TFR phase and lost MMR before the clinical data cutoff date). For the GEP analysis of TFR success vs relapse during TFR, samples were collected at Week 48 of the nilotinib consolidation treatment from patients who remained in TFR at Week 144 (good responders) and those who relapsed by Week 24 of the TFR phase (poor responders). For the transcriptome panel analysis of TFR success vs relapse during TFR, samples were collected at Week 12 post-TKI cessation from patients who remained in TFR at Week 144 (good responders) and those who relapsed by Week 36 of the TFR phase (poor responders). For each analysis, grey arrows indicate timing of sample collection and boxes represent groups of patients assessed. GEP, gene expression profiling; TFR, treatment-free remission; TKI, tyrosine kinase inhibitor; W, week.

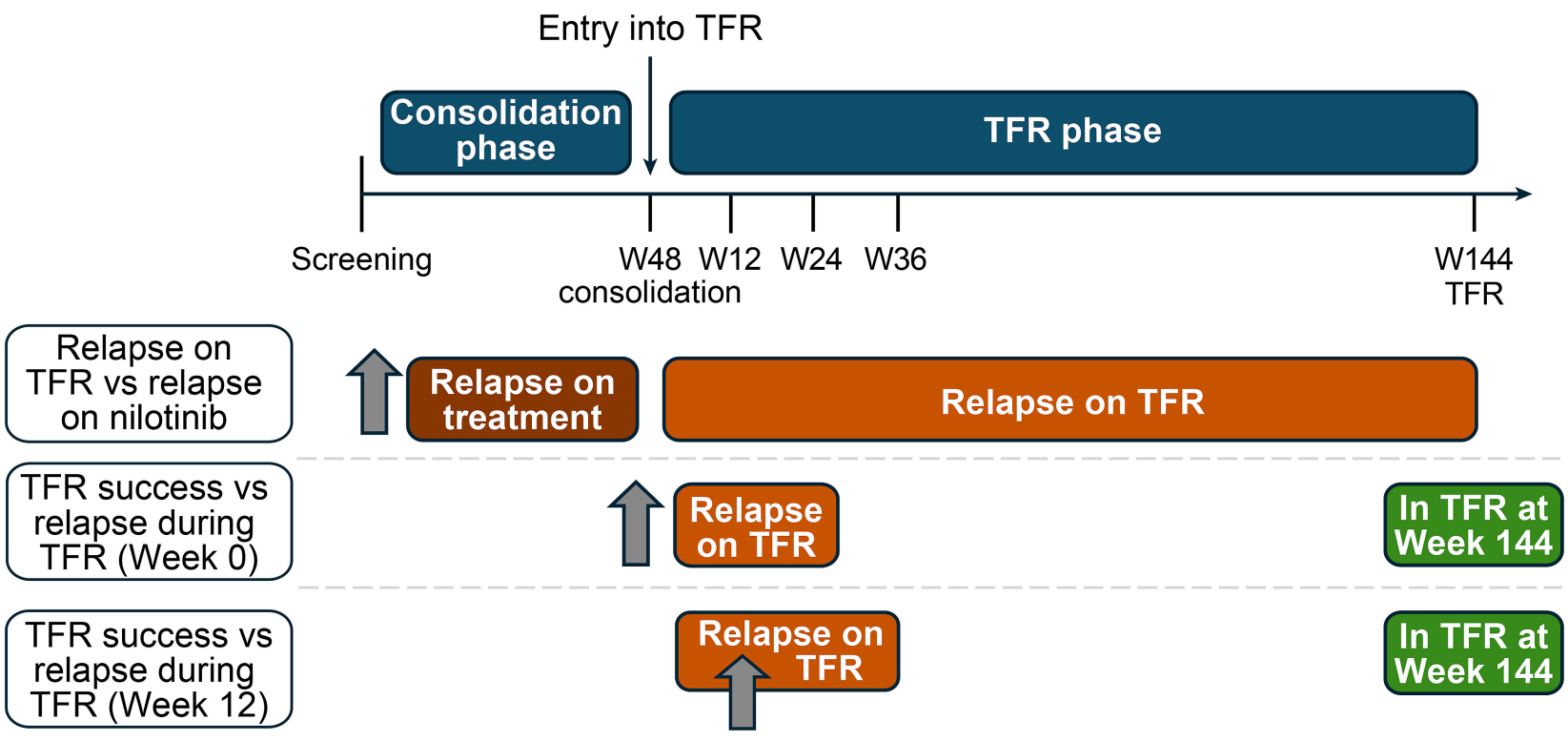
Figure 2. Differentially expressed genes between patients who relapsed on treatment and patients who relapsed during TFR. A and **B**, MA plot representing ENESTfreedom (**A**) and ENESTtop (**B**) data. The MA-plot enables the visualization of differences between GEP for the two patient groups; each gene is represented by a dot; red dots indicate differentially expressed genes based on FDR adjusted p-value <0.2 (FDR controlled by 20%). Horizontal lines at logFC=1 and logFC=-1 show fold changes of 2 and -2, respectively. The x axis represents the average expression over the mean of normalized counts (A-values); the y axis represents the $\log_2$ fold change between groups (M-values). **C** and **D**, heatmap representing top 10 DE genes in ENESTfreedom (**C**) and ENESTtop (**D**) data. Samples were ordered based on hierarchical clustering and genes within each sample were arranged in the same order for both groups. For the heatmaps, each data point (rectangle) is marked with a color that quantitatively and qualitatively reflects gene expression. Columns represent individual patients. The rows and columns of the matrix are rearranged (independently) according to a hierarchical clustering method, so that genes or groups of genes with similar expression patterns are adjacent. The computed dendrogram (tree) resulting from the clustering indicates the relationships among genes and samples. Expression values are scaled for visualization and the scales are shown in the right legend. CPM, counts per million; DE, differentially expressed; FC, fold change; FDR, false discovery rate; TFR, treatment-free remission; TRT, treatment.

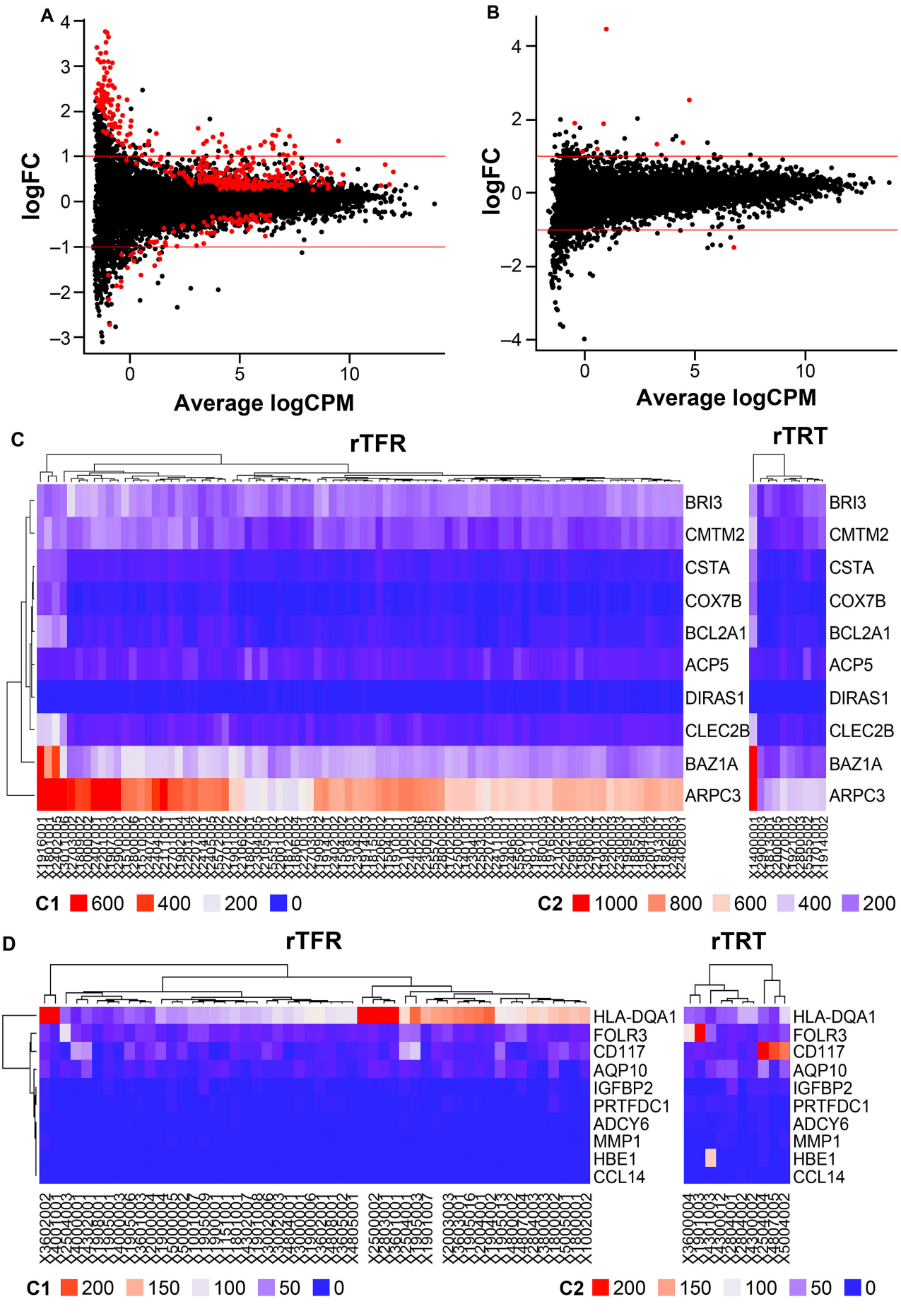
Figure 3. Principal component correlation of ENESTfreedom and ENESTop samples. A and **B**, Analysis of clinical data versus principal components in ENESTfreedom (**A**) and ENESTop (**B**) samples. The first 10 principal components (PC1 to PC10) were compared against clinical features using a Spearman correlation, Kruskal-Wallis, or Wilcoxon rank-sum test as appropriate to investigate whether these features underlie specific structure within the gene expression. The $-\log_{10}$(p-values) of associations between each clinical variable and the first 10 PCs are shown. Values greater than 1.3 (*i.e.*, pale yellow to

