**Results section:**

**Proteome and transcriptome signatures correlate in a cell-type specific manner**

While transcriptome and proteome analyses can fulfill similar roles in Systems Biology, namely the genome wide characterization of cellular responses to perturbations, it is known that mRNA copy numbers are not powerful predictors of protein abundance. Pearson Correlation numbers for the explained variation in protein concentration by mRNA levels are given around R = 0.4 (corresponding to R-squared = 0.16) [1, 2], despite high precision between biological replicates in both mRNA and protein measurements [3]. The intersection of 7285 protein groups (proteomics) and 17230 gene products (transcriptomics), totaling 4481 shared proteins across three cell types (Alpha, Beta, Delta), two treatment conditions (control, IFNγ-treated) and three biological replicates (N = 18). In this dataset, we observed high inter-replicate mean R-squared values of 0.98 and 0.91 (supplementary 20250423\_figure\_panels\_replicate\_precision\_P\_T.ai), with mean CVs of 9.3 and 21.8 % (supplementary 20250423\_figure\_panels\_CV\_across\_replicates\_P\_T.ai) for proteomics and transcriptomics, respectively. Correlation between transcriptomics and proteomics data was assessed via linear regression and Pearson correlation, where correlations were generally highest for Beta and lowest for Delta cells (figure 20250423\_figure\_panels\_figure\_5.ai, A, B). Furthermore, when exhaustively comparing all replicates with each other, proteome-transcriptome correlations were significantly lower for IFNγ-treated Delta, but not Beta cells when compared to controls (figure 20250423\_figure\_panels\_figure\_5.ai, C and supplementary 20250423\_figure\_panels\_Fig\_5\_1\_replicate\_scatter.ai). We speculate that this observed difference may be due to shorter-term transcriptional responses to IFNγ signaling, causing a discrepancy between the cell’s proteomic inventory and synthesized mRNAs. Subsequently, we asked whether cell type and/or treatment could be recovered from cellular proteomics and/or transcriptomics signatures. Indeed, we found that proteomics signatures allowed for perfect separation by cell type, but not treatment (figure 20250423\_figure\_panels\_figure\_5.ai, D). Conversely, transcriptomics signatures were only able to discern Beta cells, but failed to capture the difference between Alpha and Delta cells (figure 20250423\_figure\_panels\_figure\_5.ai, E). Taken together, these findings suggest that proteomics data offers a valuable and highly descriptive avenue of analysis beyond transcriptome sequencing. Furthermore, observed discrepancies in proteome-transcriptome correlation suggest cell-fate specific dynamics for messenger RNA and protein copy numbers. A question not investigated by this study is, whether longer IFNγ treatments might allow for closer synchronization of mRNA and protein levels upon perturbation.

[1]: Vogel, Christine, and Edward M. Marcotte. "Insights into the regulation of protein abundance from proteomic and transcriptomic analyses." Nature reviews genetics 13.4 (2012): 227-232.

[2]: Schwanhäusser, Björn, et al. "Global quantification of mammalian gene expression control." Nature 473.7347 (2011): 337-342.

[3]: Sharma, Kirti, et al. "Cell type–and brain region–resolved mouse brain proteome." Nature neuroscience 18.12 (2015): 1819-1831.

**20250423\_figure\_panels\_figure\_5.ai caption:**

**A**, **B**, Beta and Delta cell specific comparison of triplicate median values between proteomics (x-axis) and transcriptomics (y-axis) data. Simple linear regressions are shown as red dashed lines, with equations and R-squared values indicated in the respective boxes. **C**, All-versus-all replicate comparison between two treatment conditions. Each dot represents the Pearson correlation value from a unique pairing of one proteomics and one transcriptomics replicate, in their respective conditions. P-values from two-sided independent t-tests are shown. **D**, Proteomics sample-sample Pearson correlation coefficient heatmap. Each tile is colored by the R-value of comparing two samples, rows and columns are sorted by single linkage Euclidean clustering. **E**, Transcriptomics sample-sample Pearson correlation coefficient heatmap.

