Data in Brief

Use of multiple time points to model parotid differentiation

Melissa A. Metzler [a,b], Savitri Appana [c], Guy N. Brock [c], Douglas S. Darling [a,b,*]

[a] Department of Oral Immunology and Infectious Diseases, University of Louisville, Louisville, KY 40202, United States
[b] Department of Biochemistry & Molecular Biology, University of Louisville, Louisville, KY 40202, United States
[c] Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY 40202, United States

ARTICLE INFO

ABSTRACT

In order to understand the process of terminal differentiation in salivary acinar cells, mRNA and microRNA expression was measured across the month long process of differentiation in the parotid gland of the rat. Acinar cells were isolated at either nine time points (mRNA) or four time points (microRNA) in triplicate using laser capture microdissection (LCM). One of the values of this dataset comes from the high quality RNA (RIN > 7) that was used in this study, which can be prohibitively difficult to obtain from such an RNaseI-rich tissue. Global mRNA expression was measured by rat genome microarray hybridization (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65586), and expression of microRNAs by qPCR array (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65324). Comparing expression at different ages, 2656 mRNAs and 64 microRNAs were identified as differentially expressed. Because mRNA expression was sampled at many time points, clustering and regression analysis were able to identify dynamic expression patterns that had not been implicated in acinar differentiation before. Integration of the two datasets allowed the identification of microRNA target genes, and a gene regulatory network. Bioinformatics R code and additional details of experimental methods and data analysis are provided.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications

Organism/cell line/tissue	Rat (Sprague Dawley); parotid acinar cells isolated by LCM
Sex	N/A
Sequencer or array type	Affymetrix microarray Rat 230 2.0; microRNA PCR mouse and rat panel I, V1.M
Data format	Microarray: raw .CEL files; RMA normalized Log$_2$ values in Excel
	qPCR array: raw CT values in Excel; normalized deltaCT in Excel
Experimental factors	Age of animal: E18, E20, P0, P2, P5, P9, P15, P20, P25
Experimental features	Global changes of mRNAs were measured in triplicate at 9 time points of salivary gland differentiation, and microRNAs were measured in triplicate at 4 time points. Pattern analysis across time allowed inference of a transcriptional regulatory network.
Consent	N/A
Sample source location	N/A

1. Direct link to deposited data

Microarray (mRNA data): http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65586.

qPCR array (microRNA data): http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65324.

* Corresponding author.
E-mail address: douglas.darling@louisville.edu (D.S. Darling).

2. Experimental design, materials and methods

2.1. Objective and design

Differentiation requires enduring changes in gene expression. Our objective was to use mRNA and microRNA expression measurements to identify a transcriptional regulatory network driving expression of terminal differentiation markers in parotid acinar cells. Using laser capture microdissection (LCM), parotid acinar cells could be isolated during the process of differentiation. The timing of parotid terminal differentiation is well known in rats [1,2], and occurs largely postnatally. For mRNA measurements, RNA was extracted at nine time points (embryonic days 18, and 20; and postnatal days 0, 2, 5, 9, 15, 20, and 25) in triplicate, and expression was measured by microarray (Affymetrix Rat 230 2.0). For microRNA, RNA was isolated at four time points (embryonic day 20; and postnatal days 5, 15, and 25) in triplicate, and expression was measured by qPCR array (microRNA PCR mouse and rat panel I, V1.M) (Exiqon, Vedbaek, Denmark).