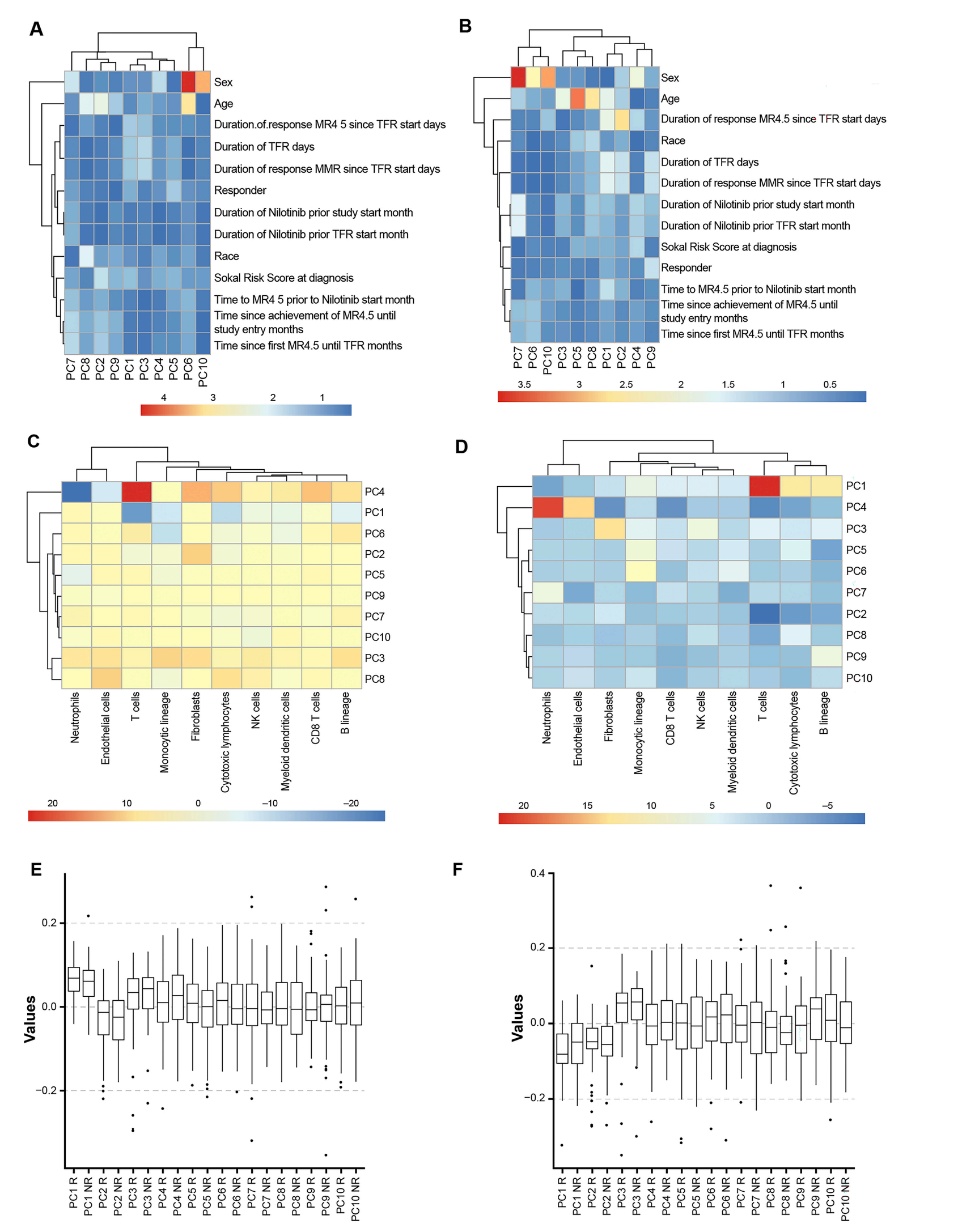
red) denote significant relationships ($p<0.05$). **C** and **D**, Analysis of immune cell infiltration data versus principal components in ENESTfreedom (**C**) and ENESTop (**D**) samples. The signed $-\log_{10}$(p-values) of correlations between inferred immune cell proportions and the first 10 PCs are shown. All entries that are not pale green correspond to significant correlations ($p<0.05$). Positive values are correlated, and negative values are anticorrelated. **E** and **F**, Analysis of PCs versus responder status in ENESTfreedom (**E**) and ENESTop (**F**) samples. The rotated values of the first 10 principal components are shown for good responders (R) and poor responders (NR). MMR, major molecular response; MR, molecular response; PC, principal component; TFR, treatment-free remission.

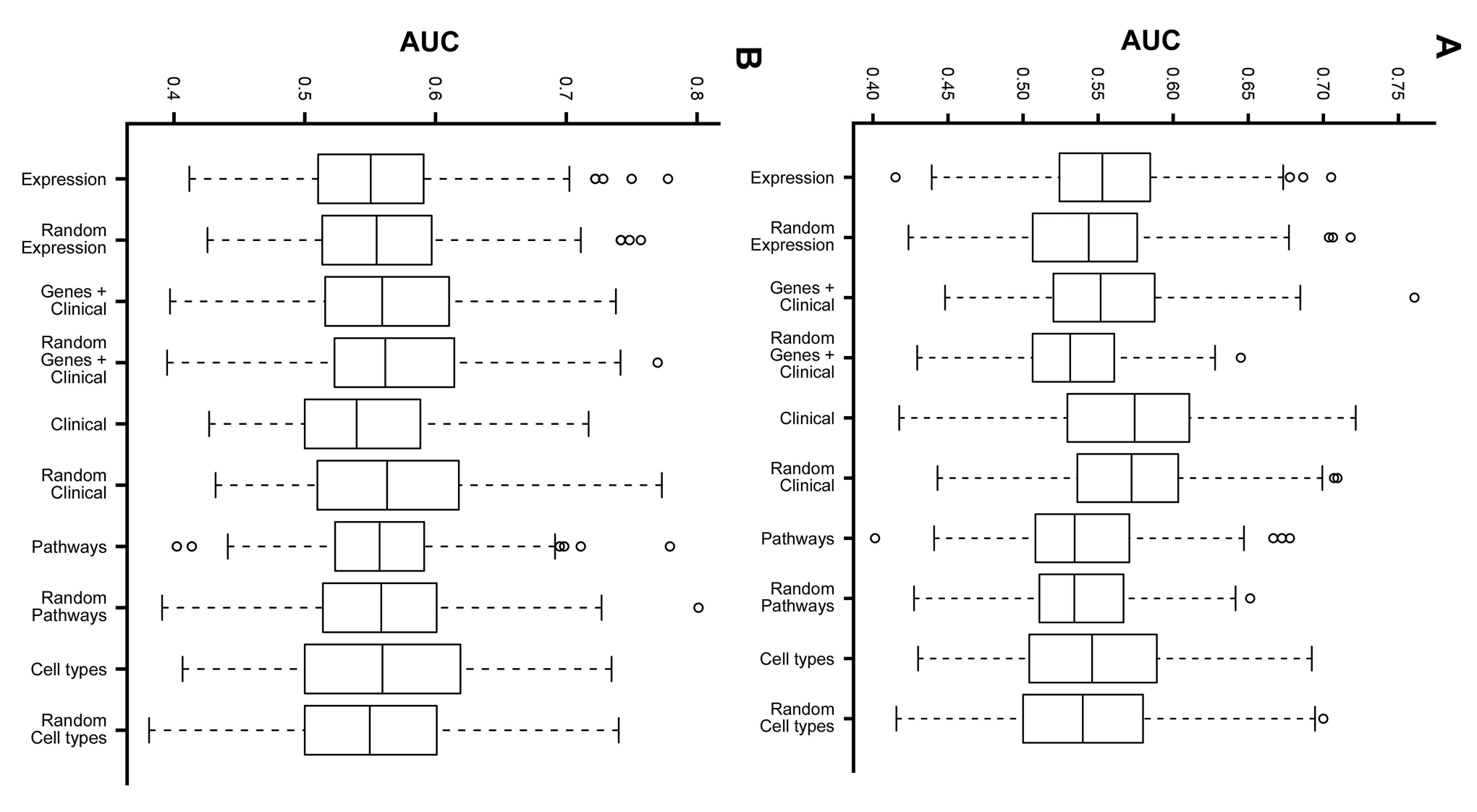
Figure 4. Regression models for gene expression and clinical variables. Bootstrap AUCs of Ridge predictors for ENESTfreedom (**A**) and ENESTop (**B**). AUC, area under the receiver operating characteristic curve.

