Biochimica et Biophysica Acta 1860 (2016) 2613–2618

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

pySAPC, a python package for sparse affinity propagation clustering: Application to odontogenesis whole genome time series gene-expression data☆

Huojun Cao [a], Brad A. Amendt [a,b,*]

[a] Iowa Institute for Oral Health Research, College of Dentistry, The University of Iowa, Iowa City, IA 52244, USA
[b] Department of Anatomy and Cell Biology and Craniofacial Anomalies Research Center, Carver College of Medicine, The University of Iowa, Iowa City, IA 52244, USA

ARTICLE INFO

Article history:
Received 26 February 2016
Received in revised form 3 June 2016
Accepted 5 June 2016
Available online 8 June 2016

Keywords:
Sparse affinity propagation clustering
Time series microarray
Dental anomalies
pySAPC

ABSTRACT

Background: Developmental dental anomalies are common forms of congenital defects. The molecular mechanisms of dental anomalies are poorly understood. Systematic approaches such as clustering genes based on similar expression patterns could identify novel genes involved in dental anomalies and provide a framework for understanding molecular regulatory mechanisms of these genes during tooth development (odontogenesis).
Methods: A python package (pySAPC) of sparse affinity propagation clustering algorithm for large datasets was developed. Whole genome pair-wise similarity was calculated based on expression pattern similarity based on 45 microarrays of several stages during odontogenesis.
Results: pySAPC identified 743 gene clusters based on expression pattern similarity during mouse tooth development. Three clusters are significantly enriched for genes associated with dental anomalies (with FDR <0.1). The three clusters of genes have distinct expression patterns during odontogenesis.
Conclusions: Clustering genes based on similar expression profiles recovered several known regulatory relationships for genes involved in odontogenesis, as well as many novel genes that may be involved with the same genetic pathways as genes that have already been shown to contribute to dental defects.
General significance: By using sparse similarity matrix, pySAPC use much less memory and CPU time compared with the original affinity propagation program that uses a full similarity matrix. This python package will be useful for many applications where dataset(s) are too large to use full similarity matrix. This article is part of a Special Issue entitled "System Genetics" Guest Editor: Dr. Yudong Cai and Dr. Tao Huang.

1. Introduction

Developmental dental anomalies are among the most common forms of congenital defects. Types of dental anomalies include anodontia (absence of all teeth) or hypodontia (absence of some teeth), supernumerary teeth, amelogenesis imperfecta and dentinogenesis imperfect. Many dental anomalies are part of a developmental syndrome that is caused by defective genes, which not only regulate tooth development but also control development of other organs and tissues such as palate, eye and heart [2,15,21,22]. In many congenital syndrome, tooth appears to be a very sensitive organ to a reduced activity of a specific genetic pathway [18]. Given its easy accessibility and relative simplicity, the tooth presents an ideal model organ to study genetic contributions to changes in developmental pathways.

To understand the complex regulatory mechanisms of genetics pathway, a popular systematic approach is clustering of whole genome time series gene-expression data [16]. Based on the principle of "guilt by association", genes that are co-expressed are very often co-regulated in the same pathway [20,33]. Affinity propagation (AP) is a new clustering method proposed by Frey and Dueck [10]. Compared with classical clustering methods such as k-means, AP has several advantages such as a lower clustering error, automatic determination of number of clusters, identification of exemplars (cluster centers), support of similarities that are not symmetric and deterministic clustering result (k-means clustering result depends on initialization, and hence requires multiple runs to achieves global optimization). AP has been successfully applied to many domains such as microarray expression data analysis [13,17,30], image clustering [12,35], structural biology [5,23] and network analysis [26,32,34]. Current implementation of Affinity propagation in Python (http://scikit-learn.org) requires full dense similarity matrix (all pair-wise similarity), which grows quadratically. In bioinformatics,

☆ This article is part of a Special Issue entitled "System Genetics" Guest Editor: Dr. Yudong Cai and Dr. Tao Huang.

* Corresponding author at: Department of Anatomy and Cell Biology, Carver College of Medicine and College of Dentistry, The University of Iowa, 51 Newton Rd, Iowa City, IA 52244, USA.
E-mail address: amendt@uiowa.edu (B.A. Amendt).

http://dx.doi.org/10.1016/j.bbagen.2016.06.008
0304-4165/

as technologies advance, more and more data are generated. In many situations, more memory and computational efficiency is required. For example, a typical mouse whole genome microarray chip has 45,000 probes and the full dense similarity matrix encoded in float data (45 k × 45 K) requires about 8 GB memory (16 GB if use double data type for similarity), a high requirement that limits the application of AP in bioinformatics. To address this limitation, we developed a python package, pySAPC, which uses a sparse similarity matrix as input. The rationale behind this approach is that in many situations, especially when dealing with large datasets, for each data point, only a small set of similarity is relevant to identify its cluster center, the majority of similarity is dispensable.

