Supplementary Information

Table S1. Filtering steps, numbers and percentages per analysis

Filtering step	OUTRIDER Gene untreated	Gene CHX	Exon untreated	Exon ratio	FRASER Splice Sites[1]	Junctions[1]
1.FEATURECOUNTS raw read counts	62755	62755	432619	432619		
2.Raw read counts - protein coded	20070	20070	306152	306152		
3.Pre-OUTRIDER filtering (exon ratio)	-	-	-	130079 (42.5%)		
4.OUTRIDER passed filterExpression	11977 (59.7%)	12184 (60.7%)	130079 (42.5%)	130038 (99.9%)		
FRASER counting					302054	2339536
FRASER passed filterExpressionAndVariability					156709 (51.9%)	86957 (3.7%)

[1]Split reads supporting exon-exon junctions as well as non-split reads overlapping splice sites are counted.

Table S2. Cut-off steps and filtering of OUTRIDER and FRASER results per method

Cut-off steps	OUTRIDER Gene level	Exon level	Exon ratio	FRASER Splicing
Before cut-off	299425 (11977*25)	3251975 (130079*25)	3250950 (130038*25)	21739265 (86957*25)
A: OUTRIDER aberrant gene expression results, untreated samples				
abs(z-score) > 2.5 & p-value < 0.01	2017 (0.67%)			
B: OUTRIDER aberrant exon expression results, untreated samples				
Abs(z-score) > 2.5 & p-value < 0.01 & mean corrected > 100		6883 (0.21%)		
C: OUTRIDER exon ratio results, untreated samples[1]				
1. gene: norm counts > 200			2917577 (89.7%)	
2. multiple exons per gene with z-score < -2 filtered out			2586582 (79.6%)	
3. exon: z-score < -2.5 & p-value < 0.01			7075 (0.22%)	
4. exon-level mean corrected > 100			1933 (0.06%)	
D: FRASER aberrant splicing results, untreated samples				
p-value < 0.001 & |deltaPsi| > 0.05				2163 (0.01%)

[1]Step 1 filters out exons corresponding *gene*-sample combinations with *gene*-level normalized counts < 200.
Step 2 In addition, exons of genes with multiple aberrant expressed exons: z-score < -2 are filtered out.
Step 3 applies cut-off values of z-score < -2.5 and p-value < 0.01 on the remaining *exon ratio* results. (**Figure 4.A**)
Step 4 applies a cut-off value of > 100 mean corrected counts. (**Figure 4.B**)

Csv file Tables:
Table S3. A Top 20 results aberrant gene expression, ranked by p-value (ascending).
Table S3. B Top 20 results aberrant exon expression, ranked by p-value (ascending).
Table S3. C Top 20 results exon ratio, ranked by p-value (ascending).
Table S3. D Top 20 results exon ratio, after filtering step 4, ranked by p-value (ascending).
Table S3. E Top 20 results aberrant splicing, ranked by p-value (ascending).

Table S3. can be found online at:
https://github.com/lonnekevanbrussel/Final-Master-Project/blob/main/FMP_supplementary_tableS3.xlsx

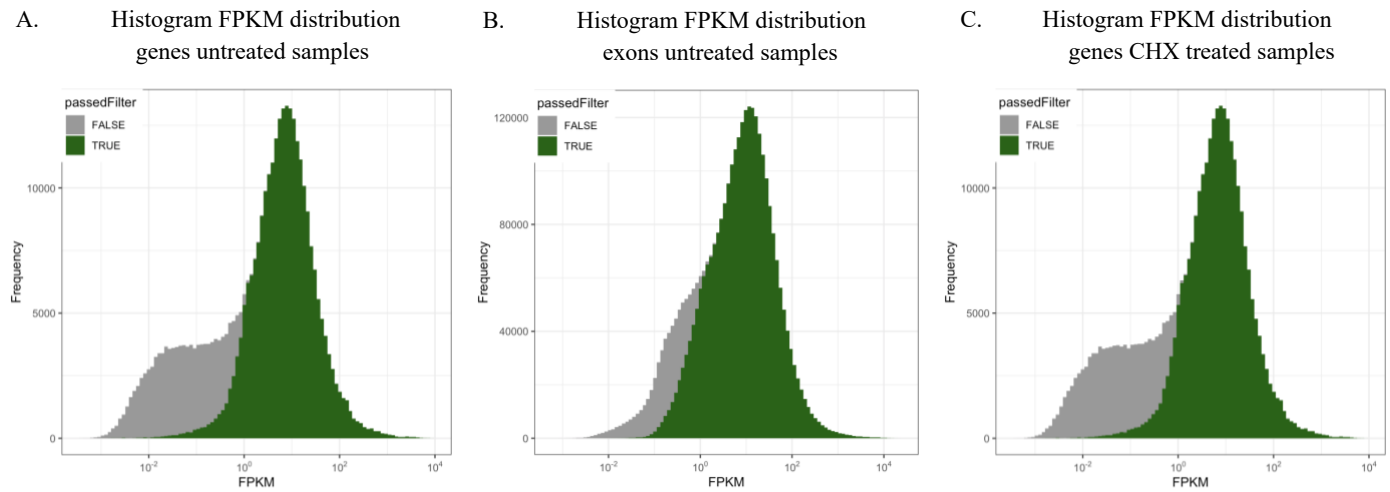


Figure S1. Histograms of filtering of low expressed genes and exons. Consecutively untreated samples on gene (A) and exon (B) level, and CHX treated samples on gene level (C). Frequency of FPKM values is depicted, passing of determined filter of the 95 percentiles of samples with FPKM>1 are depicted in green.