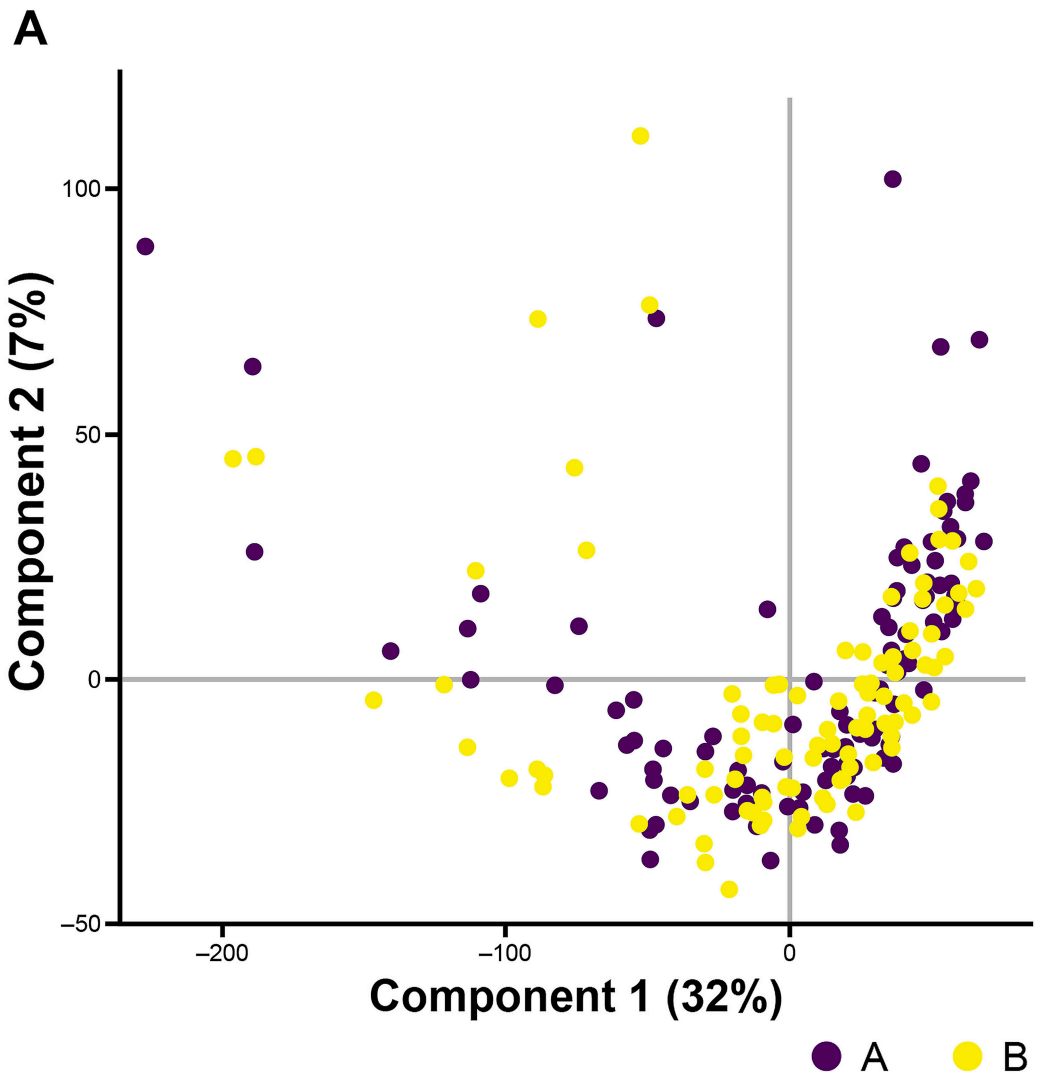
Figure 5. Principal component and differential expression analyses of ENESTfreedom and ENESTop samples from HTG probe-based expression analysis. A, Graphical representation of PCA plot of non-responders (patients who relapsed by week 36 of TFR, Group A) as compared to responders (patients who remained in TFR by week 144, Group B). Absence of separation in the PCA plot means the principal components do not align with the group label (e.g., responders vs non-responders). **B**, Volcano plot of differentially expressed probes between non-responders (Group A) and responders (Group B) for combined ENESTfreedom and ENESTop sample sets. In the volcano plot, y-axis shows the negative logarithm of the p-value and x-axis shows the logarithm of the fold change between the tested variables. Each gene is represented as a dot on the graph. Absence of points with large fold-change and low statistical significant means there are no strong, consistent transcriptional differences between group labels (e.g., responders vs non-responders). PCA, principal component analysis; TFR, treatment-free remission.

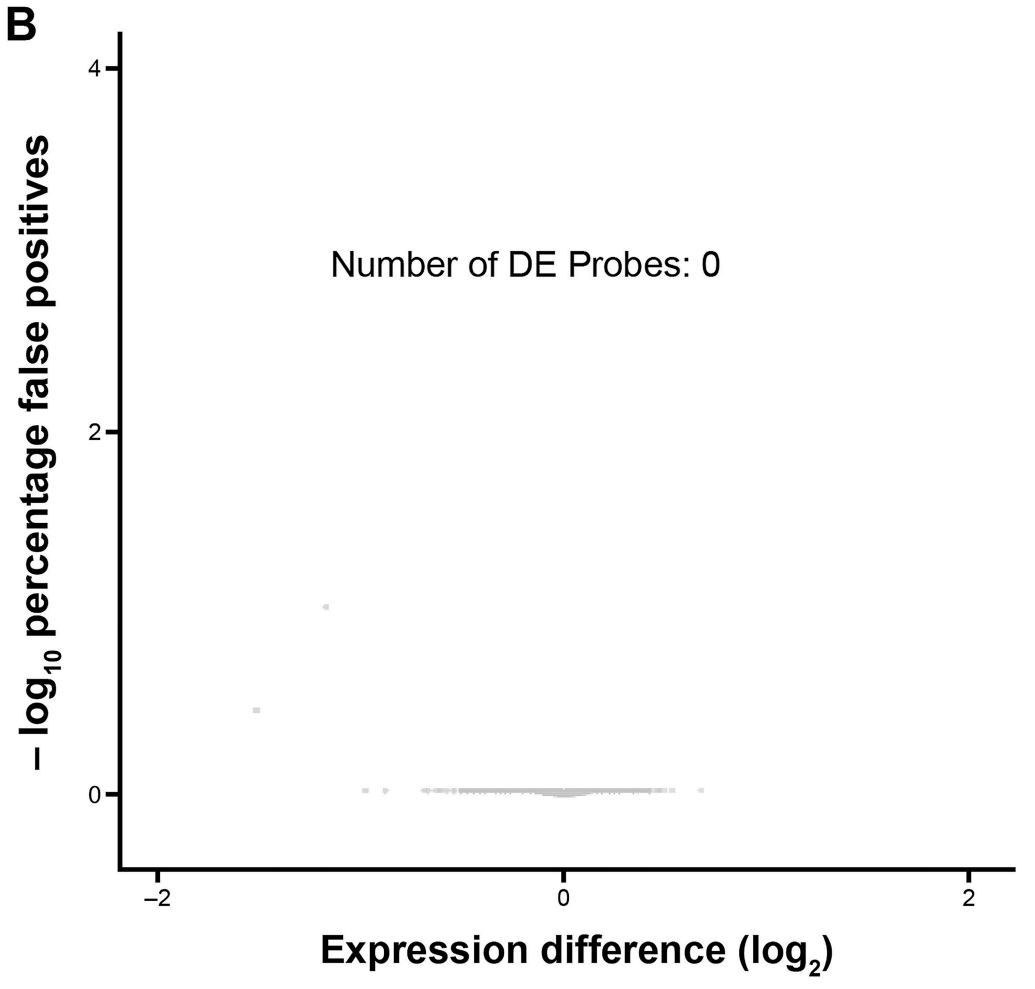

Entry into TFR
Consolidation phase
TFR phase
Screening
W48 consolidation
W12
W24
W36
W144 TFR
Relapse on TFR vs relapse on nilotinib
Relapse on treatment
Relapse on TFR
TFR success vs relapse during TFR (Week 0)
Relapse on TFR
In TFR at Week 144
TFR success vs relapse during TFR (Week 12)
Relapse on TFR
In TFR at Week 144


A
logFC
Average logCPM
B
logFC
Average logCPM
C
rTFR
rTRT
BRI3
CMTM2
CSTA
COX7B
BCL2A1
ACP5
DIRAS1
CLEC2B
BAZ1A
ARPC3
C1 600 400 200 0
C2 1000 800 600 400 200
D
rTFR
rTRT
HLA-DQA1
FOLR3
CD117
AQP10
IGFBP2
PRTFDC1
ADCY6
MMP1
HBE1
CCL14
C1 200 150 100 50 0
C2 200 150 100 50 0


