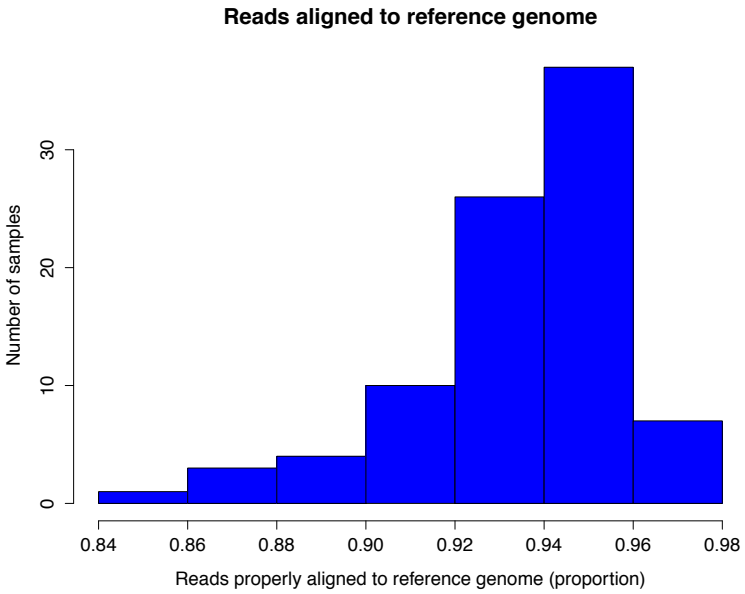
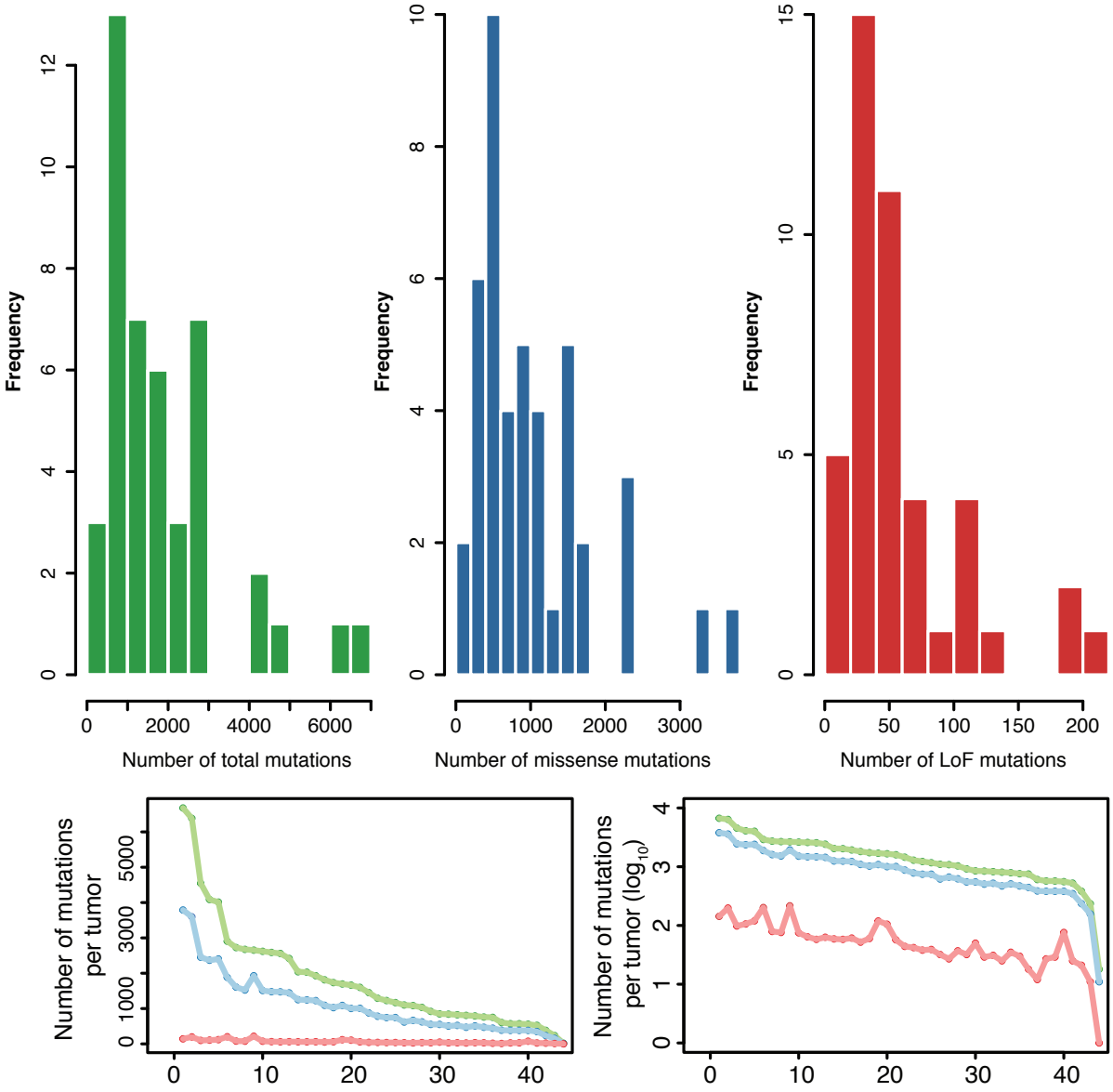
