Supplementary Figures
Detection of aberrant events in RNA sequencing data

Vicente A. Yépez[1,2,*], Christian Mertes[1,*], Michaela F. Müller[1], Daniela S. Andrade[1], Leonhard Wachutka[1], Laure Frésard[3], Mirjana Gusic[4,5], Ines Scheller[1], Patricia F. Goldberg[1], Holger Prokisch[4,5], Julien Gagneur[1,2,*]

[1] Department of Informatics, Technical University of Munich, Munich, Germany
[2] Quantitative Biosciences Munich, Department of Biochemistry, Ludwig-Maximilians Universität, Munich, Germany
[3] Department of Pathology, School of Medicine, Stanford University, Stanford, CA, USA
[4] Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
[5] Institute of Human Genetics, Technical University of Munich, Munich, Germany
[*] Corresponding authors

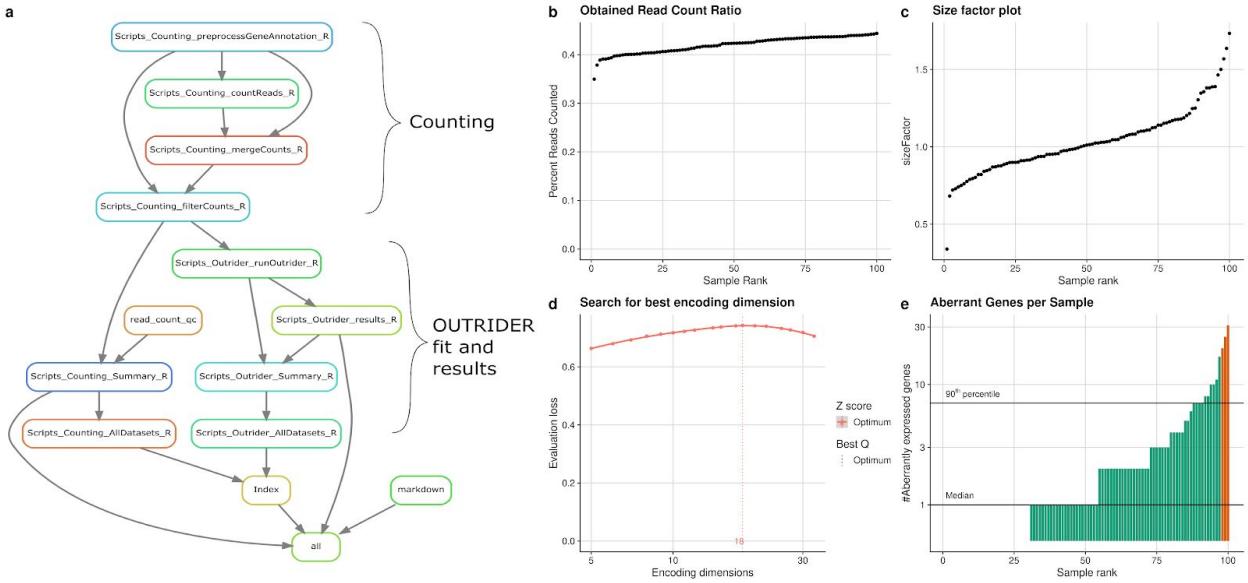

Fig S1 | Counting and OUTRIDER Summary. **a,** Aberrant expression workflow. The two main steps are counting and running the OUTRIDER fit and results. **b,** Percentage of counted reads per sample. **c,** Sorted size factors. Size factors represent the relative sequencing depth of a certain sample. **d,** OUTRIDER evaluation loss for different encoding dimensions *q* and showing the optimal value (vertical dotted line). **e,** Number of aberrantly expressed genes per sample. Aberrant samples (orange) are samples with too many outlier genes.