**Method section:**

**Multi-Omics Data Analysis:**

In total, there were 18 samples consisting of ‘control’ or ‘IFNγ’ treated Alpha, Beta or Delta cells, in biological triplicates. Proteomics and transcriptomics datasets were aggregated using custom Python code. Proteomics, transcriptomics and phosphoproteomics data were were obtained as raw reads and log2 scaled prior to all other calculations. Numerical operations were performed using the Numpy package (version 1.26.2, [4]). ENSEMBL-IDs from transcriptomics data were mapped to canonical UniProt IDs (biomart 0.9.2), which were then matched to UniProt IDs from transcriptomics data and the intersection kept for further analysis. Correlation between proteomics and transcriptomics data was assessed for each possible replicate-combination between the two analysis methods. To assess inter-replicate correlation, simple linear regression (sklearn, version 1.3.0 [5]), Pearson correlation (R) and coefficient of determination (R-squared) values were calculated for each comparison and the results compared between cell types using two-sided two-sample t-tests (scipy, version 1.11.4 [6]), (figure 20250423\_figure\_panels\_figure\_5.ai and supplementary 20250423\_figure\_panels\_Fig\_5\_1\_replicate\_scatter.ai). To assess sample-sample correlation via clustering & heatmaps proteomics and transcriptomics datasets were processed as follows: Initial Gaussian imputation to eliminate missing values: missing values for each feature are imputed by randomly drawing from a normal distribution N(μ, σ), where σ is the respective feature’s standard deviation **s** \* 0.3, and μis the feature’s mean **x̄** - 3\***s**. Then, features were scaled and centered by subtracting **x̄** from each value and dividing by **s** (z-scoring). Lastly, the Pearson correlation between each sample was computed using Pandas (version 2.1.4, [7]). Clustering and dendrograms were computed via scipy’s ‘linkage’ function, with the standard settings (linkage = ‘single’, metric = ‘euclidean’). Seaborn (version 0.14.0.dev0, [8]) was used to visualize correlation heatmaps and annotate samples (figure 20250423\_figure\_panels\_figure\_5.ai). Datasets and code for this analysis are shared on GitHub: [https://github.com/MannLabs/IsletsOmics.git].

[4]: Harris, Charles R., et al. "Array programming with NumPy." Nature 585.7825 (2020): 357-362

[5]: Pedregosa, Fabian. "Scikit‐learn: Machine learning in python Fabian." Journal of machine learning research 12 (2011): 2825.