2.2. Animals and tissue

Sprague Dawley rats were obtained from Harlan Laboratories. The study was carried out in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Institutional Animal Care
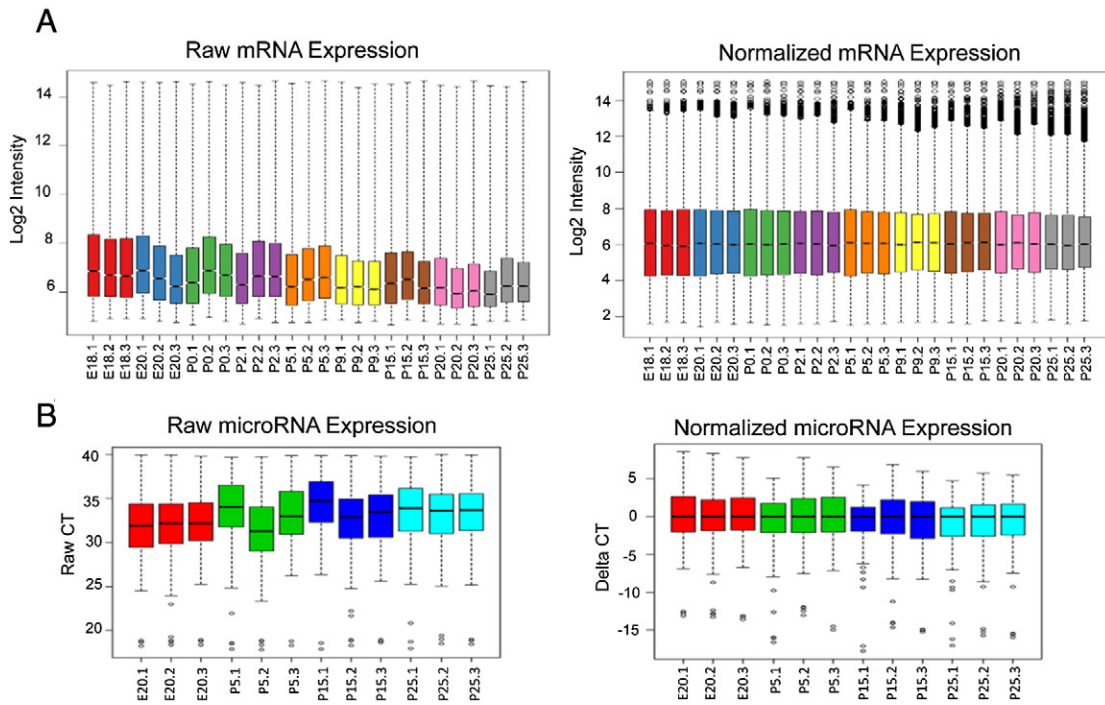
Fig. 1. Raw and normalized expression measurements. (A) Box plots generated from raw and RMA normalized microarray measurements. X-axis: individual replicate samples grouped by age; Y-axis: Log_2 intensity. Independent samples of the same age group are color coded. (B) Box plots generated from raw and median normalized CT values from a qPCR array measuring microRNA expression. X-axis: individual replicate samples grouped by age; Y-axis: Log_2 intensity. Samples of the same age group are color coded.

and Use Committee of the University of Louisville (#11059). Timed pregnant females were used for embryonic and early postnatal time points. For animals P9 and older, the parotid was removed, and immediately placed in Tissue-Tek CRYO-OCT Compound (Fisher Scientific). For younger animals the head was bisected and placed in OCT. Biological triplicates for embryonic time points were obtained

from 3 separate litters. Blocks were immediately frozen in a bath of dry ice and methyl butanol. Blocks were stored at −80°, and warmed to −30 °C for cryosectioning. H&E was used to identify acinar cells which were captured using an Arcturus PixCell IIe LCM System (Life Technologies/Thermo Fisher Scientific). Because extracting intact RNA is difficult in RNaseI rich tissues, the aqueous staining
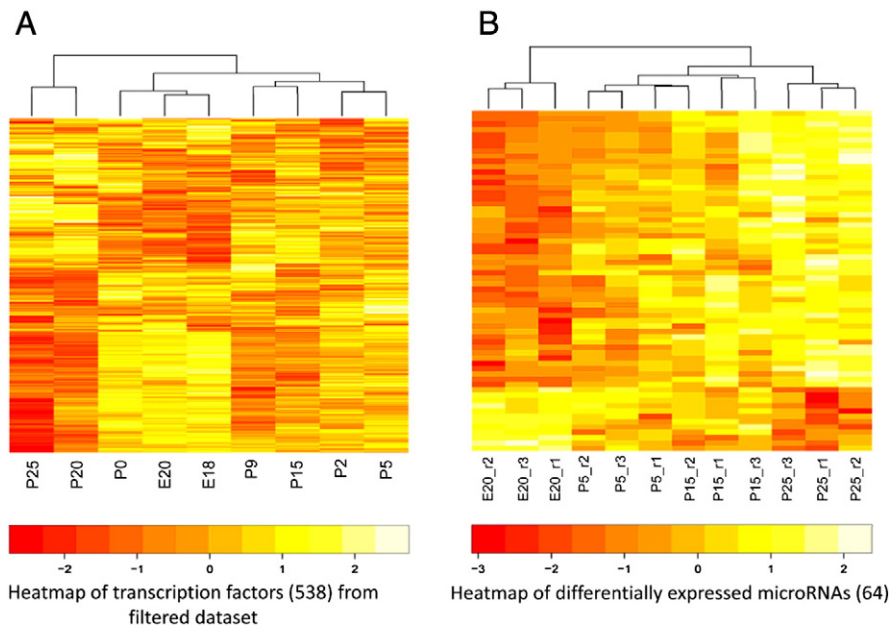
Fig. 2. Hierarchical clustering of mRNA and microRNA expression across acinar differentiation. (A) Heatmap of RMA normalized mRNA expression of transcription factors (538) in the filtered dataset from rat genome microarrays. Columns: Average expression of triplicates at that age. Rows: Affymetrix probe sets (not listed). Samples were clustered in R, and gene expression was row normalized. (B) Heatmap of differentially expressed microRNAs (64). Columns: individual samples. Rows: microRNAs. −ΔCT values of median normalized and processed expression were used to cluster samples in R, gene expression was row normalized.