To demonstrate its application, we use pySAPC to analyze whole genome time series gene-expression data of tooth development (odontogenesis). We use pySAPC to cluster genes based on their expression pattern similarity across several stages of tooth development. After clustering, we tested for enrichment of genes that are known to be associated with dental anomalies. This clustering results can provide a new way to identify novel genes that have critical roles during tooth development and also shed light on regulatory mechanisms of these genes.

2. Methods

2.1. Implementation of sparse affinity propagation clustering (pySAPC)

We implemented pySAPC using the same steps as the original affinity propagation algorithm [10]. In pySAPC, the input is a sparse similarity matrix, which makes many matrix computations impossible. Internally, the sparse similarity matrix is converted to a numpy array representation (data, row index, column index) and all computations are based on the numpy arrays. To speed up computation, time limited steps are identified by profiling using a line profiler (https://github.com/rkern/line_profiler); these steps are then convert to fast C code using cython (http://cython.org/). The detail steps of AP algorithm are detailed in the original affinity propagation paper [10]. In brief, given similarity s(i,k) between data points i and k, real-valued messages are exchanged between data points in each iteration until convergence. There are two real-valued messages: first is the r(i,k), which is how well data point k can serve as the exemplar (cluster center) for data point i; the second is a(i,k), which represents how appropriate it is for i to choose k as its exemplar. These values are updated iteratively as Eqs. (1) and (2) until algorithm convergence:

$$r(i,k) \leftarrow s(i,k) - \max_{k\prime \text{s.t.} k\prime \neq k} \{a(i,k\prime) + s(i,k\prime)\} \tag{1}$$

$$a(i,k) \leftarrow \begin{cases} \min\left\{0, r(k,k) + \sum_{i\prime \notin \{i,k\}} \max\{0, r(i\prime,k)\}\right\} & i \neq k \\ \sum_{i\prime \neq k} \max\{0, r(i\prime,k)\} & i = k \end{cases} \tag{2}$$

Upon convergence, the exemplar (cluster center) of all data points will not change any more for defined iterations (default 15 iterations), and the exemplar for each data point i is the data point k which maximizes the following criterion:

$$\arg\max_k a(i,k) + r(i,k) \tag{3}$$

Source code of pySAPC is public available at Github (https://github.com/bioinfocao). pySAPC can be installed with pip (https://pypi.python.org/pypi/pysapc) or conda (https://anaconda.org/bioinfocao).

2.2. Odontogenesis whole genome time series gene-expression data clustering

Microarray data of laser capture micro-dissection and hand dissected mouse tooth epithelium and mesenchyme from E10.0 to E13.5 at 12- or 24-hour intervals was downloaded from GEO(GSE32321) [24]. These microarray data were generated using an Illumina MouseWG-6 v2.0 expression beadchip. Log2 transformation and robust spline normalization were applied to the expression value using lumi R package and the batch effect was removed with R package combat. The pair-wise similarity (Pearson Correlation Coefficient) matrix was computed with R (https://www.r-project.org/) function "*cov*". The full similarity matrix was filtered with several thresholds ranging from 0.2 to 0.5 to form a sparse similarity matrix. pySAPC was applied to sparse similarity matrix with default parameters. All computation was done in a computer with 2.2 GHz Intel Core i7 and 16GB Memory.

2.3. Enrichment of known dental anomalies associated genes

We manually curated a list of genes that either have been linked with syndromes involving dental anomalies in the OMIM database (http://www.ncbi.nlm.nih.gov/omim) or, when mutated in mice, show dental defects. After clustering with pySAPC, for each cluster we tested the enrichment of known genes that are associated with dental anomalies. A one side fisher exact test was used to compute odd ratios and statistical significance. The Benjamini–Hochberg procedure was used to control the false discovery rate[4]. For clusters that are enriched for genes associated with dental anomalies, DAVID (https://david.ncifcrf.gov/) was used for pathway enrichment annotation.