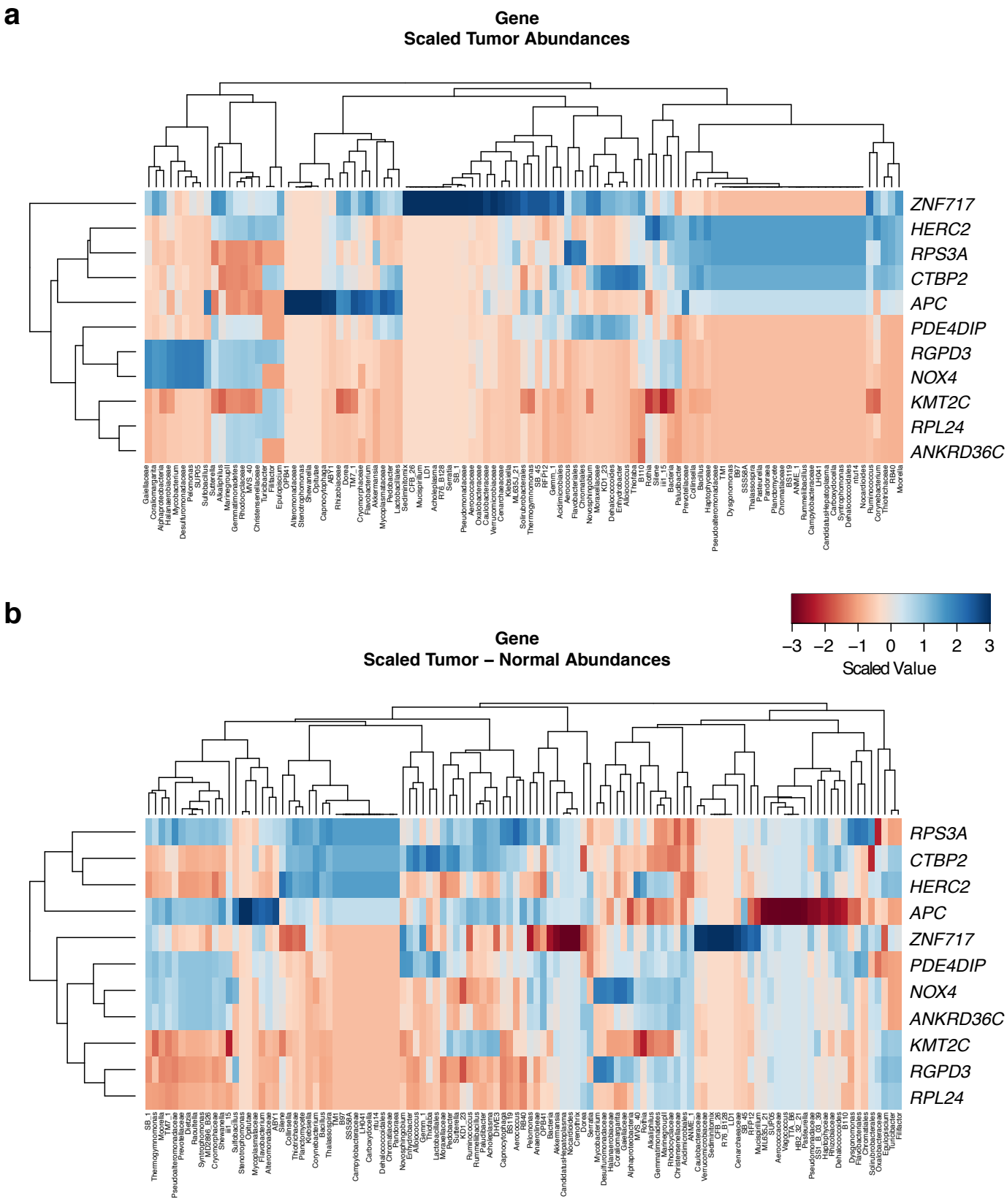


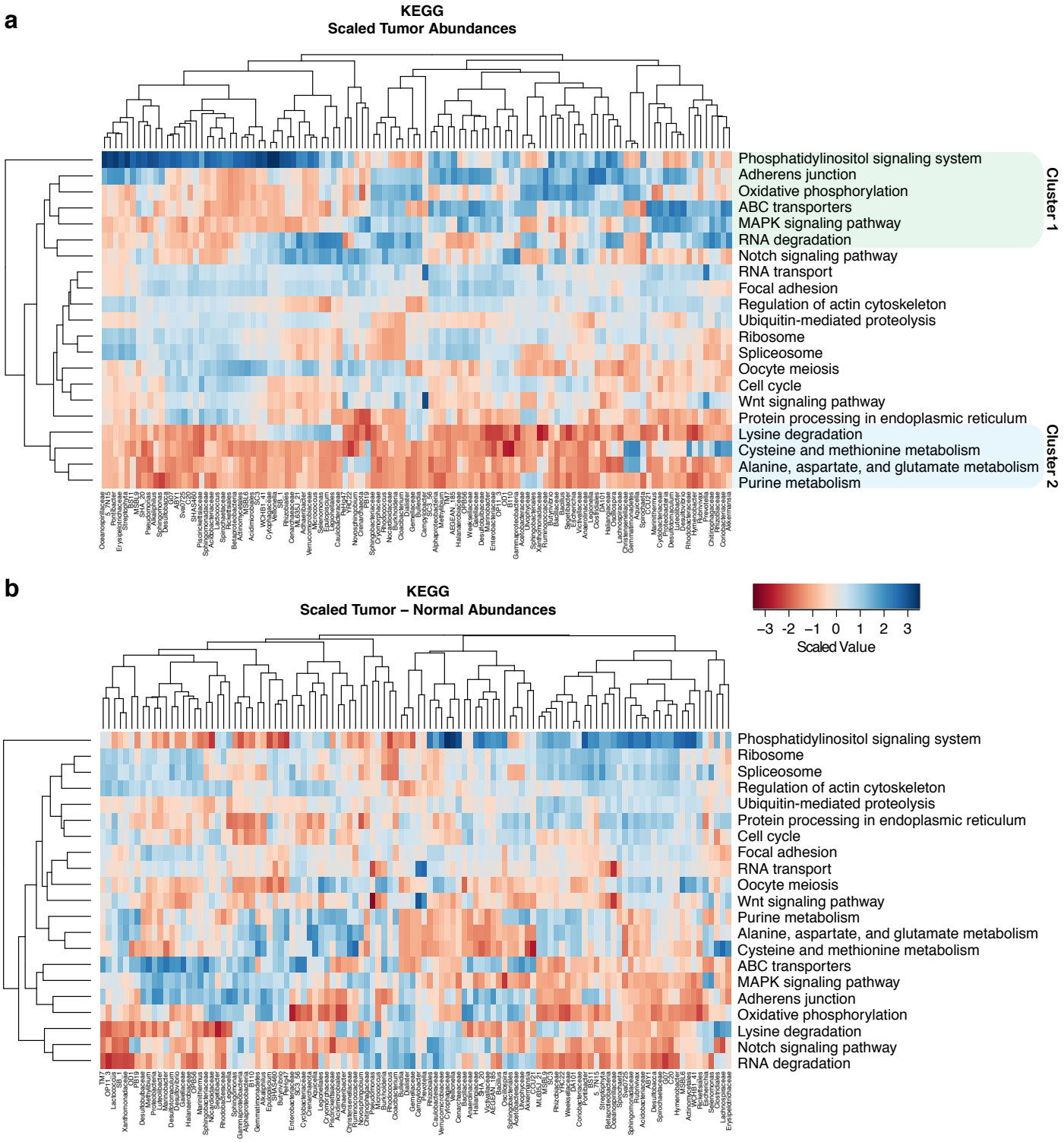
Supplementary Fig. 1. Histogram of the proportion of reads passing quality filter that aligned to the references genome.