Table 1
Differentially expressed mRNAs. Pairwise comparisons were made between adjacent time points, as well as the first (E18) vs. the last (P25) time point. FDR corrected p-value <0.05 was considered significant.

Comparison	Up-regulated mRNAs	Down-regulated mRNAs
E18 vs. E20	2	0
E20 vs. P0	2	1
P0 vs. P2	31	17
P2 vs. P5	0	0
P5 vs. P9	2	3
P9 vs. P15	6	6
P15 vs. P20	32	12
P20 vs. P25	7	83
E18 vs. P25	794	1763

Table 2
Differentially expressed microRNAs. Pairwise comparisons were made between time points. FDR corrected p-value <0.05 was considered significant.

Comparison	Up-regulated microRNA	Down-regulated microRNA
E20 vs. P5	1	0
E20 vs. P15	14	2
E20 vs. P25	52	12
P5 vs. P15	0	0
P5 vs. P25	7	0
P15 vs. P25	0	0

steps were kept short (15–30 s), and no more than 150 laser pulses were used when capturing tissue in the older animals when RNaseI expression is high.

### 2.2.3. RNA extraction and expression measurement

2.2.3.1. mRNA

PICO Pure RNA isolation kit from Life Technologies was used per the manufacturer's instructions. Quantity and quality was assessed by Bioanalyzer (Agilent) using the RNA 6000 *Pico* kit. Only samples with a RIN of 7 or above were used for expression measurement. For the microarray, biotinylated cRNA was prepared according to the standard protocol for NuGEN Ovation Pico WTA System V2, cRNA was hybridized for 16 h at 45 °C to Affymetrix Rat 230 2.0 Arrays, GeneChips were scanned using GeneChip Scanner 3000 7G (Affymetrix) and the GeneChip Command Console 1.0 (Affymetrix).

2.2.3.2. microRNA

Ambion's RNaqueous micro-kit was used per the manufacturer's instructions. Quantity and quality were assessed by Bioanalyzer (Agilent) using the RNA 6000 Pico kit. Samples with a RIN of 7 or above were used for expression measurement. For qPCR, Exiqon's Universal cDNA Synthesis Kit was used per the manufacturer instructions and combined with their SYBR green master mix. Samples were applied to miRCURY LNA Universal RT microRNA PCR mouse and rat panel I, V1.M, qPCR was run on an ABI 7900.