3. Results

3.1. Implementation of sparse affinity propagation clustering (pySAPC)

In the original affinity propagation algorithm, real-values messages were exchanged between all data points by using a full dense similarity matrix. Because the output of the affinity propagation is exemplars for each data points, we reasoned that message exchanges between two data points that are not going to be exemplar for each other are unnecessary. However, in many cases, the majority of messages exchanged are these unnecessary messages if using full dense similarity matrix. Removing these unnecessary messages saves memory and CPU time. We implemented pySAPC, a python sparse affinity propagation package, based on this idea. To use pySAPC, users only need provide a sparse similarity matrix and preferences. Preference is a parameter used to set the likelihood for each data point to be an exemplar (cluster center). Frey and Dueck suggested using the minimum or median of similarities based on their experiments [10]. Lower value (same unit of similarity) of preferences will result in fewer numbers of clusters.

We tested pySAPC on clustering images of faces, the Olivetti face data set used in the original affinity propagation paper [10]. We used several thresholds to generate sparse similarity matrix and then ran them with pySAPC (default parameters). The results, a cluster label for each face, were compared with the cluster label using a full dense similarity matrix. As shown in Fig. 1B, with only top 5% of similarity data, over half of all faces found the same cluster label as full dense similarity matrix. The top 20% of similarity data allowed pySAPC to find the exact same clustering results as the full dense similarity matrix, revealing that 80% of similarity data are dispensable and does not have any effects on the clustering result.

3.2. Applications on odontogenesis whole genome time series gene-expression data

Tooth development, termed odontogenesis, requires sequential and reciprocal signaling between epithelial and mesenchymal cells.

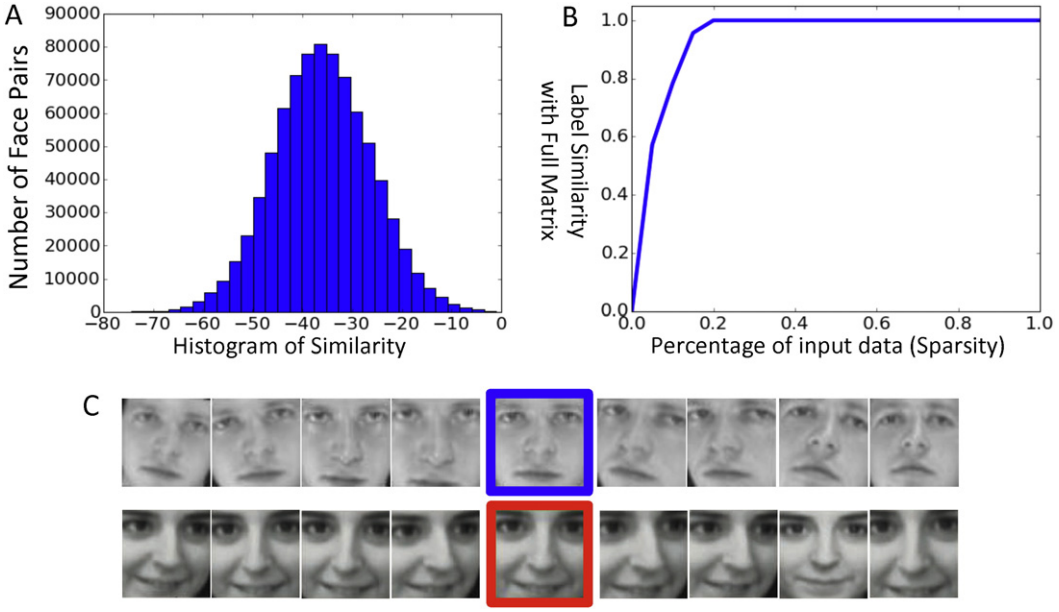

Fig. 1. Clustering faces images with pySAPC. (A) Distribution of 900 faces images pair-wise similarity scores. (B) Similarity of clustering result at different cutoff compared with full similarity matrix. Top 20% of similarity data result in exact same clustering result as using full similarity matrix. (C) Examples of clustering result, boxed face images are cluster center (exemplar) for its cluster.