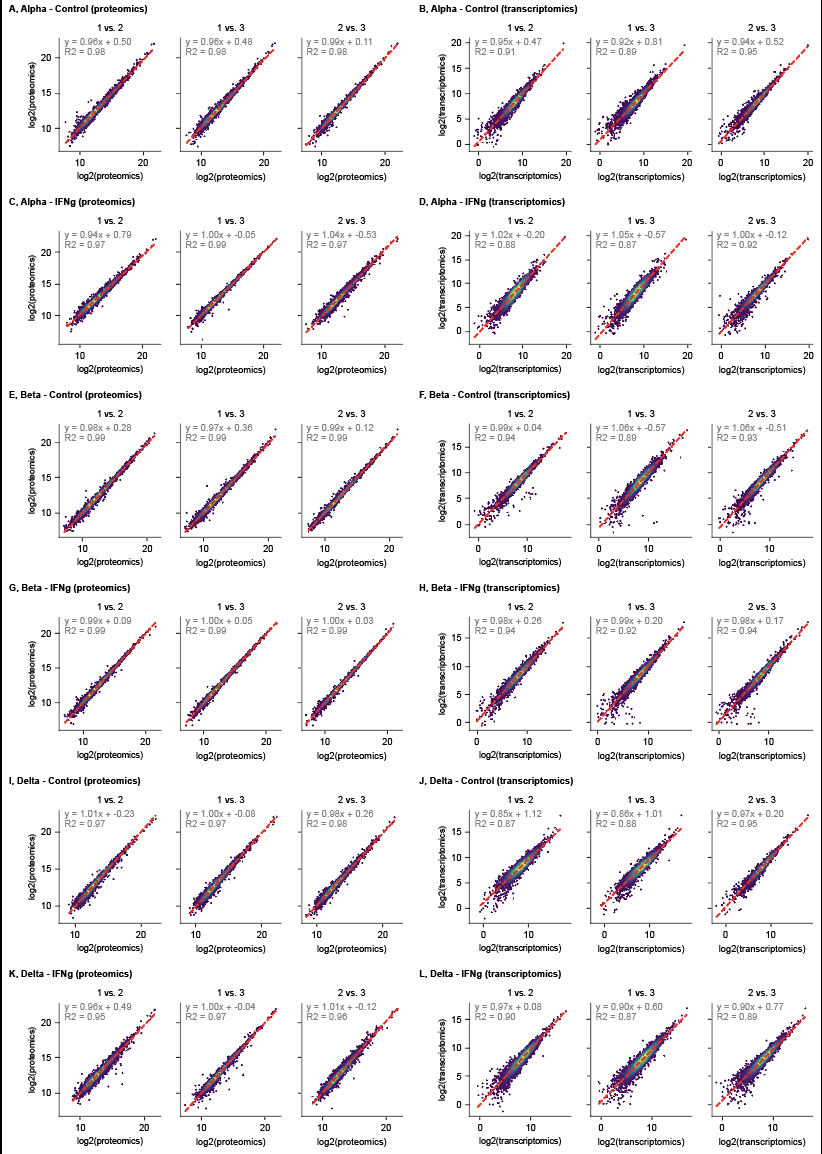
[6]: Virtanen, Pauli, et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." Nature methods 17.3 (2020): 261-272.

[7]: McKinney, Wes. "Data structures for statistical computing in Python." SciPy. Vol. 445. No. 1. 2010

[8]: Waskom, Michael L. "Seaborn: statistical data visualization." Journal of Open Source Software 6.60 (2021): 3021.