2.2.4. Normalization and filtering

The processing and analysis of the data was carried out in R, and the source code is available in the Appendix.
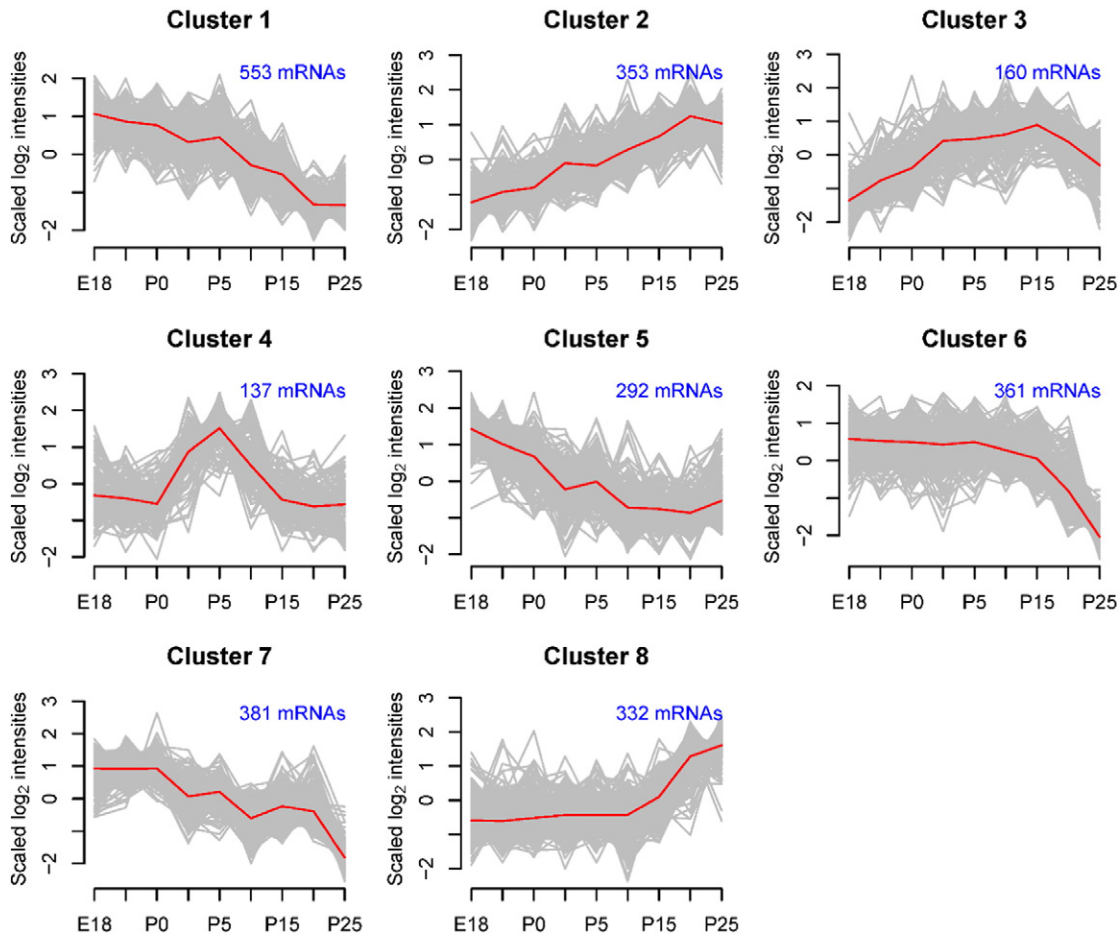
Fig. 3. Expression patterns of differentially expressed mRNAs. The 2569 differentially expressed mRNAs identified by ANOVA were clustered by k-means clustering using the per-gene scaled expression data. Eight clusters were identified. For each cluster, average expression at each age is plotted in red, and each individual microRNA is plotted in gray. Y-axis: scaled Log$_2$ intensity (gene expression for each gene was scaled to mean = 0 and stdev = 1).
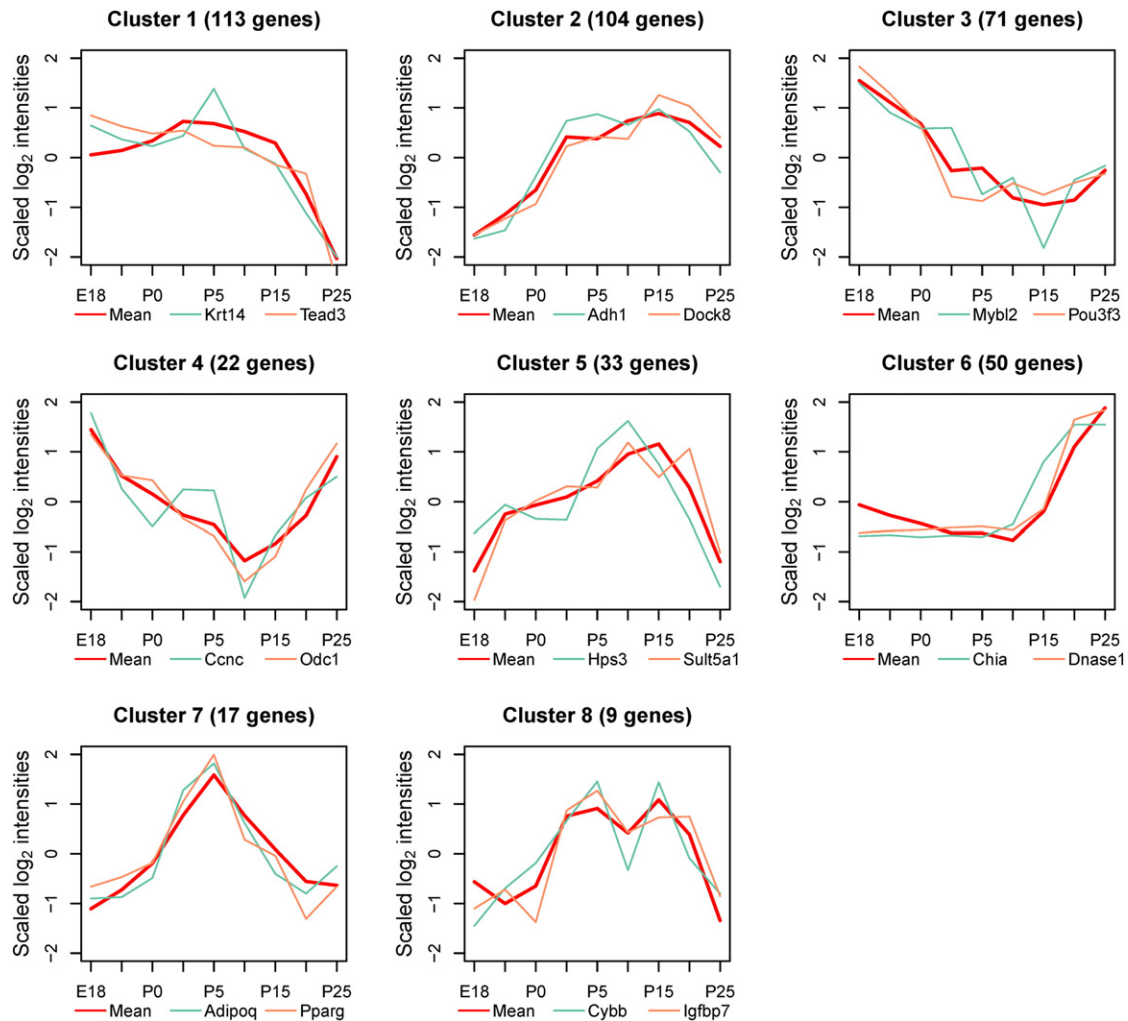
Fig. 4. Regression analysis and quadratic trends in mRNA expression. Regression analysis was used to identify mRNAs with a significant quadratic trend. These genes were then clustered. Eight clusters were identified. For each cluster, average expression at each age is plotted in red, along with two example genes that are members of the cluster. Y-axis: scaled Log$_2$ Intensity (gene expression was scaled to mean = 0 and stdev = 1 within each gene).

