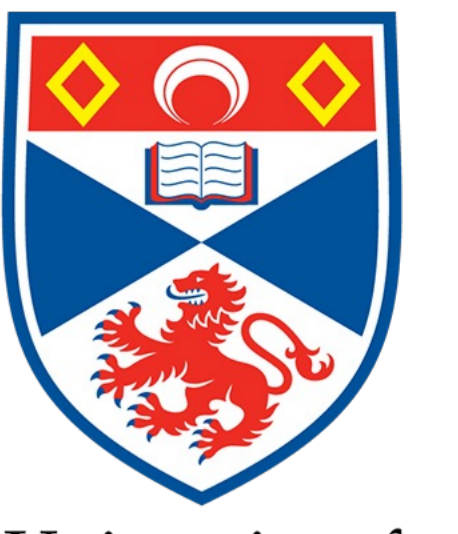


Does differential mRNA expression predict better protein-mRNA correlation? A meta-analysis

Lillian Bates, Supervised by Dr V Anne Smith

School of Biology, University of St Andrews

Contact: lgb9@st-andrews.ac.uk



University of St Andrews

Introduction

Central Dogma of Molecular Biology:



- Describes the **flow of genetic information from the DNA genome to protein** called **gene expression**. Several regulatory mechanisms at each step.
 - Messenger RNA (mRNA)**: blueprint copied from DNA to construct the protein.
- Protein expression**:
 - has important applications in **medicine, drug discovery, & disease research**.
 - Level often inferred by mRNA expression, assuming a correlation between mRNA & protein.
 - However, poor mRNA-protein correlation has been, which calls into question the use of mRNA to infer protein expression levels.
- Cells can control the amount of a protein by regulating the expression of its encoding gene during transcription & translation.
- Differential Expression:
 - Differentially Expressed Gene (DEG)** → a gene that is regulated differently between two conditions
 - Non-differentially Expressed Gene (NDEG)** → Opposite of DEG; gene's regulation does not change between two conditions.

Research Question: Do DEGs have a tighter mRNA-protein correlation compared to NDEGs genes?

- Two studies have found DEGs have tighter mRNA-protein correlation (Koussounadis et al., 2015; Erdmann et al., 2018). These studies only assessed this question in single species.
- Meta-analysis was used to assess to this question across 7 independent datasets.**

Methods

Stage 1: Finding datasets

Found 7 eligible datasets.

Stage 2: Reading the data into R

Importing data into R and restructuring the data for analysis.

Stage 3: Identification of differentially expressed genes (Figure 1)

Classified all genes as a DEG or NDEG by comparing mRNA expression under control and experimental conditions.

Stage 4: mRNA-protein correlation

Found the correlation between mRNA and protein expression for each gene, using Spearman's rank correlation (Spearman, 1904).

Stage 5: Significance testing (Table 1)

Testing if DEGs have significantly higher mRNA-protein correlation.

Stage 6: Meta-analysis (Table 1)

Combining the *P*-value result from Stage 5 from each dataset to generate a combined *P*-value. Used Stouffer's Meta-analysis method (Stouffer et al., 1949).