A
Sex
Age
Duration.of.response MR4 5 since TFR start days
Duration of TFR days
Duration of response MMR since TFR start days
Responder
Duration of Nilotinib prior study start month
Duration of Nilotinib prior TFR start month
Race
Sokal Risk Score at diagnosis
Time to MR4 5 prior to Nilotinib start month
Time since achievement of MR4.5 until study entry months
Time since first MR4.5 until TFR months
PC7 PC8 PC2 PC9 PC1 PC3 PC4 PC5 PC6 PC10
4 3 2 1
B
Sex
Age
Duration of response MR4.5 since TFR start days
Race
Duration of TFR days
Duration of response MMR since TFR start days
Duration of Nilotinib prior study start month
Duration of Nilotinib prior TFR start month
Sokal Risk Score at diagnosis
Responder
Time to MR4.5 prior to Nilotinib start month
Time since achievement of MR4.5 until study entry months
Time since first MR4.5 until TFR months
PC7 PC6 PC10 PC3 PC5 PC8 PC1 PC2 PC4 PC9
3.5 3 2.5 2 1.5 1 0.5
C
PC4
PC1
PC6
PC2
PC5
PC9
PC7
PC10
PC3
PC8
Neutrophils
Endothelial cells
T cells
Monocytic lineage
Fibroblasts
Cytotoxic lymphocytes
NK cells
Myeloid dendritic cells
CD8 T cells
B lineage
20 10 0 −10 −20
D
PC1
PC4
PC3
PC5
PC6
PC7
PC2
PC8
PC9
PC10
Neutrophils
Endothelial cells
Fibroblasts
Monocytic lineage
CD8 T cells
NK cells
Myeloid dendritic cells
T cells
Cytotoxic lymphocytes
B lineage
20 15 10 5 0 −5
E
Values
0.2
0.0
−0.2
PC1 R
PC1 NR
PC2 R
PC2 NR
PC3 R
PC3 NR
PC4 R
PC4 NR
PC5 R
PC5 NR
PC6 R
PC6 NR
PC7 R
PC7 NR
PC8 R
PC8 NR
PC9 R
PC9 NR
PC10 R
PC10 NR
F
Values
0.4
0.2
0.0
−0.2
PC1 R
PC1 NR
PC2 R
PC2 NR
PC3 R
PC3 NR
PC4 R
PC4 NR
PC5 R
PC5 NR
PC6 R
PC6 NR
PC7 R
PC7 NR
PC8 R
PC8 NR
PC9 R
PC9 NR
PC10 R
PC10 NR


A
AUC
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
Expression
Random Expression
Genes + Clinical
Random Genes + Clinical
Clinical
Random Clinical
Pathways
Random Pathways
Cell types
Random Cell types
B
AUC
0.4
0.5
0.6
0.7
0.8
Expression
Random Expression
Genes + Clinical
Random Genes + Clinical
Clinical
Random Clinical
Pathways
Random Pathways
Cell types
Random Cell types


A
100
50
0
−50
Component 2 (7%)
−200
−100
0
Component 1 (32%)
A
B


B
4
2
0
$-\log_{10}$ percentage false positives
Number of DE Probes: 0
−2
0
2
Expression difference ($\log_2$)

Supplementary Data

Predictive biomarkers of sustained treatment-free remission in chronic myeloid leukemia: Gene expression analyses from the ENESTfreedom and ENESTop studies

Jerald P. Radich[1], Shalini Chaturvedi[2,*], Islam Sadek[2], Vanessa Obourn[3, †]

[1]Fred Hutchinson Cancer Center, Seattle, WA, United States

[2]Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States

[3]Novartis, Cambridge, MA, United States

Supplementary Methods 2

Supplementary Table S1. Analysis of clinical variables versus responder status. 6

Supplementary Table S2. Differential analysis results for ENESTfreedom study. 7

Supplementary Table S3. Differential analysis results for ENESTop study. 14

Supplementary Table S4. Pathway analysis in samples from ENESTfreedom. 15

Supplementary Table S5. Pathway analysis in samples from ENESTop. 16

Supplementary Table S6. Post-sequencing QC summary. 17

Supplementary Figure S1. Violin plots of differentially expressed genes in samples from the ENESTop study. 18

Supplementary Figure S2. Kernel density plots for samples from ENESTfreedom (A) and ENESTop (B). 19

Supplementary Figure S3. Quality control of ENESTfreedom (A, C) and ENESTop (B, D) samples. 20

Supplementary Figure S4. Univariate p-value distribution of gene expression versus responder status in samples from ENESTfreedom (A) and ENESTop (B). 22

Supplementary Figure S5. Graphical representation of PCA for responder and non-responder groups within the ENESTfreedom and ENESTop studies. 23

Supplementary Figure S6. Graphical representation of differentially expressed probes between non-responder and responder groups within the ENESTfreedom and ENESTop studies and between studies within responder groups. 24

Supplementary Methods

Clinical Studies

ENESTfreedom (ClinicalTrials.gov identifier: NCT01784068) was a phase 2, single-arm study evaluating patients with CML-CP who had achieved $MR^{4.5}$ after at least 2 years of frontline treatment with nilotinib. Patients must have achieved a sustained DMR for at least 1 year of consolidation nilotinib therapy within the study before stopping treatment.(1) Patients re-initiated nilotinib treatment if they experienced loss of MMR while in the TFR phase.

ENESTop (ClinicalTrials.gov identifier: NCT01698905) was a phase 2, single-arm study in patients with CML-CP who achieved sustained DMR following a switch from imatinib to nilotinib. Patients in $MR^{4.5}$ after at least three years of TKI treatment (>4 weeks with imatinib, then ≥2 years with nilotinib) attempted TFR after one year of nilotinib consolidation therapy within the trial.(2) Nilotinib treatment was re-initiated in patients who lost MMR or had a confirmed loss of MR^4 in two consecutive assessments within 4 weeks.

For both studies, successful TFR was defined as no loss of MMR and no re-starting of nilotinib therapy following cessation of nilotinib. Patients were monitored for up to 528 weeks (approximately 10 years) after the last patient entered the TFR phase. Final trial data from ENESTfreedom and ENESTop have been published.(3,4)

Gene expression profiling (GEP) of patients who relapsed during treatment-free remission (TFR) vs relapsed on nilotinib therapy

The data from the ENESTfreedom and ENESTop studies were analyzed separately, as the patient populations and pre-treatment regimens used were different. A total of 15,036 genes in the ENESTfreedom study and 14,320 genes in the ENESTop study were used for downstream analysis and normalized after gene filtration.

Prior to the downstream analysis, gene filtering was performed on the data sets to remove red blood cell-derived hemoglobin RNA (by removing RNA from three hemoglobin genes: *HBA1*, *HBA2* and *HBB*), followed by additional gene filtration to remove low-expression genes based on counts per million (CPM), to increase detection power.(5) To perform low-expression gene filtration, the number of patients with CPM >1 for each gene was counted; if this count was ≥2, the genes were not filtered. Following the pre-processing steps, the RNAseq data were normalized using the Trimmed Mean of M-values (TMM) normalization method.(6) The edgeR library of the Bioconductor package was used to normalize the RNAseq counts via the TMM method and to perform a $\log_2$ (counts per million +1) transformation. Genes

with low expression (*i.e.,* $\log_2$ (cpm+1) <1.1) were removed. The $\log_2$ (cpm) data was standardized by a Z-score transformation, such that the mean of each gene across all samples =0 and the standard deviation =1.

The gene-set enrichment analysis (GSEA) based on Gene Ontology (GO)(7) and Kyoto encyclopedia of genes and genomes (KEGG)(8) pathway analysis was performed using the linear models for microarray data (limma) package(9) in R software. Due to the small number of differentially expressed (DE) identified genes, a relaxed threshold was used to select DE input genes. For the ENESTop study, input DE genes were selected if nominal p-value was <1 and the absolute log2 fold change was >1. For the ENESTfreedom study, input DE genes were selected if nominal p-value was <0.05 and the absolute log2 fold change was >1.

Once the input genes were selected, the analysis was performed using this set of genes: input genes were split into two groups (up-regulated or down-regulated) and the same analysis was performed using each group as an input. GO gene sets were filtered based on size: gene sets with greater than 25 and less than 500 genes were used.

GEP of patients who relapsed during TFR vs those with sustained TFR following nilotinib therapy

Gene expression in counts format as the processed output of RNASeq were obtained. Filtration to remove RNA from a selected subset of highly expressed hemoglobin transcripts was performed as previously described. Clinical data other than *BCR::ABL1* ratio data (MMR/DMR, yes or no) were filtered for predictive analysis to include only variables that were reported at study baseline. The clinical features were compared against responder status using either a chi-square (for categorical variables) or Wilcoxon rank-sum test (for continuous variables); clinical variables which were significantly associated with responder status are presented in **Supplementary Table S1**. Inference of relative cell type abundance in each sample was performed using the MCPcounter algorithm applied to the $\log_2$(CPM+1) gene expression data. MCPcounter assigns a score for 10 different cell types that is proportional to the inferred amount of that cell type in a given sample.

Penalized logistic regression models were constructed from gene expression, clinical variables, inferred immune cell type compositions, and combinations thereof using the glmnet library in the R programming language. LASSO and Ridge regression models were constructed for all input data matrices and evaluated by bootstrapped area under the receiver operating characteristic curve (AUC). For each iteration of predictive analysis, bootstrapping (i.e., random sampling with replacement) of the input samples was performed to create a random subset of data (approximately 2/3 of total samples) on which

the model was trained. The remaining samples (approximately 1/3 of total samples) were held out as a test set to evaluate the model performance via AUC. Each input dataset was subject to 250 iterations of this procedure. Random models were generated for comparison by the standard practice of randomly shuffling the labels of good and poor responder and performing the same bootstrap AUC analysis. GSEA was used to interpret results with continuous output, such as the principal component rotation values, and hypergeometric tests (Enrichr) were used to interpret results in the form of gene sets, such as the output of LASSO regression. Biological pathways and gene ontology databases were used as references for these tests.

Transcriptome panel analyses of patients who relapsed during TFR vs those with sustained TFR following nilotinib therapy

Samples were processed as a single batch, with a total of 3 HTG EdgeSeq (HTG Molecular Diagnostics, Inc., Tucson, AZ) processing runs, resulting in 12 sequencing runs. The HTG EdgeSeq Reveal Application was utilized to quality check and normalize data, and to complete output analysis of annotated responder groups form blinded clinical outcome data. Post-sequencing quality control (QC) metrics (QC0, QC1, QC2, and QC3) were analyzed by the reveal analysis pipeline.