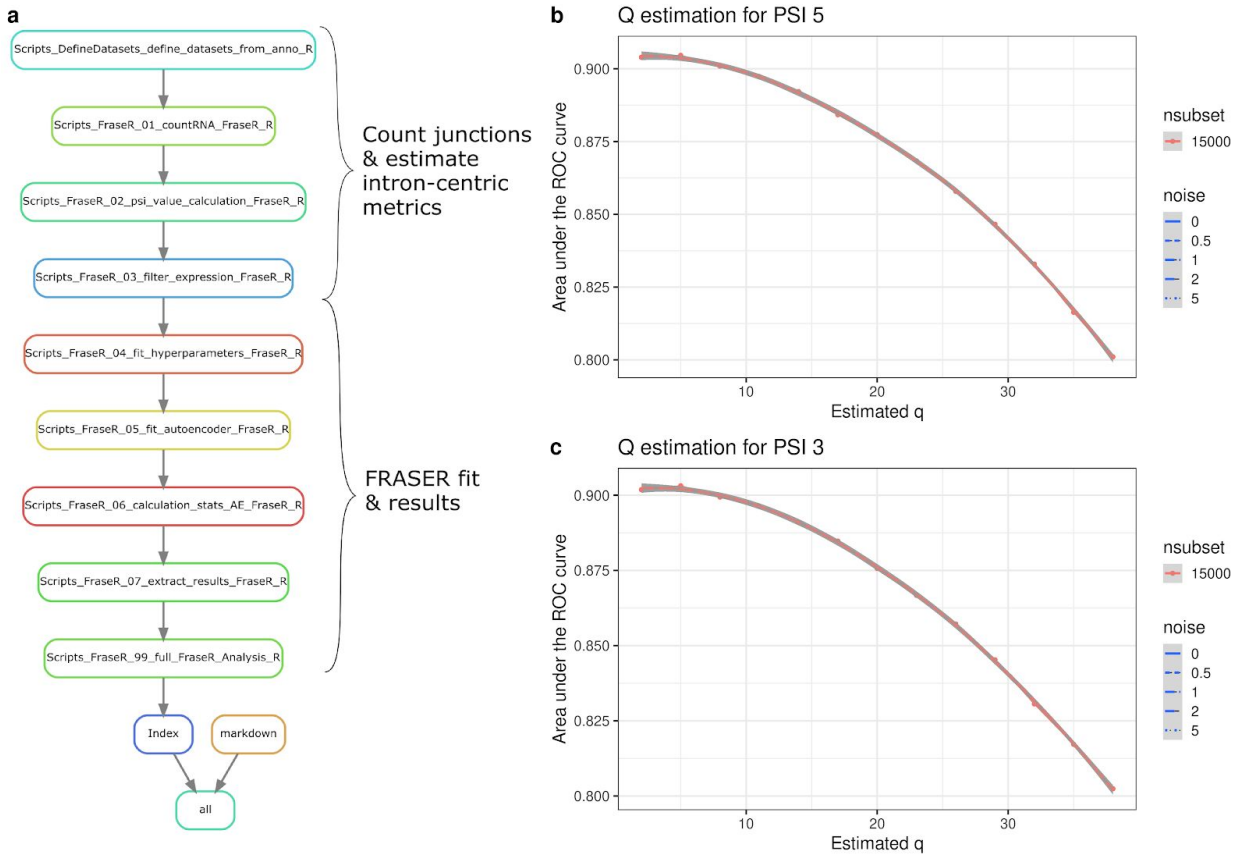

Fig S2 | Aberrant Splicing Module. **a,** Aberrant splicing workflow. The two main steps are counting the junctions and running the FraseR fit and results. **b-c,** Area under the ROC curve (y-axis) after different encoding dimensions (x-axis) and noise level injections for PSI 3 (b) and PSI 5 (c). The ranking of outliers was bootstrapped to yield 95% confidence bands.

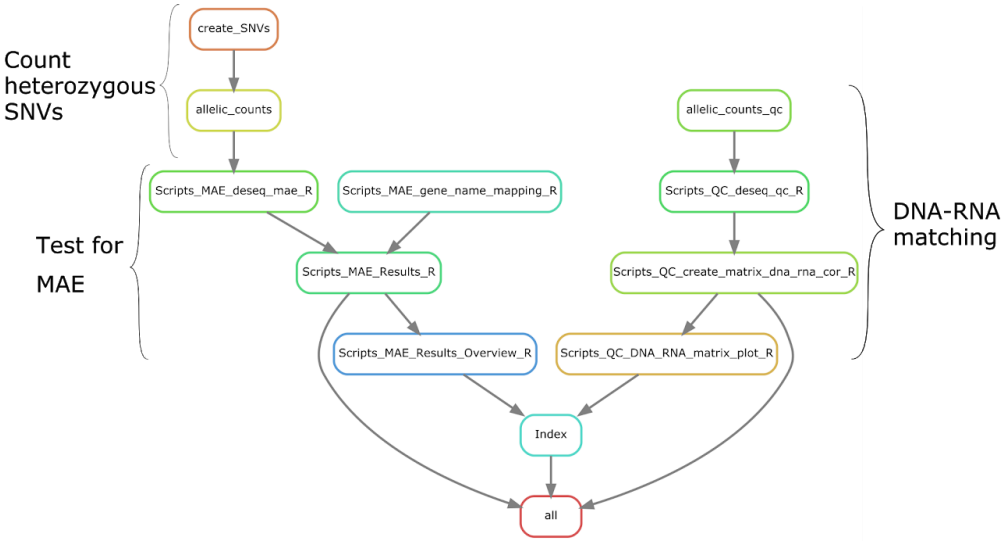

Fig S3 | Mono-allelic expression workflow. It is composed of two parts, the first one tests for heterozygous SNVs that are mono-allelically expressed and the second one matches VCF with BAM files.