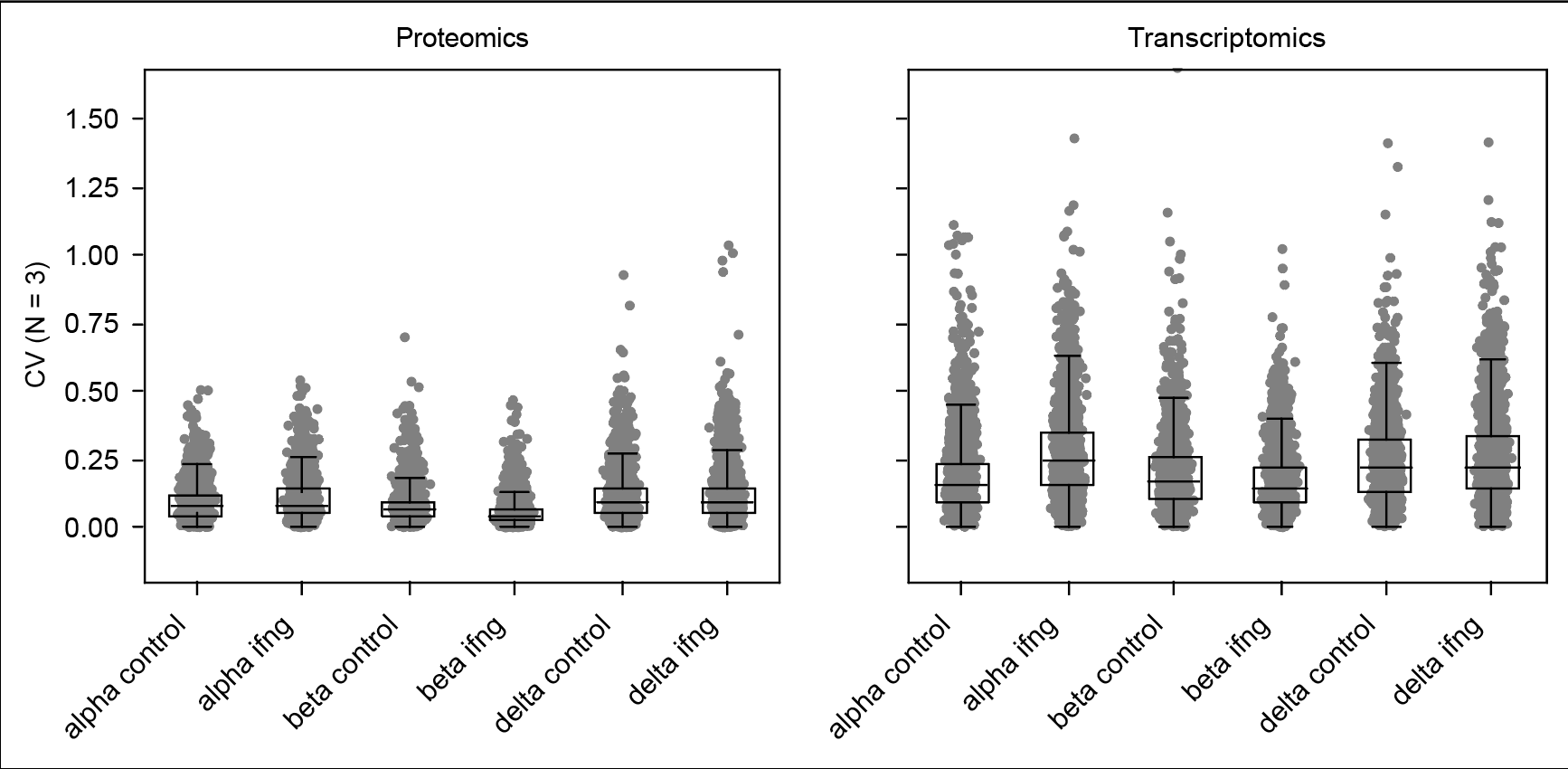
20250423\_figure\_panels\_replicate\_precision\_P\_T.ai (screenshot as placeholder)

*Caption:* Supplementary Fig. [], [] Inter-replicate comparison: Comparison of log2-transformed expression values of either proteomics (A, C, E, G, I, K) and transcriptomics (B, D, F, H, J, L) triplicates against each other. Facet titles “1 vs. 2”, “1 vs. 3”, “2 vs. 3” indicate the respective replicate comparison. Simple linear regression (red dashed lines) was fitted to each comparison, with the equation shown in the respective tile.