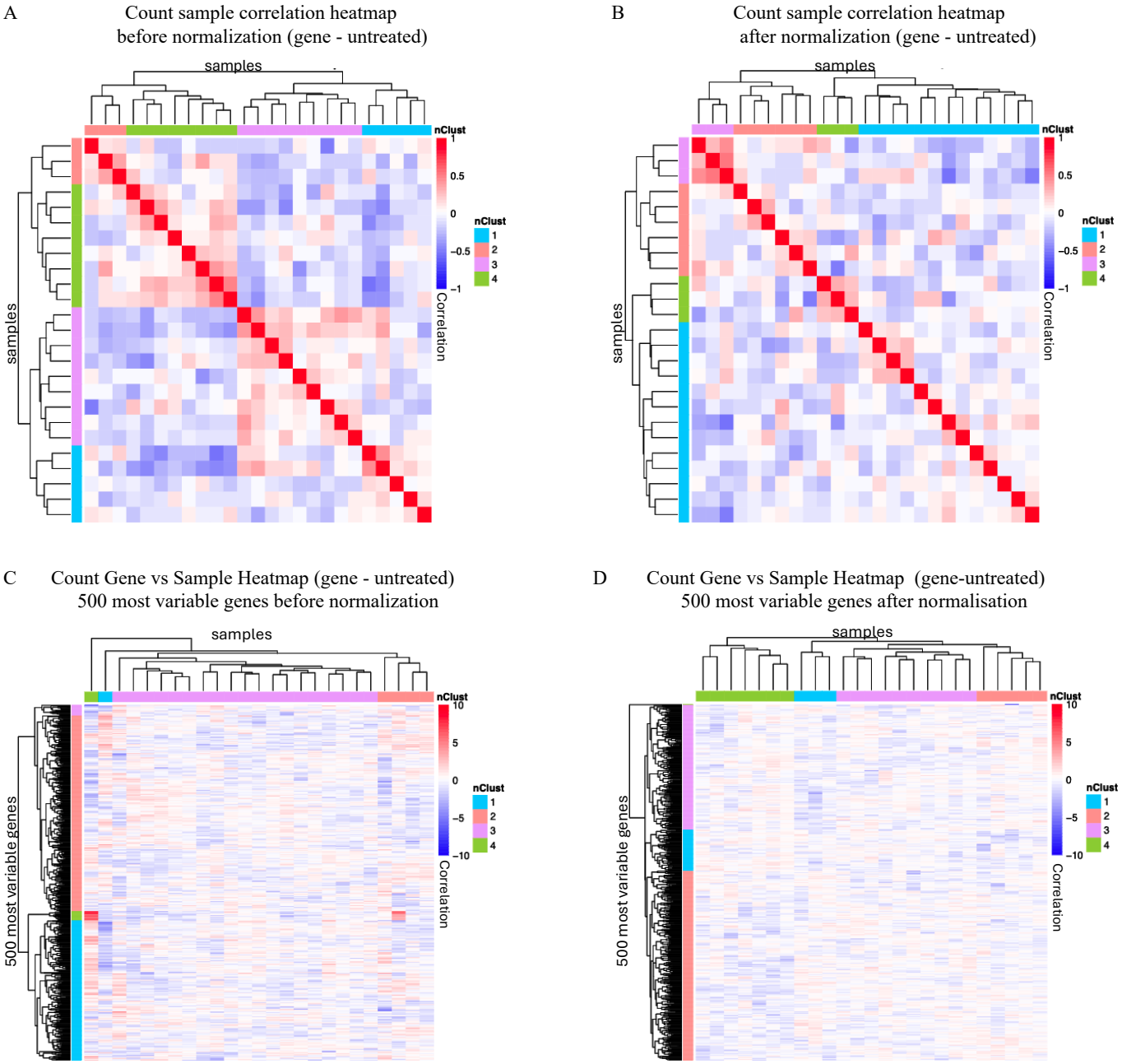


Figure S2 sample-sample (A and B) and sample-gene (500 most variable genes, C and D) count heatmap before and after autoencoder normalization at gene level for untreated samples. Samples and genes are clustered in four groups. After this normalization step, in (B) and (D) less batches are visible compared to (A) and (C). In (B) and (D), lighter color reflects decreased correlation - closer to zero - after normalization. This indicates the normalization succeeded.

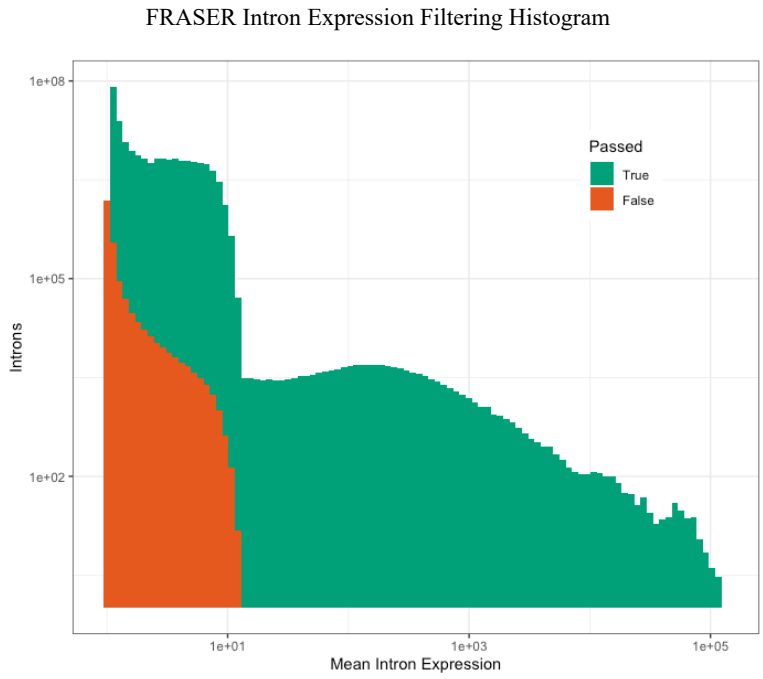


Figure S3. Histogram of filtering out low expressed introns in FRASER module. In orange introns not passing *filterExpressionAndVariability* function: at least one sample has 20 (or more) reads, at least 5% of the samples per junction need to have at least 10 total read counts (N). The minimal delta psi variation for an intron to pass the filter is set to 0.05.

