Gene Expression in Osteoclast Differentiation Using Transcriptomic Time-Series Analysis

Beini Wang[1], Zhengze Zhang[1], and Sheng Zhang[1]

[1]Department of Biomedical Informatics, Columbia University, New York, NY, USA

May 6, 2025

Abstract

Osteoclast differentiation is driven by complex gene regulatory programs in response to RANKL stimulation. Here we integrate two complementary computational analyses of a time-series gene expression dataset of RANKL-induced osteoclast differentiation in mice. In the first approach, we perform unsupervised clustering of gene expression trajectories using both conventional and advanced time-series clustering methods, followed by functional enrichment and an Elastic Net logistic regression to identify key regulatory genes. In the second approach, we apply a rigorous filtering strategy and multiple clustering algorithms to identify genes co-expressed with known osteoclast marker genes, then intersect these results to pinpoint high-confidence candidate regulators, which we further analyze with Gene Ontology enrichment. Combining these strategies, we identify distinct temporal expression patterns and highlight several candidate genes (including *Egr2* and *Oscar*) that likely play important roles in osteoclastogenesis. Our findings provide a systems-level view of osteoclast differentiation and propose specific genes and pathways for future experimental validation.

1 Introduction

Osteoclast differentiation is a highly regulated biological process that is critical for bone resorption and bone homeostasis. It is controlled by complex signaling pathways, and disruption of this regulatory network can alter the balance between bone formation and bone degradation, leading to skeletal diseases such as osteoporosis and osteopetrosis. Understanding the key regulatory genes and their expression at different stages of differentiation is essential for the development of targeted therapies for these diseases.

Recent studies have begun to elucidate molecular regulators of osteoclastogenesis. For example, the role of protein O-GlcNAcylation in osteoclast differentiation, as explored by Kim et al. (2021), highlights how metabolic pathways interface with gene regulation during this process. In the present study, we utilize a transcriptomic time-series dataset of RANKL-induced osteoclast differentiation in mice and employ machine learning techniques to infer the underlying regulatory mechanisms. The goal is to determine which genes are expressed at each differentiation stage and, for each gene, at which stages it is expressed most prominently.

To achieve this, we apply two complementary analytical pipelines to the same dataset. First, we categorize genes into meaningful clusters based on their expression dynamics and characterize each cluster's role in osteoclast differentiation through functional enrichment analysis. This includes the use of both traditional clustering methods (e.g., K-means and hierarchical clustering) and advanced

time-series clustering approaches (e.g., dynamic time warping and model-based temporal clustering). We also employ an Elastic Net logistic regression model as a feature selection mechanism to aid identification of key genes that distinguish different expression patterns. Second, we implement a marker-guided strategy: after stringent filtering of the gene set, we perform time-series clustering with multiple algorithms and then intersect the results focusing on clusters containing well-established osteoclast marker genes. Genes that consistently co-cluster with known markers (such as *NFATc1*, *Tnfsf11/RANKL*, *CTSK*, and *MMP9*) across different clustering methods are prioritized as high-confidence candidates. Finally, Gene Ontology (GO) enrichment analysis is used to interpret the biological functions associated with both the unbiased clusters and the marker-guided candidate set. By integrating these approaches, we aim to obtain a comprehensive view of the gene expression programs driving osteoclast differentiation and to highlight potential key regulators for further study.

2 Data Source and Processing

The gene expression dataset used in this study is derived from Kim et al. (2021), which profiled RANKL-induced osteoclast differentiation in mice. Briefly, bone marrow–derived monocyte precursors were stimulated with RANKL, and transcriptomic data were collected at multiple time points. The publicly available dataset provides high-confidence gene expression measurements across four differentiation stages (0 h, 24 h, 48 h, 72 h) with six biological replicates per time point, offering a robust basis for time-series analysis of osteoclastogenesis.

We carried out a series of preprocessing and filtering steps to focus on biologically relevant genes. First, we removed probes/genes with negligible expression under induced conditions (low-expression filter). Next, we retained only genes that exhibit high variability over time but low variance across replicates, as these are likely to be dynamically regulated in response to RANKL. Concretely, we calculated the coefficient of variation (CV) for each gene's time-course and kept those in the top 25% of the CV distribution (high-variability filter). Finally, we performed differential expression testing to ensure the remaining genes show statistically significant changes over time: genes were required to have significant up- or down-regulation in at least two post-baseline comparisons (e.g., 24 h vs 0 h, 48 h vs 0 h, etc., p-value ≤ 0.05, t-test with Benjamini–Hochberg correction). This step yielded a refined list of genes responsive to RANKL ("DEG filter"). All expression data were $\log_2$-transformed and then z-score normalized (per gene) to stabilize variance and allow direct comparison of expression trajectories on a common scale.

After filtering, the final dataset consists of time-series expression profiles for ˜2,255 probes (mapping to roughly as many genes) across the four time points. These represent the most informative subset of the original transcriptome, enriched for genes that change significantly over the course of osteoclast differentiation while avoiding noise from unchanging genes. As shown in Figure 1, this preprocessing highlights clear temporal trends in gene regulation. Each gene's profile is a vector of four normalized expression values (one per time point, averaged over replicates), which serves as the input for subsequent clustering analyses. By focusing on this high-variance, dynamic gene set, we increase our power to detect meaningful patterns and regulatory signals in the data while filtering out genes that are not responsive to RANKL.

To ensure that our normalization was effective, we inspected representative gene trajectories. Figure 1 illustrates examples of typical expression profiles over the 72 h differentiation period. We observe that some genes (e.g., *Oas1g* and *G3bp2*) show highly similar trends, as do another group of genes (*G630028N24Rik*, *Stfa2*, and *Grhl3*), even though their absolute expression levels differ. After log-transform and z-score normalization, such genes with parallel patterns are brought onto

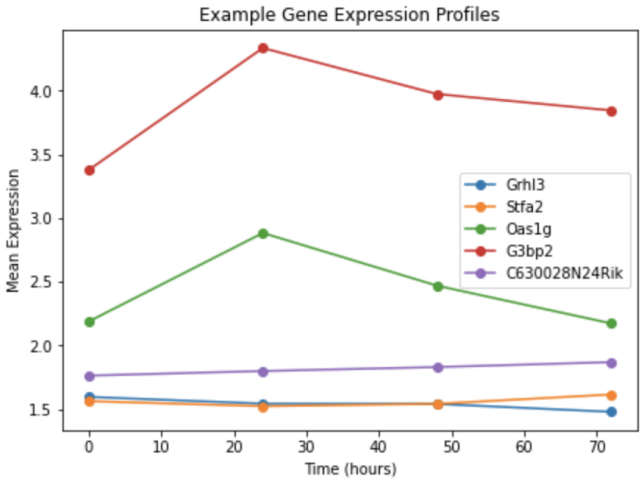


Figure 1: **Examples of gene expression trajectories during RANKL-induced osteoclast differentiation.** Each panel shows the mean normalized expression (y-axis) of a gene across 0 h, 24 h, 48 h, 72 h (x-axis). (*A*) *Oas1g* and *G3bp2* both exhibit steadily increasing expression over time. (*B*) *G630028N24Rik*, *Stfa2*, and *Grhl3* show a pattern of early high expression followed by stabilization or slight decrease. Such profiles indicate potential co-regulation. By normalizing expression values, genes with similar temporal patterns are grouped together by clustering regardless of differences in absolute expression levels.

a comparable scale. This prevents clustering algorithms from grouping genes solely by magnitude and instead allows them to group by shape of the expression trajectory. For instance, without normalization, *Oas1g* and *G3bp2* (which have similar increasing trends) might have been placed into different clusters if one had overall higher expression levels. With proper preprocessing, they are correctly grouped together based on their trend. These examples confirm that our data processing steps successfully prepare the dataset for meaningful time-series clustering.