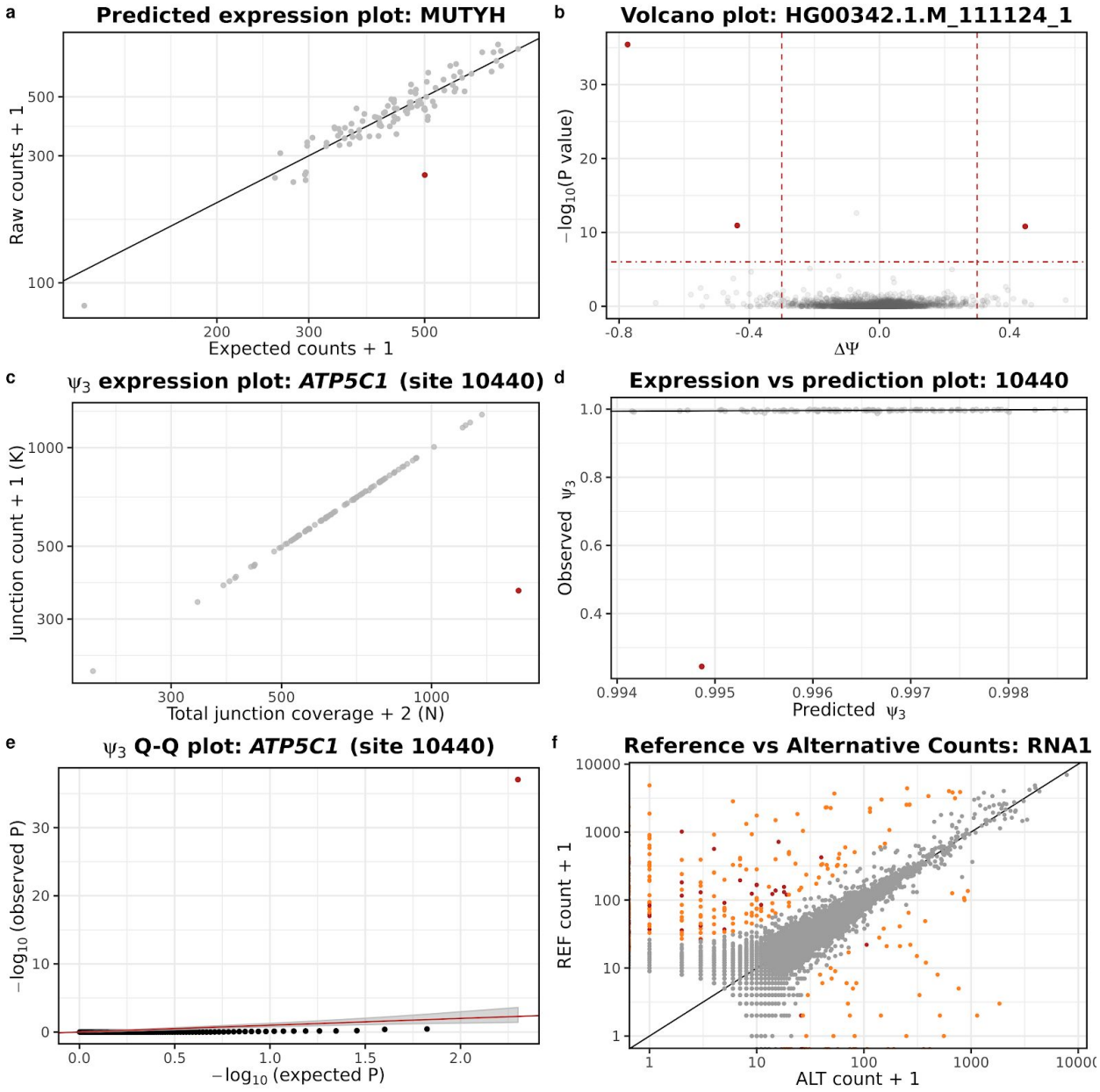

Fig. S4 | a, Raw (y-axis) versus expected (x-axis) counts of gene *MUTYH* showing one outlier (red). **b,** Negative log-transformed nominal *P* values (y-axis) versus $\Delta\psi_3$ values (x-axis) derived from all the splice sites (aggregated by gene) of sample HG00342.1.M_111124_1. Outliers are marked in red. **c,** Junction read counts (*K*, y-axis) plotted against the total split read coverage at the acceptor site (*N*, x-axis), of one junction in gene *ATP5C1*, which is the most severe outlier of panel (b). **d,** Observed (y-axis) versus expected (FRASER-predicted, x-axis) ψ_3 values of the same junction. **e,** Quantile-quantile plot of observed *P* values (-$\log_{10}$ scale, y-axis) against expected *P* values (-$\log_{10}$ scale, x-axis), with a 95% confidence band (gray) of the same junction. **f,** Counts assigned to the alternative (y-axis) versus reference allele (x-axis), highlighted by significance (orange) and significance and rarity (red), of sample HG00096.