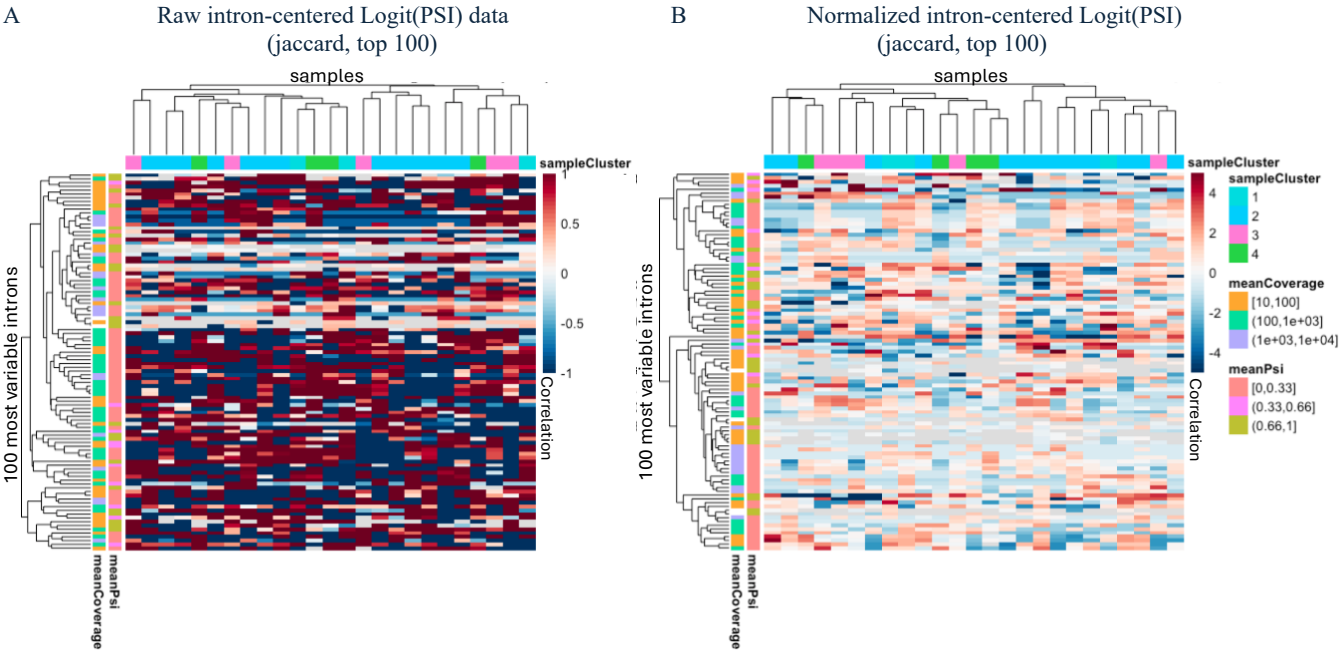
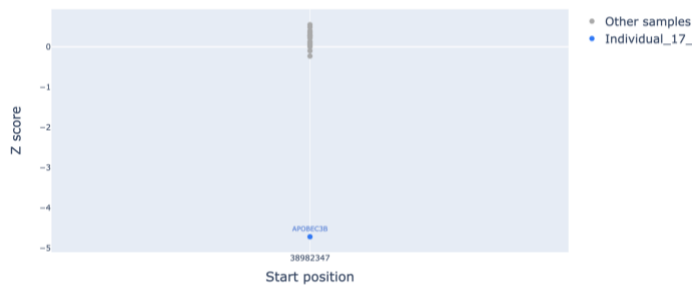
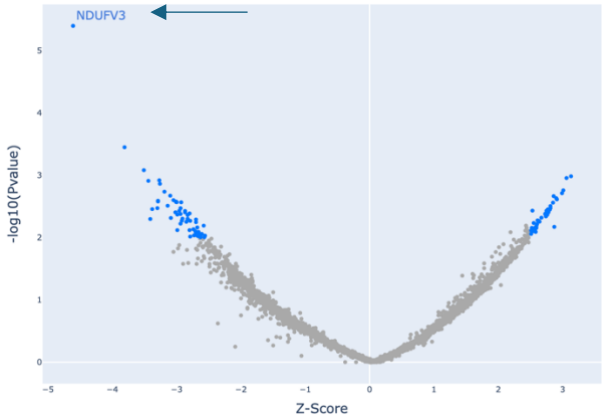


Figure S4: sample-intron correlation heatmaps of samples versus the top 100 most variable introns (junctions) by logit jaccard psi values before and after normalization in FRASER. Correlation between samples is reduced (much closer to zero) after normalization, which indicates the normalization succeeded.

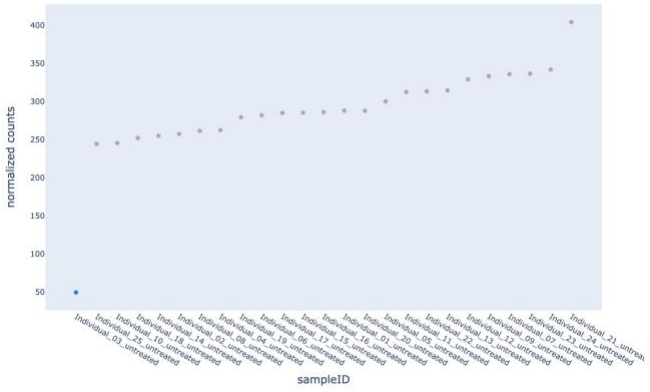
A. APOBEC3B z-scores (gene - untreated samples)



B. Volcano plot Individual 4 (gene - untreated sample)



C. Normalized counts NPHP1 untreated samples



D. IGV coverage SLC38A5

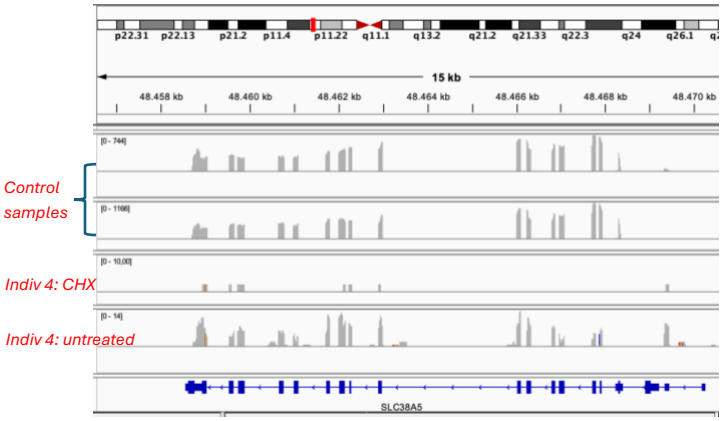
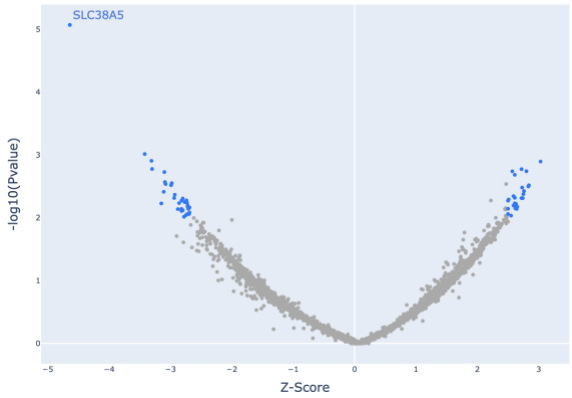
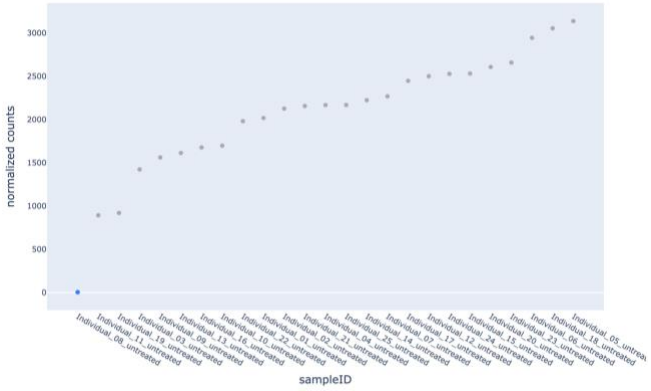