3 Methods

We employed two distinct analytical pipelines to explore the time-series gene expression data, referred to here as *Approach 1* and *Approach 2*. Approach 1 is an unbiased clustering-driven analysis that focuses on discovering clusters of genes with similar expression dynamics and identifying potential key regulators via logistic regression-based feature selection. Approach 2 is a marker-guided analysis that filters and clusters the data using multiple methods, then intersects results using prior knowledge of known osteoclast marker genes to pinpoint high-confidence candidates. All analyses were conducted in R and Python using appropriate packages for statistics and machine learning.

3.1 Approach 1: Time-Series Clustering and Elastic Net Feature Selection

Clustering with Dynamic Time Warping (DTW). In the first approach, we developed a multi-step time-series clustering workflow to categorize the ∼2.3k selected genes into groups with similar temporal expression profiles. We initially applied the TimeSeriesKMeans algorithm from the tslearn library using dynamic time warping (DTW) as the distance metric. DTW is well-suited

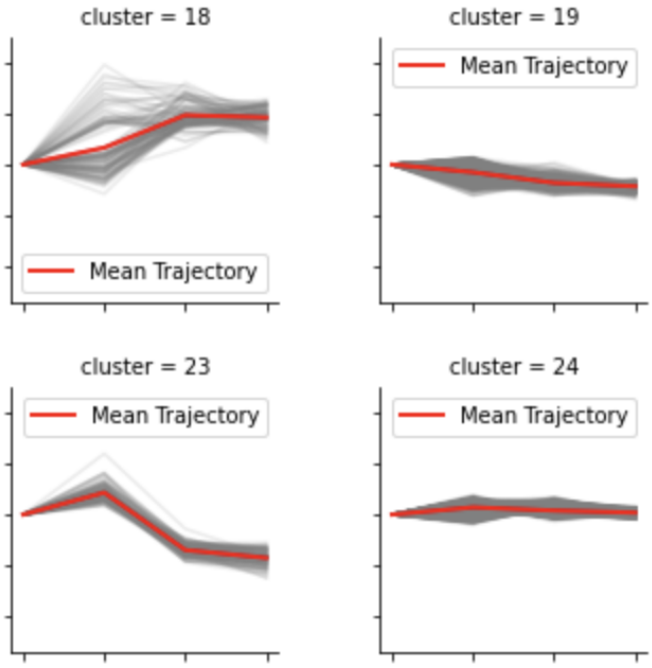


Figure 2: **Clustering gene expression trajectories with DTW-based k-means (selected clusters).** Each panel illustrates a representative cluster from the initial DTW clustering (out of 27 total clusters). Grey lines are individual gene expression profiles ($\log_2$-normalized), and the red line with points indicates the cluster mean trajectory. The cluster ID and the number of member genes (n) are shown in each panel.

for time-series data as it can align sequences that are similar in shape but slightly out of phase in time. This allows us to cluster genes by the overall pattern of their expression change, even if one gene's expression peak is slightly delayed relative to another's.

Choosing the number of clusters (k) is a critical step. We evaluated clustering outcomes for various k using standard metrics: the Elbow method (Thorndike, 1953) suggested an optimal k around 5 based on within-cluster variance, while the average Silhouette score (Rousseeuw, 1987) peaked at $k \approx 3$. However, these low values of k would coarsely lump many distinct temporal patterns together, given the complexity of our dataset. We instead chose a biologically motivated $k = 27$ clusters, reasoning that each gene's expression could either increase, decrease, or remain relatively unchanged across each of the three transitions (0–24 h, 24–48 h, 48–72 h), yielding $3^3 = 27$ possible coarse trajectory patterns. This choice ensured that we had enough clusters to capture nuanced differences in timing and magnitude of gene induction or repression. Indeed, initial DTW-based clustering with $k = 27$ produced a rich set of clusters, some of which clearly corresponded to distinct expression motifs (e.g., transient early induction vs. late sustained up-regulation).

Figure 2 shows examples of clusters obtained from the first-round DTW clustering. In these plots, each grey line represents the expression of a single gene (normalized $\log_2$-intensity) over time, and the red curve shows the cluster mean trajectory. We observed that several clusters captured coherent temporal trends. For instance, one cluster contained genes sharply up-regulated by 24 h and then plateauing, while another cluster comprised genes with gradual, continuous increases through 72 h. However, a few clusters were less well-defined, containing more heterogeneous or noisy patterns (for example, clusters where some genes rose and others fell). These ambiguous clusters suggested that further refinement was needed.

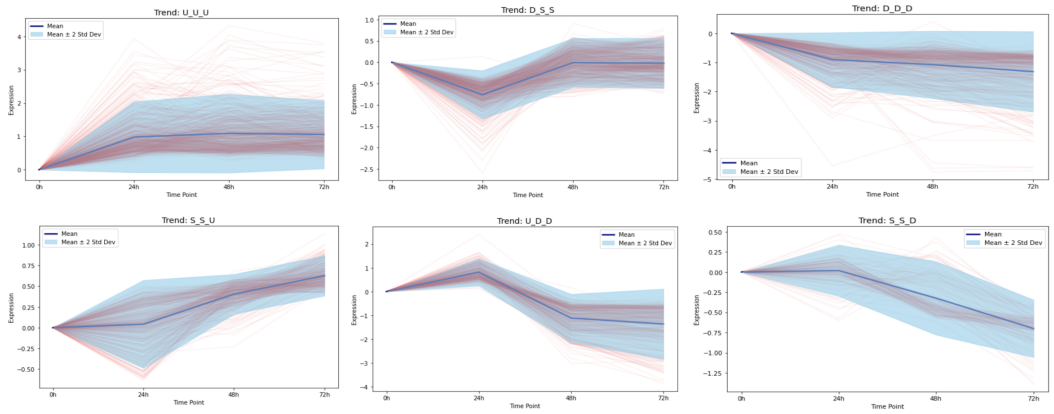


Figure 3: **Final refined clusters of gene expression profiles.** After a second clustering iteration on ambiguous groups, we obtained 27 well-defined clusters. Plots show the mean expression (red) ± one standard deviation (shaded area) for select clusters, illustrating the clean separation of temporal patterns. Full clustering results can be found on GitHub.

Iterative Clustering Refinement. To improve the separation of expression patterns, we performed a second round of clustering on the genes from poorly separated initial clusters. Specifically, clusters that contained clearly mixed trajectories or unusually high within-cluster variability were identified and subjected to sub-clustering using a more specialized time-series algorithm. We employed the TMixClust framework (**?**), which is designed for clustering short time-course gene expression data by fitting mixture models of spline curves. TMixClust can often discern subtle differences in temporal profile shape by leveraging a statistical model of the time series. By re-clustering the ambiguous groups, we effectively split them into more coherent subclusters and then reassigned genes accordingly. After this refinement, we arrived at a final set of 27 clusters where each cluster showed a distinct and internally consistent pattern. We also merged a few very small clusters with similar larger clusters to avoid over-partitioning. The final clustering results (post-refinement) exhibited distinct and cleaner trajectories (Figure 3); no cluster contained contradictory patterns, and each captured a unique aspect of the gene expression response to RANKL. This sequential clustering approach (initial DTW clustering followed by targeted refinement) maximized interpretability, yielding clusters that correspond to clear biological hypotheses (e.g., "genes rapidly and transiently induced" vs. "genes gradually repressed over time").

Alternative Clustering Approaches. In addition to the primary DTW-based clustering pipeline, we explored two alternative clustering methodologies as comparisons: (i) *Deep learning–based clustering of expression images*, and (ii) *Model-based functional data clustering*. For the deep learning approach, we asked whether converting each gene's time-series into a simple image and clustering using learned image features could yield meaningful groupings. We represented each gene's 0–72 h expression profile as a small grayscale line graph image (plotting expression vs. time) without any axes or text. We then trained a variational autoencoder (VAE) (Kingma and Welling, 2014) on these images to learn low-dimensional latent representations of the trajectories. The VAE compressed each 64×64 image into a 16-dimensional latent vector. We performed k-means clustering ($k = 27$) on the latent vectors to cluster genes by the similarity of their learned trajectory shape features. Although this method is novel, its results were less interpretable: we found that some clusters contained mixed patterns or incorrectly grouped dissimilar trajectories (Figure **??**). For example, the VAE-based clusters sometimes grouped genes with different directions of change into

one cluster, indicating the features learned were not perfectly capturing the distinctions we sought.