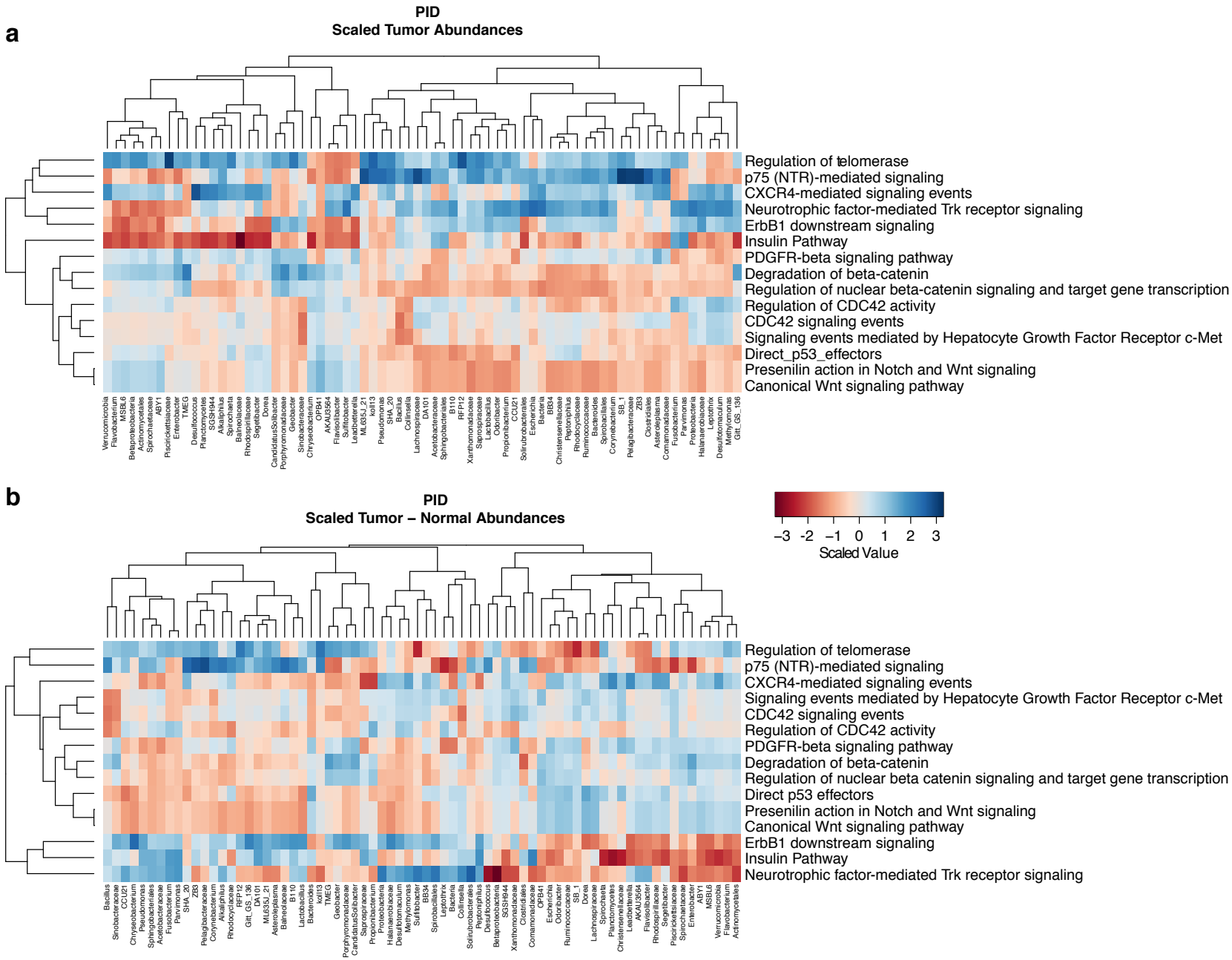
Supplementary Fig. 2. Tumors harbored a wide range of mutations. The histograms at top indicate the frequency distribution of the total (green), missense (blue), and LoF (red) mutations found in the tumors from our patient cohort. The plots at bottom show the absolute numbers of tumor-specific mutations detected in the tumors on a flat scale (bottom left) as well as on a log scale (bottom right).