Figure S5: Jupyter web app and IGV visualisation results of top 4 results most aberrantly expressed genes (see Table S3A top 20).
(A) Z-scores untreated gene-level rank 1: APOBEC3B for individual 17 is depicted in blue, showing a z-score > -4, compared to the other samples of the sample set, depicted in grey.
(B) Volcano plot for individual 4, untreated gene level result rank 2: NDUFV3. All results for individual 4 are depicted in the volcano plot, outliers in blue. NDUFV3 in the upper left cornder.
(C) normalized counts plot, for all untreated samples at gene level for result rank 3: NPHP1
(D) IGV analysis for gene level result rank 4: SLC38A. Coverage for individual 4 is loaded in the lower two bars for CHX treated and untreated samples, compared to untreated samples of two control samples in the upper two bars. Coverage for individual 4 (max 10; 14) is much lower compared to the other two samples (744; 1166)

A. Volcano plot genes individual 8 untreated sample



B. Normalized counts SLC38A5 untreated samples



C. Z-scores SLC38A5 untreated samples

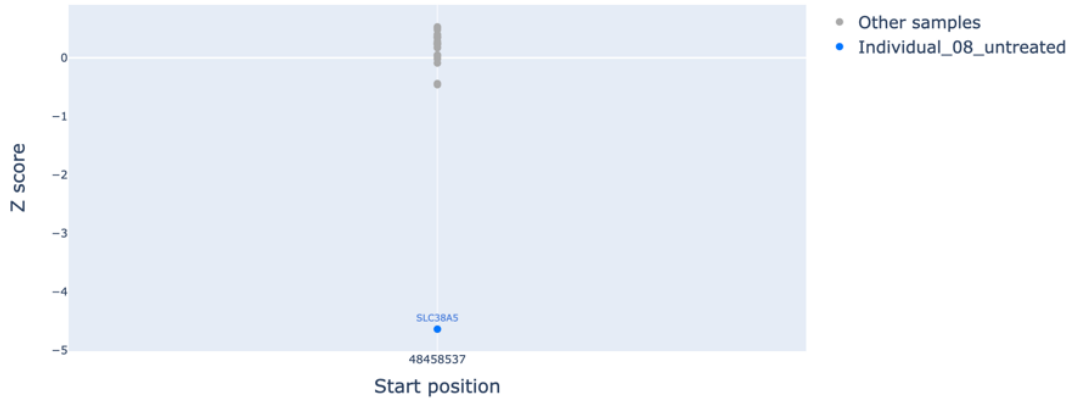
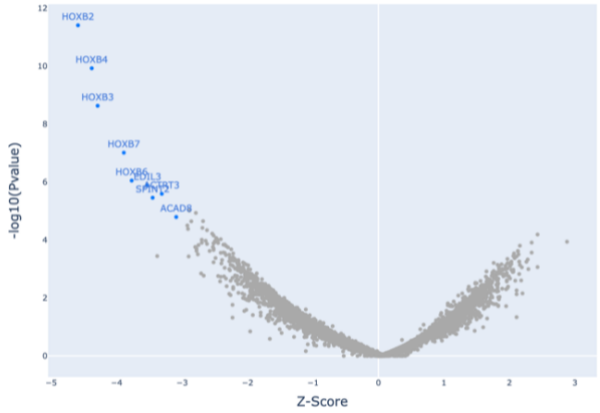
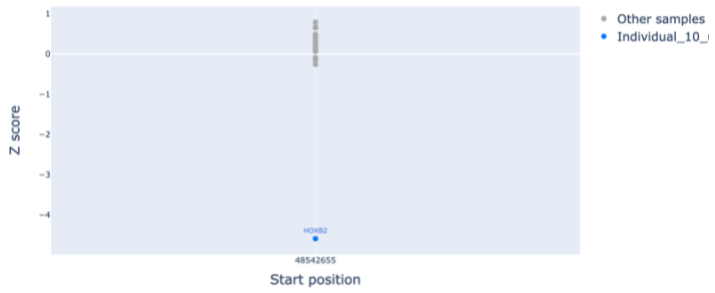


Figure S6. Jupyter Notebook app plots for candidate gene SLC38A5 regarding individual 8
(A) Volcano plot, showing all untreated, gene-level OUTRIDER results for individual 8, with SLC38A5 is depicted in the upper left corner, demonstrating SLC38A5 is most downregulated outlier for individual 8. Genes depicted in blue are within cut-off values for gene-level analysis.
(B) Normalized counts plot for SLC38A5, regarding all untreated samples counted at gene-level, confirming downregulation (zero normalized counts) for individual 8, depicted in blue.
(C) Z-score plot, showing SLC38A5 gene-level z-scores for all untreated samples of the sample set, individual 8 depicted in blue and all other samples of the sample set in grey. Demonstrating downregulation of SLC38A5 for individual 8.

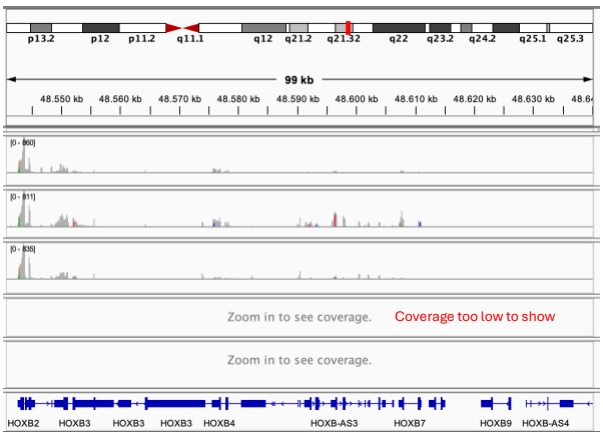
A. Volcano plot genes individual 10 untreated samples.