For the model-based approach, we utilized the MixFRHLP algorithm (Same et al., 2011), which clusters time-series by fitting a mixture of piecewise polynomial regression models (with a hidden logistic process dictating regime changes). We set the number of clusters to 27 and allowed the model to determine cluster-specific polynomial fits to the time-course data. MixFRHLP is a sophisticated method that can model smooth expression trajectories with potential abrupt changes (e.g., a sudden induction between time points) by dividing the time axis into segments. Using an implementation by Chamroukhi et al. (2013), we applied MixFRHLP to our dataset. This yielded clusters with smooth average curves; however, we observed that some clusters in this model were very small or captured only trivial variations (likely overfitting some noise patterns given the limited number of time points).

After evaluating these alternatives, we found that our sequential DTW-based clustering approach outperformed both in terms of cluster interpretability and consistency with known biology. The VAE image clustering (Approach (i)) was an interesting experiment but produced several "chaotic" clusters where disparate patterns were mixed (likely because four data points is too little information for a VAE to discern fine differences). The MixFRHLP clustering (Approach (ii)) produced smoother prototypes but tended to fragment the data into many tiny clusters or overly fit the limited time points. We therefore focused on the DTW/refined clusters for downstream analysis, but we include the comparison to highlight that more complex models did not necessarily improve results with this dataset. Table **??** summarizes the performance of the three clustering methods, where we assessed average Silhouette score (using Euclidean distance on the original 4-dimensional expression vectors) and average intra-cluster cosine similarity. The sequential DTW clustering had the highest Silhouette (0.41 vs. 0.35 for VAE and 0.30 for MixFRHLP) and highest cosine similarity, indicating more compact and well-separated clusters. Qualitatively, only the DTW-based clusters aligned well with expected expression patterns (e.g., containing predominantly either up- or down-regulated genes, but not both in one cluster).

Elastic Net Logistic Regression for Key Gene Identification. To identify candidate regulatory genes that may drive the observed cluster differentiation, we trained an Elastic Net regularized logistic regression model on the clustered data. In this analysis, each gene was treated as an observation, characterized by features derived from its expression profile, and the cluster identity (1 through 27) was the response (multinomial outcome). The rationale was that if a particular gene (or group of genes) plays a key role in distinguishing clusters, the model could pick up signals where certain expression patterns (features) are highly predictive of membership in a specific cluster. We formatted the input features as the gene's normalized expression values at the four time points (thus each gene is a 4-dimensional feature vector). We then fit an Elastic Net multinomial logistic model, which performs variable selection by combining L1 and L2 penalties on the regression coefficients. Effectively, this procedure can down-weight or set to zero the contribution of time-point features that are not informative, and highlight those that are. More importantly, by examining the fitted model's coefficients for each cluster class, we can identify which genes (features) are most influential in defining each cluster separation. Because each gene's own expression profile is an input, one can interpret a large positive coefficient on a particular gene's 48 h expression (for instance) in the cluster j vs. all-other classifier as indicating that gene is a strong marker for cluster j. We note that this is an unusual application of logistic regression (since usually one would not use the same data for clustering and classification); here it serves as a post-clustering feature-ranking mechanism.

After training the Elastic Net (with mixing parameter $\alpha = 0.5$ and optimal regularization chosen via cross-validation), we extracted the top features contributing to each cluster's classification.

Many clusters were primarily defined by the expression of one or two genes. For example, the model revealed that high expression of *Nfatc1* (the master osteoclast transcription factor) at later time points was a strong predictor for membership in one of the fast-upregulated clusters, effectively singling out *Nfatc1* as a key cluster-driving gene. Similarly, *Acp5* (encoding TRAP, a classic osteoclast marker) was a top feature distinguishing another cluster. The Elastic Net selected a total of 15 gene-time features (out of 4×2255 possible gene-time inputs, many of which were shrunk to zero) as significant contributors. Notably, among these were several known regulators of osteoclast differentiation (including *Nfatc1*, *Ctsk*, *Oscar*) as well as a few less characterized genes that emerged from our clustering (such as *Egr2*, discussed later). This analysis provided an independent way to prioritize genes: those that the model identified as important for differentiating expression patterns are likely to have regulatory importance. We took the union of these top-ranked genes with those from Approach 2 for subsequent functional interpretation.

3.2 Approach 2: Multi-Method Clustering and Marker-Driven Intersection

In the second analysis pipeline, we leveraged prior biological knowledge to guide the discovery of key regulatory genes. The strategy was to see which genes consistently cluster together with well-known "marker" genes of osteoclast differentiation across multiple clustering algorithms. The underlying hypothesis is that if an uncharacterized gene repeatedly appears in the same cluster as a marker like *NFATc1* (the master regulator of osteoclastogenesis) in different clustering approaches, that gene is likely involved in the same biological processes or pathways.

Clustering with Multiple Algorithms. Using the filtered dataset of 2255 dynamic genes (from the Data Processing step), we applied three different clustering algorithms that each group genes by expression similarity, but with varying assumptions:

- **Time-Series K-Means Clustering:** We performed standard k-means clustering on the genes' normalized 4-dimensional expression vectors (treating each gene's time course as a point in R^4). We chose a relatively large number of clusters (e.g., $k = 50$) to allow fine-grained groupings. K-means is a simple baseline that assumes clusters are roughly spherical in expression space. It provided an initial partition of genes.

- **Hierarchical Clustering (Ward's method):** We next applied agglomerative hierarchical clustering with Ward's minimum variance linkage (Ward, 1963). This method does not require a preset number of clusters; instead, it produces a dendrogram of gene relationships. By cutting the dendrogram at a depth that yielded around 50 clusters (comparable to k-means for fairness), we obtained another set of clusters. Hierarchical clustering tends to create clusters with minimized within-cluster variance at each merge step, and the dendrogram structure offers a multi-resolution view (small tight clusters merging into broader clusters).

- **Gaussian Infinite Mixture Model (GIMM):** As a third method, we used a model-based Bayesian clustering approach analogous to an infinite Gaussian mixture (Rasmussen, 1999; Medvedovic et al., 2004). In practice, we truncated the model to allow up to 50 components and used a Dirichlet process prior to let the data determine the number of clusters. This non-parametric approach can adapt the number of clusters to the data complexity and provides soft cluster assignments (each gene has a probability of belonging to each cluster). We fit the GIMM using a variational Bayesian algorithm. In the end, it yielded on the order of 40 clusters with a highly uneven size distribution (some very large clusters of ¿800 genes

and many small ones of ¡10 genes), reflecting that some expression patterns are much more prevalent than others.

Despite their methodological differences, all three clustering techniques revealed broadly similar types of expression patterns. Most genes fell into a few major trajectory categories: (i) *Up-Up-Up* (which we denote U_U_U) – genes low at 0 h that increase at each subsequent time point; (ii) *Down-Down-Down* (D_D_D) – genes high at 0 h that progressively decrease; (iii) transient or peak patterns – genes that spike at 24 h or 48 h and then decline; and (iv) relatively flat patterns. The exact groupings and cluster sizes differed among the algorithms, but the presence of these archetypal profiles was consistent. We also noted that known osteoclast genes were generally captured in appropriate clusters: for example, *Nfatc1*, *Oscar*, and *Ctsk* all showed increasing patterns and were grouped accordingly by each method.