The HTG Transcriptome Panel (HTP) includes probes targeting most known human mRNA transcripts. These probes were selected against the National Center for Biotechnology Information (NCBI) human genomic database and screened against the NCBI reference sequence (RefSeq) RNA database for primary accession of the gene of interest. The HTP has shown equivalent results compared with RNA sequencing,(10) considered to be the gold standard for evaluating transcriptome data, using a fraction of the sample input. Furthermore, the HTP shows strong linearity across a wide dynamic range.

References

1. Hochhaus A, Masszi T, Giles FJ, Radich JP, Ross DM, Gómez Casares MT, Hellmann A, Stentoft J, Conneally E, García-Gutiérrez V, Gattermann N, Wiktor-Jedrzejczak W, le Coutre PD, Martino B, Saussele S, Menssen HD, Deng W, Krunic N, Bedoucha V, Saglio G. Treatment-free remission following frontline nilotinib in patients with chronic myeloid leukemia in chronic phase: results from the ENESTfreedom study. Leukemia. 2017;31(7):1525-1531.
2. Mahon FX, Boquimpani C, Kim DW, Benyamini N, Clementino NCD, Shuvaev V, Ailawadhi S, Lipton JH, Turkina AG, De Paz R, Moiraghi B, Nicolini FE, Dengler J, Sacha T, Takahashi N, Fellague-Chebra R, Acharya S, Wong S, Jin Y, Hughes TP. Treatment-Free Remission After Second-Line Nilotinib Treatment in Patients With Chronic Myeloid Leukemia in Chronic Phase:

Results From a Single-Group, Phase 2, Open-Label Study. Ann Intern Med. 2018;168(7):461-470.
3. Radich JP, Hochhaus A, Masszi T, Hellmann A, Stentoft J, Casares MTG, García-Gutiérrez JV, Conneally E, le Coutre PD, Gattermann N, Martino B, Saussele S, Giles FJ, Ross DM, Aimone P, Li S, Titorenko K, Saglio G. Treatment-free remission following frontline nilotinib in patients with chronic phase chronic myeloid leukemia: 5-year update of the ENESTfreedom trial. Leukemia. 2021;35(5):1344-1355.
4. Hughes TP, Clementino NCD, Fominykh M, Lipton JH, Turkina AG, Moiraghi EB, Nicolini FE, Takahashi N, Sacha T, Kim DW, Fellague-Chebra R, Tiwari R, Bouard C, Mahon FX. Long-term treatment-free remission in patients with chronic myeloid leukemia after second-line nilotinib: ENESTop 5-year update. Leukemia. 2021;35(6):1631-1642.
5. Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. Proc Natl Acad Sci U S A. 2010;107(21):9546-51.
6. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.
7. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9.
8. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.
9. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14.
10. Ran D, Moharil J, Lu J, Gustafson H, Culm-Merdek K, Strand-Tibbitts K, Benjamin L, Navratil M. Platform comparison of HTG EdgeSeq and RNA-Seq for gene expression profiling of tumor tissue specimens. J Clin Oncol. 2020;38:3566-3566.

Supplementary Table S1. Analysis of clinical variables versus responder status.

Variable	***p*-value**
Age (continuous)	4.35E-01
Sex	2.56E-01
Race	5.50E-01
Sokal Score at diagnosis	9.45E-02
Duration of nilotinib treatment prior to study entry	2.17E-01
Duration of nilotinib treatment prior to achieving TFR	1.79E-01
Time to $MR^{4.5}$ prior to nilotinib treatment	2.67E-01
Time since $MR^{4.5}$ prior to study entry	6.35E-02
Time since $MR^{4.5}$ prior to achieving TFR	5.20E-02

Time between first achieving $MR^{4.5}$ and either entering the study or first achieving TFR was not determined at a central laboratory prior to study entry. MR, molecular response; TFR, treatment-free remission.

Supplementary Table S2. Differential analysis results for ENESTfreedom study.
Median values of TMM and CPM are provided. Genes are called to be differentially expressed if FDR adjusted p-value is less than 0.2 (FDR controlled by 20%).

Genes	logFC	logCPM	P value	FDR	rTRT.TLL.med	rTFR.TMM.med	rTRT.CPM.med	rTFR.CPM.med
RBFOX3	2.26	-0.14	<0.01	0	0.34	0.26	0.32	0.28
ASB4	3.4	-1.5	<0.01	0	0	0	0	0
HOXC8	3.74	-1.04	<0.01	0	0.06	0	0.04	0
CACNG2	3.76	-1.12	<0.01	0	0	0	0	0
CSTA	1.32	4.08	<0.01	0	14.4	11.4	14.76	11.14
CMTM2	0.88	6.1	<0.01	0	83.66	60.14	78.64	59.36
DRC1	2.12	-0.54	<0.01	0	0.26	0.18	0.26	0.18
FAM155A	3.64	-0.98	<0.01	0	0	0	0	0
FOXA2	3.28	-1.34	<0.01	0	0	0	0	0

Genes	logFC	logCPM	P value	FDR	rTRT.TLL.med	rTFR.TMM.med	rTRT.CPM.med	rTFR.CPM.med
TMEM70	0.9	3.24	<0.01	0	9.24	7.84	9.42	7.68
G0S2	1.5	1.2	<0.01	0	2.78	1.08	2.52	1.14
SOX11	2.86	-1.06	<0.01	0	0.02	0	0.02	0
CELF4	2.32	-0.5	<0.01	0.02	0.16	0.12	0.16	0.12
CADPS2	3.06	-0.82	<0.01	0.02	0	0	0	0
PRDM6	3.14	-1.44	<0.01	0.02	0	0	0	0
ONECUT2	2.96	-0.9	<0.01	0.02	0.1	0	0.1	0
RALYL	3.52	-1.1	<0.01	0.02	0	0	0	0
BRI3	0.54	6.62	<0.01	0.02	118.98	89.36	119.08	86.1
EN2	3.28	-1.1	<0.01	0.02	0	0	0	0
RND1	2.24	-0.82	<0.01	0.02	0	0	0	0
LHX2	3.1	-0.98	<0.01	0.02	0.1	0	0.1	0
LOC643802	1.44	3.78	<0.01	0.02	9.6	7.6	9.3	7.46
LY96	1.28	3.34	<0.01	0.02	6.36	6.46	6.48	6.48
DUSP23	0.6	3.3	<0.01	0.02	11.28	8.74	10.36	9.12
MSRB2	0.66	3.9	<0.01	0.02	18.46	13.26	17.34	13.16
S100A12	1.04	6.98	<0.01	0.02	169.22	92.82	155.8	94.42

Genes	logFC	logCPM	PValue	FDR	rTRT.TLL.med	rTFR.TMM.med	rTRT.CPM.med	rTFR.CPM.med
COX7B	1.62	3.1	<0.01	0.02	5.4	3.64	4.78	3.78
GMFG	0.72	7.26	<0.01	0.02	164.58	132.96	160.64	133.64
ISG15	0.88	5.62	<0.01	0.02	77.78	36.62	72.2	38.32
KIAA1456	2.56	-0.8	<0.01	0.02	0.22	0	0.22	0
MYL12B	0.48	7.6	<0.01	0.02	245.32	184.74	224.96	184.44
S100A9	0.66	11.98	<0.01	0.02	5091.78	3516.3	4876.5	3512.4
SPATS2L	0.76	2.84	<0.01	0.02	8.68	5.84	7.96	5.96
ADGRL4	2.96	-1.16	<0.01	0.02	0	0	0	0
TFAP2D	3.08	-1.18	<0.01	0.02	0	0	0	0
A1CF	2.46	-1.28	<0.01	0.02	0	0	0	0
FAM32A	0.42	5.36	<0.01	0.02	44.86	38.3	43.26	37.58
IFI6	0.98	5.6	<0.01	0.02	94.06	33.9	84.16	34.44
NDUFA2	0.52	4.56	<0.01	0.02	25.44	21.44	23.82	21.26
USP10	0.5	6.96	<0.01	0.02	132.7	115.24	134.02	118.7
ARMC9	1.78	0.04	<0.01	0.02	0.52	0.34	0.6	0.32
FAM78B	1.92	-0.76	<0.01	0.02	0.4	0.14	0.38	0.12
ZNF318	-0.42	4.98	<0.01	0.02	24.74	32.92	23.34	32.74

Genes	logFC	logCPM	PValue	FDR	rTRT.TLL.med	rTFR.TMM.med	rTRT.CPM.med	rTFR.CPM.med
BAZ1A	0.78	7.24	<0.01	0.02	129.06	126.3	126.76	128.46
MLH3	0.72	3.78	<0.01	0.02	12.44	11.76	12.14	11.68
SOGA1	-0.68	3.84	<0.01	0.02	8.66	14.48	8.26	14.2
S100A8	1.36	9.5	<0.01	0.02	910.98	384.56	851.52	376.68
BCL2A1	1.48	4.18	<0.01	0.04	14.86	8.44	12.78	8.66
LIMCH1	1.9	-0.34	<0.01	0.04	0.32	0.24	0.32	0.26
SUSD2	1.48	-0.1	<0.01	0.04	0.42	0.38	0.38	0.4
SUCLG1	0.38	4.9	<0.01	0.04	38.94	28.06	35.06	28.7
PFDN5	1.26	5.96	<0.01	0.04	41.1	36.18	38.86	35.74
ACP5	0.64	4.82	<0.01	0.04	28.52	23.96	30.64	23.46
POLE4	0.7	3.28	<0.01	0.04	8.7	8.8	8.86	8.34
CYLC2	3.08	-1.32	<0.01	0.04	0	0	0	0
MRPL46	0.64	2.44	<0.01	0.04	6	4.7	5.8	4.74
MYO1E	-0.7	3.5	<0.01	0.04	7.3	11.52	7.68	12
NKAIN2	2.7	-1.16	<0.01	0.04	0.16	0	0.16	0
NOP10	0.66	5.1	<0.01	0.04	35.64	30.46	34.42	30.38
S100P	1.2	5.42	<0.01	0.04	52.54	30.68	48.38	32.94
TMA7	1.22	4.34	<0.01	0.04	13.84	12.22	13.3	11.72