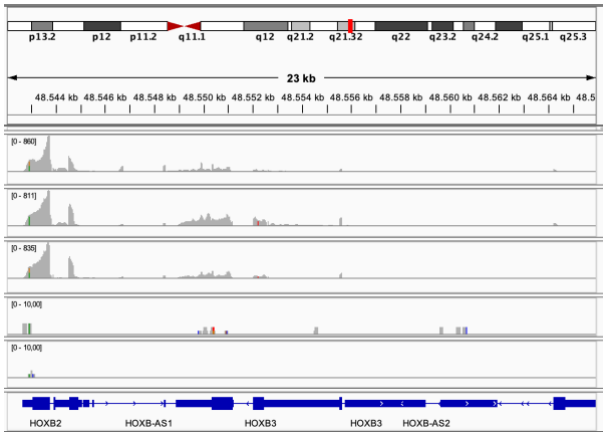
B. Z-scores HOXB2 untreated samples



C. IGV: HOXB genes chr17: 48.544-48.630



D. IGV zoomed in chr 17: 48.544 – 48.564

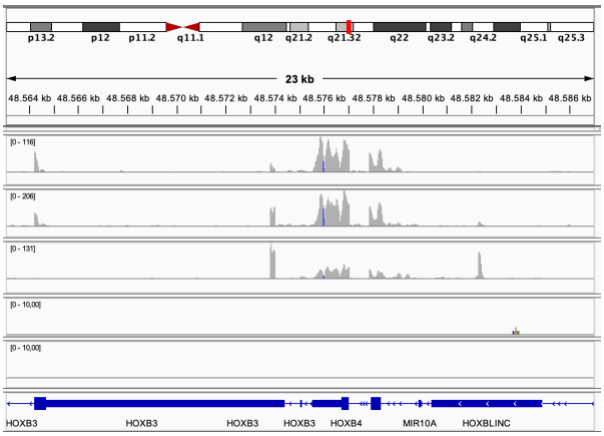


Control samples

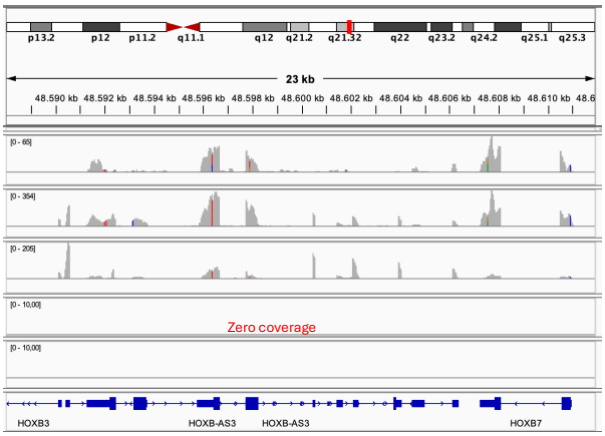
Indiv 10: CHX

Indiv 10: untreated

E. IGV zoomed in chr 17: 48.564 – 48.586



F. IGV zoomed in chr 17: 48.588 – 48.612



Control samples

Indiv 10: CHX

Indiv 10: untreated

Figure S7. Result plots for candidate genes individual 10

(A) Jupyter app Volcano plot showing all OUTRIDER results for individual 10, untreated sample at gene level. Most aberrant expressed genes are depicted in blue, HOXB genes are located in the upper left corner, demonstration most dowregulated genes for this individual.
(B) Jupyter app Z-score plot, showing HOXB2 gene z-scores for all untreated samples, with in blue z-score for individual 10, in grey z-scores for other sample set untreated samples, demonstrating downregulation of HOXB2 for individual 10.
(C) shows rna-seq coverage in IGV viewer for genes HOXB2 - HOXB-AS1 – HOXB3 – HOXB-AS2 – HOXB4 – HOXB-AS3 – HOXB6 - HOXB7 and HOXB-AS4, located consecutively on chr17 q21.32 (48.544-48.630). The lower 2 bars correspond to individual 10 CHX and untreated samples, both showing max 10 counts coverage. In contrast to three untreated control samples in the upper bars.
(D.E.F.) like (**C**), zoomed in on consecutive loci on the chromosome: (**C**): 48.544 – 48.564; (**D**): 48.564 – 48.586; (**E**): 48.588 – 48.612.
(**C – F**) demonstrate near zero expression of the HOXB genes.

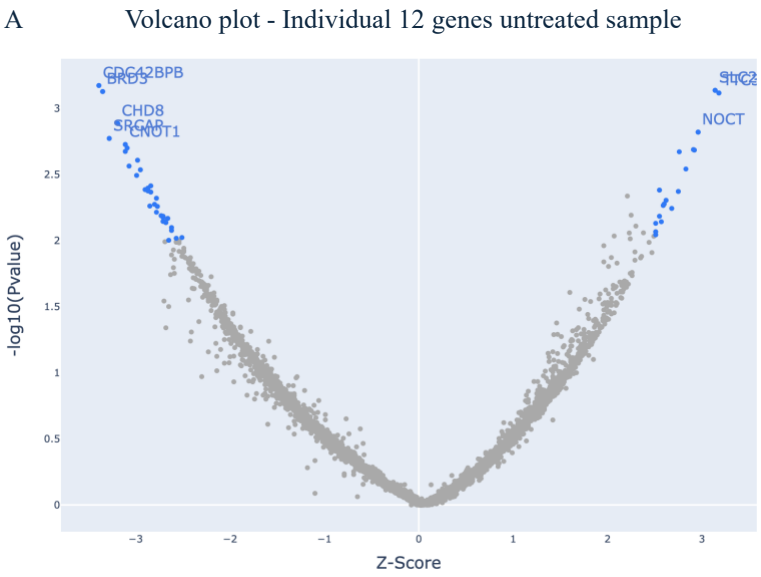
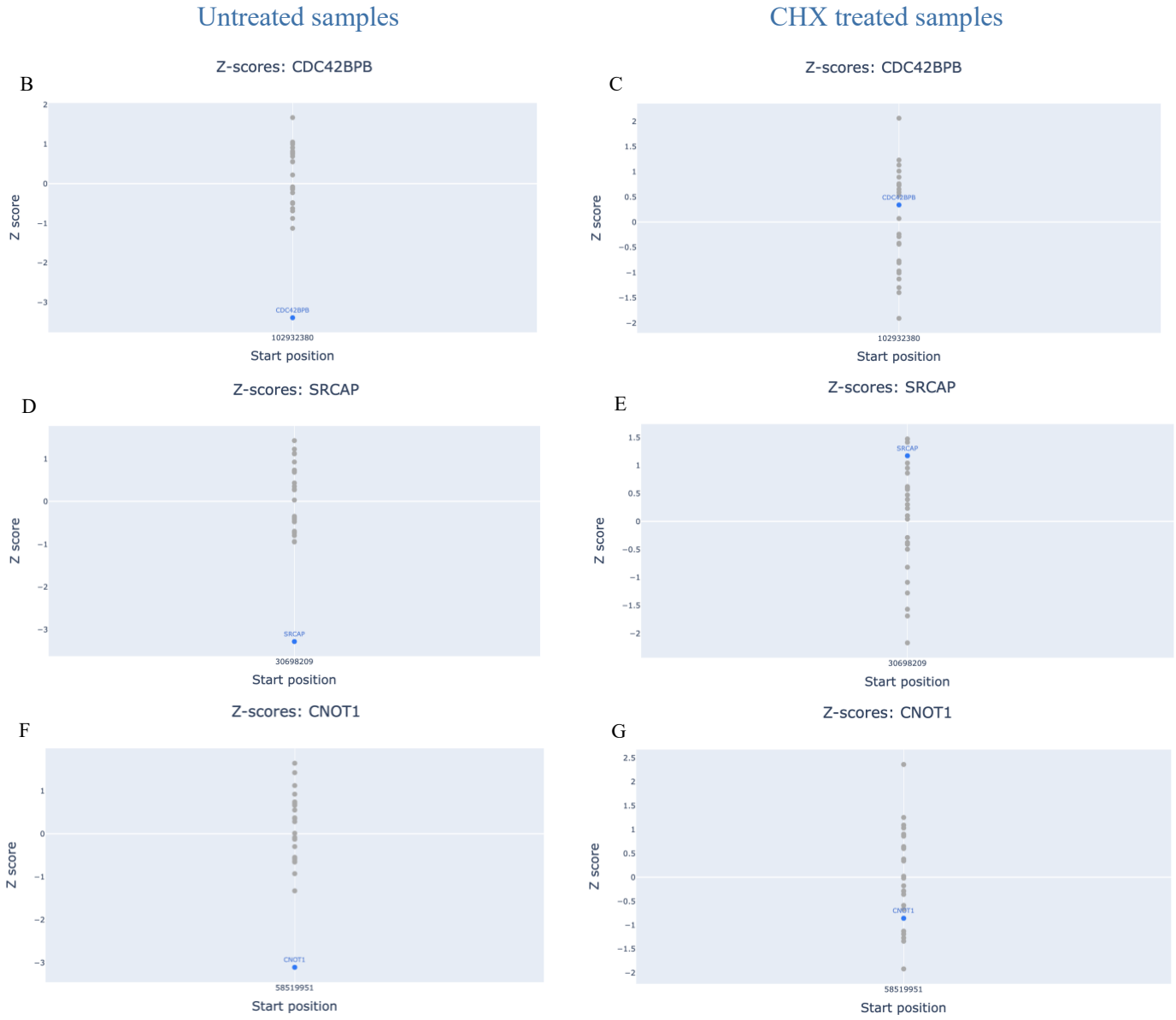