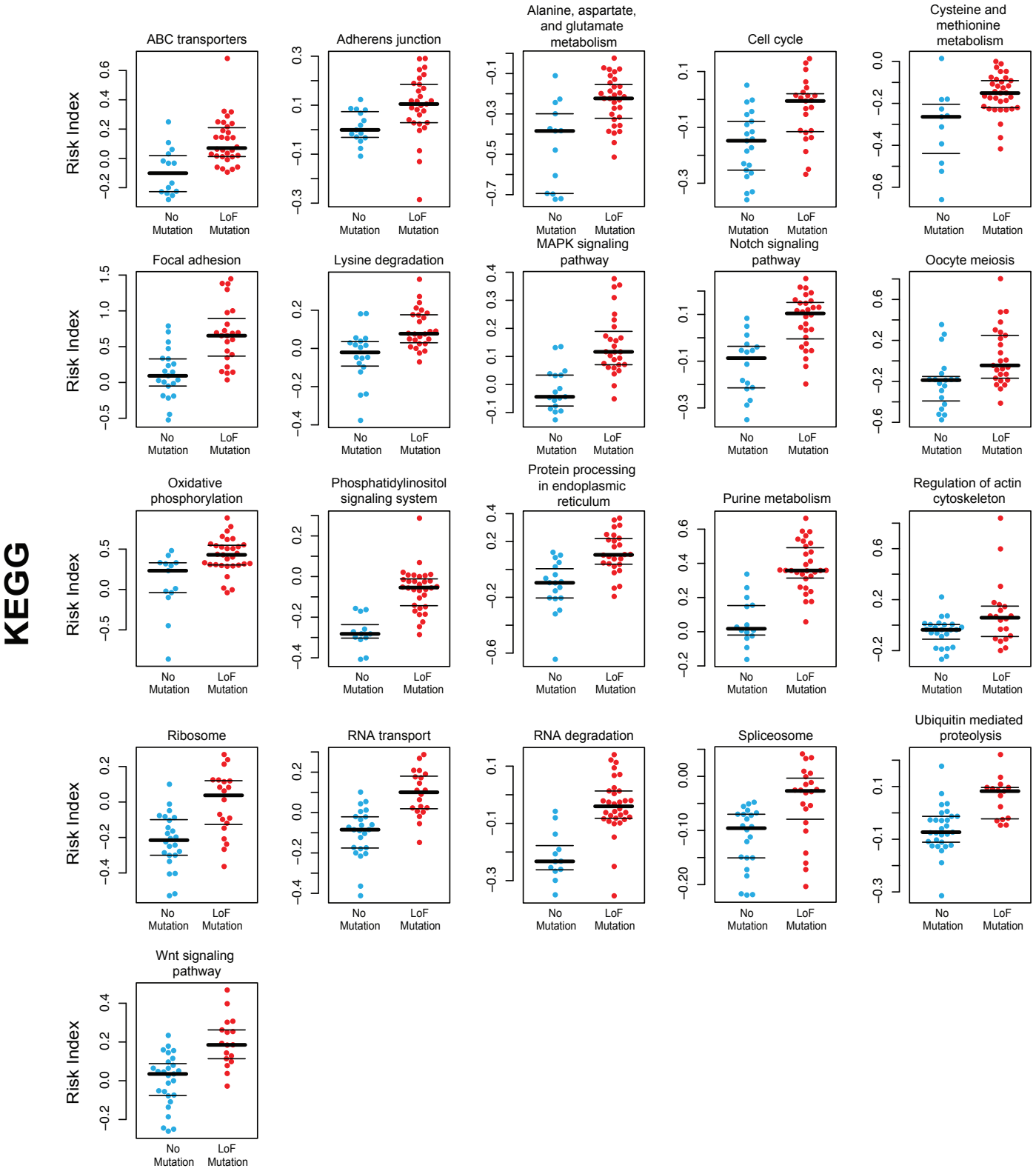
Supplementary Fig. 3. Heatmaps demonstrating the patterns of microbial abundances for patient samples with prevalent LoF mutations. **a**, Scaled taxon abundances (columns) in the tumor samples that harbor LoF mutations in the genes indicated (rows). **b**, Scaled differences (tumor abundance - matched normal abundance) patients that harbor tumor-specific LoF mutations in the genes indicated (rows).