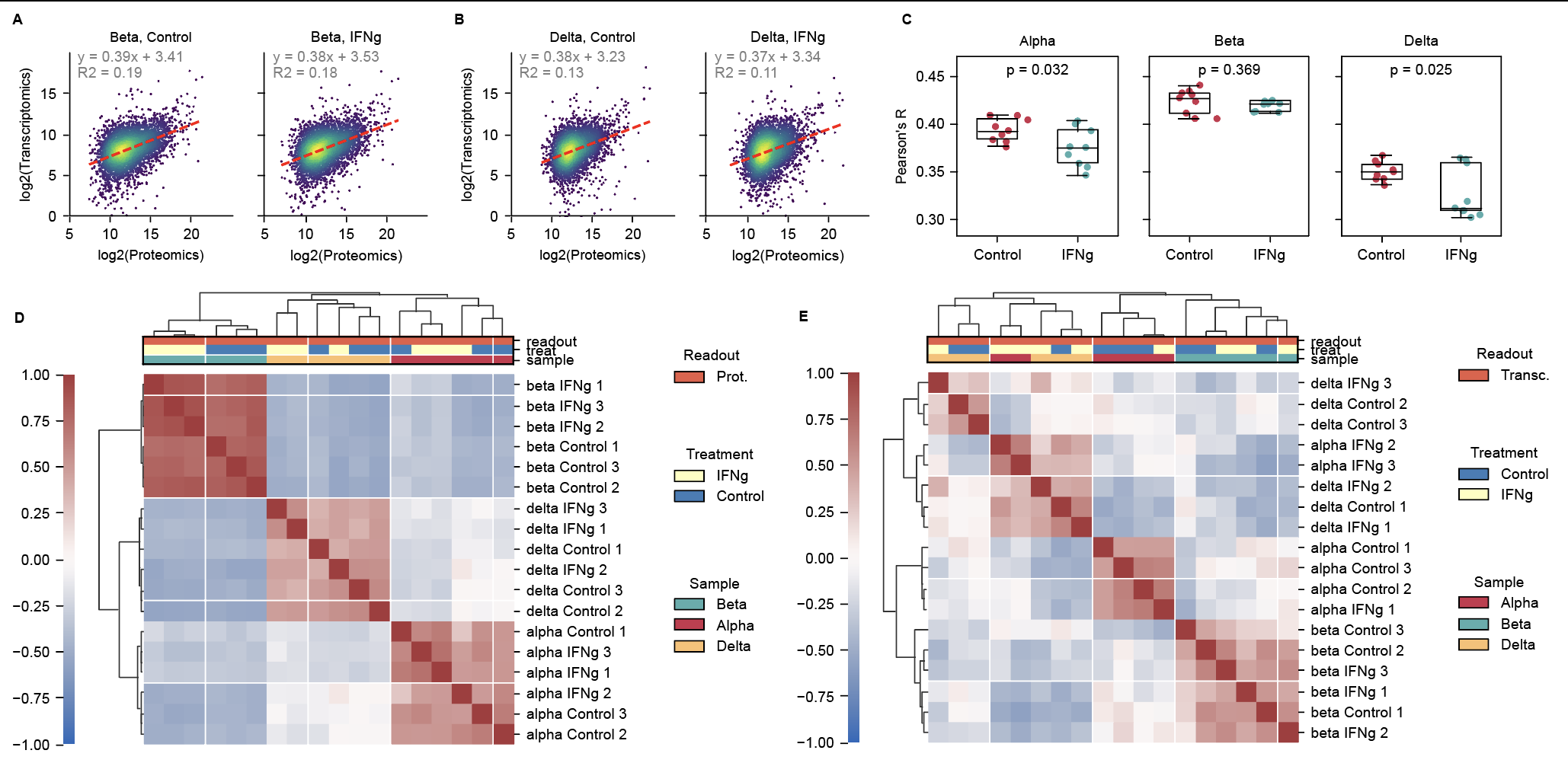


20250423\_figure\_panels\_CV\_across\_replicates\_P\_T.ai (screenshot as placeholder)

Caption: Supplementary Fig. [], [] Inter-replicate coefficients of variation for proteomics and transcriptomics data: For each respective cell type – treatment combination, each gene’s expression values across three available replicates were considered. To calculate CVs, the standard deviation of three replicates was divided by their mean, and the resulting fractions plotted. Mean CV across all proteomics groups (left panel): 0.0931 (9.31 %), mean CV across all transcriptomics groups (right panel): 0.2178 (21.78 %). The same 3273 genes with no missing values in any replicate were evaluated for both proteomics and transcriptomics.



20250423\_figure\_panels\_figure\_5.ai (screenshot as placeholder)



20250423\_figure\_panels\_Fig\_5\_1\_replicate\_scatter.ai (screenshot as placeholder)

Caption: Supplementary Fig. [], [] Proteomics versus Transcriptomics replicate comparison: Every replicate from the proteomics dataset was evaluated against every replicate from the transcriptomics dataset, yielding 9 unique comparisons (columns 1 – 1, 1 – 2, etc.) for each cell-type – treatment combination (rows A – F). Log2-transformed values are shown. Simple linear regression (red dashed lines) was fitted to each comparison, with the equation shown in the respective tile. Note that these R2 values correspond to the squared Pearson’s R values from Figure 5 C.

