Molecular response in newly diagnosed chronic-phase chronic myeloid leukemia: prediction modeling and pathway analysis

Jerald P. Radich,[1*] Matthew Wall,[2*] Susan Branford,[3] Catarina D. Campbell,[4] Shalini Chaturvedi,[5] Daniel J. DeAngelo,[6] Michael W. Deininger,[7] Justin Guinney,[2] Andreas Hochhaus,[8] Timothy P. Hughes,[9] Hagop M. Kantarjian,[10] Richard A. Larson,[11] Sai Li,[12] Rodrigo Maegawa,[5] Kaushal Mishra,[5] Vanessa Obourn,[4] Javier Pinilla-Ibarz,[13] Das Purkayastha,[12] Islam Sadek,[5] Giuseppe Saglio,[14] Alok Shrestha,[5] Brian S. White[2] and Brian J. Druker[15]

[1]Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [2]Sage Bionetworks, Seattle, WA, USA; [3]SA Pathology, Centre for Cancer Biology, Adelaide, South Australia, Australia; [4]Novartis Institutes for Biomedical Research, Cambridge, MA, USA; [5]Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA; [6]Dana-Farber Cancer Institute, Boston, MA, USA; [7]Huntsman Cancer Institute, The University of Utah, Salt Lake City, UT, USA; [8]Universitätsklinikum Jena, Jena, Germany; [9]South Australian Health and Medical Research Institute and University of Adelaide, Adelaide, South Australia, Australia; [10]MD Anderson Cancer Center, University of Texas, Houston, TX, USA; [11]Medicine-Hematology/Oncology, University of Chicago, Chicago, IL, USA; [12]Novartis Pharmaceuticals Corporation, Basel, Switzerland; [13]Moffitt Cancer Center, University of South Florida, Tampa, FL, USA; [14]University of Turin, Turin, Italy and [15]Knight Cancer Institute, Oregon Health and Science University, Portland, OR, USA

JPR and MW contributed equally as co-first authors.

Correspondence: J.P. Radich
jradich@fredhutch.org

Received: August 5, 2022.
Accepted: January 20, 2023.
Early view: February 2, 2023.

https://doi.org/10.3324/haematol.2022.281878

Supplemental Information

Molecular response in newly diagnosed chronic-phase chronic myeloid leukemia: prediction modeling and pathway analysis

Jerald P. Radich,[1,*] Matthew Wall,[2,*] Susan Branford,[3] Catarina D. Campbell,[4] Shalini Chaturvedi,[5] Daniel J. DeAngelo,[6] Michael Deininger,[7] Justin Guinney,[2] Andreas Hochhaus,[8] Timothy P Hughes,[9] Hagop M. Kantarjian,[10] Richard A. Larson,[11] Sai Li,[12] Rodrigo Maegawa,[5] Kaushal Mishra,[5] Vanessa Obourn,[4] Javier Pinilla-Ibarz,[13] Das Purkayastha,[12] Islam Sadek,[5] Giuseppe Saglio,[14] Alok Shrestha,[5] Brian S. White,[2] and Brian J. Druker[15]

Methods

Sample processing and quality control. All RNA was extracted from whole blood in PAXgene (Qiagen) tubes using the Maxwell (Promega) system and was analyzed for quality and quantity using Ribogreen (Thermo Fisher Scientific) and TapeStation (Agilent) quality control (QC) assays.

***Genomic analysis*.** Normalized aliquots of extracted RNA were treated with GLOBINclear (Invitrogen) according to the manufacturer's protocol to deplete α and β globin mRNA from the total RNA preparation. The globin-depleted RNA was evaluated for content and integrity and then used in a TruSeq RNA v2 (Illumina) preparation for RNA sequencing (RNA-seq) library construction. Constructed RNA-seq libraries were evaluated for yield and average size of library fragments. Libraries were then loaded onto the HiSeq 2500 (Illumina) rapid run for 2 × 100–base pair sequencing to a target depth of 50 million single reads per sample. Libraries were originally processed in 6 batches, each containing a process control (cell line K562) and 2 replicate samples with sufficient RNA to be included in all batches. Batches were also matched from arm and response groups, and samples from the same patient were sequenced in the same batch when possible. **Table S1** and **Figures S6-S8** show QC information for processed samples and sequenced libraries.

Sequence analysis pipeline. Raw sequence data were quality controlled using the FastQC package (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). A number of QC metrics were assessed to ensure high-quality data, including data quality and guanine-cytosine content on per-base and per-sequence levels, sequence length distribution and duplication levels, and insert size distribution. Reads were aligned to the reference human genome (build hg19) using STAR.(1) Next, mapped reads were assembled into transcripts, and transcript abundances were estimated and normalized using Cufflinks.(2) Finally, HTSeq was used to count the number of reads mapping to each gene.(3)

Quality control of RNA-seq data. We assessed key quality metrics of the sequencing data before performing downstream analyses. Almost all of the libraries achieved the target read depth of 50 million reads (**Figure S6**). We assessed the evenness of coverage across transcripts to detect 3' biases that could be indicative of RNA degradation (**Figure S7**). As expected, we observed that libraries with lower RNA integrity number (RIN) tended to have slightly more 3' bias. We observed 1 library with a strong 3' bias that seemed unrelated to RIN; this library was reconstructed and the bias corrected. We also confirmed that the proportion of reads aligning to the α- and β-globin genes was low.

Assessment for potential library construction batch effects. The library construction was performed in 7 automated batches. We included a process control and 2 replicate samples in the first 6 batches (the seventh batch consisted of library construction rework for only 3 samples). To determine whether there were large batch effects, we compared the global gene expression pattern across the replicate samples and observed high correlations (**Figure S8**). In addition, we clustered the sequenced samples based on the expression of the top 500 most-variable genes. The replicate samples cluster together suggested no strong batch effects.

Quality control of gene expression data. Kernel density plots of all samples (i.e., expression profiles from individual patients) were inspected to verify acceptable gene expression distributions (**Figure S9**). The density of undetected transcripts (i.e., expression of 0) was <20% for nearly all samples such that there were no immediate concerns regarding the sensitivity of the RNA-seq experiments. Principal components analysis (PCA) of the gene expression data was performed to confirm or deny the presence of any distinct populations (unidentified batch effects) and quantify the overall structure of the data. The remaining clinical data were filtered for predictive analysis to include only variables that were reported at baseline.

All noncontinuous variables were one-hot encoded such that they had values of 0 or 1. The clinical features were compared against responder status using either a χ^2 or Wilcoxon rank-sum test, as appropriate. Sex exhibited the most significant stratification of responder status among the clinical variables, with women significantly more likely to be responders (*P*<.001) (**Table 2**). The treatment type also influenced prognosis, with nilotinib treatment yielding better outcomes than imatinib (*P*=.056). Quality control analysis of the ENESTnd gene expression data exhibited sufficiently high quality for downstream analysis. We confirmed that the mean expression of genes on the Y chromosome were bimodally distributed and that the labels provided for sex agreed with the expectation of higher expression in men (**Figure S10**). PCA indicated that the data possessed underlying structure (i.e., not just noise) (**Figure S11**) and did not cluster into batches requiring correction (**Figure S12**). The variance explained by the first PC was larger than typically observed (observed, 28%; typical, 22%-26%) but not sufficiently high to raise concern of unidentified batch effects. The first 10 PCs were compared against clinical features using a Spearman correlation, Kruskal-Wallis, or Wilcoxon rank-sum test as appropriate to investigate whether these features underlaid specific structure within the gene expression. Significant associations between PC1 and PC2 with race, PC2 and PC3 with previous treatment with imatinib, and PC4 with Sokal risk score at diagnosis were identified (**Figure S13**). The most significant association was between race and PC1, but a labeled PCA plot did not suggest sufficient separation to warrant batch correction to the expression data (**Figure S14**). The PC were then compared against the inferred cell type scores from MCP-counter, revealing strong associations between immune cells and PC1 and PC2 (**Figure S15**). Finally, the first 10 PCs were compared against responder status using a Wilcoxon rank-sum test, revealing that PC6 (*P*<.01) and PC9 (*P*<.001) were significantly associated with response (**Figure S16**).

Bootstrapped prediction of responder status. Penalized logistic regression models were constructed from gene expression, clinical variables, normalized enrichment scores of biological pathways, and

inferred cell type compositions using the glmnet library in the R programming language. 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 on which the model was trained, and the remaining samples were held out as a test set to evaluate the model's performance via area under the curve (AUC). Each input data set was subject to 250 iterations of this procedure. The significance of the model performance was evaluated by comparing the 95% CI of the AUC distribution to AUC = 0.5, which corresponds to a random model. If the lower bound of the 95% CI was >0.5, the model performance was determined to be significant. Normal distributions were fitted to the bootstrap AUC values to estimate the proportion of bootstrap iterations expected to have random-or-worse performance. These proportions are given as empirical *P*-value estimates. A final predictive model was trained on all of the ENESTnd samples using logistic ridge regression.