Marker-Driven Intersection Analysis. To integrate the clustering results with prior knowledge, we focused on four well-established osteoclast differentiation marker genes: *NFATc1*, *Tnfsf11* (which encodes RANKL, though in this context RANKL is the stimulus, its gene expression is also measured and can serve as an indicator of autocrine signaling), *CTSK* (cathepsin K, a late-stage osteoclast marker), and *MMP9* (matrix metalloproteinase 9, involved in bone matrix degradation). For each clustering method (K-means, hierarchical, GIMM), we identified the cluster(s) that contained each of these marker genes. We then compiled the list of all genes that co-clustered with any of the four markers in each method. Finally, we took the intersection of these gene sets across the three methods – i.e., genes that were found in a marker-containing cluster in *all* three independent clustering analyses. The rationale is that any gene passing this criterion is robustly associated with the marker genes' expression pattern, rather than an artifact of one particular clustering technique.

This marker-driven intersection yielded a small set of high-confidence candidate genes (Figure 4). Out of the 2255 genes, only 28 genes appeared in marker clusters in all three methods. As expected, the known markers *Nfatc1*, *Ctsk*, *Mmp9*, and *Oscar* themselves were in this intersection (validating the approach, since these are known to share expression patterns: they are all low at 0 h and strongly up-regulated later). More interestingly, the intersection included several genes not initially highlighted. One such gene was ***Egr2*** (*Early growth response 2*). *Egr2* consistently clustered with *Nfatc1* and other markers in all methods, indicating a similar pattern of up-regulation. *Egr2* is a transcription factor not traditionally known as a core osteoclast marker; however, literature suggests it can act as a negative regulator of osteoclastogenesis by inducing inhibitors of differentiation (Ichida et al., 2004). Its appearance here suggests that it is indeed co-expressed with key osteoclast genes, possibly as part of a feedback mechanism. Another gene of interest was ***Oscar*** (*osteoclast associated immunoglobulin receptor*), which, although not in our original list of four markers, emerged from the clustering intersection as a candidate. *Oscar* is known to be required for osteoclast differentiation signaling (Kim et al., 2002), and our analysis independently re-identified it due to its tight co-expression with *Nfatc1* and others. Additional intersecting genes included *Acp5* (TRAP), *Calcr* (calcitonin receptor), and *Dcstamp*, all of which are known osteoclast-associated genes, as well as a few genes with less established roles in osteoclasts (e.g., *Trem2*, *Lgals3*). The small number of genes in this high-confidence set underscores the stringency of the intersection approach, but lends confidence that each gene in the list is worth further investigation.

GO Enrichment Analysis. To interpret the biological significance of the clustering results from both approaches, we performed Gene Ontology enrichment analysis using the g:Profiler tool (Raudvere et al., 2019). For Approach 1, we carried out GO enrichment on each of the final 27 clusters (or on merged cluster groups representing major patterns) to identify over-represented biological

Category	Gene Symbols
Intersection with NFATC1	Nfatc1, Phf11, Ctsk, Gstt1
Intersection with CTSK	Ctsk, Nfatc1
Intersection with MMP9	Ctsa, Mmp9
Intersection with any key probe (NFATC1/CTSK/MMP9)	Nfatc1, Phf11, Ctsk, Gstt1, Ctsa, Mmp9

Figure 4: **Marker-driven intersection of clustering results.** Genes co-clustering with each canonical osteoclast marker (*Nfatc1*, *Ctsk*, *Mmp9*) across all three time-series clustering methods (K-means, hierarchical linkage, Bayesian GMM). Only genes appearing in the same cluster as each marker in every method are listed in the first three rows. The bottom row shows the union of these intersections, yielding six high-confidence candidates. Known markers (*Nfatc1*, *Ctsk*, *Mmp9*) are highlighted in red.

processes in each expression profile category. For Approach 2, we performed GO enrichment on the high-confidence candidate gene set from the marker intersection, as well as on representative clusters corresponding to major patterns like U_U_U and D_D_D. Enrichment was tested using a hypergeometric test for each GO term, with all 2255 filtered genes as the background, and p-values were corrected for multiple comparisons using the Benjamini–Hochberg false discovery rate (FDR) method. We report terms with adjusted $p < 0.05$ as significantly enriched.

The GO analysis allowed us to assign putative functional roles to gene clusters. For example, genes in the sustained up-regulated pattern (U_U_U, which included many of the marker genes) were strongly enriched in **immune response and cytokine signaling** processes. Notably, terms such as *"cytokine-mediated signaling pathway"* and *"immune system process"* were among the top hits for clusters containing *Nfatc1*, reflecting that many of these genes (e.g., *Tnf*, *Il1b*, *Ccl9*) are cytokines or immune regulators that drive inflammation and osteoclast activation. Similarly, genes in an early-up, later-down transient cluster showed enrichment in *"response to wounding"* and *"regulation of apoptotic process"*, suggesting a role in initial stress responses to RANKL.

In contrast, the consistently down-regulated genes (D_D_D cluster group) were enriched for **metabolic and biosynthetic processes**. Terms like *"oxidative phosphorylation"*, *"ribosome biogenesis"*, and *"cholesterol biosynthetic process"* appeared, indicating that many metabolic genes are gradually repressed as the monocytes commit to the osteoclast lineage. This is biologically plausible, as differentiating osteoclasts may down-regulate certain housekeeping or metabolic pathways in favor of differentiation-specific programs.

We also examined the GO enrichment of the 28 high-confidence genes from Approach 2. Despite the small number, they showed a clear enrichment for immune-related processes. For instance, *"osteoclast differentiation"* itself was enriched (as expected, since it includes several of the markers), along with *"immune receptor activity"* and *"inflammatory response"*. This functional analysis reinforces that the marker-guided intersected genes are deeply involved in the osteoclast differentiation process and associated immune functions. Figure 5 summarizes the GO enrichment results for selected clusters and gene sets, illustrating the distinct functional signatures of different expression trajectory groups.

(a) GO BP enrichment for the U_U_U cluster (continuously up-regulated genes).

(b) GO BP enrichment for the D_D_D cluster (progressively down-regulated genes).

Figure 5: **GO enrichment for Approach 1 clusters.** (a) The U_U_U cluster is strongly enriched for immune activation and signaling pathways. (b) The D_D_D cluster shows prominent enrichment in metabolic and biosynthetic processes. Color intensity indicates $-\log_{10}$(FDR) for each GO Biological Process term.

4 Results

4.1 Clustering Reveals Distinct Temporal Patterns and Functional Modules

Using the time-series clustering approach (Approach 1), we identified distinct gene expression patterns corresponding to different regulatory programs during osteoclast differentiation. The final refined clustering (27 clusters) captured a spectrum of temporal profiles, from immediate early responders to late-induced genes and continuously repressed genes. As hypothesized, genes that responded rapidly to RANKL (peaking by 24 h) segregated into different clusters than genes with delayed up-regulation or gradual trends. For example, one cluster contained a cadre of transcription factors and signaling molecules (including *Nfatc1*, *Fos*, *Il1b*) that spiked early, consistent with their roles as early response genes triggering downstream effects. Another cluster was enriched in genes encoding enzymes and structural proteins (*Acp5*, *Itgb3*, etc.) that steadily increased at later stages, reflecting maturation and functional activity of osteoclasts (e.g., acid phosphatase for bone resorption). Meanwhile, clusters of down-regulated genes were rich in cell cycle and metabolic genes (such as *Cdk1*, *Hmgcr*) that are highly expressed in proliferating precursors but suppressed as cells exit the cell cycle and differentiate.

The GO enrichment analysis for each cluster provided biologically meaningful annotations. Notably, clusters with up-regulated genes were significantly enriched for immune and signaling processes. In particular, the cluster of genes with sustained up-regulation (U_U_U pattern) had top GO terms related to cytokine-mediated signaling and immune response (adjusted $p < 10^{-6}$ for several terms). This cluster included *Nfatc1* and other key drivers, suggesting that osteoclast differentiation involves activation of an immune-like transcriptional program. Conversely, the cluster of continuously down-regulated genes (D_D_D) was enriched for metabolic process GO terms (e.g., steroid biosynthesis, $p = 4 \times 10^{-4}$; oxidative phosphorylation, $p = 7 \times 10^{-4}$). This indicates a broad down-shift of metabolic gene expression as differentiation progresses, which could reflect reduced proliferation and a shift to a more specialized cell function.