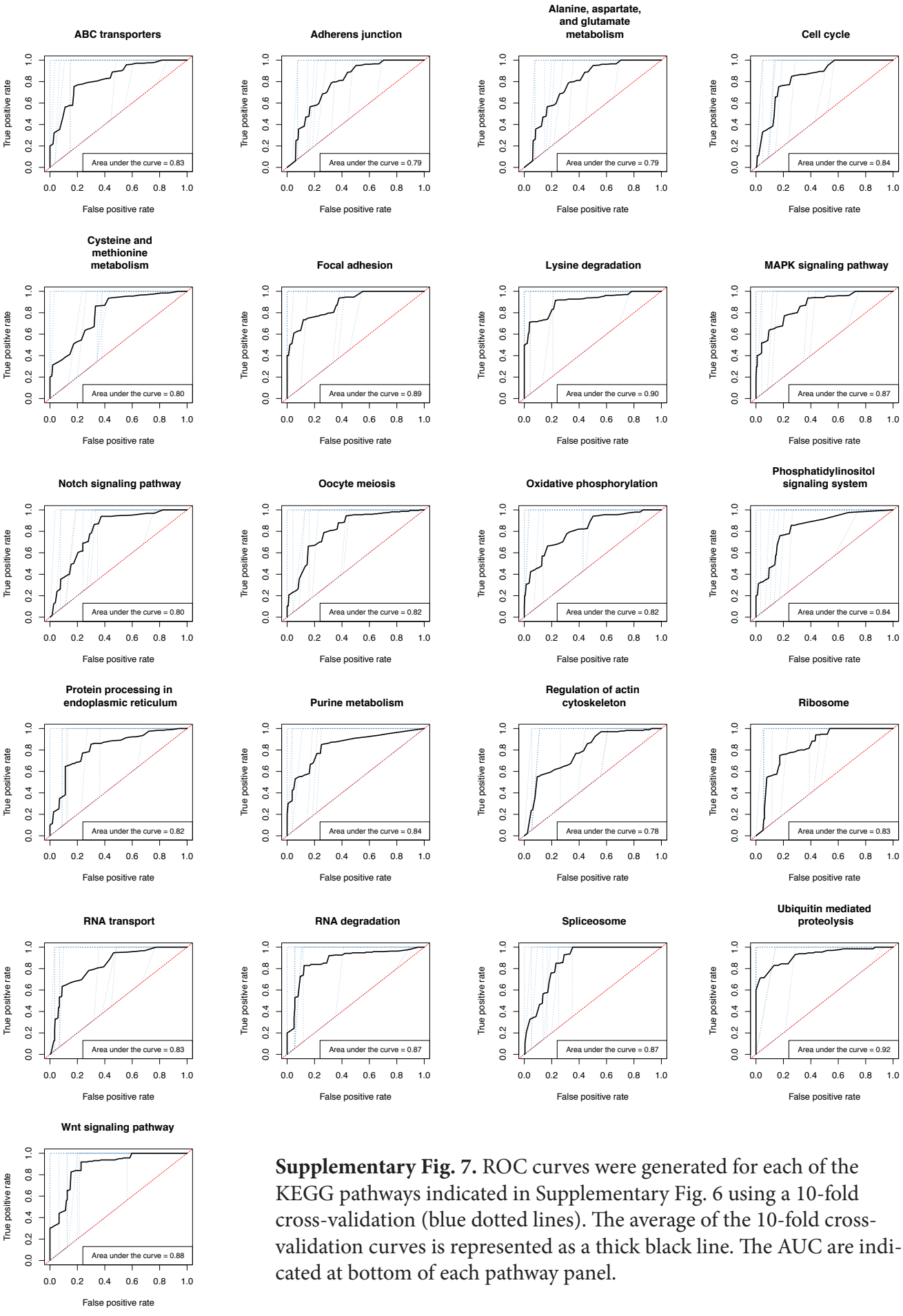
Supplementary Fig. 4. Heatmaps demonstrating the patterns of microbial abundances for patient samples with prevalent LoF mutations in KEGG pathways. **a**, Scaled taxon abundances (columns) in the tumor samples that harbor LoF mutations in the KEGG pathways indicated (rows). Clusters 1 and 2 are labeled to facilitate discussion in main text. **b**, Scaled differences (tumor abundance - matched normal abundance) patients that harbor tumor-specific LoF mutations in the KEGG pathways indicated (rows).



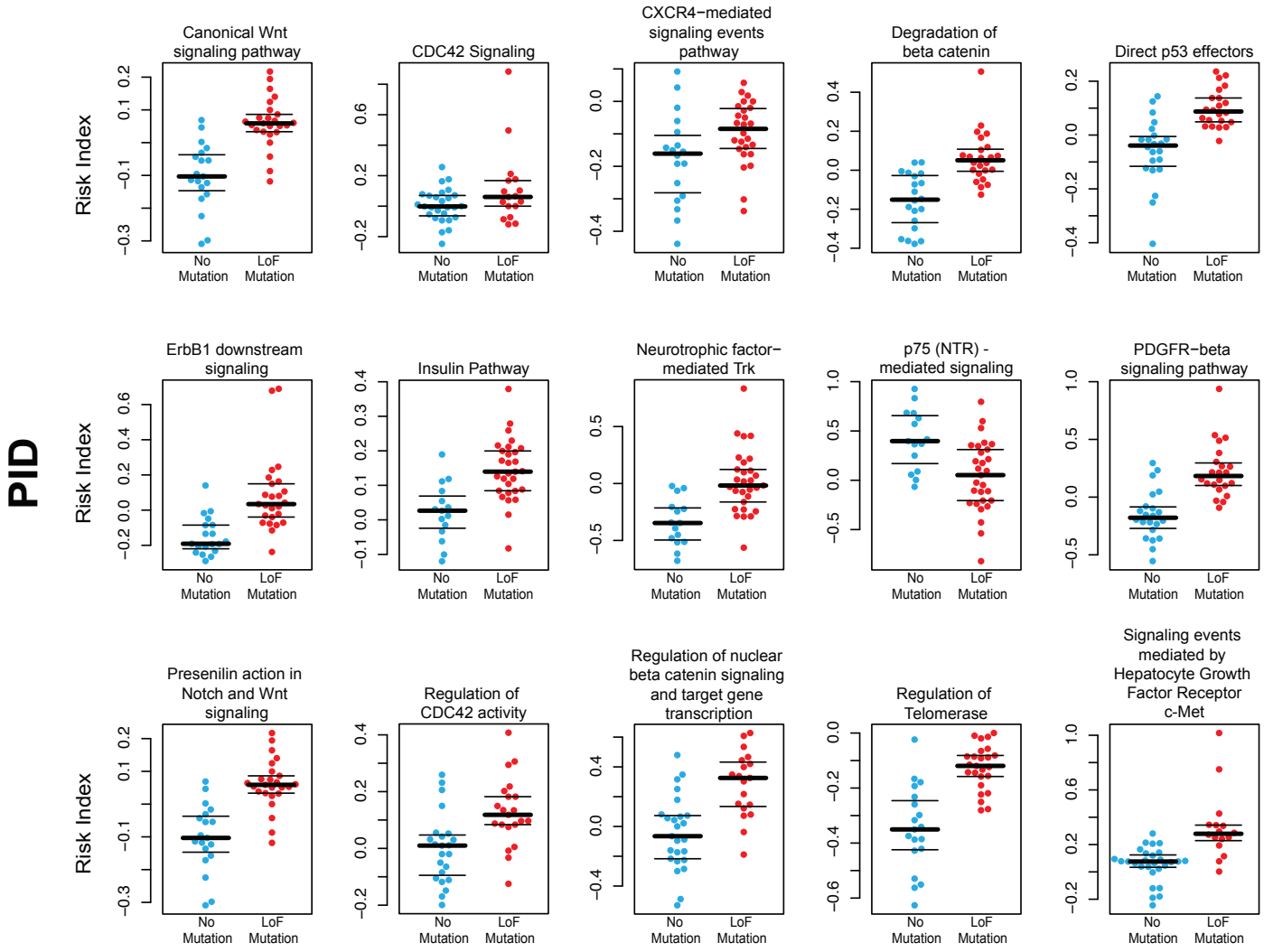
Supplementary Fig. 5. Heatmaps demonstrating the patterns of microbial abundances for patient samples with prevalent LoF mutations in PID pathways. **a**, Scaled taxon abundances (columns) in the tumor samples that harbor LoF mutations in the PID pathways indicated (rows). **b**, Scaled differences (tumor abundance - matched normal abundance) patients that harbor tumor-specific LoF mutations in the PID pathways indicated (rows).