Prediction of responder status in validation cohort. A final logistic ridge regression model was trained on the gene expression data of all 112 ENESTnd samples. The Branford et al. validation data set was batch-corrected via Combat to align with ENESTnd, then the ENESTnd-trained model was applied to the batch-corrected validation data. A random AUC distribution was generated by repeatedly shuffling the responder labels of the validation data, then applying the ENESTnd-trained model. The performance of the ENESTnd-trained model on the validation data was determined by AUC, and an empirical *P* value was calculated as the proportion of random AUC values that performed as well or better than the observed AUC.

References

1. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
2. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-578.

3. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-169.

Supplemental Tables and Figures

Table S1. RNA-seq quality control metrics summary. Max, maximum; min, minimum; UTR, untranslated region.

	Mean	SD	Min	Max
Total reads, n	9.96E+07	2.83E+07	4.56E+07	2.34E+08
Proportion coding bases	0.49	0.03	0.43	0.62
Proportion UTR bases	0.32	0.02	0.27	0.40
Proportion intronic bases	0.12	0.03	0.03	0.20
Proportion intergenic bases	0.08	0.01	0.05	0.12
Proportion of reads mapping to globin genes	4.88E−03	6.66E−03	1.82E−04	6.05E−02

Table S2:
https://drive.google.com/file/d/1FASei9Cl6j3iH8xtg0P0aIJvJkEgtndO/view?usp=sharing

Table S3:
https://drive.google.com/file/d/1ugTG2eqGuWHqo7kQG8Inm0cKqypgO40Z/view?usp=sharing

Table S4:
https://drive.google.com/file/d/1Ah67P7RzMDflN-DL0f6Tjs4lkWMKk9iT/view?usp=sharing

Table S5. Summary statistics of blood cell count data versus responder status.

Cell Type	Welch's t-test statistic	Welch's t-test p-value	Wilcoxon p-value	Wilcoxon FDR
Absolute Bands	-1.725	0.092	0.095	0.212
Absolute Basophils	-0.515	0.608	0.301	0.387
Absolute Blasts	-2.011	0.047	0.042	0.212
Other Absolute	-0.589	0.558	0.343	0.387
Absolute Eosinophils	-1.434	0.155	0.668	0.668
Absolute Lymphocytes	-1.725	0.088	0.568	0.590
Absolute Metamyelocytes	-1.990	0.050	0.061	0.212
Absolute Monocytes	-3.659	0.0004	0.006	0.149
Absolute Myelocytes	-2.273	0.026	0.065	0.212
Absolute Neutrophils	-2.530	0.013	0.151	0.262
Absolute Promyelocytes	-1.669	0.100	0.040	0.212
Bands (%)	-1.760	0.083	0.098	0.212
Basophils (%)	0.844	0.401	0.313	0.387
Blasts (%)	-0.508	0.613	0.098	0.212
Other (%)	-0.651	0.517	0.343	0.387
Platelets	1.610	0.112	0.070	0.212
Eosinophils (%)	0.692	0.491	0.305	0.387
Hematocrit	1.027	0.307	0.332	0.387
Hemoglobin	0.948	0.346	0.312	0.387
Lymphocytes (%)	1.452	0.151	0.184	0.299
Metamyelocytes (%)	-1.085	0.281	0.148	0.262
Monocytes (%)	-2.906	0.005	0.089	0.212
Myelocytes (%)	-1.630	0.106	0.108	0.217
Neutrophils (%)	0.844	0.401	0.467	0.506
Promyelocytes (%)	-1.517	0.133	0.042	0.212
WBC Total	-3.069	0.003	0.063	0.212

Table S6:
https://drive.google.com/file/d/1SOCrQHAa1z8635Dwc3fre1yOYnPJPlJi/view?usp=sharing

Table S7:
https://drive.google.com/file/d/1VHn4h2ucFZydoF3dUS6zOQAlVRAJe_s1/view?usp=sharing

Table S8:
https://drive.google.com/file/d/1qMHPwmA5IJvZ252lUeJQzQlZOFbJdN4m/view?usp=sharing

Table S9:
https://drive.google.com/file/d/1y74dnz0iVOptZcsC8MtlRc0ZEdtRHLMH/view?usp=sharing

Table S10:
https://drive.google.com/file/d/1s3Vjb5EeUr3tZNkVjfgrnts7DhLFEi44/view?usp=sharing

Table S11:
https://drive.google.com/file/d/17SqOH5IGMTx3-oEMIdxsbnJQE3Izxw2b/view?usp=sharing

Table S12:
https://drive.google.com/file/d/1wxlJ-JD6ZuAK16cIhPa8g8qVupNdHADH/view?usp=sharing

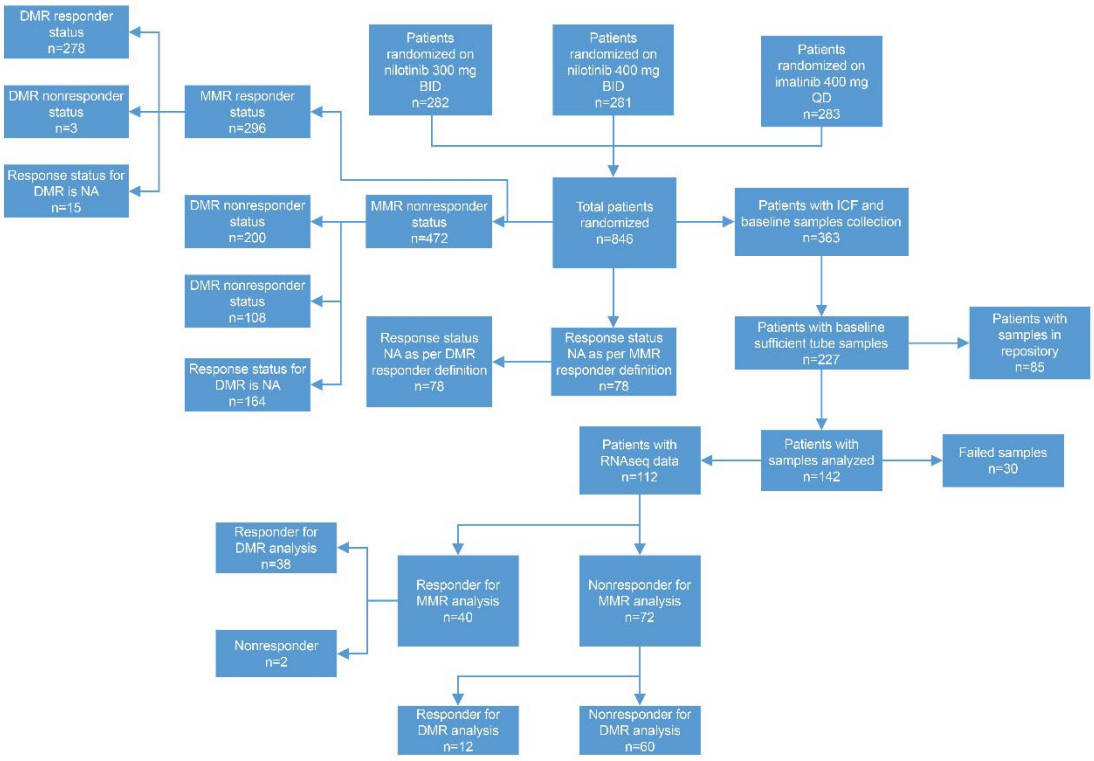
DMR responder status n=278
DMR nonresponder status n=3
Response status for DMR is NA n=15
MMR responder status n=296
Patients randomized on nilotinib 300 mg BID n=282
Patients randomized on nilotinib 400 mg BID n=281
Patients randomized on imatinib 400 mg QD n=283
DMR nonresponder status n=200
DMR nonresponder status n=108
Response status for DMR is NA n=164
MMR nonresponder status n=472
Total patients randomized n=846
Patients with ICF and baseline samples collection n=363
Response status NA as per DMR responder definition n=78
Response status NA as per MMR responder definition n=78
Patients with baseline sufficient tube samples n=227
Patients with samples in repository n=85
Patients with RNAseq data n=112
Patients with samples analyzed n=142
Failed samples n=30
Responder for DMR analysis n=38
Nonresponder n=2
Responder for MMR analysis n=40
Nonresponder for MMR analysis n=72
Responder for DMR analysis n=12
Nonresponder for DMR analysis n=60

Figure S1. Summary of patient randomization in the ENESTnd trial. BID, twice daily; DMR, deep molecular response; ICF, informed consent form; MMR, major molecular response; NA, not available; QD, once daily; RNA-seq, RNA sequencing.