Results

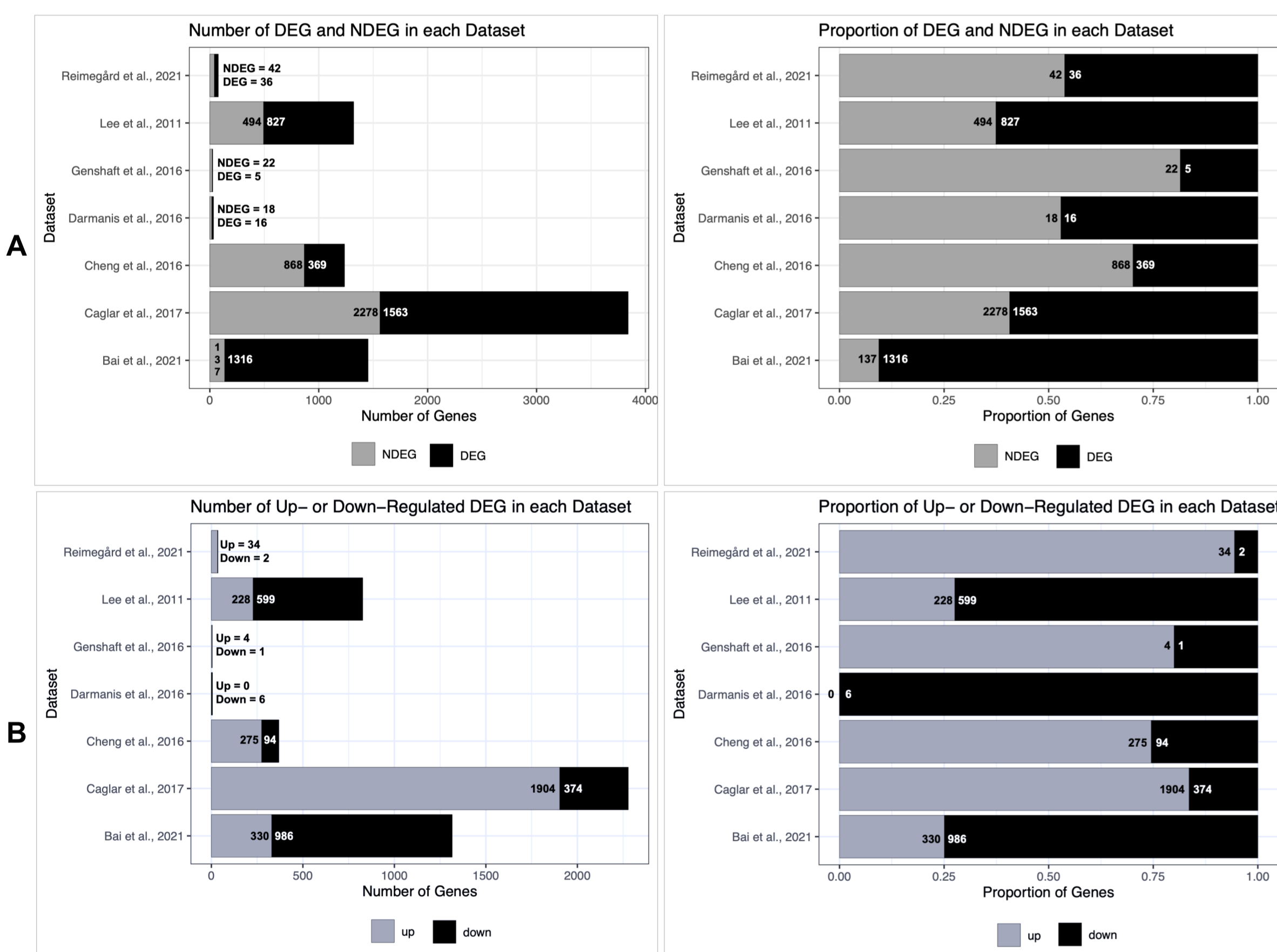


Figure 1. Identification of DEGs for each dataset. For all the figures, black or white text on or adjacent to the bars states the number of genes the bar represents. Figure 2A (top) shows the breakdown of DEG (black bars) and NDEG (grey bars) for each dataset. 2A (left) compares the breakdown of DEGs/NDEGs scaled by the number of genes. 2A (right) compares the ratio of DEG (black) to NDEG (grey). Figure 2B (bottom) compares the breakdown of up- (grey) and down-regulated (black) DEG for each dataset. 2B (left) compares the breakdown of up-/down-regulated genes scaled by the number of genes. 2B (right) compares the ratio of up-regulated DEG (grey) to down-regulated DEG (black). Plots were made in R using ggplot2 (Wickham, 2016).

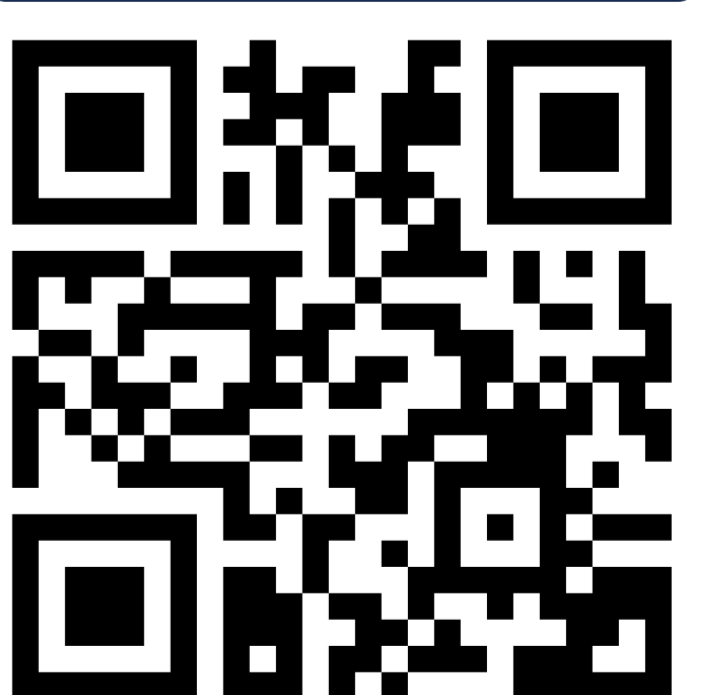
Table 1. Description of datasets used in analysis, and individual (Stage 5) and meta-analysis (Stage 6) results from testing if DEGs have significantly higher mRNA-protein correlations. *P* < 0.05 were considered significant.