Figure S8. Candidate genes for individual 12
(A) Volcano plot showing all OUTRIDER gene-level results for individual 12, untreated sample. Aberrant expressed genes within cut-off values are depicted in blue, including the three most downregulated candidate genes CDC42BPB, SRCAP and CNOT in the untreated sample, annotated in the left upper corner.
(B-G) Jupyter Notebook app z-score plots for the three most downregulated candidate genes, for untreated samples (left: **B-D-F**) and CHX treated samples (right: **D-E-G**) of individual 12. The sample of interest (individual 12) is depicted in blue in all figures, other samples of the sample set are depicted in grey. These plots show downregulation for these three candidate genes in the untreated sample and normal expression for these genes in the CHX treated sample, suggesting the transcripts are subject to NMD.



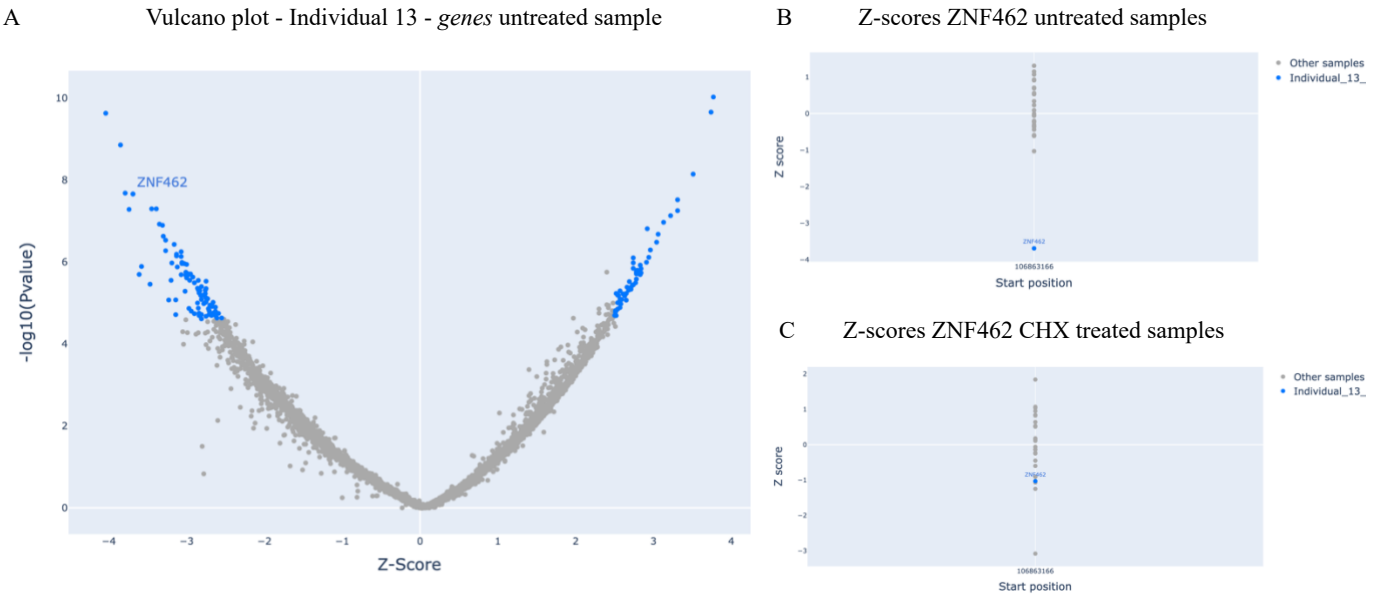


Figure S9. Jupyter app charts regarding candidate gene ZNF462 for individual 13

(A) Volcano plot, gene expression analysis regarding the untreated sample of individual 13, demonstrating ZNF462 is one of the most downregulated genes for this individual.
(B) and (C) gene expression z-scores for ZNF462 , respectively for the untreated (B) and CHX treated (C) samples. Individual 13 is depicted in blue in both figures, the other refererence set samples are depicted in grey. These figures show downregulation for the untreated sample and more upregulated gene expression for the CHX treated sameple. This suggests ZNF462 is subject to NMD.