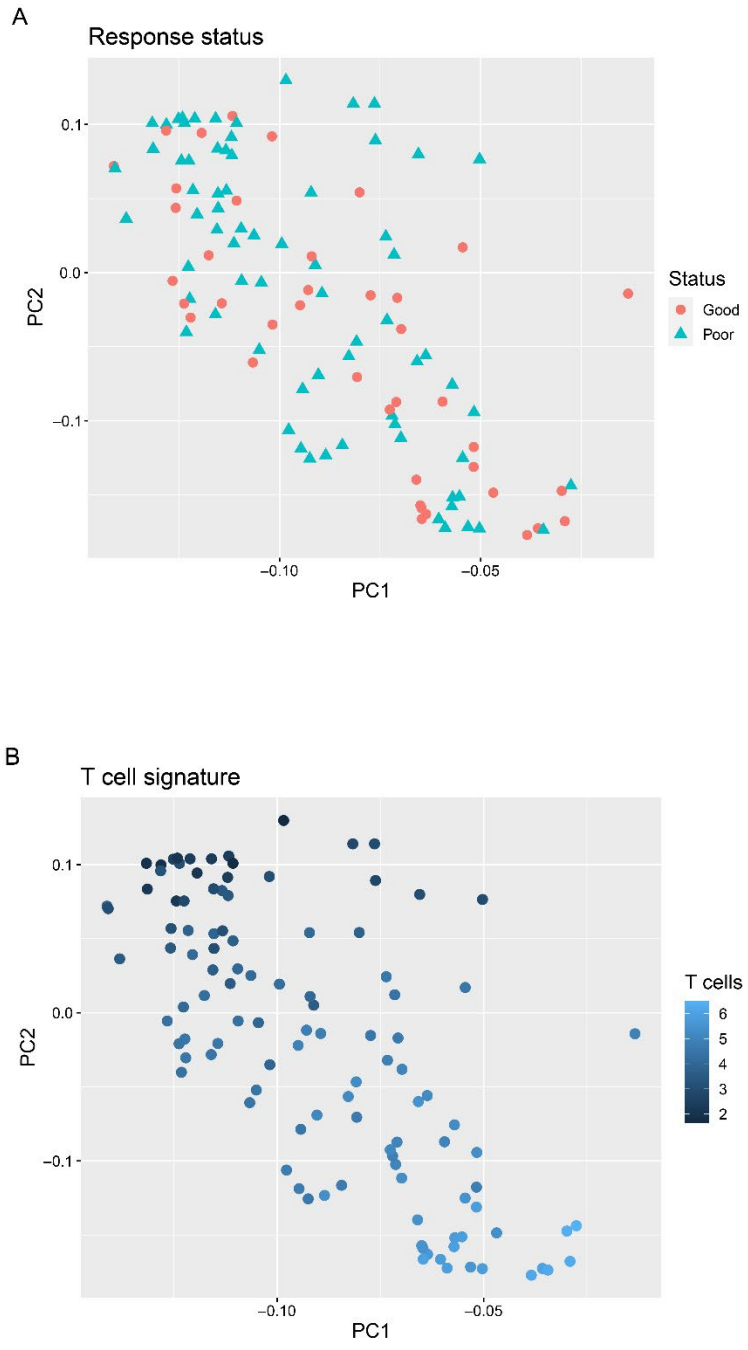

Figure S2. Labeled PCA plots. A-B, PCA plots labeled by (A) responder status and (B) inferred T cell abundance via MCP-counter.

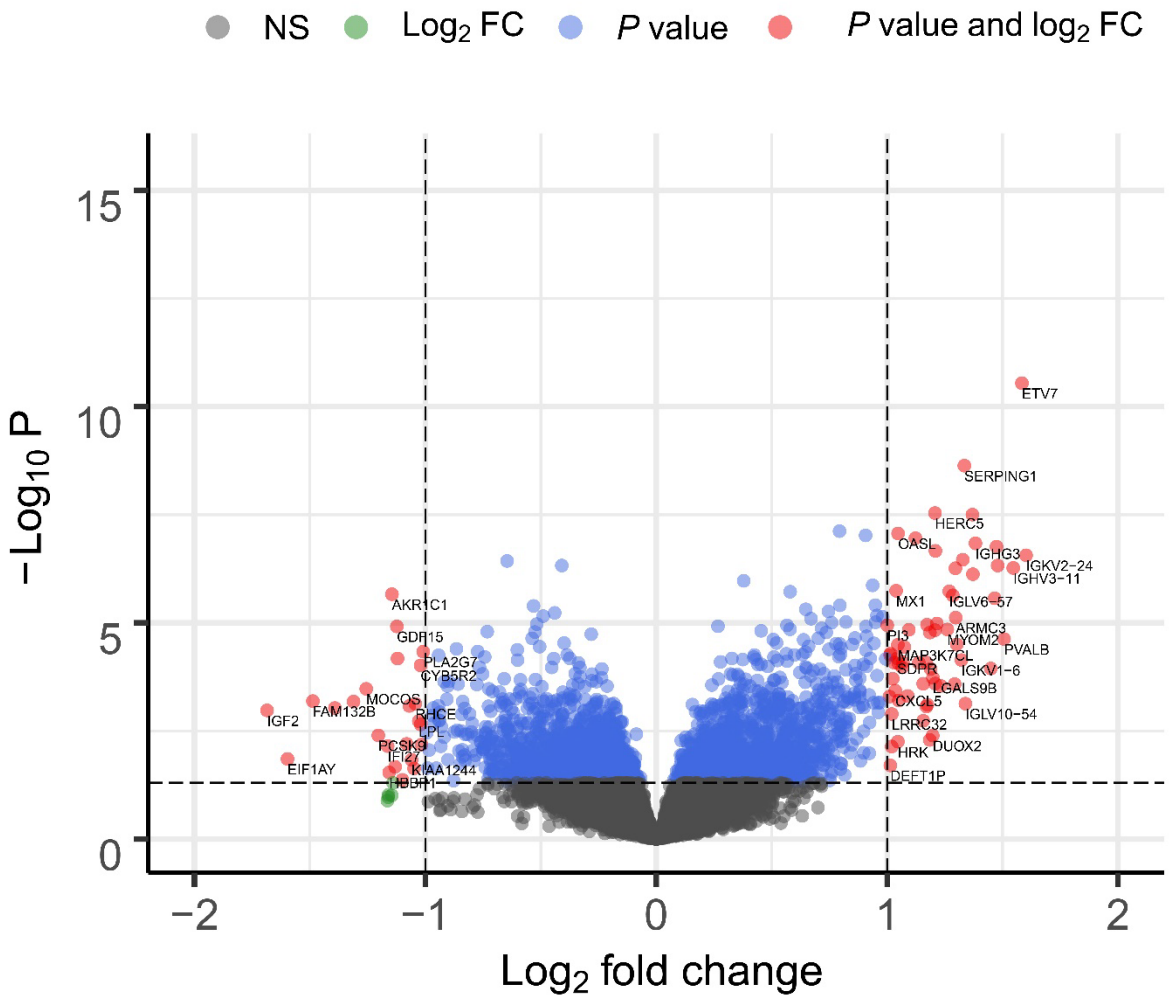

Figure S3. **Volcano plot of differentially expressed genes from ENESTnd.** Genes that were differentially expressed at a level satisfying P<.05 and $|\log_2(FC)|$ >1 are highlighted by red circles. FC, fold change.