Genes	logFC	logCPM	PValue	FDR	rTRT.TLL.med	rTFR.TMM.med	rTRT.CPM.med	rTFR.CPM.med
TMEM119	1.34	0.86	<0.01	0.04	1.3	1.04	1.26	0.98
TRPM3	2.42	-1.16	<0.01	0.04	0.06	0	0.06	0
UBL5	0.62	5.18	<0.01	0.04	34.54	32.26	34.5	30.82
HEBP2	0.56	4.96	<0.01	0.04	32.86	28.96	32.38	28.04
ZNRD1	0.46	3.86	<0.01	0.04	17.26	13.36	15.58	13.18
ARPC3	0.48	8.36	<0.01	0.04	402.26	297.22	379.58	307.68
NECTIN4	2.68	-0.92	<0.01	0.04	0.1	0	0.08	0
ZFHX4	2.52	-0.7	<0.01	0.04	0.06	0	0.04	0
S100A8	1.36	9.5	<0.01	0.02	910.98	384.56	851.52	376.68
BCL2A1	1.48	4.18	<0.01	0.04	14.86	8.44	12.78	8.66
LIMCH1	1.9	-0.34	<0.01	0.04	0.32	0.24	0.32	0.26
SUSD2	1.48	-0.1	<0.01	0.04	0.42	0.38	0.38	0.4
SUCLG1	0.38	4.9	<0.01	0.04	38.94	28.06	35.06	28.7
PFDN5	1.26	5.96	<0.01	0.04	41.1	36.18	38.86	35.74

Genes	logFC	logCPM	PValue	FDR	rTRT.TLL.med	rTFR.TMM.med	rTRT.CPM.med	rTFR.CPM.med
ACP5	0.64	4.82	<0.01	0.04	28.52	23.96	30.64	23.46
POLE4	0.7	3.28	<0.01	0.04	8.7	8.8	8.86	8.34
CYLC2	3.08	-1.32	<0.01	0.04	0	0	0	0
MRPL46	0.64	2.44	<0.01	0.04	6	4.7	5.8	4.74
MYO1E	-0.7	3.5	<0.01	0.04	7.3	11.52	7.68	12
NKAIN2	2.7	-1.16	<0.01	0.04	0.16	0	0.16	0
NOP10	0.66	5.1	<0.01	0.04	35.64	30.46	34.42	30.38
S100P	1.2	5.42	<0.01	0.04	52.54	30.68	48.38	32.94
TMA7	1.22	4.34	<0.01	0.04	13.84	12.22	13.3	11.72
TMEM119	1.34	0.86	<0.01	0.04	1.3	1.04	1.26	0.98
TRPM3	2.42	-1.16	<0.01	0.04	0.06	0	0.06	0
UBL5	0.62	5.18	<0.01	0.04	34.54	32.26	34.5	30.82
HEBP2	0.56	4.96	<0.01	0.04	32.86	28.96	32.38	28.04
ZNRD1	0.46	3.86	<0.01	0.04	17.26	13.36	15.58	13.18
ARPC3	0.48	8.36	<0.01	0.04	402.26	297.22	379.58	307.68
NECTIN4	2.68	-0.92	<0.01	0.04	0.1	0	0.08	0

Genes	logFC	logCPM	PValue	FDR	rTRT.TLL.med	rTFR.TMM.med	rTRT.CPM.med	rTFR.CPM.med
ZFHX4	2.52	-0.7	<0.01	0.04	0.06	0	0.04	0
AMBN	2.34	-1.28	<0.01	0.04	0.08	0	0.08	0
BVES	2.36	-1.08	<0.01	0.04	0	0	0	0
HIST1H2AE	1.14	1.02	<0.01	0.04	1.38	1.36	1.38	1.4
SNX29	-0.42	5.4	<0.01	0.04	32.38	42.82	30.54	42.24
SYNPO	-0.92	3.36	<0.01	0.04	6.34	9.18	6.4	8.84

CPM, counts per million; FDR, p-value adjusted for multiple testing; logCPM, average log 2 count per million (abundance); logFC, log 2 fold change (log-abundance ratio); P value, two sided p-value for differential expression using the negative binomial model; rTFR.CPM.med, median CPM normalized value of relapse on TFR group; rTRT.CPM.med, median CPM normalized value of relapse on treatment group; rTFR.TMM.med, median TMM normalized value of relapse on TFR group; rTRT.TMM.med, median TMM normalized value of relapse on treatment group; TMM, Trimmed Mean of M-values normalization method; TFR, treatment-free remission.

Supplementary Table S3. Differential analysis results for ENESTop study.
Median values of TMM and CPM are provided. Genes are called to be differentially expressed if FDR adjusted p-value is less than 0.2 (FDR controlled by 20%).

Genes	logFC	logCPM	P value	FDR	rTRT.TLL.med	rTFR.TMM.med	rTRT.CPM.med	rTFR.CPM.med
HBE1	6.2	0.98	<0.01	0	0.2	0	0.2	0
CD177	2.52	4.74	<0.01	0	11.7	7.7	10.08	8.3
FOLR3	1.36	4.46	<0.01	0.02	17.34	14.9	14.1	14.34
IGFBP2	1.9	0.86	<0.01	0.04	2.34	0.76	2.1	0.72
HLA-DQA1	-1.48	6.76	<0.01	0.06	34.34	106.54	31.1	106.34
ADCY6	1.12	-0.1	<0.01	0.12	1.22	0.5	1.14	0.52
AQP10	1.32	3.28	<0.01	0.12	15.2	5.62	11.9	6.34
MMP1	1.9	-0.44	<0.01	0.12	0.78	0.16	0.64	0.16
PRTFDC1	1.2	0.58	<0.01	0.14	2.68	0.92	2.36	1.02

CPM, counts per million; FDR, p-value adjusted for multiple testing; logCPM, average log 2 count per million (abundance); logFC, log 2 fold change (log-abundance ratio); P value, two sided p-value for differential expression using the negative binomial model; rTFR.CPM.med, median CPM normalized value of relapse on TFR group; rTRT.CPM.med, median CPM normalized value of relapse on treatment group; rTFR.TMM.med, median TMM normalized value of relapse on TFR group; rTRT.TMM.med, median TMM normalized value of relapse on treatment group; TMM, Trimmed Mean of M-values normalization method; TFR, treatment-free remission.

Supplementary Table S4. Pathway analysis in samples from ENESTfreedom. Differentially expressed genes with nominal p-value <0.1 and absolute log2 fold-change >1 were used. Fold change was calculated as FC = relapse on TRT/relapse on TFR. P value is provided for over-representation of the pathway in the set.

ID	Pathway	N	DE	P value
path:hsa03010	Ribosome	153	13	<0.01
path:hsa02010	ABC Transporters	44	5	<0.01
path:hsa05033	Nicotine Addiction	40	4	<0.01
path:hsa05133	Pertussis	76	5	<0.01
path:hsa00591	Linoleic Acid Metabolism	29	3	0.01
path:hsa04727	Gabaergic Synapse	88	5	0.01
path:hsa05322	Systemic Lupus Erythematosus	133	6	0.02
path:hsa05034	Alcoholism	180	7	0.02
path:hsa04975	Fat Digestion and Absorption	41	3	0.03
path:hsa05412	Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC)	72	4	0.03
path:hsa05202	Transcriptional Misregulation in Cancer	186	7	0.03
path:hsa04723	Retrograde Endocannabinoid Signaling	148	6	0.03
path:hsa04940	Type I Diabetes Mellitus	43	3	0.03
path:hsa05410	Hypertrophic Cardiomyopathy (HCM)	83	4	0.04

DE, number of genes in the DE set; N, number of genes in the pathway; TFR, treatment-free remission; TRT, treatment.

Supplementary Table S5. Pathway analysis in samples from ENESTop.
Differentially expressed genes with nominal p-value <0.1 and absolute log2 fold-change >1 were used. Fold change was calculated as FC = relapse on TRT/relapse on TFR. P value is provided for over-representation of the pathway in the set.