O'Connell et al. have systematically profiled transcriptome dynamics during mouse tooth development from embryonic day 10.0 (E10.0) to embryonic day 14.5 (E14.5) [24]. Time points and cell types of these whole genome microarrays are summarized in Table 1. Genes that have similar expression pattern are very often co-regulated in the same developmental pathway. To identify groups of genes that are important for tooth development, we first calculated pair-wise Pearson Correlation Coefficient (PCC) for all 45 K probes and then perform clustering based on this expression pattern similarity with pySAPC. The full similarity matrix has 2050 million data points. Fig. 2A shows the distribution of PCC similarity. The full similarity matrix was filtered with several thresholds range from 0.2 to 0.5 to form a sparse similarity matrix. Similar to the face data set, the odontogenesis gene clustering result reaches a plateau very quickly (Fig. 2C). With cutoff set to 0.2, the sparse similarity matrix has 356 million data points (17% of full data). It took pySAPC a little over 10 h to cluster these data in a laptop with 2.2 GHz Intel Core i7 and 16 GB Memory.

3.3. Identification dental anomalies associated clusters

Through genome-wide association studies (GWAS), mouse genetics and other approaches, many genes associated with dental anomalies have already been identified. We manually curated a list of 42 genes that either have been linked with syndromes have dental anomalies in OMIM database (http://www.ncbi.nlm.nih.gov/omim) or when mutated in mouse show dental defects (Full list of genes can be found in Supplemental Table 1).

Table 1
Summary of tooth development time series whole genome microarrays.

GEO accession #	GSE32321
Platform	Illumina MouseWG-6 v2.0 expression beadchip
Time series of dental epithelium	E10.0, E11.0, E11.5, E12.0, E12.5, E13.0, E13.5, E14.5 (enamel knot)
Epithelium number of arrays	24 arrays
Time series of dental mesenchyme	E10.0, E11.0, E11.5, E12.0, E12.5, E13.0, E13.5
Mesenchyme number of arrays	21 arrays

pySAPC identified 743 groups of genes based on expression pattern during mouse tooth development with default parameters (preferences are set to minimum of similarities).

To identify groups of genes that might have important roles during tooth development, we tested enrichment of known dental anomalies associated genes in each of these 743 groups using Fisher exact test. As shown in Fig. 2D and Table 2, there are three groups of genes that are significantly enriched for known dental anomalies associated genes (with FDR <0.1). These three groups of genes have distinct expression pattern as shown in Fig. 3. Cluster 1 genes, which include *Pitx2*, *Ctnnb1* (*Beta-Catenin*), *Fgfr2* and other genes are highly expressed in the epithelium, show reduced expression in the enamel knot and are not expressed in mesenchyme. Functional annotation of genes in cluster 1 by DAVID(https://david.ncifcrf.gov/) shown enrichment of "regulation of cell proliferation", "epithelial cell differentiation" and other pathways (Supplemental Table 3). In contrast, genes of cluster 2 and cluster 3 are expressed predominantly in enamel knot and mesenchyme respectively. Pathways enriched in cluster 2 are "cell motion", "localization of cell" and others (Supplemental Table 3). Pathways enriched in cluster 3 are "sequence-specific DNA binding", "activator" and others (Supplemental Table 3). Many genes in the same cluster act in the same pathway. For example, *Pitx2* and *Beta-Catenin* are known to work cooperatively to regulate Wnt, Fgf signaling [3,6,14,27,31]. Interestingly, a few Wnt signal ligands are in the same group as *Pitx2* and *Beta-Catenin* (Supplemental Table 2). It has also been shown that during mouse lung development *Fgfr2* expression is reduced in *Pitx2* mutant [9]. However, whether Pitx2 also regulates *Fgfr2* expression during tooth development is unknown. In the second group, *Lrp4* and *Shh* are known regulators for each other. It has been shown that *Shh* expression is significantly reduced in *Lrp4* mutant tooth [25] and ectopic *Shh* expression in *Lrp4* mutants promotes the survival of rudimentary tooth germs [11].

4. Discussion and conclusion

Affinity propagation (AP) is a new unsupervised clustering algorithm. It has several advantages over classical clustering algorithms such as k-means. For example, it can automatically determine the number of clusters given a similarity matrix. The clustering result is