Dataset	Cell/Tissue	Treatment	KS-Test <i>P</i> -value (<i>P</i> < 0.05)	KS-Test Significance	BM-Test <i>P</i> -value (<i>P</i> < 0.05)	BM-Test Significance
Lee et al., 2011	<i>Saccharomyces cerevisiae</i> (BY4741 Strain)	0.7 M NaCl	0.904	Not Significant	0.999999998	Not Significant
Darmanis et al., 2016	<i>Homo sapiens</i> (U3035MG grade IV glioblastoma cell line)	10 ng/μl bone morphogenetic protein 4 (BMP4)	0.051	Not Significant	6.83 × 10 ⁻³	Significant
Genshaft et al., 2016	<i>Homo sapiens</i> (Human Breast Adenocarcinoma cell line)	1 μM phorbol-12-myristate-13-acetate (PMA)	0.251	Not Significant	0.296	Not Significant
Reimegård et al., 2021	<i>Homo sapiens</i> (embryonic stem cell line HS181)	Neural Induction by neural induction medium (NIM)	0.126	Not Significant	0.183	Not Significant
Cheng et al., 2016	<i>Homo sapiens</i> (HeLa cell line)	2.5 mM dithiothreitol (DTT)	0.167	Not Significant	0.355	Not Significant
Caglar et al., 2017	<i>Escherichia coli</i> (REL606)	variable carbon source, variable Mg concentrations, variable Na concentrations	< 2.2 × 10 ⁻¹⁶	Significant	1.27 × 10 ⁻⁷	Significant
Bai et al., 2021	Maturing <i>Arabidopsis thaliana</i> seeds	Seed germination	< 2.2 × 10 ⁻¹⁶	Significant	7.99 × 10 ⁻²⁹	Significant
Stouffer's Meta-analysis: Combined <i>P</i>-value			1.23 × 10⁻¹³	Significant	1.49 × 10⁻⁸	Significant

Conclusions

- The meta-analysis combined *P*-values were significant, supporting the hypothesis that DEGs have a tighter mRNA-protein correlation compared to NDEGs.
- Suggests mRNA expression data can more reliably be used to infer protein expression for DEGs than NDEGs, which an impact future research.
- Suggests that mRNA expression regulation (differential expression) does have downstream impact on the amount of protein expression.

Essay QR Code



Acknowledgements

I would like to thank my research advisor, Dr V. Anne Smith, for her invaluable support, guidance, and mentorship on my project. I would also like to thank the Laidlaw Foundation for funding my project through my participation in the Laidlaw Scholars Leadership and Research Programme.

References

- Bai, B. et al., 2021. Delayed Protein Changes During Seed Germination. *Frontiers in Plant Science*, Volume 12, pp. 1-11.
- Caglar, M. U. et al., 2017. The *E. coli* molecular phenotype under different growth conditions. *Scientific Reports*, 7(1), p. 45303.
- Cheng, Z. et al., 2016. Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. *Molecular Systems Biology*, 12(1), p. 855.
- Darmanis, S. et al., 2016. Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells. *Cell Reports*, 14(2), pp. 380-389.
- Erdmann, J. et al., 2018. Environment-driven changes of mRNA and protein levels in *Pseudomonas aeruginosa*. *Environmental Microbiology*, 20(11), pp. 3952-3963.
- Genshaft, A. S. et al., 2016. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biology*, 17(1), p. 188.
- Koussounadis, A. et al., 2015. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Scientific Reports*, 5(1), p. 10775.
- Lee, V. M. et al., 2011. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Molecular Systems Biology*, 7(1), p. 514.
- Reimegård, J. et al., 2021. A combined approach for single-cell mRNA and intracellular protein expression analysis. *Communications Biology*, 4(1), p. 624.
- Spearman, C., 1904. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1), p. 72.
- Stouffer, S. et al., 1949. *The American Soldier: Adjustment during Army Life*. Princeton(NJ): Princeton University Press.