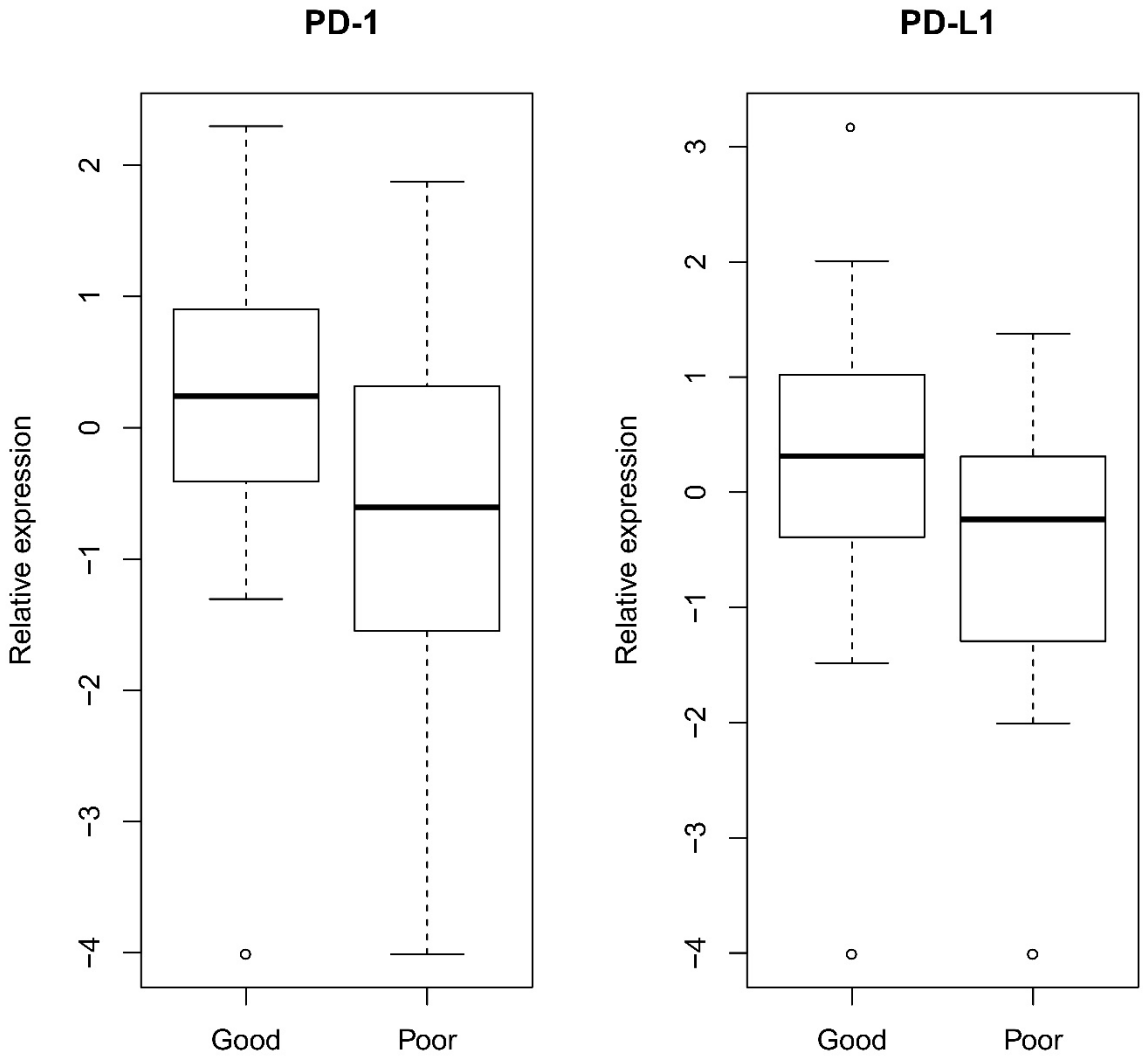

Figure S4. PD-1 and PD-L1 were overexpressed in good responders. Good responders exhibited significantly higher expression of PD-1 (P=8.4 × 10^{-4}) and PD-L1 (P=1.4 × 10^{-3}) than did poor responders.

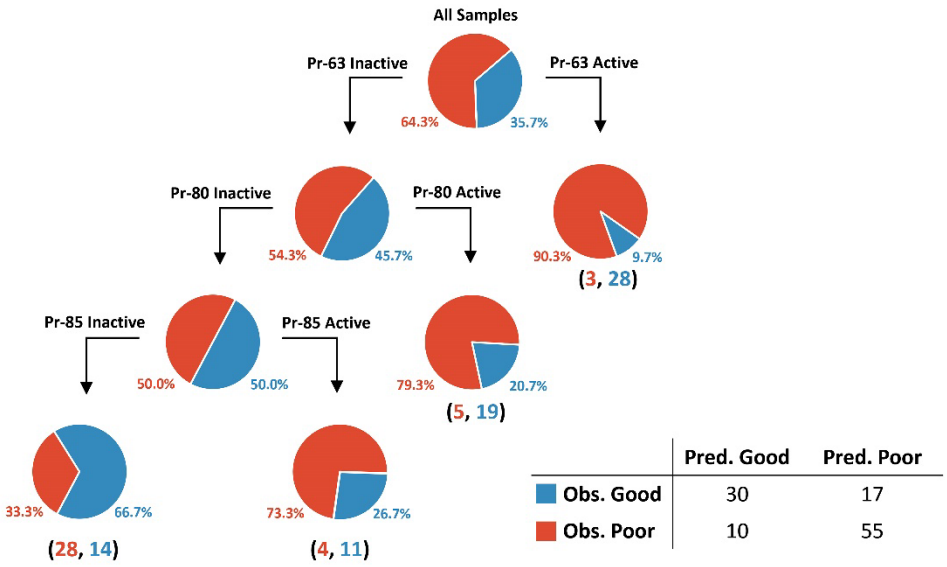

Figure S5. Decision tree predictor of response based on genetic program activity. The activity status (*i.e.*, active or inactive) of the genetic programs Pr-63, Pr-80, and Pr-85 can be used to predict responder status. The leaf node on the bottom left (all three programs inactive) predicts good response, all other leaf nodes predict poor response. Parentheses denote observed number of good and poor responders (Good, Poor).

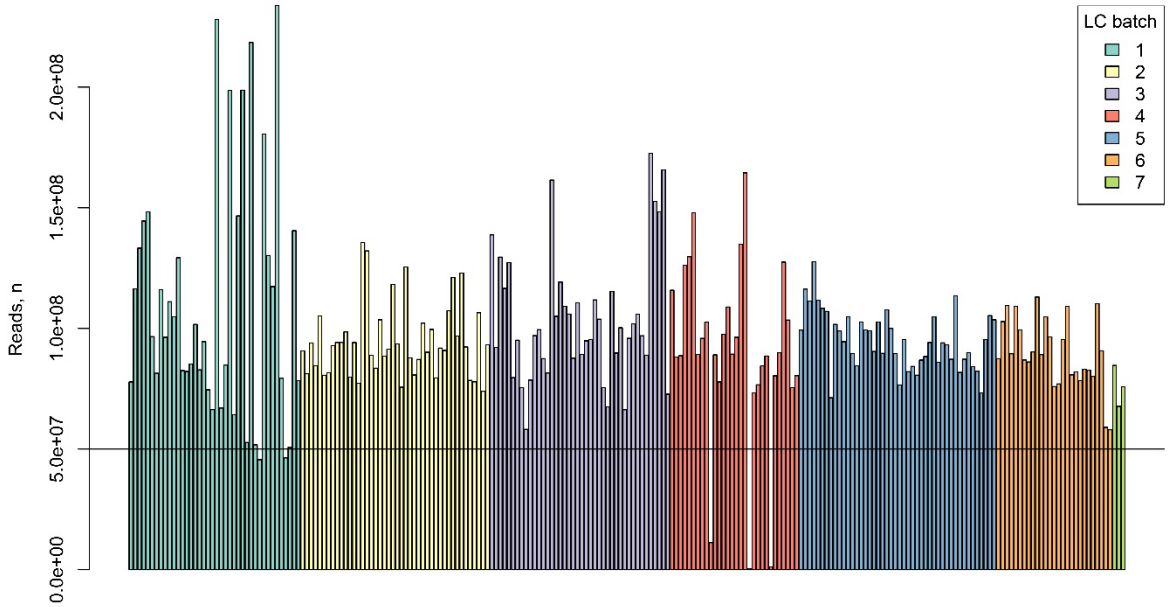

Figure S6. Total number of sequencing reads per library. Each vertical bar is a library (including controls). Libraries are colored by library construction (LC) batch. The horizontal line represents 50 million reads.