2.2.4.1. mRNA

The data were normalized using the *rma* function in the Bioconductor package *affy* [3], with the default settings. Boxplots for each sample before and after normalization can be seen in Fig. 1A. Quality was also checked using PCA analysis (Partek). Probes were then filtered according to the following criteria: 1) required mapping to an Entrez Gene ID (11,660 probes removed), 2) removal of duplicate probes which map to the same Entrez Gene ID (5749 probes removed, probe with the largest variability in each case retained). To focus on genes involved in the process of differentiation, probes with variability below the 50th percentile across all samples (as measured by the interquartile range) were removed. Data filtering was done using the *nsFilter* function in the Bioconductor package *genefilter* [4]. 6842 probes were carried forward for statistical analysis. Note that the number of filtered probes differs slightly from what we previously reported [5] due to the updating of reference genomes and annotation packages since our previous analysis of this data. The supplemental R files provide version numbers of all R packages used to produce the analysis presented here (via the *sessionInfo* function). However, the main results and conclusions from the analysis did not change.

2.2.4.2. microRNA

microRNAs were removed that had >50% missing data (i.e. expression was not detected in seven or more samples, which excluded 100 microRNAs), missing values from remaining probes were imputed using R function *imputeKNN* in Bioconductor package *MmPalateMiRNA* [6](14.5% of values were imputed) [7]. A median sweep was performed to normalize delta Ct values by subtracting the global median for each array. Box plots of raw and normalized values are shown in Fig. 1B. Two-hundred and eighty microRNAs were carried forward for statistical analysis.