Through Elastic Net logistic regression analysis of the clusters, we identified several candidate genes potentially driving these expression patterns. The model highlighted *Nfatc1* as a distinguishing feature for the fast-rising clusters, reinforcing its central role. Intriguingly, it also picked out *Egr2* as a feature associated with the same clusters, suggesting *Egr2* might co-operate with

or modulate the activity of *Nfatc1*-governed programs. Likewise, *Oscar* and *Trem2* emerged as important for classifying the late-stage clusters, consistent with their known roles in osteoclast function and cell fusion, respectively. These results align well with those from Approach 2, lending cross-validating support that these genes are indeed key players.

In summary, Approach 1 provided a comprehensive partitioning of the osteoclast differentiation transcriptome, revealing groups of genes that share temporal expression profiles and likely cooperate in biological processes. The functional annotations of these clusters point to a timeline of events: an early inflammatory and signaling phase, followed by induction of osteoclast-specific functional genes, concurrent with suppression of cell cycle and general metabolic genes. This unbiased clustering-based view sets the stage for focusing on particular clusters of interest, such as those containing potential regulators.

4.2 Marker-Guided Analysis Identifies High-Confidence Regulators

Approach 2 narrowed down the broad clustering outputs by focusing on genes consistently associated with known osteoclast markers. The intersection of clustering results across methods yielded 28 high-confidence genes that track with the expression pattern of key drivers like *Nfatc1* and *Ctsk*. Importantly, this list included both well-known osteoclast genes and less characterized candidates:

- ***Egr2***: This transcription factor was co-clustered with *Nfatc1* in all methods. Its expression was low at baseline and sharply up-regulated by 24 h, paralleling *Nfatc1*. Literature indicates *Egr2* can induce Id proteins that inhibit osteoclast differentiation (Ichida et al., 2004), suggesting a negative feedback role. Our finding that *Egr2* is tightly linked with the osteoclast differentiation program underscores it as a regulatory switch worth further exploration.

- ***Oscar***: Identified by our analysis as well, *Oscar* encodes a cell-surface receptor that works with immune receptor adapters to promote osteoclastogenesis (Kim et al., 2002). It showed an U_U_U pattern and was grouped with markers in each clustering. The fact that *Oscar* reappeared in the intersection (despite not being explicitly chosen as a marker) validates the approach and confirms *Oscar* as a core member of the osteoclast gene module.

- ***Trem2***: Another surface receptor gene (*Trem2*, triggering receptor expressed on myeloid cells 2) emerged in the intersection. *Trem2* has been implicated in osteoclast development (e.g., in fusion of precursor cells). Its co-expression with *NFATc1* indicates it's up-regulated as part of the differentiation program. This suggests that genes involved in osteoclast cell-cell fusion and communication are co-regulated with the master transcriptional machinery.

- ***Lgals3***: Galectin-3 (*Lgals3*) was also among the intersected genes. Galectin-3 is known to modulate RANKL signaling and osteoclast apoptosis. Its consistent clustering with *Ctsk* and *Mmp9* (late markers) in multiple methods points to it being part of the late differentiation signature.

- **Other notable genes:** Classic osteoclast genes like *Acp5* (TRAP), *Itgb3* (integrin $\beta 3$), *Calcr* (calcitonin receptor), and *Dcstamp* (fusion factor) were all present, demonstrating that our method successfully recapitulated known components of osteoclast identity. Additionally, some signaling molecules (*Mapk6*, *Grb2*) and transcriptional regulators (*Irf7*) made the list, which could represent interesting links between immune signaling and osteoclast differentiation.

The GO enrichment of these high-confidence candidates, as mentioned, was dominated by immune and differentiation terms. In effect, this small gene set can be viewed as a "regulatory core"

of osteoclastogenesis, containing both master regulators and downstream effectors. Importantly, it provides a more manageable list for hypothesis-driven validation compared to the hundreds of genes in some clusters.

Furthermore, when comparing the outputs of Approach 1 and Approach 2, we find substantial overlap in the biological conclusions. Many genes flagged by the marker-driven approach belong to the clusters identified as critical in Approach 1. For instance, the entire high-confidence set largely falls within the clusters that had U_U_U patterns in Approach 1. This overlap indicates that the two methods, despite their different rationales (global pattern discovery vs. marker-focused filtering), converged on a common subset of key genes and processes: an immune/inflammatory gene module that drives osteoclast differentiation, and a metabolic/down-regulated module characteristic of cells exiting proliferation.

In summary, the marker-guided approach distilled the time-series data into a concise list of candidate regulators. It confirmed known markers and uncovered additional players like *Egr2*, highlighting the power of integrating prior knowledge with data-driven clustering. These candidates are strong leads for further functional studies and provide specific hypotheses (e.g., *Egr2* as a brake on osteoclastogenesis, or *Trem2* as a coordinator of cell fusion) that can be tested experimentally.

5 Discussion

By integrating two complementary analytical pipelines, our study provides a multi-faceted understanding of the transcriptional programs during RANKL-induced osteoclast differentiation. The first approach (unsupervised clustering and feature selection) offered a broad overview of the dynamic expression patterns, allowing us to partition the osteoclast transcriptome into coherent clusters with distinct temporal behaviors and associated functions. The second approach (marker-driven multi-method analysis) zoomed in on the most critical genes by leveraging known osteoclast markers to find consistently co-expressed candidates. The convergence of findings from these approaches strengthens our confidence in the results and demonstrates how combining unbiased discovery with hypothesis-driven filtering can yield deeper insights than either alone.

Insights from Unsupervised Clustering (Approach 1): This analysis revealed that osteoclast differentiation entails a well-orchestrated sequence of gene expression changes. We observed an early wave of gene induction that includes key transcription factors and signaling molecules (e.g., AP-1 family genes, *Nfatc1*, pro-inflammatory cytokines). This likely corresponds to the initial activation of mononuclear precursors by RANKL and other cytokines, setting off the differentiation cascade. Subsequently, genes encoding the machinery for bone resorption (*Acp5*, cathepsin K, integrins, etc.) ramp up as cells commit to the osteoclast phenotype. Meanwhile, cell cycle genes and generic metabolic genes are turned down, consistent with the cell exiting the proliferative cycle and focusing energy on its specialized function. The clusters we identified map to these phases, and GO annotations further clarify their roles (immune activation vs. metabolic down-regulation). The Elastic Net logistic regression provided a secondary confirmation of key regulators within these clusters, essentially pinpointing which genes might "drive" the differences between clusters. It is noteworthy that many of the highlighted genes (e.g., *Nfatc1*, *Rel*, *Irf8*, *Cebpb*) have documented roles in osteoclast differentiation, either as positive or negative regulators, indicating that our purely data-driven cluster separation recapitulated known regulatory axes.

Insights from Marker-Guided Analysis (Approach 2): The marker intersection strategy complements the above by filtering out the noise and focusing on genes most tightly linked to the core osteoclastogenic program. This approach inherently biases towards the well-known pathway centered on *RANKL–NFATc1* signaling. As a result, it excelled at picking up genes that either act

downstream of *NFATc1* or modulate its activity. For instance, *Egr2* emerging from this analysis is intriguing — as a negative regulator, it might be part of a feedback loop that ensures osteoclasts do not over-activate. Traditional clustering alone might have flagged *Egr2* as part of a cluster, but combining multiple methods and seeing it consistently with *Nfatc1* across all gave us greater certainty to highlight it. Similarly, the re-identification of *Oscar* (which was not one of the initially "given" markers) showcases the ability of this approach to expand the set of markers through the data itself. An interesting observation is that the high-confidence gene set was largely a subset of one of the major clusters from Approach 1 (the immune/inflammatory, up-regulated cluster). This overlap underscores that the fundamental difference between the approaches is scale and focus: Approach 1 casts a wide net to categorize everything, whereas Approach 2 zeroes in on the key components of one category (the osteoclast activation module). Together, they give a comprehensive picture where broad patterns are identified and then narrowed to critical genes.