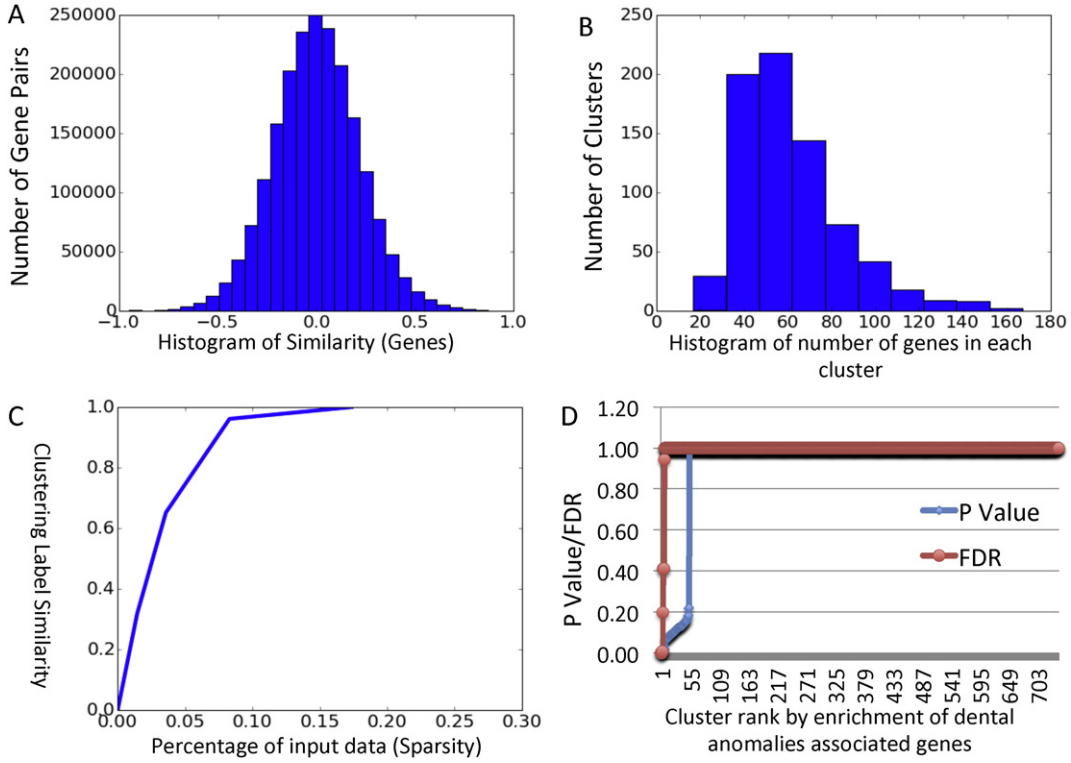

Fig. 2. Clustering genes based on expression pattern similarity with pySAPC. (A) Distribution of 45 K probes pair-wise similarity scores. (B) Histogram of number of genes in each cluster. (C) Similarity of clustering result at different cutoff compared with similarity matrix use 0.2 as cutoff. (D) Enrichment of known dental anomalies associated genes in each cluster.

deterministic, which means AP only need runs one time for a dataset, where in contrast k-means needs to run many times to find global optimization solution. The input for affinity propagation is a pair-wise similarity matrix and it grows quadratically, which limits the application of AP in bioinformatics. We implemented a python package pySAPC, which takes sparse similarity matrix as input. As we shown in the face clustering result, many similarities are dispensable, and the removal these similarities does not affect the clustering result. In bioinformatics, more and more data are generated at low cost and high speed. pySAPC will enable many applications of affinity propagation when the full similarity matrix is just too big to fit into computer memory.

Developmental dental anomalies are very common forms of congenital defects. Normal tooth development arises from complex and progressive interactions between oral epithelium and mesenchyme, a process shared by many epithelial appendages such as hair follicles, lungs, kidneys and other organs. For this reason, dental anomalies are often associated with defects in other organs that share developmental pathways. To understand the molecular mechanisms of genes and pathways that involved in tooth development, we took a systematic approach by clustering genes based on expression pattern similarity across several stages during odontogenesis using pySAPC. The clustering result helped us identified several known and several novel regulatory mechanisms that are critical during tooth development.

pySAPC clustered *Pitx2*, *Beta-Catenin*, *Fgfr2* and few other genes together based on expression pattern similarity (Supplemental Table 2). *Pitx2* is one of the earliest transcription factors expressed in dental epithelium and acts as a critical regulator of the transcriptional hierarchy in several stages of odontogenesis [7,8,18]. Knock out of the *Pitx2* gene in mice cause tooth development arrest [19]. Mutations of the *PITX2* gene have been associated with Axenfeld-Rieger Syndrome (ARS, OMIM 180500). ARS is a rare, autosomal-dominant genetic disorder occurring in about 1:200,000 human [1,18,29]. *Pitx2* and *Beta-Catenin* has been shown work together to regulate Wnt and Fgf signaling [3,6,14,28,31]. Genes in this same cluster such as *Fgfr2* could act in the same pathway as *Pitx2* and *Beta-Catenin* during tooth development. Interestingly, during mouse lung development *Fgfr2* expression is reduced in *Pitx2* mutant [9]. It will be interesting to test whether *Pitx2* also regulates *Fgfr2* expression during tooth development. As shown in Supplemental Table 2, many other genes also belong to the same group as *Pitx2* and *Beta-Catenin*. Based on their expression pattern similarities with *Pitx2* and *Beta-Catenin*, it will not be a surprise if some of these genes have important roles during tooth development and associate with dental anomalies.