2.2.5. Hierarchical clustering

2.2.5.1. mRNA

Broad patterns of gene expression were initially examined by hierarchical clustering to test for changes across differentiation. We have previously reported that clustering of all differentially expressed mRNAs indicates the presence of 4 stages of differentiation [5]. We tested whether these changes may be supported by changes of transcription factor expression. After normalization and filtering, 538 transcription factor mRNAs were clustered into a heatmap (Fig. 2A). The transcription factor data cluster into the 4 stages which were previously observed using all differentially expressed mRNAs [5]. The 4 stages guided choice of time points for microRNA samples. The embryonic ages (E18 and E20) cluster with the day of birth (P0) distinct from other ages. Similarly, a distinct separation is observed between younger ages and the clustered pair of P20 and P25.
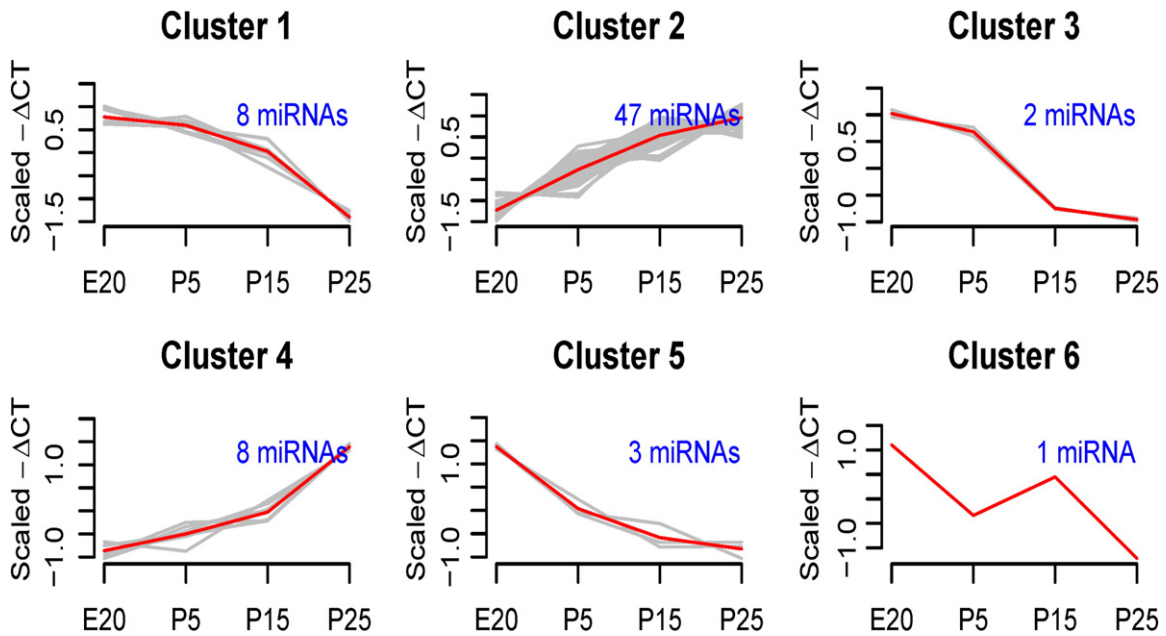
Fig. 5. Regression analysis and linear trends in microRNA expression. Regression analysis was used to identify microRNAs with a significant linear trend. These genes were then clustered into six clusters. For each cluster, average expression is plotted in red, and each individual microRNA is plotted in gray. This emphasizes patterns of progressive increases or decreases. Y-axis: scaled Log_2 intensity (gene expression for each gene was scaled to mean = 0 and stdev = 1).

2.2.5.2. microRNA

Differentially expressed microRNAs were clustered using the *hclust* function (hierarchical clustering) in the *stats* package in R. Distance *d* was calculated based on the correlation coefficient *r*, with $d = 1 - r$. A heatmap was generated from the microRNA clusters using the *heatmap_2* function in the Bioconductor package *Heatplus* (expression was scaled by rows) (Fig. 2B). There is a clear division between embryonic and postnatal microRNA samples. The embryonic time points form a distinct cluster. P5 samples cluster close together, but the P15 samples cluster with both P5 and P25.