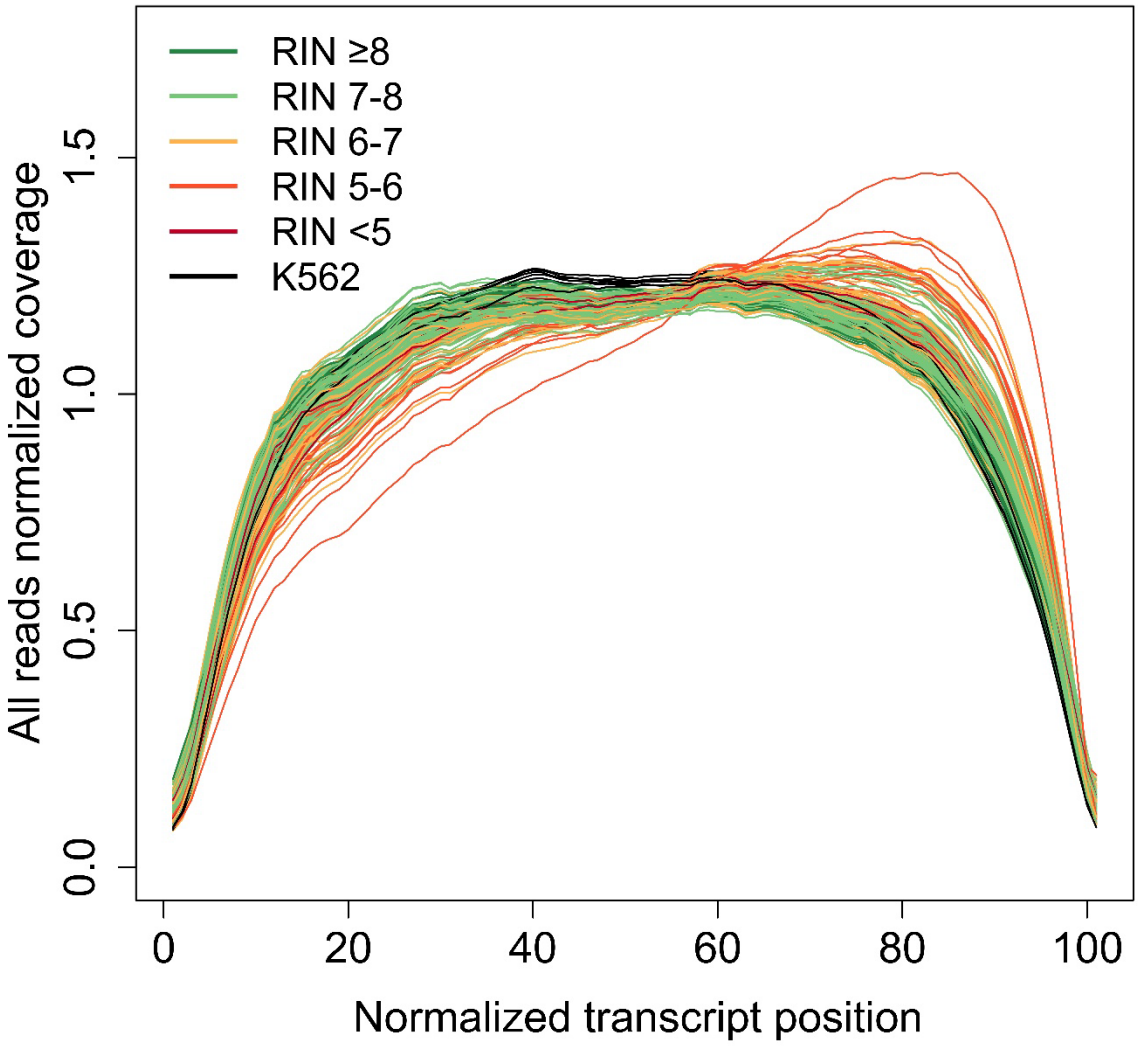

Figure S7. Coverage across transcripts for sequenced libraries. Normalized coverage (y axis) was plotted vs percentile normalized positions (5' to 3') across transcripts for each library generated and sequenced. Lines are colored by RIN score; we observed a slightly stronger 3' bias in samples with lower RIN, as would be expected. One library showed a strong 3' bias, but this bias was corrected by library reconstruction. RIN, RNA integrity number.

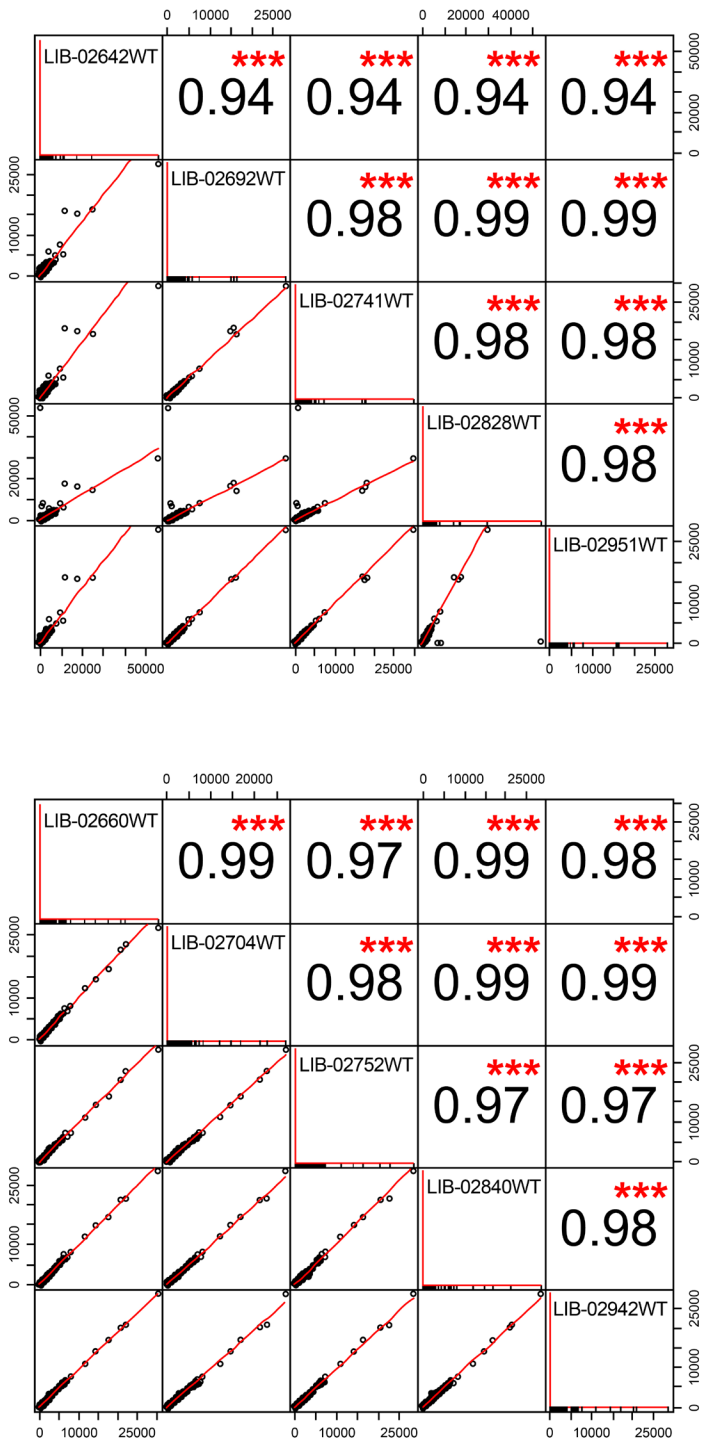

Figure S8. Correlation of expression values in replicate samples. Correlation of gene expression between across-batch replicates as measured by fragments per kilobase per million reads is shown for both replicated samples. Below the diagonal, the correlations are plotted for each pair of samples. Above the

diagonal are the correlation coefficients. There are 5 replicates per sample due to library construction failures in batch 3. Stars denote significance level: *** $P<.001$; ** $P<.01$; * $P<.05$.