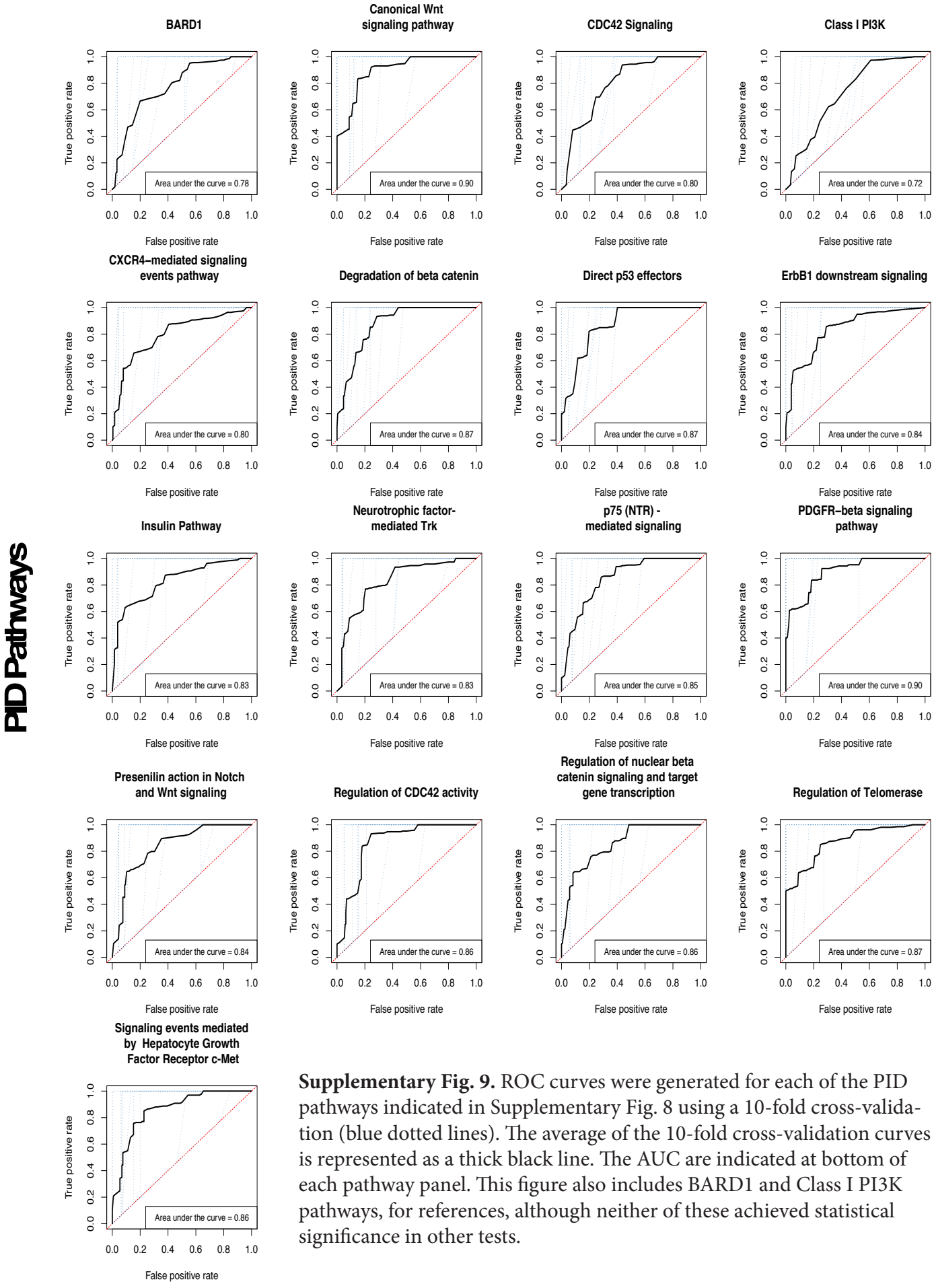
Supplementary Fig. 6. LoF mutations in KEGG pathways can be predicted using a risk index as a classifier (y-axis).



Supplementary Fig. 7. ROC curves were generated for each of the KEGG pathways indicated in Supplementary Fig. 6 using a 10-fold cross-validation (blue dotted lines). The average of the 10-fold cross-validation curves is represented as a thick black line. The AUC are indicated at bottom of each pathway panel.

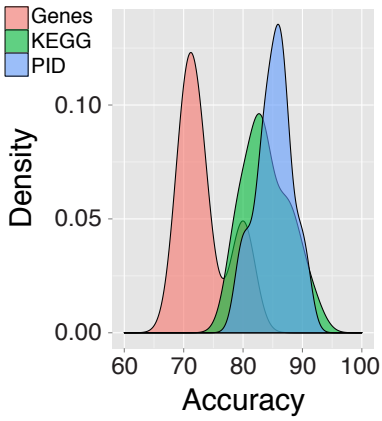


Supplementary Fig. 8. LoF mutations in PID pathways can be predicted using a risk index as a classifier (y-axis).



Supplementary Fig. 9. ROC curves were generated for each of the PID pathways indicated in Supplementary Fig. 8 using a 10-fold cross-validation (blue dotted lines). The average of the 10-fold cross-validation curves is represented as a thick black line. The AUC are indicated at bottom of each pathway panel. This figure also includes BARD1 and Class I PI3K pathways, for references, although neither of these achieved statistical significance in other tests.

Supplementary Fig. 10. (see separate supplementary file) Large set of abundance plots for the taxa from each of the genes, KEGG pathways, and PID pathways that harbored prevalent LoF mutations and could be showed significance as a means of predicting the interaction between the mutation and the microbiota. Abundances are plotted as both column dot plots as well as horizontal bar plots of the differences in the mean abundances of a subset of the taxa predicted to interact differentially with tumors with a LoF mutation relative to those without the indicated mutation. This subset represents those taxa that had a mean difference in abundance of greater than 0.1%, proportionally.



Supplementary Fig. 11. Interaction prediction accuracies increase when assessing biological pathways. A density histogram showing the distribution of prediction accuracies for individual genes (red), KEGG pathways (green), and PID pathways (blue). Interaction prediction accuracies are highest for human cancer-specific pathways from PID.