In current study we found that genes associated with dental anomalies have many different expression patterns. Only a few genes such as *Pitx2*, *Beta-Catenin* and *Fgfr2* cluster together. There are a few possible explanations. First, tooth is a very complex organ despite its simple appearance. Teeth have many cell types their development takes many years. For this reason the tooth is very sensitive to genetics perturbation. Mutations in various pathways could lead to dental defects. Second, the list of known dental anomalies associated genes probably only represents a small percentage of genes that could lead to dental anomalies. Third, the microarray data used in current study only span few days (E10.0 to E14.5). In human, tooth development can take years, so

Table 2
Dental anomalies associated clusters.

Cluster	Known dental anomalies associated genes	Transcription factors in this group	Odds ratio	P value	FDR
1	Ctnnb1; Fgfr2; Irf6; Pitx2; Trp63	Foxo1; Irf6; Irx5; Lass4; Ovol2; Pitx2; Sox21; Trp63; Zbtb8b; Zfp64	48.04	1.06E-13	7.91E-11
2	Lrp4; Shh; Sp6	Pou4f1; Prdm1; Prox1; Sp6; Zbtb32; Zfp710	53.13	8.78E-08	3.26E-05
3	Barx1; Fgf10; Spry4	Barx1; Dlx5; Dlx6; Ebf3; Elk3; Fli1; Foxc2; Nfatc2; Prdm16; Preb; Zeb1	25.48	2.78E-05	0.006895

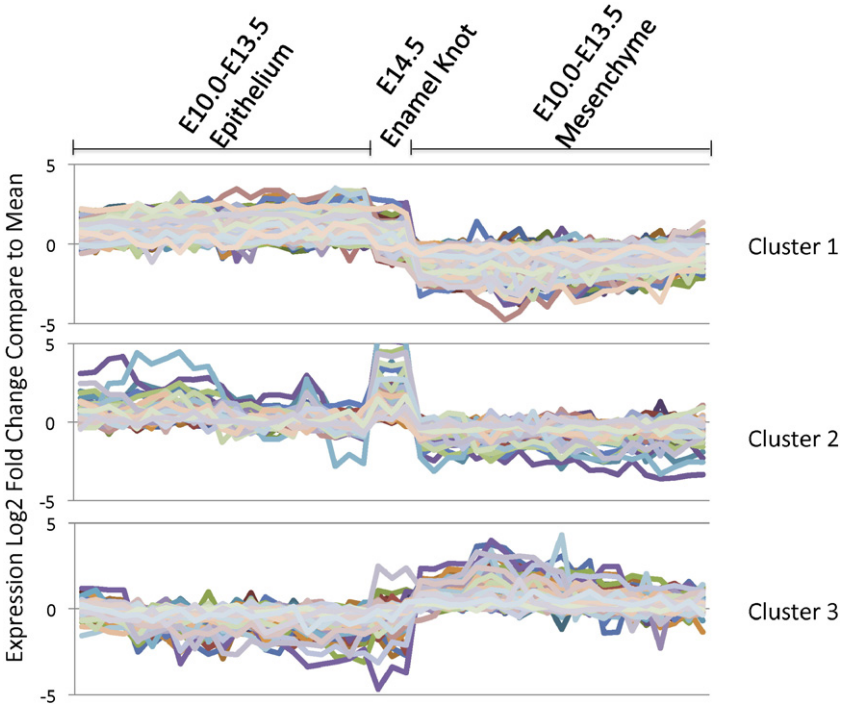

Fig. 3. Expression pattern for dental anomalies associated clusters. Expression pattern (mean adjusted) for probes in clusters that have known dental anomalies significantly enriched. 45 microarrays samples are arranged by their cell types (Epithelium, Enamel Knot and Mesenchyme) and time points (from E10.0 to E14.5).

more expression profiles at different time points should be useful in future studies.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2016.06.008.

Transparency Document

The Transparency document associated with this article can be found, in online version.