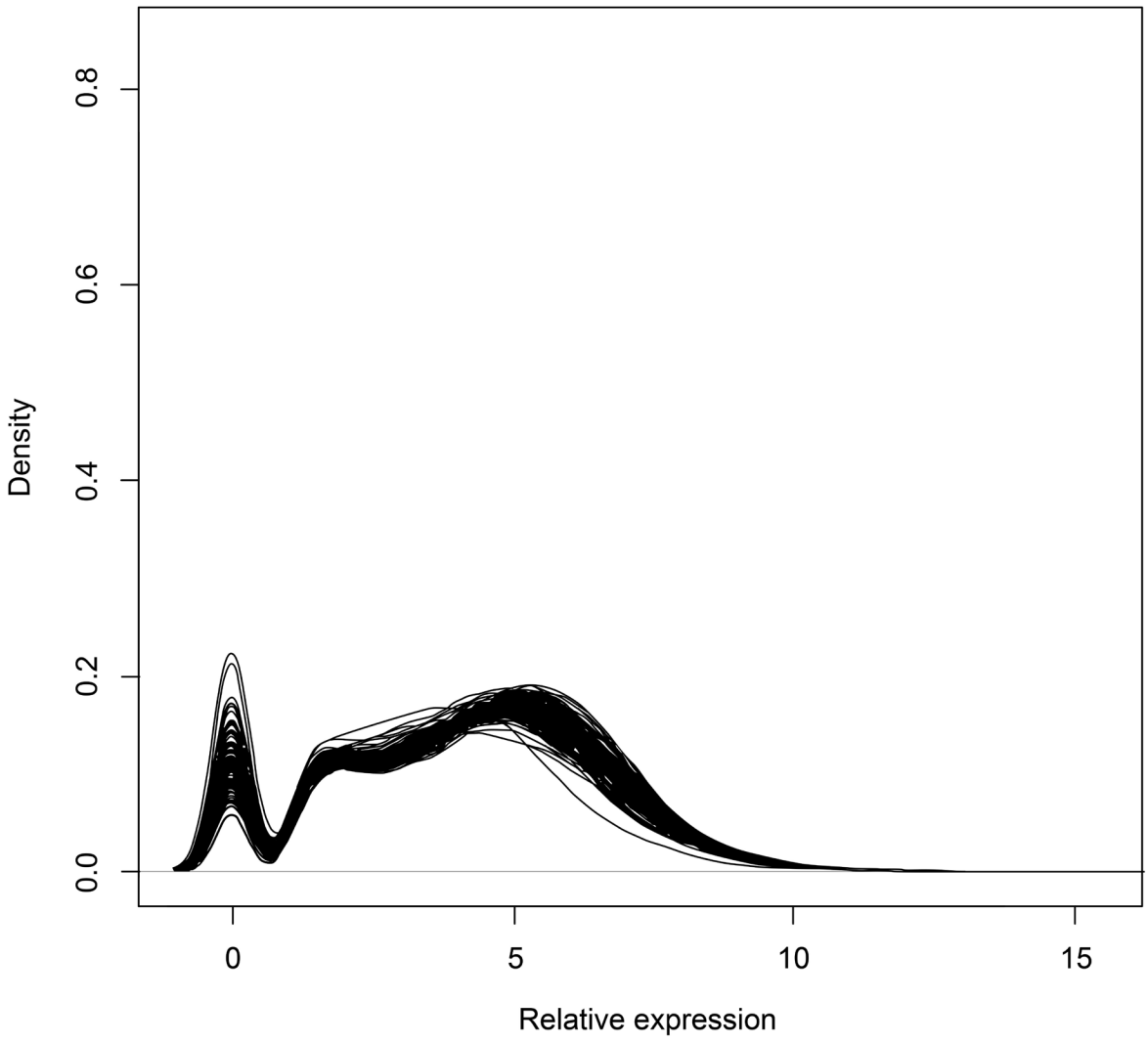

Figure S9. Kernel density plot of ENESTnd log_2(cpm + 1) gene expression. The gene expression profiles overlapped across samples. cpm, counts per million.

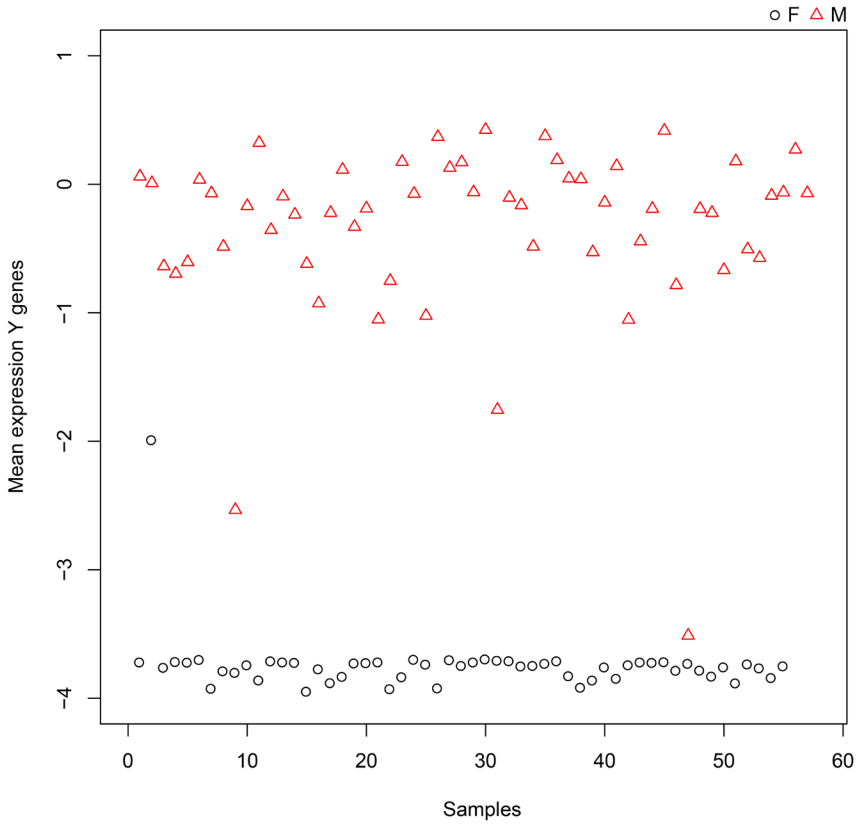

Figure S10. Mean Y chromosome gene expression vs sex. The mean expression of genes on the Y chromosome clearly separated men and women, as expected, although 2 male-labeled samples exhibited unusually low expression.

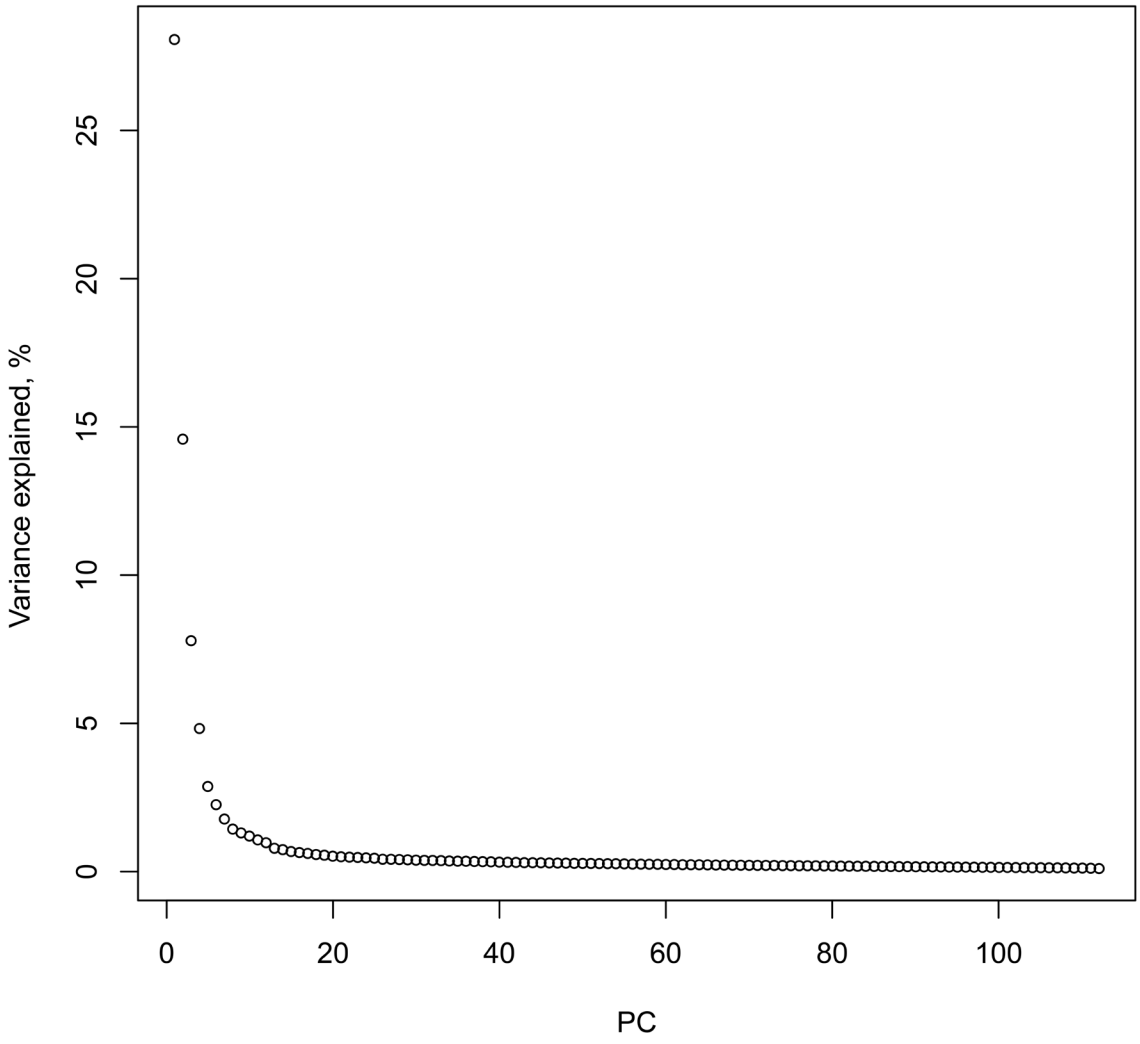

Figure S11. Variance explained by PC of gene expression. The variance explained by each principal component (PC) showed that most of the signal in the data was represented in the first 10 PC.