2.2.6. Differential expression analysis

2.2.6.1. mRNA

Normalized Log_2 values were used for analysis. One-way ANOVA was used to identify differential expression of mRNA, with a false discovery rate (FDR) correction to account for multiple tests [8]. A *p*-value <0.05 was considered significant. An analysis of all nine time points identified 2569 differentially expressed mRNAs [5]. Of these, 1763 genes have a net decrease in expression (from E18 to P25) and 794 have a net increase. Comparisons were also made between adjacent
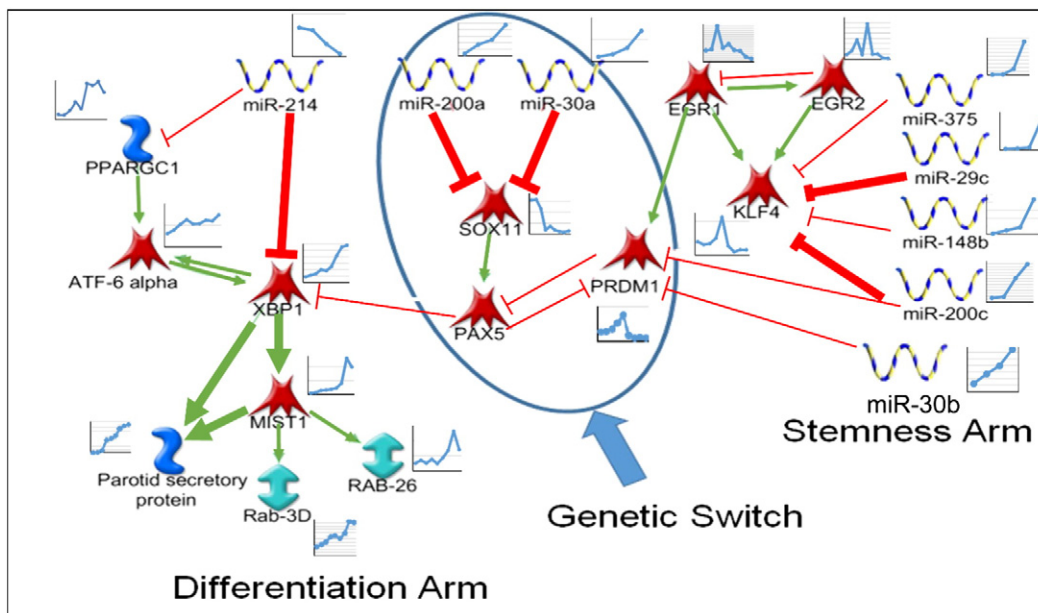
Fig. 6. Gene regulatory network during parotid acinar cell differentiation. Regulatory network derived from the network analysis of differentially expressed mRNAs and microRNAs. Edges between each node are either activating (green arrow) or repressing (red lines). Beside each node, a small graph depicts its relative expression across the time points. The network can be divided into three broad arms, on the right a pro-stemness arm, in the middle a genetic switch that reduces repression of Xbp1, and a pro-differentiation arm on the left.

time points by empirical Bayes t-tests using the Bioconductor package *limma* [9](Table 1). At the early stages, relatively few mRNAs are differentially expressed (DE) between time points within a stage, with more changes occurring between time points that span stages. The only exception being P20 vs. P25 which are in the same stage but also have a high number of DE genes, indicating that additional inputs of regulation occur in the adolescent pup.

2.2.6.2. microRNA

Normalized −ΔCT values were used for analysis of microRNA results. Differential expression was analyzed by one-way ANOVA and also by FDR corrected empirical Bayes t-tests comparing pairs of time points (Table 2). 64 unique differentially expressed microRNAs were identified by empirical Bayes t-tests across all comparisons. The E20 vs. P25 comparison contained all significant DE miRNAs identified by other comparisons. 52 microRNAs increased in expression while 12 decreased. These data are consistent with relatively gradual and relatively linear changes of microRNA expression. This was further supported by linear trend analysis of microRNAs (below).

2.2.7. Expression patterns and regression analysis

2.2.7.1. mRNA

Having nine time points of measurements allows the identification of dynamic and complex patterns of mRNA expression over time. Hierarchical clustering of the mRNA dataset indicated several patterns of expression (Fig. 2 in [5]). We extended this by the independent approach of k-means clustering of the per-gene scaled expression data (function *kmeans* in R package *stats*). The 2569 differentially expressed mRNAs were separated into eight clusters (Fig. 3). Many mRNAs are either progressively increasing (cluster 2) or decreasing (clusters 1 and 7). However, cluster 4 identifies a group of 137 genes that are transiently activated midway through differentiation. This cluster contains the transcription factor *Pparg* along with many of its downstream targets, due to which gene ontology (GO) analysis shows significant ($p \leq 0.05$) enrichment in pathways related to adipogenesis and RXR signaling. Clusters 6 and 8 identified genes that were repressed (cluster 6) or activated (cluster 8) only during the last stage of differentiation (Fig. 3). In addition, regression analysis was used to identify genes that significantly fit into either linear, quadratic, or cubic trends. The trends were then grouped into clusters. Quadratic regression in particular was able to group a relatively large number of genes (419) into dynamic expression patterns (Fig. 4). This analysis also identified the group of transiently activated genes that were found with the k-means clustering (Fig. 4, cluster 7). Regression also identified the group of genes activated only in the last stage of differentiation (Fig. 4, cluster 6), which contains many parotid secreted proteins, emphasizing that parotid differentiation is not a linear process.

2.2.7.2. microRNA

Normalized expression values of microRNAs exhibiting a significant linear trend were clustered using hierarchical clustering (function *hclust* in R package *stats*) and between 2 and 6 clusters were evaluated for validity using R package *clValid* [10]. Expression largely fit into only two clusters: increasing expression or decreasing expression. Fig. 5 demonstrates the cluster profiles for six clusters. None of the microRNAs had a significant quadratic trend.