Acknowledgements

We would like to thank members of the Amendt laboratory for helpful discussions. We would like to thank our reviewers for their invaluable comments and suggestions. We would like to thank Steven Eliason and Mason Sweat for proof reading and language editing. This work was supported by the National Institutes of Health (NIH grant DE13941) and College of Dentistry, The University of Iowa.

References

[1] B.A. Amendt, The molecular and biochemical basis of Axenfeld-Rieger syndrome, The Molecular Mechanisms of Axenfeld-Rieger Syndrome, Springer, US 2005, pp. 32–53.
[2] T.N. Bartzela, C.E. Carels, E.M. Bronkhorst, E. Rønning, S. Rizell, A.M. Kuijpers-Jagtman, Tooth agenesis patterns in bilateral cleft lip and palate, Eur. J. Oral Sci. 118 (2010) 47–52.
[3] M. Basu, S.S. Roy, Wnt/β-catenin pathway is regulated by PITX2 homeodomain protein and thus contributes to the proliferation of human ovarian adenocarcinoma cell, SKOV-3, J. Biol. Chem. 288 (2013) 4355–4367.
[4] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B Methodol. 57 (1995) 289–300.
[5] U. Bodenhofer, A. Kothmeier, S. Hochreiter, APCluster: an R package for affinity propagation clustering, Bioinformatics 27 (2011) 2463–2464.
[6] P. Briata, C. Ilengo, G. Corte, C. Moroni, M.G. Rosenfeld, C.-Y. Chen, R. Gherzi, The Wnt/β-catenin → Pitx2 pathway controls the turnover of Pitx2 and other unstable mRNAs, Mol. Cell 12 (2003) 1201–1211.
[7] H. Cao, S. Florez, M. Amen, T. Huynh, Z. Skobe, A. Baldini, B.A. Amendt, Tbx1 regulates progenitor cell proliferation in the dental epithelium by modulating Pitx2 activation of p21, Dev. Biol. 347 (2010) 289–300.
[8] H. Cao, A. Jheon, X. Li, Z. Sun, J. Wang, S. Florez, Z. Zhang, M.T. McManus, O.D. Klein, B.A. Amendt, The Pitx2:miR-200c/141:noggin pathway regulates Bmp signaling and ameloblast differentiation, Dev. Camb. Engl. 140 (2013) 3348–3359.
[9] S.P. De Langhe, G. Carraro, D. Tefft, C. Li, X. Xu, Y. Chai, P. Minoo, M.K. Hajihosseini, J. Drouin, V. Kaartinen, et al., Formation and differentiation of multiple mesenchymal lineages during lung development is regulated by β-catenin signaling, PLoS One 3 (2008), e1516.
[10] B.J. Frey, D. Dueck, Clustering by passing messages between data points, Science 315 (2007) 972–976.
[11] Z. Hardcastle, R. Mo, C.C. Hui, P.T. Sharpe, The Shh signalling pathway in tooth development: defects in Gli2 and Gli3 mutants, Development 125 (1998) 2803–2811.
[12] Y. Jia, J. Wang, C. Zhang, X.-S. Hua, Finding image exemplars using fast sparse affinity propagation, Proceedings of the 16th ACM International Conference on Multimedia, (ACM) 2008, pp. 639–642.
[13] S.J. Kiddle, O.P.F. Windram, S. McHattie, A. Mead, J. Beynon, V. Buchanan-Wollaston, K.J. Denby, S. Mukherjee, Temporal clustering by affinity propagation reveals transcriptional modules in *Arabidopsis thaliana*, Bioinformatics 26 (2010) 355–362.
[14] C. Kioussi, P. Briata, S.H. Baek, D.W. Rose, N.S. Hamblet, T. Herman, K.A. Ohgi, C. Lin, A. Gleiberman, J. Wang, et al., Identification of a Wnt/Dvl/β-catenin → Pitx2 pathway mediating cell-type-specific proliferation during development, Cell 111 (2002) 673–685.
[15] O.D. Klein, S. Oberoi, A. Huysseune, M. Hovorakova, M. Peterka, R. Peterkova, Developmental disorders of the dentition: an update: American Journal Of Medical Genetics Part C (Seminars in Medical Genetics), Am. J. Med. Genet. C Semin. Med. Genet. 163 (2013) 318–332.
[16] L. Kuenzel, Gene clustering methods for time series microarray data, Biochemistry (Mosc) 218 (2010).
[17] M. Leone, Sumedha, M. Weigt, Clustering by soft-constraint affinity propagation: applications to gene-expression data, Bioinformatics 23 (2007) 2708–2715.
[18] X. Li, S.R. Venugopalan, H. Cao, F.O. Pinho, M.L. Paine, M.L. Snead, E.V. Semina, B.A. Amendt, A model for the molecular underpinnings of tooth defects in Axenfeld-Rieger syndrome, Hum. Mol. Genet. 23 (2014) 194–208.
[19] W. Liu, J. Selever, M.-F. Lu, J.F. Martin, Genetic dissection of Pitx2 in craniofacial development uncovers new functions in branchial arch morphogenesis, late aspects of tooth morphogenesis and cell migration, Development 130 (2003) 6375–6385.
[20] R. Månsson, P. Tsapogas, M. Åkerlund, A. Lagergren, R. Gisler, M. Sigvardsson, Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor, J. Biol. Chem. 279 (2004) 17905–17913.
[21] E. Matalova, J. Fleischmannova, P.T. Sharpe, A.S. Tucker, Tooth agenesis: from molecular genetics to molecular dentistry, J. Dent. Res. 87 (2008) 617–623.
[22] I. Miletich, P.T. Sharpe, Normal and abnormal dental development, Hum. Mol. Genet. 12 (2003) R69–R73.
[23] B. North, A. Lehmann, R.L. Dunbrack, A new clustering of antibody CDR loop conformations, J. Mol. Biol. 406 (2011) 228–256.
[24] D.J. O'Connell, J.W.K. Ho, T. Mammoto, A. Turbe-Doan, J.T. O'Connell, P.S. Haseley, S. Koo, N. Kamiya, D.E. Ingber, P.J. Park, et al., A Wnt-bmp feedback circuit controls Intertissue signaling dynamics in tooth organogenesis, Sci Signal 5 (2012) (ra4–ra4).