Comparative Evaluation of Methods: The two approaches also differ in the techniques and assumptions, which is worth reflecting upon. Approach 1 employed advanced time-series clustering techniques (DTW, etc.) and even attempted deep learning and complex statistical models. Interestingly, the most straightforward method (DTW k-means with some manual refinement) produced superior results to the more complex ones (VAE clustering, MixFRHLP) for this dataset. This might be due to the limited number of time points and relatively high noise in biological replicates, where overly complex models overfit or fail to generalize. It highlights an important point: in practical data analysis, simpler methods with domain-informed adjustments can outperform sophisticated models if those models' assumptions do not align with the data characteristics. Approach 2, on the other hand, relied on repeating simple clustering with different algorithms and applying an intersection logic. This consensus approach mitigated the biases or random errors of any single clustering algorithm — if a gene clusters with *NFATc1* just by chance in one method, it is unlikely to do so in all three by chance. Thus, taking the intersection was an effective way to ensure robustness. It effectively increased specificity (at the cost of some sensitivity, as any gene not picked up by all methods was dropped).

The overlap between the two approaches' results provides mutual validation. Genes like *Nfatc1*, *Oscar*, *Acp5*, *Ctsk*, and *Egr2* appear as significant in both analyses, either as cluster-defining in Approach 1 or intersection-picked in Approach 2. This convergence gives us high confidence that these genes truly are central to the differentiation process. On the other hand, each approach also brought unique contributions: Approach 1 identified patterns (like metabolic down-regulation) that the marker-based approach did not emphasize (because the markers chosen were all up-regulated genes). This highlights a limitation of Approach 2: processes not linked to the chosen markers could be overlooked. Indeed, the down-regulated cluster (rich in metabolic genes) did not have a known osteoclast "marker" to pull it into focus in Approach 2, yet biologically it is an important aspect of differentiation. Conversely, Approach 2's strict filtering found *Egr2*, which Approach 1's cluster lists might have buried among dozens of other up-regulated genes in the same cluster. Thus, the approaches are complementary: one ensures no stone is unturned, the other zooms in on the most promising stones.

Limitations: Despite the insights gained, our analyses have limitations. First, the data itself has only four time points, which restricts the temporal resolution of patterns we can detect. Short, transient changes might be missed if they occur between sampled times. Future studies with more frequent sampling (e.g., adding a 6 h, 12 h, or 96 h time point) could refine the clustering and capture early transient waves or late outcomes more clearly. Second, our clustering and marker intersection approaches assume that co-expression implies co-regulation or functional relatedness, which is generally true but not guaranteed. Some genes might cluster together due to co-regulation by a common factor, but others might do so by coincidence or because they respond to different

pathways that happen to have similar timing. Experimental perturbation would be needed to confirm whether the genes we identified actually influence osteoclast differentiation or are merely correlated. Third, the marker-driven approach is inherently biased toward known biology; it might miss entirely novel regulatory modules that do not involve the chosen markers. For example, if there were an unknown pathway active in a subset of osteoclast genes independent of *NFATc1*, our Approach 2 would likely ignore it. We attempted to mitigate this by also doing the unbiased Approach 1, but it is possible that we, as analysts, focused discussion on what we recognized (immunity, metabolism) and could have underexplored a cluster of unknown function. In the future, integrating additional prior knowledge (e.g., transcription factor binding motifs or signaling network information) could help link such clusters to known pathways or reveal new ones.

Future Directions: The findings from this computational study generate several testable hypotheses. High on the priority list is experimental validation of the novel candidate regulators like *Egr2*. We propose to perform knockdown or overexpression of *Egr2* in osteoclast precursor cells in vitro to see if it indeed modulates differentiation (e.g., does overexpressing *Egr2* attenuate osteoclast formation, as its known function would predict?). Similarly, *Oscar* and *Trem2* are cell-surface proteins that could be targeted with antibodies or knocked out in precursor cultures to assess their impact on differentiation efficiency and osteoclast activity. Our results suggest these are important co-factors; validating that would improve understanding of how osteoclasts form and might highlight therapeutic targets (for instance, blocking OSCAR might inhibit osteoclast formation in diseases with excessive bone resorption).

Another future direction is to investigate the interplay between the immune-signaling module and the metabolic-repression module that we observed. One hypothesis is that as the immune/inflammatory program (driven by NFATc1 and others) ramps up, it actively suppresses the metabolic genes, perhaps via transcriptional repressors or chromatin remodeling. It would be enlightening to examine the promoters of the down-regulated metabolic genes for binding sites of factors induced by NFATc1, or to perform a time-course ATAC-seq or ChIP-seq to see if the chromatin accessibility at metabolic gene loci diminishes as NFATc1 binding increases at osteoclast genes. Such mechanistic experiments could connect the dots between the clusters we identified.

Finally, translating these findings in vivo is an important step. Our analysis was based on an in vitro-like system (cells stimulated with RANKL). In an actual organism, osteoclast differentiation is influenced by a complex environment (other cell types, varying ligand concentrations, mechanical forces). We plan to collaborate with a rheumatology research group to conduct in vivo experiments in a mouse model. As outlined in our experimental plan, we would induce osteoclastogenesis in mice (for example, by RANKL injection or an osteoporosis model) and then isolate bone marrow cells at key time points. We will validate that the expression patterns of the identified key genes (*Nfatc1*, *Egr2*, *Oscar*, etc.) in vivo mirror our in vitro findings using RT-qPCR. Furthermore, we can test functional outcomes: e.g., mice lacking *Egr2* specifically in osteoclast lineage (conditional knockout) might be examined for abnormal bone density or osteoclast numbers to directly see if *Egr2* affects osteoclast activity in vivo.

In conclusion, our dual-approach analysis not only charts the gene expression landscape of osteoclast differentiation but also provides concrete leads and a framework for subsequent experimental work. By combining unsupervised discovery with knowledge-driven filtering, we ensured that our conclusions are robust and biologically pertinent. This integrative strategy can be applied to other systems where a well-known core pathway exists alongside a broader context – for instance, studying myogenic differentiation with MyoD as a marker, or B-cell activation with CD markers – to derive both expected and novel insights.

6 Conclusion

We have presented a comprehensive analysis of RANKL-induced osteoclast differentiation in mice, integrating two complementary computational approaches on the same transcriptomic dataset. The unsupervised clustering approach delineated the major temporal patterns of gene expression, revealing an early immune activation program and a later osteoclast functional program, while highlighting potential key regulators through cluster-specific feature selection. The marker-driven approach distilled the results to a concise set of high-confidence genes co-regulated with established osteoclast markers, effectively pinpointing critical components of the osteoclastogenic transcriptional network. Both analyses converged on the importance of an immune/inflammatory gene module (including *Nfatc1*, cytokines, and receptors) and identified modulators such as *Egr2* and *Oscar* that merit further investigation.

Our findings deepen the understanding of osteoclast differentiation by connecting dynamic gene expression patterns to functional themes and candidate regulators. They suggest that successful osteoclastogenesis requires not only the activation of key transcription factors and enzymes for bone resorption but also the coordinated down-regulation of metabolic programs and involvement of feedback inhibitors to fine-tune the process. The integration of multiple analytical methods provided a robust cross-validation, increasing confidence in the results.

This work lays a strong foundation for future studies. It charts specific genes and pathways that can be targeted in follow-up experiments to validate their roles in osteoclast differentiation and activity. In practical terms, the candidate genes highlighted (for example, *Egr2*, *Trem2*, *Lgals3*) could serve as new therapeutic targets or biomarkers in diseases of bone loss if their roles are confirmed in vivo. More broadly, our integrative analytical approach can be applied to other developmental or differentiation systems to combine broad discovery with focused, knowledge-based filtering.