2.2.8. Integration of mRNA and microRNA data

All differentially expressed microRNAs were used to identify prospective target mRNAs by TargetScan Mouse 6.2 [11–13]. Well conserved target sites were considered for further analysis by integrating mRNA and microRNA expression patterns. Inverse correlations were calculated using the four time points in which both mRNA and microRNA were measured (E20, P5, P15, and P25). Only expression patterns for significantly differentially expressed mRNAs and microRNAs were considered. For mRNA expression Log$_2$ values were used and for microRNA the −ΔCT. Expression of biological replicates at each time point was averaged before correlations were calculated using the *correl* function in Excel.

2.2.9. Network analysis

All differentially expressed mRNAs were uploaded into Metacore (Thomson Reuters Inc., Carlsbad, CA). Markers of terminal differentiation with increasing expression (i.e. PSP, amylase) were used as initial nodes. The neighborhood around each of these nodes was explored using the *expand* function to identify possible regulating factors. DE genes in the neighborhood were kept for another round of expanding only if their expression pattern over time was consistent with the reported interaction (i.e., activating vs. repressing) and the pattern of the target gene. The *expand* function was used iteratively until no further DE genes were identified. microRNAs predicted to target any nodes were incorporated into the network when their expression patterns had an inverse correlation with the pattern of the target mRNA. The resulting network (Fig. 6) can be divided into three arms: on the right, pro-stemness factors such as Egr1 and Klf4 decrease as differentiation proceeds, in the middle a genetic switch mediated by miR-200a and miR-30a, Sox11, Prdm1, and Pax5 removes repression of the transcription factor Xbp1, and on the left pro-acinar differentiation factors stimulate the expression of terminal differentiation markers such as PSP and Connexin32. Several of the proposed interactions in this network were validated in vitro using luciferase assays [5].

3. Discussion

In this study, we present both mRNA and microRNA expression profiling from high quality samples during acinar cell differentiation in the parotid gland. Sampling multiple time points across in vivo differentiation allows the identification of dynamic, and sometimes transient, expression changes, which can facilitate our understanding of gene expression regulation during salivary gland development. The use of multiple time points is essential for integration of mRNA and microRNA expression, allowing identification of mRNA:microRNA interactions which are supported by in vivo changes of expression.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2015.05.005.

References

[1] S. Sivakumar, L. Mirels, A.J. Miranda, A.R. Hand, Secretory protein expression patterns during rat parotid gland development. Anat. Rec. 252 (3) (1998) 485–497.
[2] R.S. Redman, L.M. Sreebny, Morphologic and biochemical observations on the development of the rat parotid gland. Dev. Biol. 25 (2) (1971) 248–279.
[3] L. Gautier, L. Cope, B.M. Bolstad, R.A. Irizarry, *affy*—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20 (3) (2004) 307–315.
[4] R. Gentleman, V. Carey, F. Hahne, Genefilter: methods for filtering genes from high throughput experiments, R package version 1.50.0. 2015.
[5] M.A. Metzler, S.G. Venkatesh, J. Lakshmanan, A.L. Carenbauer, S.M. Perez, S.A. Andres, S. Appana, G.N. Brock, J.L. Wittliff, D.S. Darling, A systems biology approach identifies a regulatory network in parotid acinar cell terminal differentiation. PLoS One 10 (4) (2015) e0125153.
[6] G.N. Brock, P. Mukhopadhyay, V. Pihur, C. Webb, R.M. Greene, M.M. Pisano, MmPalateMiRNA, an R package compendium illustrating analysis of miRNA microarray data. Source code for biology and medicine 8 (1) (2013) 1–20.

[7] S. Oh, D.D. Kang, G.N. Brock, G.C. Tseng, Biological impact of missing-value imputation on downstream analyses of gene expression profiles. Bioinformatics 27 (1) (2011) 78–86.

[8] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological) (1995) 289–300.

[9] G.K. Smyth, Limma: Linear Models for Microarray Data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, 2005. 397–420.

[10] G. Brock, V. Pihur, S. Datta, S. Datta, clValid, an R package for cluster validation. Journal of Statistical Software 25 (4) (2008).

[11] B.P. Lewis, C.B. Burge, D.P. Bartel, Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120 (1) (2005) 15–20.

[12] A. Grimson, K.K.-H. Farh, W.K. Johnston, P. Garrett-Engele, L.P. Lim, D.P. Bartel, MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell 27 (1) (2007) 91–105.

[13] B.P. Lewis, I.H. Shih, M.W. Jones-Rhoades, D.P. Bartel, C.B. Burge, Prediction of mammalian microRNA targets. Cell 115 (7) (2003) 787–798.