[25] A. Ohazama, T. Porntaveetus, M.S. Ota, J. Herz, P.T. Sharpe, Lrp4: a novel modulator of extracellular signaling in craniofacial organogenesis, Am. J. Med. Genet. A 152A (2010) 2974–2983.
[26] G.A. Pavlopoulos, S.I. O'Donoghue, V.P. Satagopam, T.G. Soldatos, E. Pafilis, R. Schneider, Arena3D: visualization of biological networks in 3D, BMC Syst. Biol. 2 (2008) 104.
[27] T. Sharp, J. Wang, X. Li, H. Cao, S. Gao, M. Moreno, B.A. Amendt, A pituitary homeobox 2 (Pitx2):microRNA-200a-3p:β-catenin pathway converts mesenchymal cells to amelogenin-expressing dental epithelial cells, J. Biol. Chem. 289 (2014) 27327–27341.
[28] T. Sharp, J. Wang, X. Li, H. Cao, S. Gao, M. Moreno, B.A. Amendt, A pituitary homeobox 2 (Pitx2):microRNA-200a-3p:β-catenin pathway converts mesenchymal cells to amelogenin-expressing dental epithelial cells, J. Biol. Chem. 289 (2014) 27327–27341.
[29] M.B. Shields, E. Buckley, G.K. Klintworth, R. Thresher, Axenfeld-Rieger syndrome, A Spectrum of Developmental Disorders, Surv. Ophthalmol., 29 1985, pp. 387–409.
[30] D. Tang, Q. Zhu, F. Yang, A Poisson-based adaptive affinity propagation clustering for SAGE data, Comput. Biol. Chem. 34 (2010) 63–70.
[31] U. Vadlamudi, H.M. Espinoza, M. Ganga, D.M. Martin, X. Liu, J.F. Engelhardt, B.A. Amendt, PITX2, β-catenin and LEF-1 interact to synergistically regulate the LEF-1 promoter, J. Cell Sci. 118 (2005) 1129–1137.
[32] J. Vlasblom, S.J. Wodak, Markov clustering versus affinity propagation for the partitioning of protein interaction graphs, BMC Bioinf. 10 (2009) 99.
[33] C.J. Wolfe, I.S. Kohane, A.J. Butte, Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks, BMC Bioinf. 6 (2005) 227.
[34] M. Woźniak, J. Tiuryn, J. Dutkowski, MODEVO: exploring modularity and evolution of protein interaction networks, Bioinformatics 26 (2010) 1790–1791.
[35] X. Zhang, J.C. Lv, Sparse affinity propagation for image analysis, J. Softw. 9 (2014).