

RNA-Seq Differential Expression Analysis

Tom O Project

2025-01-08

Contents

1 Experimental Design	1
1.1 Samples and Conditions	1
1.2 Reference Level	1
1.3 Design Matrix	1
1.4 Single Comparisons	2
1.5 Double Comparisons (DNAPKi Effects)	2
1.6 Interpretation	2
1.6.1 Single Comparison vs. NTC_FALSE:	2
1.6.2 Double Comparison (DNAPKi effect):	2

1 Experimental Design

1.1 Samples and Conditions

Each sample has two key attributes: - Gene Target (e.g. GAL4, PRDX1, NTC, etc.) - DNAPKi Treatment (TRUE/FALSE)

We combine these into a single factor called condition. For example: - GAL4_FALSE = “GAL4 with no DNAPKi treatment” - GAL4_TRUE = “GAL4 with DNAPKi treatment” - PRDX1_FALSE = “PRDX1 with no DNAPKi treatment” - PRDX1_TRUE = “PRDX1 with DNAPKi treatment” - NTC_FALSE = “Negative control (no DNAPKi)”

1.2 Reference Level

We make NTC_FALSE the reference (baseline) level. This ensures that any coefficient for “conditionXYZ” in the linear model is interpreted as:

(XYZ) - (NTC_FALSE)

1.3 Design Matrix

We set up a linear model with the formula `~ replicate + condition`, where: - replicate accounts for batch or biological replicate effects - condition captures all combinations of gene target and DNAPKi status

Concretely, our design matrix columns include: - Intercept (which corresponds to

NTC_FALSE) - Replicate terms - conditionGAL4_FALSE, conditionGAL4_TRUE, condition-PRDX1_FALSE, conditionPRDX1_TRUE, etc.

1.4 Single Comparisons

For any given conditionX, we do a single comparison X vs NTC_FALSE by extracting that specific coefficient in the model. This tells us whether X is significantly different from the negative control (no DNAPKi).

1.5 Double Comparisons (DNAPKi Effects)

We define contrasts to compare (GeneTarget_TRUE - GeneTarget_FALSE).

For example: - GAL4_DNAPKi_effect = conditionGAL4_TRUE - conditionGAL4_FALSE - PRDX1_DNAPKi_effect = conditionPRDX1_TRUE - conditionPRDX1_FALSE

Each such contrast tests whether DNAPKi treatment (for a specific gene target) significantly changes expression compared to the same gene target without DNAPKi.

1.6 Interpretation

1.6.1 Single Comparison vs. NTC_FALSE:

- $\log FC > 0 \Rightarrow$ Genes are up-regulated relative to the negative control
- $\log FC < 0 \Rightarrow$ Genes are down-regulated relative to the negative control

1.6.2 Double Comparison (DNAPKi effect):

- $\log FC > 0 \Rightarrow$ Genes are further increased by DNAPKi treatment (for that gene target)
- $\log FC < 0 \Rightarrow$ Genes are decreased by DNAPKi for that gene target