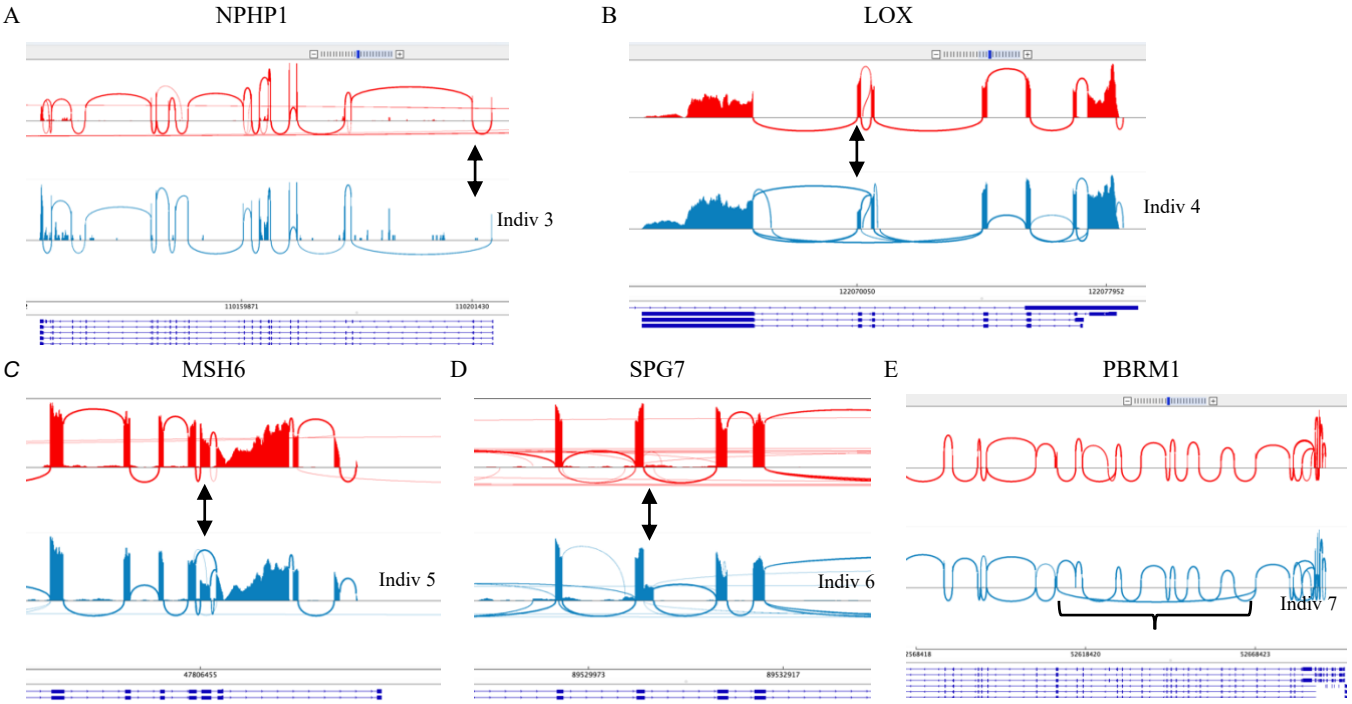


Figure S10. IGV sashimi plots demonstrate the splicing events for all positive control samples, selected for exon-level analyses. All positive controls (untreated sample, depicted in blue) are compared to a random sample from the sample set in red, showing normal splicing. Arrows and bracket (in **E**) indicate the position of the aberrant splice event. (**A**), (**B**) and (**C**) show downregulated exon expression and exon skips regarding the NPHP1, LOX and MSH6 gene for individuals 3,4 and 5 respectively in blue, compared to a control sample with normal exon expression and splicing. (**D**) depicts an alternative splice site and partial intron retention for individual 6 (blue) in the SPG7 gene and (**E**) shows a multiple exonskip for individual 7 (blue) in the PBRM1 gene.