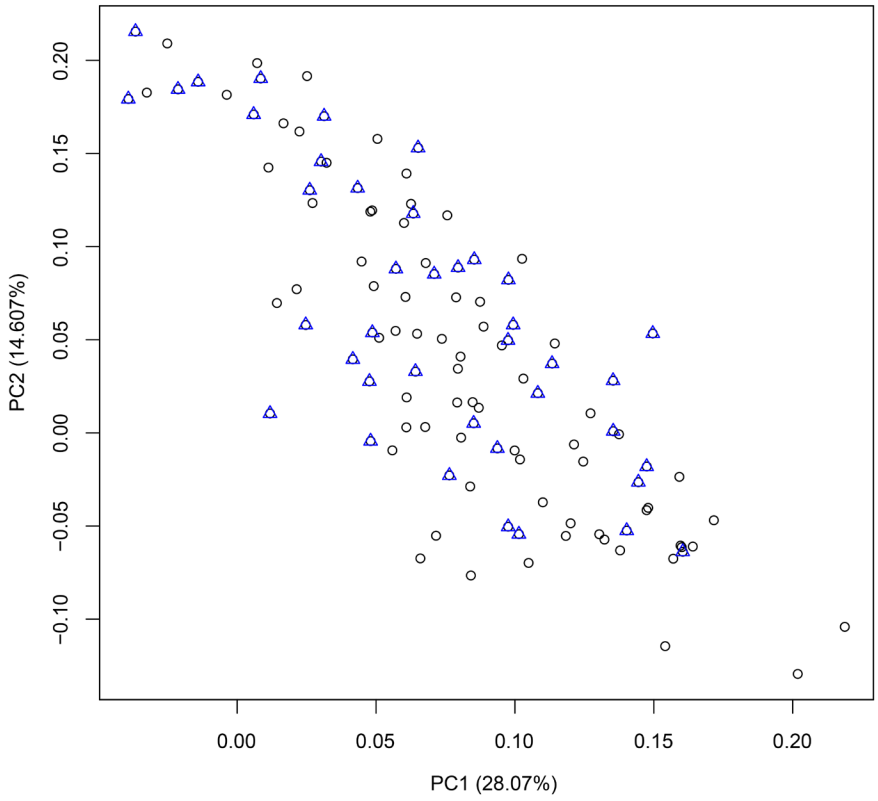

Figure S12. PCA plot of ENESTnd gene expression. The PCA plot did not suggest the presence of batch effects. Good responders are labeled by blue triangles, and poor responders are labeled by black circles. PCA, principal components analysis.

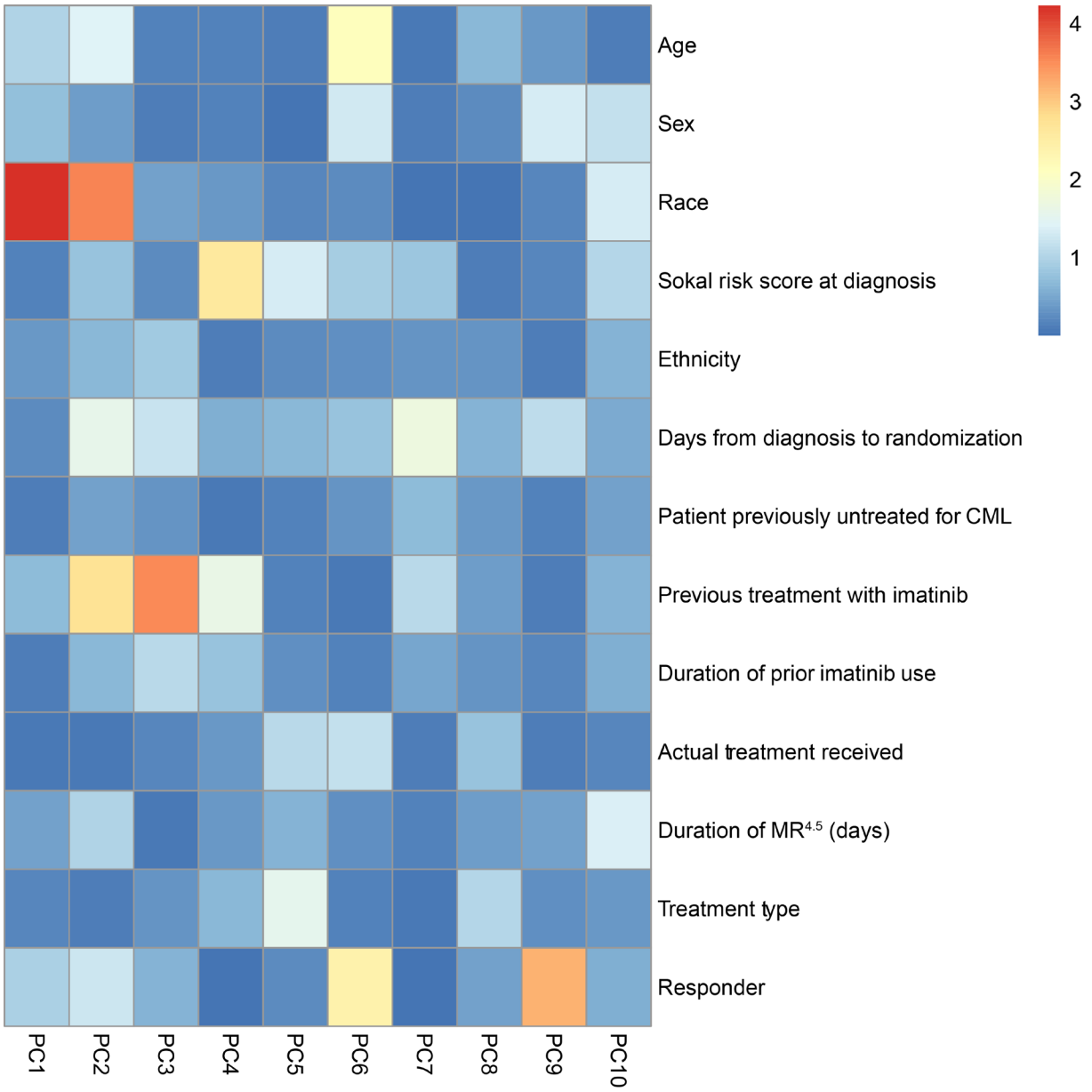

Figure S13. Clinical data vs PC. The $-\log_{10}(p)$ of associations between each clinical variable and the first 10 PCs are shown. Values >1.3 correspond to $P<.05$. CML, chronic myeloid leukemia; MMR, major molecular response; MR4.5, *BCR::ABL1*^{IS} ≤ 0.0032%; PC, principal component.