ID	Pathway	N	DE	P value
path:hsa04080	Neuroactive Ligand-Receptor Interaction	277	8	<0.01
path:hsa00360	Phenylalanine Metabolism	17	2	0.01
path:hsa05030	Cocaine Addiction	49	3	0.01
path:hsa04972	Pancreatic Secretion	96	4	0.01
path:hsa04923	Regulation of Lipolysis in Adipocytes	54	3	0.01
path:hsa03010	Ribosome	153	5	0.02
path:hsa00340	Histidine Metabolism	23	2	0.02
path:hsa03320	PPAR Signaling Pathway	74	3	0.03
path:hsa04610	Complement and Coagulation Cascades	79	3	0.04
path:hsa04742	Taste Transduction	83	3	0.05
path:hsa00350	Tyrosine Metabolism	36	2	0.05

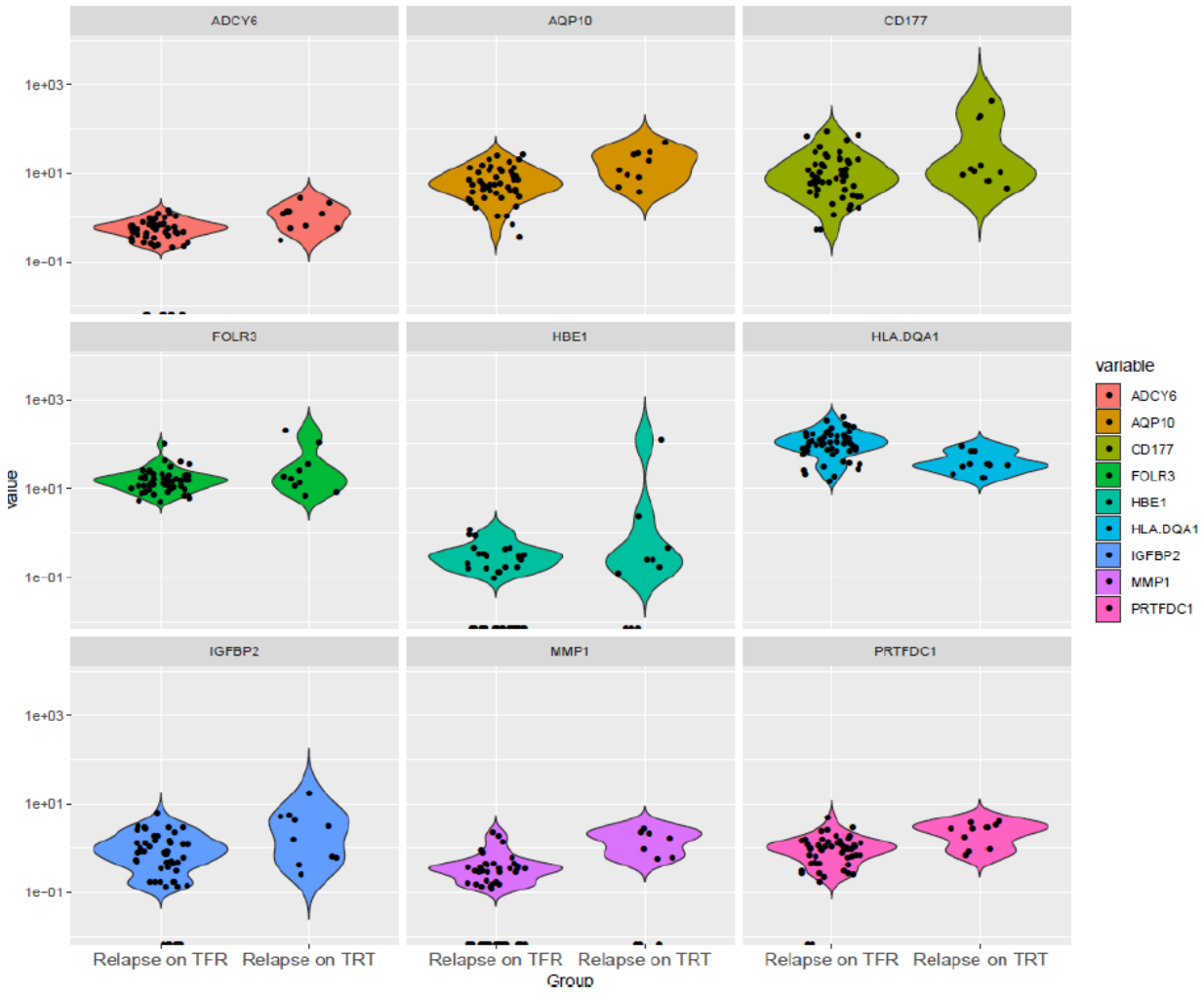
DE, number of genes in the DE set; N, number of genes in the pathway; TFR, treatment-free remission; TRT, treatment.

Supplementary Table S6. Post-sequencing QC summary.

QC Status	ENESTfreedom	ENESTop
QC0 failures	29	17
QC1 failures	1	2
QC2 failures	5	5
QC3 failures	13	5
Pass	128	69
Total	165 (78% pass)	109 (63% pass)

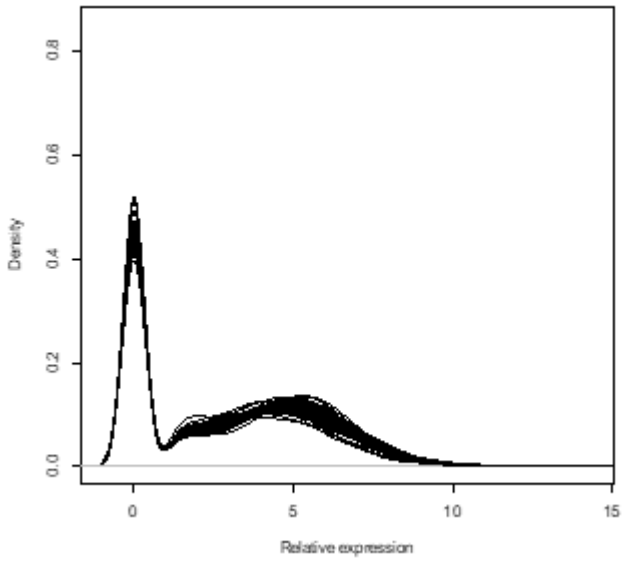
QC, quality control.

Supplementary Figure S1. Violin plots of differentially expressed genes in samples from the ENESTop study.
Width of each plot shows the distribution of the probability density. Black dots represent individual samples. TFR, treatment-free remission; TRT, treatment.

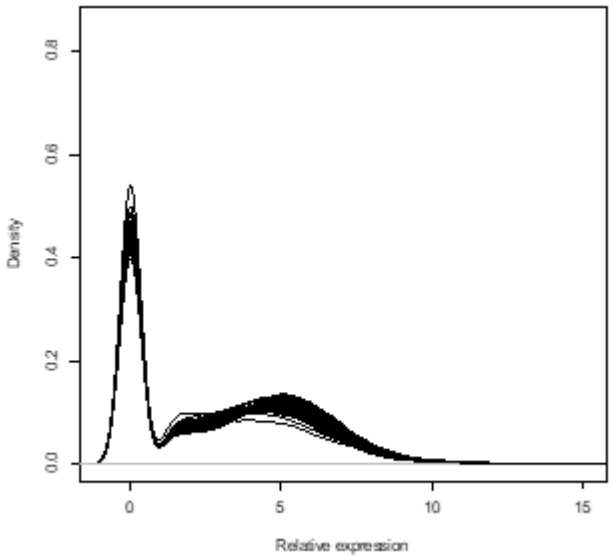


Supplementary Figure S2. Kernel density plots for samples from ENESTfreedom (**A**) and ENESTop (**B**).