In summary, by merging two analytical perspectives, we achieved a more comprehensive and reliable characterization of the osteoclast differentiation process than either approach alone. This underscores the value of integrating multiple methodologies in systems biology to unravel complex biological phenomena. We conclude that osteoclast differentiation is driven by a tightly regulated gene network with discernible temporal modules, and we provide a curated list of key genes for this process. Moving forward, experimental validation of these findings will be crucial, and could ultimately inform strategies to modulate osteoclast activity in clinical settings such as osteoporosis, rheumatoid arthritis, and other osteolytic conditions.

References

Chamroukhi, F., Huynh, B., and Nguyen, S. (2013). Model-based clustering and classification of functional data. In *Advances in Data Analysis and Classification*, 7(3), pages 231–255.

Ichida, F., Nakajima, T., Tanioka, K., et al. (2004). *Egr2* is a negative regulator of osteoclastogenesis and is induced by the transcription factor NFATc1. *Journal of Biological Chemistry*, 279(36): 37219–37222.

Kim, N., Takami, M., Rho, J., et al. (2002). A novel member of the leukocyte receptor complex regulates osteoclast differentiation. *Journal of Experimental Medicine*, 195(2): 201–209.

Kim, M. J., Kim, H. S., Lee, S., Min, K. Y., Choi, W. S., and You, J. S. (2021). Hexosamine biosynthetic pathway-derived O-GlcNAcylation is critical for RANKL-mediated osteoclast differentiation. *International Journal of Molecular Sciences*, 22(16): 8888.

Kingma, D. P. and Welling, M. (2014). Auto-encoding variational Bayes. *arXiv:1312.6114*.

Liu, Z. and Barahona, M. (2021). Similarity measure for sparse time course data based on Gaussian processes. In *Proc. 37th Conf. Uncertainty in AI (UAI)*, PMLR 161: 1332–1341.

McDowell, I. C., Manandhar, D., Vockley, C. M., et al. (2018). Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLoS Computational Biology*, 14(1): e1005896.

Medvedovic, M. and Sivaganesan, S. (2002). Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics*, 18(9): 1194–1206.

Medvedovic, M., Yeung, K. Y., and Bumgarner, R. E. (2004). Bayesian mixture model based clustering of replicated microarray data. *Bioinformatics*, 20(8): 1222–1232.

Rasmussen, C. E. (1999). The infinite Gaussian mixture model. In *Advances in Neural Information Processing Systems (NIPS)*, pages 554–560.

Raudvere, U., Kolberg, L., Kuzmin, I., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 47(W1): W191–W198.

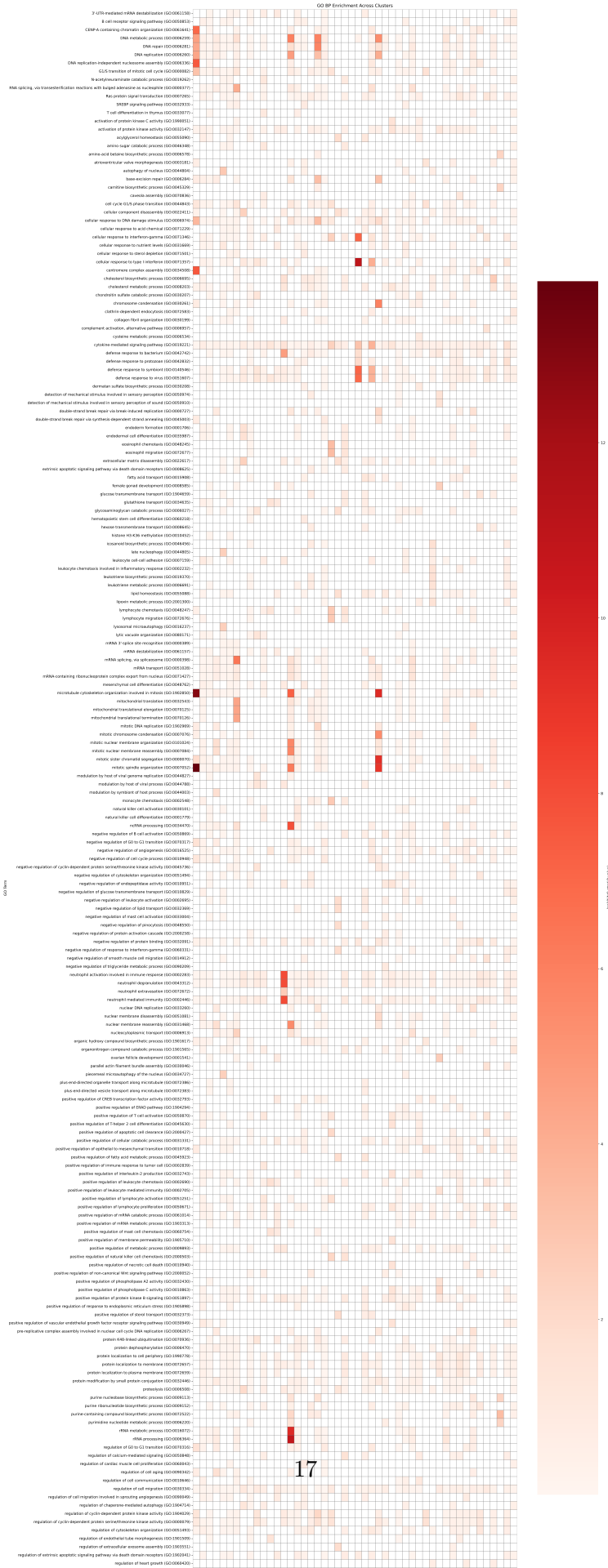
Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20: 53–65.

Samé, A., Chamroukhi, F., Govaert, G., and Aknin, P. (2011). A model-based clustering and classification for time series with changes in regime. *Advances in Data Analysis and Classification*, 5(4): 301–321.

Singhal, A. (2001). Modern information retrieval: a brief overview. *IEEE Data Engineering Bulletin*, 24(4): 35–43.

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4): 267–276.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301): 236–244.