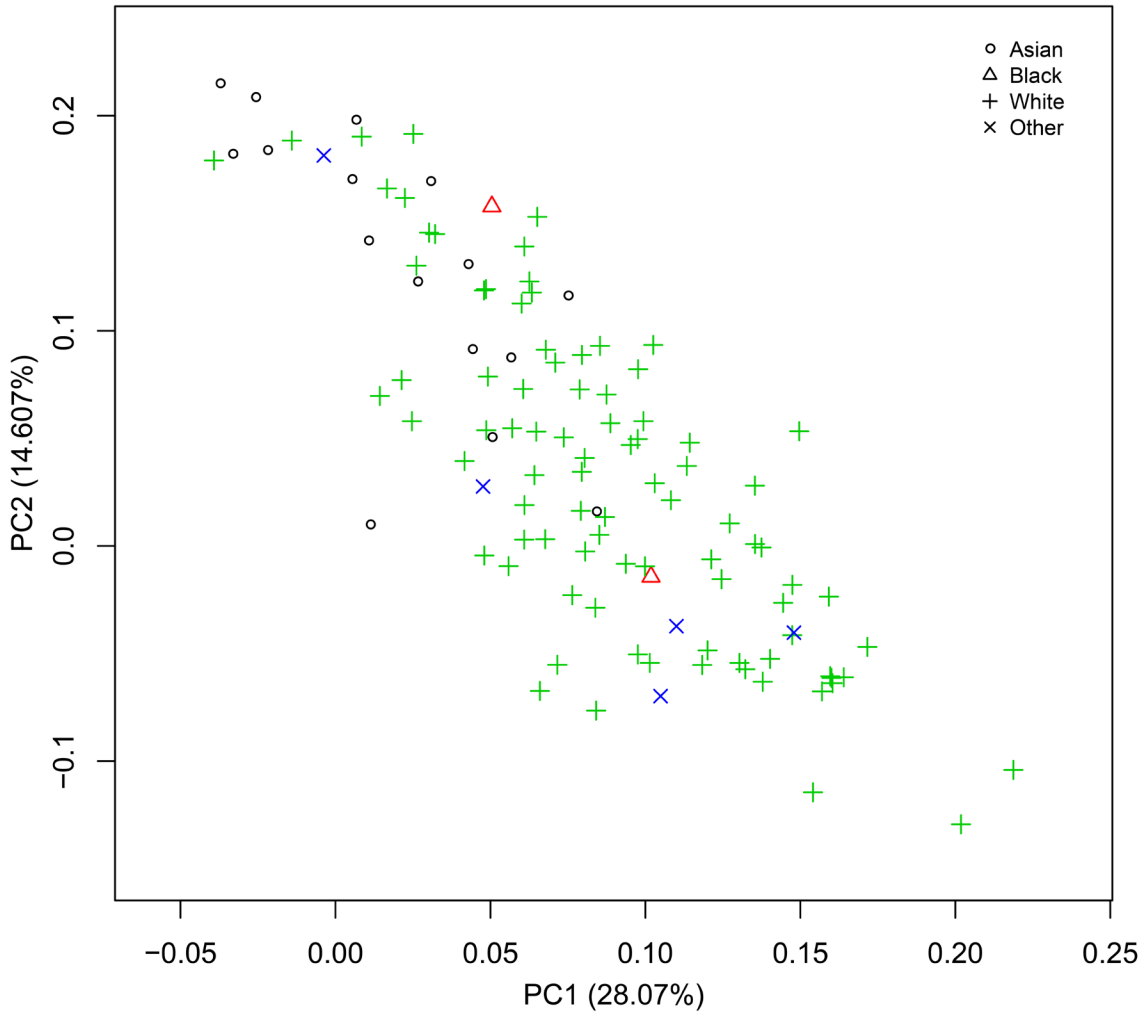

Figure S14. Race vs PC1. Race significantly associated with the first PC but did not generate sufficient differences to warrant batch correction. PC, principal component.

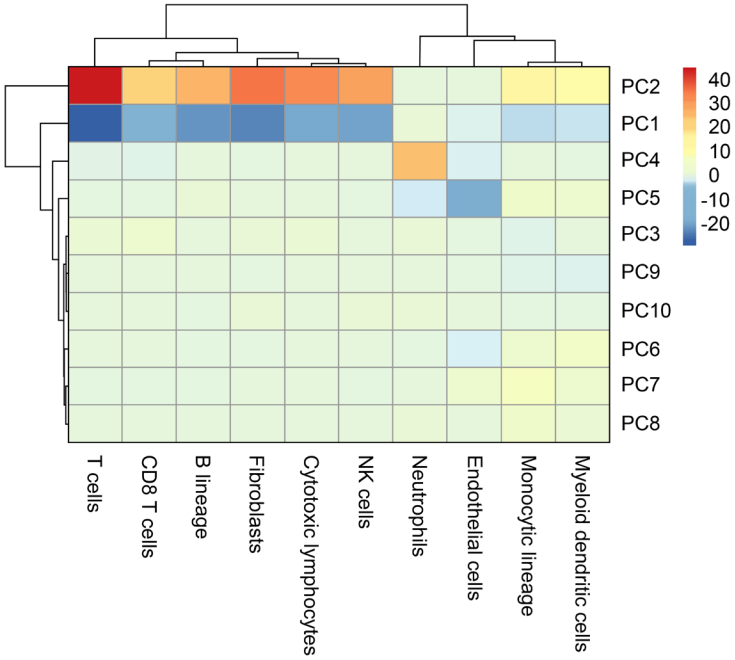

Figure S15. Inferred cell populations vs PC. The signed $-\log_{10}(p)$ of associations between each of the inferred proportions of various cell types and the first 10 PCs are shown. Values with a magnitude >1.3

correspond to *P*<.05. Positive values are correlated, and negative values are inversely correlated. NK, natural killer; PC, principal component.

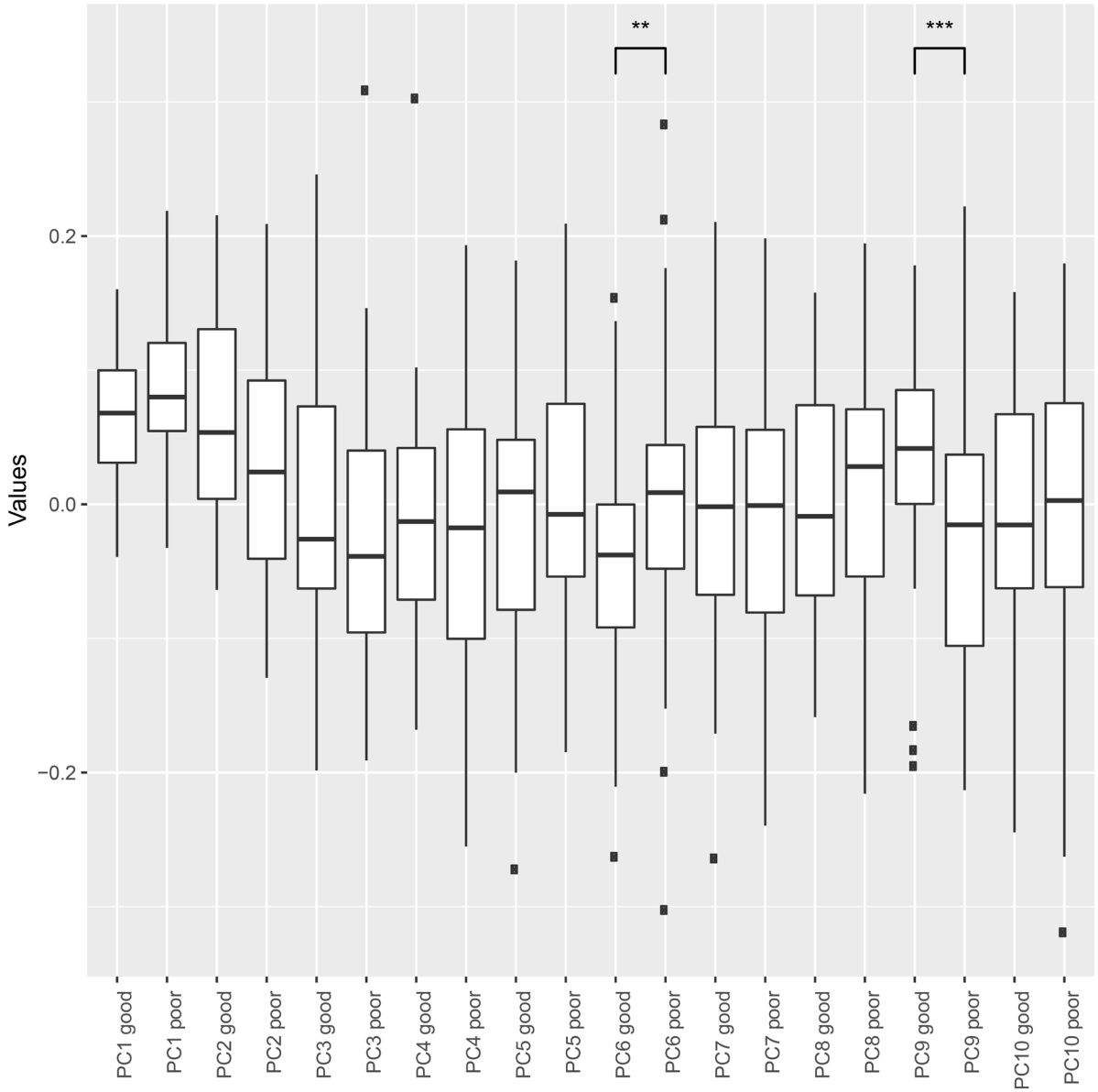

Figure S16. PC vs responder status. The rotated values of the first 10 PCs are shown for good and poor responders. PC6 and PC9 showed significant differences. ** *P*<.01; *** *P*<.001. PC, principal component.