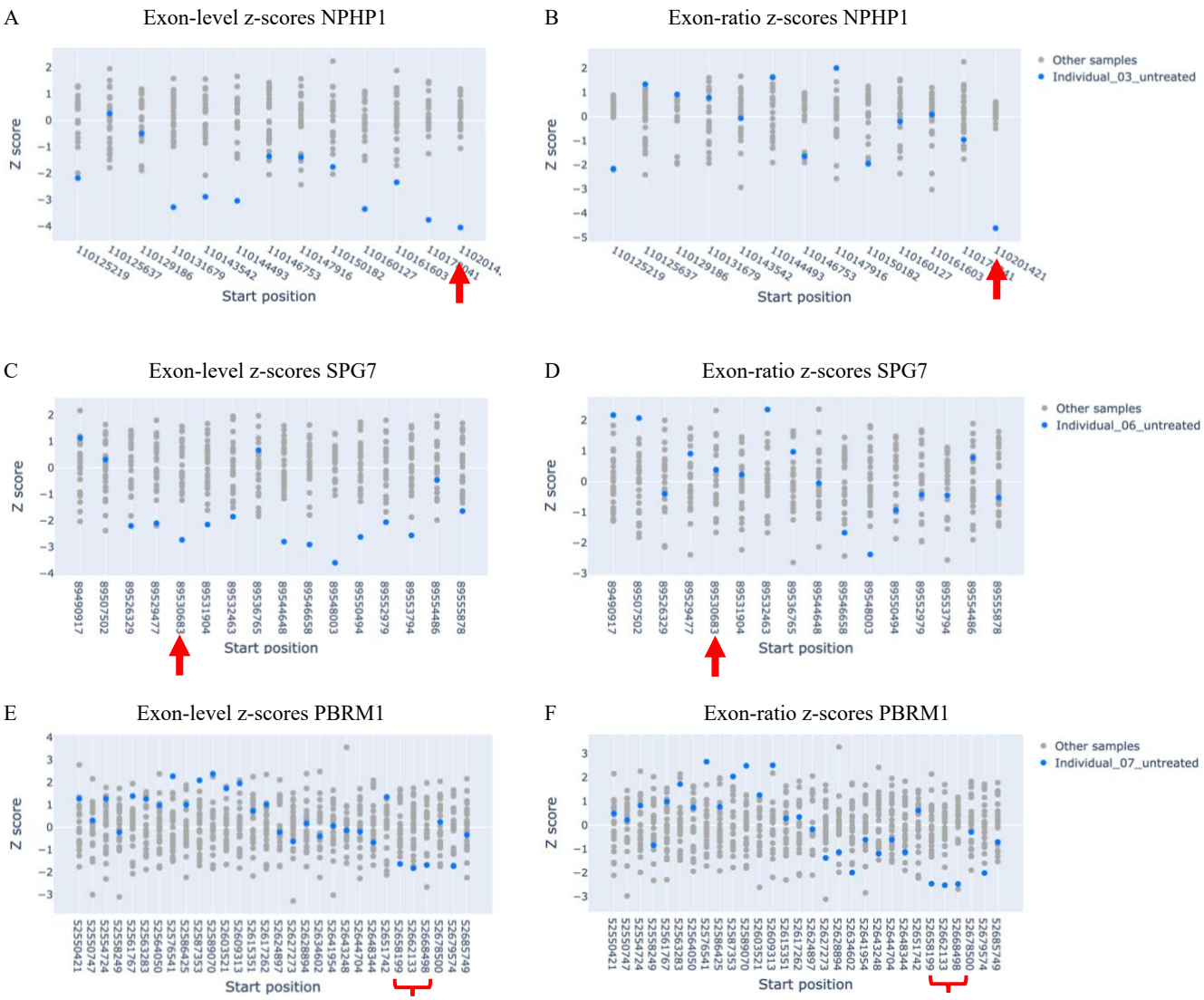


Figure S11 Exonic z-score differences between OUTRIDER exon-level (**A, C, E**) and exon-ratio (**B, D, F)** results show improved detection of exon skips in positive control samples with the exon-ratio method, but not for retained intron detection. All figures depict exonic z-scores for untreated samples of the NPHP1, SPG7 and PBRM1 gene for individual 3,6 and 7 respectively (depicted in blue), compared to all other samples of the sample set (depicted in grey).
(A) and (B) For individual 3, the NPHP1 gene is downregulated (**Error! Reference source not found.**). Furthermore, almost all exons showing downregulated z-scores as well for the OUTRIDER exon-level results for this individual in A. In contrast, the exonic exon-ratio z-scores in B demonstrate aberrant exon expression is exclusively detected at the expected position of the exon skip in NPHP1 for individual 3 (see arrow and sashimi plot Figure 10A).
(C) and (D) Individual 6 has an alternative splice site in the SPG7 gene (IGV sashimi plot Figure S10D), therefore upregulated exon expression is expected at the position of the arrow. However, the OUTRIDER exon-level z-score at this position shows exonic downregulation for individual 6 in blue (**C**), as well as for multiple other exons of SPG7. In (**D**) normal exonic expression is detected with the OUTRIDER exon-ratio method (z-score +0.38, Table 3). For the CHX treated sample of this individual, an exon-ratio z-score of +0.82 is detected (not depicted), which is still not an aberrant z-score.
(E) and (F) F shows aberrant exon-ratio z-scores for three of the skipped exons in the PBRM1 gene for individual 7 untreated sample (see also sashimi plot **Figure S10E**), whereas no aberrant z-scores with the exon-level analysis (**E**) at these positions.

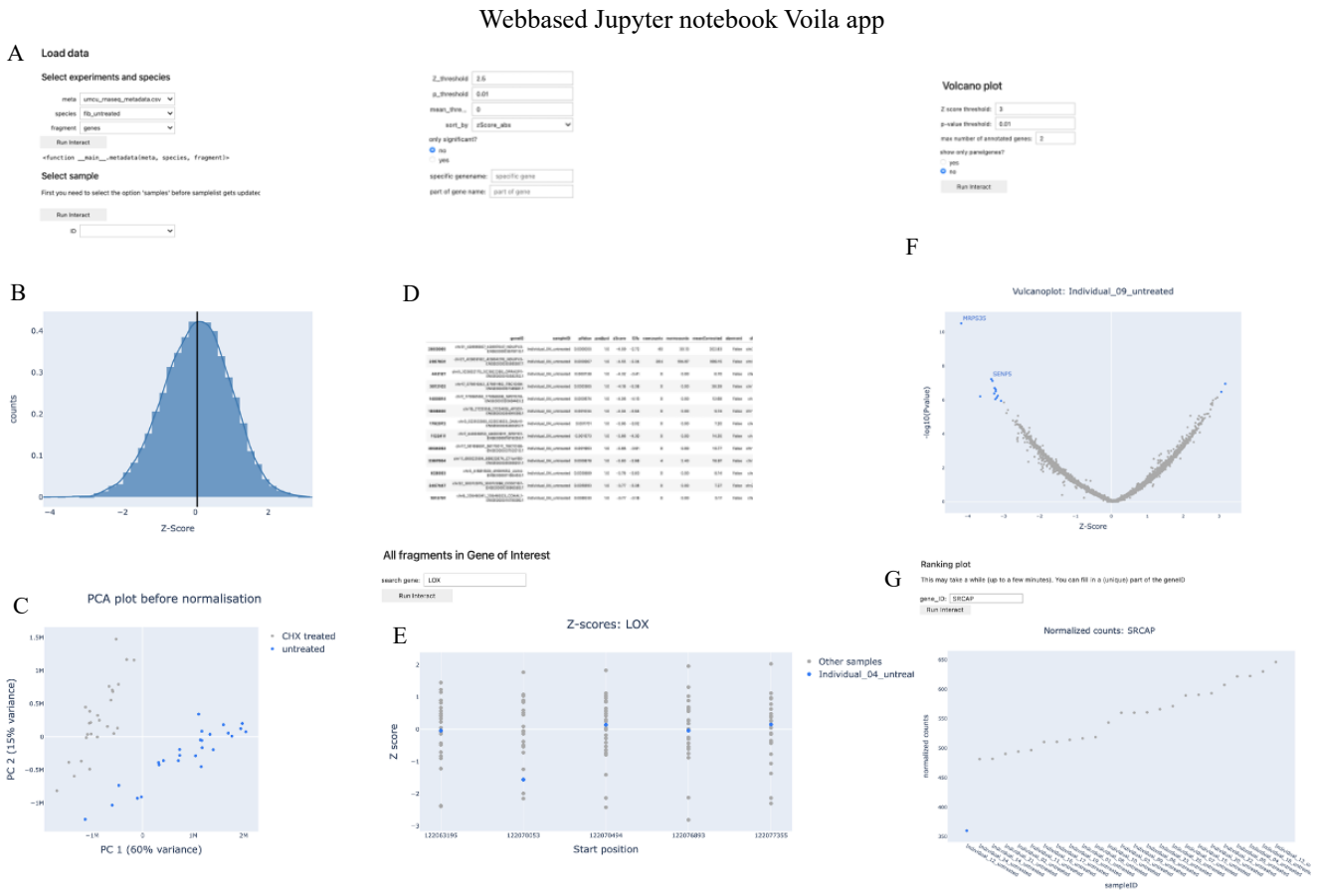


Figure S12. Diverse functionalities of webbased Jupyter notebook Voila app (Erasmus Medical Center[6]) and additional functionalities, colorings and annotations created.

- **(A)** Filtering panels to select sample, treatment, gene, exon and diverse settings per chart
- **(B)** Z-score distribution plot for selected individual, feature (gene/exon) and treatment (untreated/CHX)
- **(C)** PCA plot, added functionality: comparing normalized counts untreated and CHX samples
- **(D)** list of OUTRIDER result scores for selected individual, feature (gene/exon) and treatment (untreated/CHX)
  Added functionality: associated clinical phenotype, filtering on minimum mean corrected (mean normalized counts per gene)
- **(E)** z-scores on selected gene- or exon-level compared to all other samples.
  Added functionality: comparison z-scores sample of interest versus all other samples in grey
- **(F)** Vulcanoplot: added functionality: clickable gene names. coloring of genes, sorted by p-value
  Added functionality: clickable genes, opens window with
- **(G)** Normalized counts plot