GO BP Enrichment Across Clusters
3'-UTR-mediated mRNA destabilization (GO:0061158)
B cell receptor signaling pathway (GO:0050853)
CENP-A containing chromatin organization (GO:0061641)
DNA metabolic process (GO:0006259)
DNA repair (GO:0006281)
DNA replication (GO:0006260)
DNA replication-independent nucleosome assembly (GO:0006336)
G1/S transition of mitotic cell cycle (GO:0000082)
N-acetylneuraminate catabolic process (GO:0019262)
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)
Ras protein signal transduction (GO:0007265)
SREBP signaling pathway (GO:0032933)
T cell differentiation in thymus (GO:0033077)
activation of protein kinase C activity (GO:1990051)
activation of protein kinase activity (GO:0032147)
acylglycerol homeostasis (GO:0055090)
amino sugar catabolic process (GO:0046348)
amino-acid betaine biosynthetic process (GO:0006570)
atrioventricular valve morphogenesis (GO:0003181)
autophagy of nucleus (GO:0044804)
base-excision repair (GO:0006284)
carnitine biosynthetic process (GO:0045329)
caveola assembly (GO:0070836)
cell cycle G1/S phase transition (GO:0044843)
cellular component disassembly (GO:0022411)
cellular response to DNA damage stimulus (GO:0006974)
cellular response to acid chemical (GO:0071229)
cellular response to interferon-gamma (GO:0071346)
cellular response to nutrient levels (GO:0031669)
cellular response to sterol depletion (GO:0071501)
cellular response to type I interferon (GO:0071357)
centromere complex assembly (GO:0034508)
cholesterol biosynthetic process (GO:0006695)
cholesterol metabolic process (GO:0008203)
chondroitin sulfate catabolic process (GO:0030207)
chromosome condensation (GO:0030261)
clathrin-dependent endocytosis (GO:0072583)
collagen fibril organization (GO:0030199)
complement activation, alternative pathway (GO:0006957)
cysteine metabolic process (GO:0006534)
cytokine-mediated signaling pathway (GO:0019221)
defense response to bacterium (GO:0042742)
defense response to protozoan (GO:0042832)
defense response to symbiont (GO:0140546)
defense response to virus (GO:0051607)
dermatan sulfate biosynthetic process (GO:0030208)
detection of mechanical stimulus involved in sensory perception (GO:0050974)
detection of mechanical stimulus involved in sensory perception of sound (GO:0050910)
double-strand break repair via break-induced replication (GO:0000727)
double-strand break repair via synthesis-dependent strand annealing (GO:0045003)
endoderm formation (GO:0001706)
endodermal cell differentiation (GO:0035987)
eosinophil chemotaxis (GO:0048245)
eosinophil migration (GO:0072677)
extracellular matrix disassembly (GO:0022617)
extrinsic apoptotic signaling pathway via death domain receptors (GO:0008625)
fatty acid transport (GO:0015909)
female gonad development (GO:0008585)
glucose transmembrane transport (GO:1904659)
glutathione transport (GO:0034635)
glycosaminoglycan catabolic process (GO:0006027)
hematopoietic stem cell differentiation (GO:0060218)
hexose transmembrane transport (GO:0008645)
histone H3-K36 methylation (GO:0140938)
icosanoid biosynthetic process (GO:0046456)
late nucleophagy (GO:0044805)
leukocyte cell-cell adhesion (GO:0007159)
leukocyte chemotaxis involved in inflammatory response (GO:0002232)
leukotriene biosynthetic process (GO:0019370)
leukotriene metabolic process (GO:0006691)
lipid homeostasis (GO:0055088)
lipoxin metabolic process (GO:2001300)
lymphocyte chemotaxis (GO:0048247)
lymphocyte migration (GO:0072676)
lysosomal microautophagy (GO:0016237)
lytic vacuole organization (GO:0080171)
mRNA 3'-splice site recognition (GO:0000389)
mRNA destabilization (GO:0061157)
mRNA splicing, via spliceosome (GO:0000398)
mRNA transport (GO:0051028)
mRNA-containing ribonucleoprotein complex export from nucleus (GO:0071427)
mesenchymal cell differentiation (GO:0048762)
microtubule cytoskeleton organization involved in mitosis (GO:1902850)
mitochondrial translational elongation (GO:0070125)
mitochondrial translational termination (GO:0070126)
mitotic DNA replication (GO:1902969)
mitotic chromosome condensation (GO:0007076)
mitotic nuclear membrane organization (GO:0010320)
mitotic nuclear membrane reassembly (GO:0007084)
mitotic sister chromatid segregation (GO:0000070)
mitotic spindle organization (GO:0007052)
modulation by host of viral genome replication (GO:0044827)
modulation by host of viral process (GO:0044788)
modulation by symbiont of host process (GO:0044033)
monocyte chemotaxis (GO:0002548)
natural killer cell activation (GO:0030101)
natural killer cell differentiation (GO:0001779)
ncRNA processing (GO:0034470)
negative regulation of B cell activation (GO:0050869)
negative regulation of G0 to G1 transition (GO:0070317)
negative regulation of angiogenesis (GO:0016525)
negative regulation of cell cycle process (GO:0010948)
negative regulation of cyclin-dependent protein serine/threonine kinase activity (GO:0045736)
negative regulation of cytoskeleton organization (GO:0051494)
negative regulation of endopeptidase activity (GO:0010951)
negative regulation of glucose transmembrane transport (GO:0010829)
negative regulation of leukocyte activation (GO:0002695)
negative regulation of lipid transport (GO:0032369)
negative regulation of mast cell activation (GO:0033004)
negative regulation of pinocytosis (GO:0048550)
negative regulation of protein activation cascade (GO:2000258)
negative regulation of protein binding (GO:0032091)
negative regulation of response to interferon-gamma (GO:0060333)
negative regulation of smooth muscle cell migration (GO:0014912)
negative regulation of triglyceride metabolic process (GO:0090209)
neutrophil activation involved in immune response (GO:0002283)
neutrophil degranulation (GO:0043312)
neutrophil extravasation (GO:0072672)
neutrophil mediated immunity (GO:0002446)
nuclear DNA replication (GO:0033260)
nuclear membrane disassembly (GO:0051081)
nuclear membrane reassembly (GO:0031468)
nucleocytoplasmic transport (GO:0006913)
organic hydroxy compound biosynthetic process (GO:1901617)
organonitrogen compound catabolic process (GO:1901565)
ovarian follicle development (GO:0001541)
parallel actin filament bundle assembly (GO:0030046)
piecemeal microautophagy of the nucleus (GO:0034727)
plus-end-directed organelle transport along microtubule (GO:0072386)
plus-end-directed vesicle transport along microtubule (GO:0072383)
positive regulation of CREB transcription factor activity (GO:0032793)
positive regulation of ERAD pathway (GO:1904294)
positive regulation of T cell activation (GO:0050870)
positive regulation of T-helper 2 cell differentiation (GO:0045630)
positive regulation of apoptotic cell clearance (GO:2000427)
positive regulation of cellular catabolic process (GO:0031331)
positive regulation of epithelial to mesenchymal transition (GO:0010718)
positive regulation of fatty acid metabolic process (GO:0045923)
positive regulation of immune response to tumor cell (GO:0002839)
positive regulation of interleukin-2 production (GO:0032743)
positive regulation of leukocyte chemotaxis (GO:0002690)
positive regulation of leukocyte mediated immunity (GO:0002705)
positive regulation of lymphocyte activation (GO:0051251)
positive regulation of lymphocyte proliferation (GO:0050671)
positive regulation of mRNA catabolic process (GO:0061014)
positive regulation of mRNA metabolic process (GO:1903313)
positive regulation of mast cell chemotaxis (GO:0060754)
positive regulation of membrane permeability (GO:1905710)
positive regulation of metabolic process (GO:0009893)
positive regulation of natural killer cell chemotaxis (GO:2000503)
positive regulation of necrotic cell death (GO:0010940)
positive regulation of non-canonical Wnt signaling pathway (GO:2000052)
positive regulation of phospholipase A2 activity (GO:0032430)
positive regulation of phospholipase C activity (GO:0010863)
positive regulation of protein kinase B signaling (GO:0051897)
positive regulation of response to endoplasmic reticulum stress (GO:1903898)
positive regulation of sterol transport (GO:0032373)
positive regulation of vascular endothelial growth factor receptor signaling pathway (GO:0030949)
pre-replicative complex assembly involved in nuclear cell cycle DNA replication (GO:0006267)
protein K48-linked ubiquitination (GO:0070936)
protein dephosphorylation (GO:0006470)
protein localization to cell periphery (GO:1990778)
protein localization to membrane (GO:0072657)
protein localization to plasma membrane (GO:0072659)
protein modification by small protein conjugation (GO:0032446)
proteolysis (GO:0006508)
purine nucleobase biosynthetic process (GO:0009113)
purine ribonucleotide biosynthetic process (GO:0009152)
purine-containing compound biosynthetic process (GO:0072522)
pyrimidine nucleotide metabolic process (GO:0006220)
rRNA metabolic process (GO:0016072)
rRNA processing (GO:0006364)
regulation of G0 to G1 transition (GO:0070316)
regulation of calcium-mediated signaling (GO:0050848)
regulation of cardiac muscle cell proliferation (GO:0060043)
regulation of cell aging (GO:0090342)
regulation of cell communication (GO:0010646)
regulation of cell migration (GO:0030334)
regulation of cell migration involved in sprouting angiogenesis (GO:0060430)
regulation of chaperone-mediated autophagy (GO:1904714)
regulation of cyclin-dependent protein kinase activity (GO:1904029)
regulation of cyclin-dependent protein serine/threonine kinase activity (GO:0000079)
regulation of cytoskeleton organization (GO:0051493)
regulation of endothelial tube morphogenesis (GO:1903099)
regulation of extracellular exosome assembly (GO:1903551)
regulation of extrinsic apoptotic signaling pathway via death domain receptors (GO:1902041)
regulation of heart growth (GO:0060420)
GO Term
-log10(Adjusted P-value)
12
10
8
6
4
2