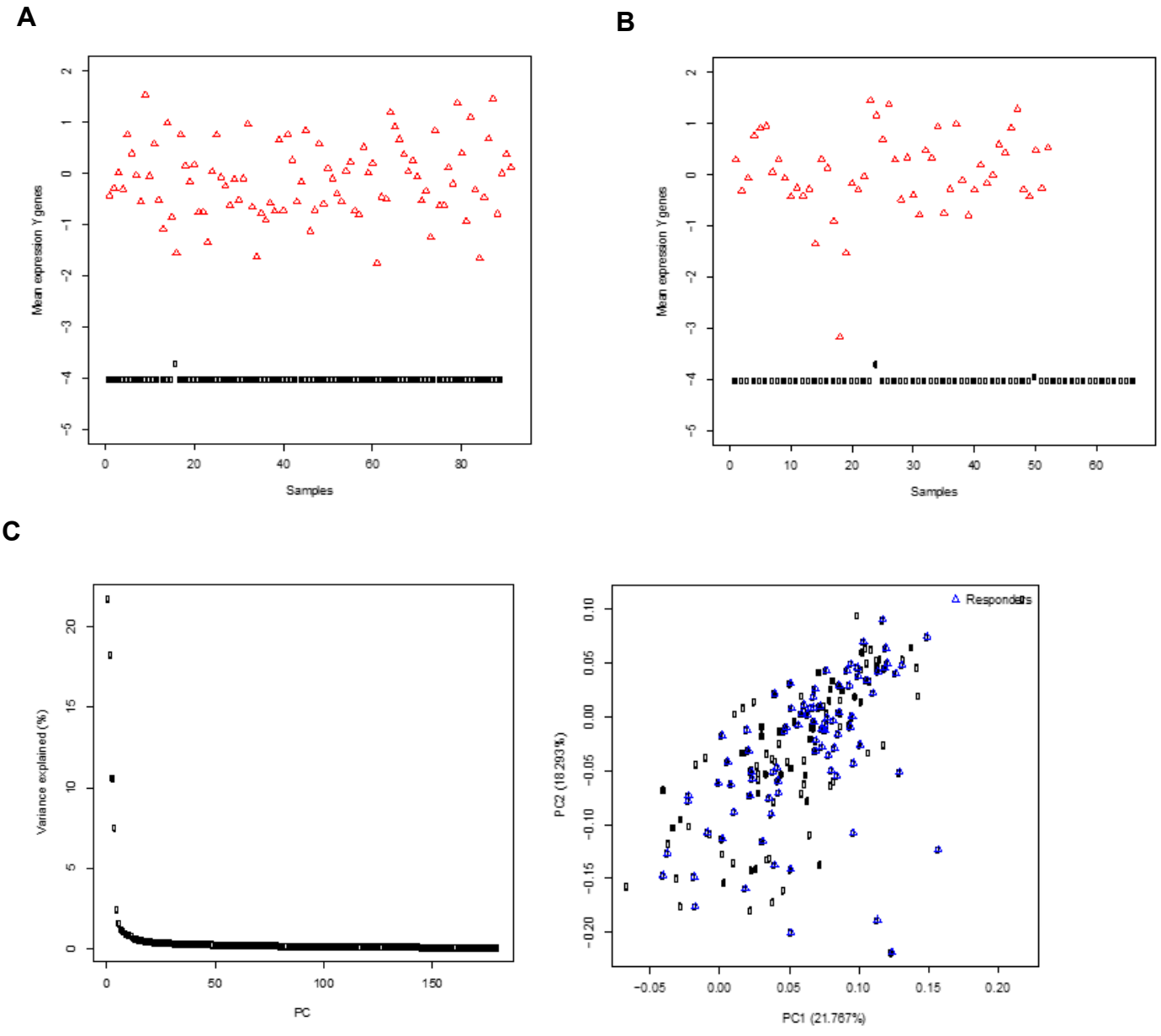
A



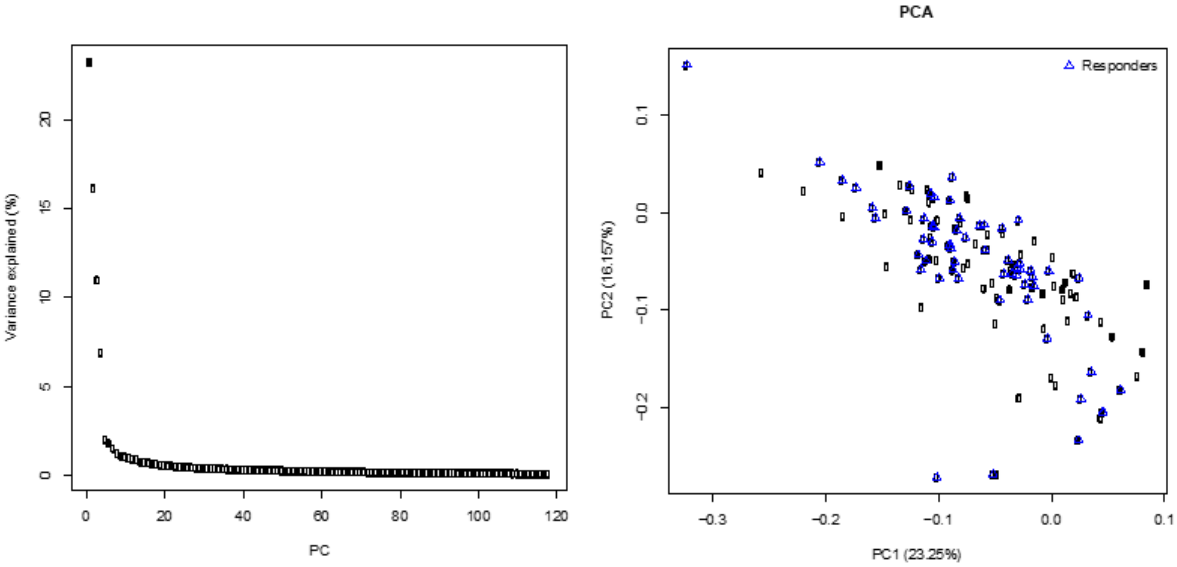
B



Supplementary Figure S3. Quality control of ENESTfreedom (**A, C**) and ENESTop (**B, D**) samples.
A and **B**, mean Y-chromosome expression versus sex. Red triangles indicate male patients and black circles are female patients. **C** and **D**, variance explained by principal components of gene expression (left panel) and PCA plot (right panel). The variance explained plot shows that all of the meaningful information is in the first strongest signals in the data. In the PCA plot, good responders are shown as blue triangles and poor responders are shown as black circles; no separation was observed between groups. PCA, principal component analysis.

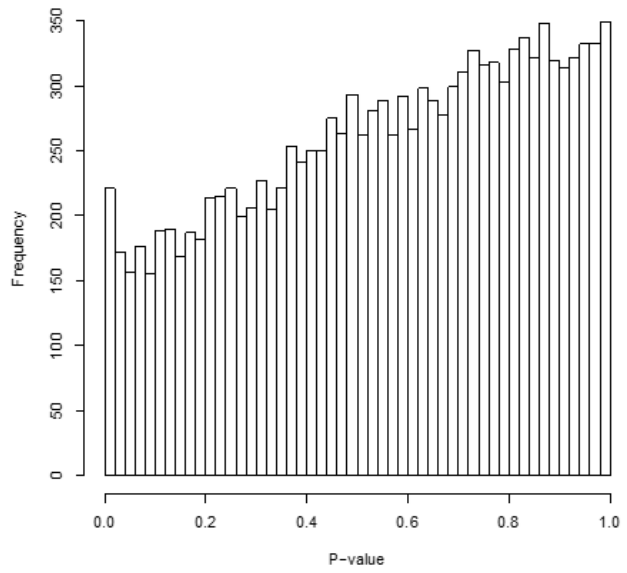


D

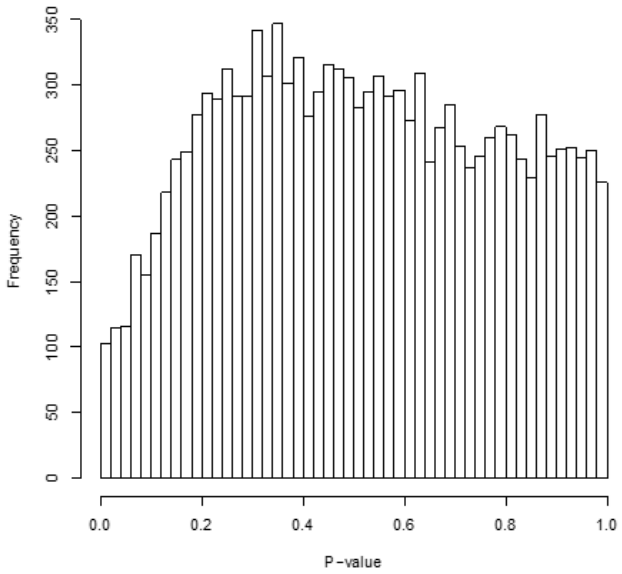


Supplementary Figure S4. Univariate p-value distribution of gene expression versus responder status in samples from ENESTfreedom (A) and ENESTop (B).
The Wilcoxon rank-sum test was performed for the expression of each gene in good versus poor responders. The p-values of these histograms (corresponding to the differential expression of each gene) were not enriched for small values and did not show a uniform distribution of insignificant p-values; this suggests that the FDR would be unacceptably high. FDR, false discovery rate.

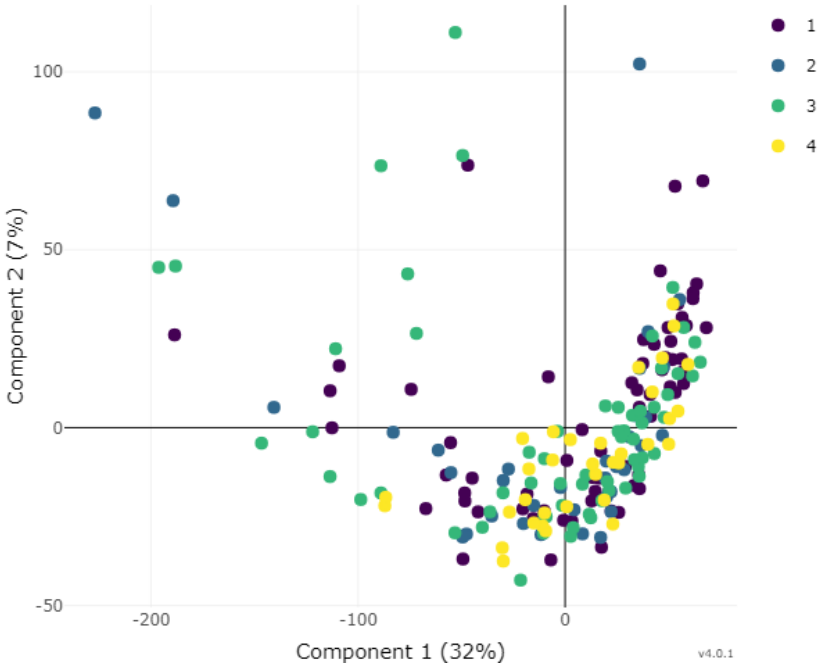
A



B



Supplementary Figure S5. Graphical representation of PCA for responder and non-responder groups within the ENESTfreedom and ENESTop studies.
Group 1, ENESTfreedom non-responders; group 2, ENESTop non-responders; group 3, ENESTfreedom responders; group 4, ENESTop responders. PCA, principal component analysis.



Supplementary Figure S6. Graphical representation of differentially expressed probes between non-responder and responder groups within the ENESTfreedom and ENESTop studies and between studies within responder groups.
In the volcano plot, y-axis shows the negative logarithm of the p value and x-axis shows the logarithm of the fold change between the tested variables. Each gene is represented as a dot on the graph. Group 1, ENESTfreedom non-responders; group 2, ENESTop non-responders; group 3, ENESTfreedom responders; group 4, ENESTop responders.