



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Gene expression profile data for mouse facial development

Sonia M. Leach^a, Weiguo Feng^{b,c}, Trevor Williams^{b,c,*}^a Department of Biomedical Research, Center for Genes, Environment and Health, National Jewish Health, 1400 Jackson Street, Denver, CO 80206, USA^b Department of Craniofacial Biology, University of Colorado School of Dental Medicine, 12801 E 17th Avenue, Aurora, CO 80045, USA^c Department of Cell and Developmental Biology, University of Colorado School of Medicine, 12801 E 17th Avenue, Aurora, CO 80045, USA

ARTICLE INFO

Article history:

Received 29 March 2017

Received in revised form

29 April 2017

Accepted 3 May 2017

Available online 6 May 2017

Keywords:

Transcriptome

Functional genomics

Ectodermal

Mesenchymal

Embryonic mouse face

Expression profiles

ABSTRACT

This article contains data related to the research articles "Spatial and Temporal Analysis of Gene Expression during Growth and Fusion of the Mouse Facial Prominences" (Feng et al., 2009) [1] and "Systems Biology of facial development: contributions of ectoderm and mesenchyme" (Hooper et al., 2017 In press) [2]. Embryonic mammalian craniofacial development is a complex process involving the growth, morphogenesis, and fusion of distinct facial prominences into a functional whole. Aberrant gene regulation during this process can lead to severe craniofacial birth defects, including orofacial clefting. As a means to understand the genes involved in facial development, we had previously dissected the embryonic mouse face into distinct prominences: the mandibular, maxillary or nasal between E10.5 and E12.5. The prominences were then processed intact, or separated into ectoderm and mesenchyme layers, prior analysis of RNA expression using microarrays (Feng et al., 2009, Hooper et al., 2017 in press) [1,2]. Here, individual gene expression profiles have been built from these datasets that illustrate the timing of gene expression in whole prominences or in the separated tissue layers. The data profiles are presented as an indexed and clickable list of the genes each linked to a graphical image of that gene's expression profile in the ectoderm, mesenchyme, or intact prominence. These data files

DOI of original article: <http://dx.doi.org/10.1016/j.ydbio.2017.03.025>

* Corresponding author at: Department of Craniofacial Biology, University of Colorado School of Dental Medicine, 12801 E 17th Avenue, Aurora, CO 80045, USA.

E-mail address: trevor.williams@ucdenver.edu (T. Williams).<http://dx.doi.org/10.1016/j.dib.2017.05.003>2352-3409/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

will enable investigators to obtain a rapid assessment of the relative expression level of any gene on the array with respect to time, tissue, prominence, and expression trajectory.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	<i>Developmental Biology</i>
More specific subject area	<i>Mouse craniofacial development</i>
Type of data	<i>html pages</i>
How data was acquired	<i>Affymetrix Mouse430v2.0 and Affymetrix MoGene-1.0-st-v1 microarrays</i>
Data format	<i>Analyzed</i>
Experimental factors	<i>Samples for microarray analysis were pooled microdissected facial prominence as described in [1,2]. Ectoderm was separated from mesenchyme by 'peeling' after dispase treatment, as described in Li and Williams [3].</i>
Experimental features	<i>Gene expression data was compared across age, prominence and tissue layer.</i>
Data source location	<i>University of Colorado School of Dental Medicine, Aurora, CO</i>
Data accessibility	<i>The raw data (.cel files) and normalized expression data are available at GEO (www.ncbi.nlm.nih.gov/geo) under accession numbers GSE7759 [1] and GSE62214 [2] and at FaceBase (www.facebase.org) under accession FB00000803.</i>
Related research article	<i>[1,2]</i>

Value of the data

- The data is a statistically robust and extensively verified multidimensional gene expression resource for mouse facial development.
- By comparing the transcriptome across age, prominence and tissue layer, the data provide a valuable tool for studying the complex process of craniofacial development.
- The data could contribute to interpretation of gene expression data in mouse mutants.
- The data could contribute to interpretation of clinical genetic data pertaining to facial development, morphology and orofacial cleft pathogenesis.

1. Data

The data are contained in the folder "Trev HTMLs" and within this folder are three items: two folders and an Index.html (see [Supplementary material](#)). The data are designed to be uncompressed in a single location, where hyperlinks use relative file path names to navigate the set of files. Clicking on the top-level Index.html link will provide information concerning the two datasets. This sheet provides an overview, using color-coded boxes for each named gene, to illustrate the gene expression profile found within each of the datasets (Fig. 1). Hyperlinks are also available to access details for each gene in the Whole Prominence or Ectoderm/Mesenchyme datasets. The two folders, Whole-Prominence and EctoMesen, contain expression profiles and database annotations for every named gene available as.html pages, indexed by gene name for the two studies. Again, within each folder

Summary of Whole Prominence and Ectoderm/Mesenchyme Data

Legend:

Color Scale: Gene data are presented per time point in a given prominence relative to the average expression level across all replicates and all transcript ids representing the gene. Colored boxes indicate gene expression level relative to the mean:

Red =over-expressed relative to mean, **Blue** =under-expressed relative to mean, **White** =expressed at same level as mean.

Abbreviations: **WholeProm**=link to data in Whole Prominence dataset (if available)
EctoMesen=link to data in Ectoderm/Mesenchyme dataset (if available)
b=E10.5, **c**=E11.0, **d**=E11.5, **e**=E12.0, **f**=E12.5

Gene	Whole Prominence			Ectoderm			Mesenchyme			Description
	MdP	MxP	FNP	MdP	MxP	FNP	MdP	MxP	FNP	
	bcdef	bcdef	bcdef	bdf	df	df	bdf	df	bdf	
0610005C13Rik [WholeProm]				-	-	-	-	-	-	RIKEN cDNA 0610005C13 gene
0610006L08Rik [WholeProm]				-	-	-	-	-	-	RIKEN cDNA 0610006L08 gene
0610007N19Rik [WholeProm]				-	-	-	-	-	-	RIKEN cDNA 0610007N19
0610007P14Rik [WholeProm] [EctoMesen]										RIKEN cDNA 0610007P14 gene
0610009B22Rik [WholeProm]										RIKEN cDNA 0610009B22 gene

Fig. 1. Top-level summary index: An overview of the expression profiles is provided for each named gene in both the whole prominence and ectoderm/mesenchyme datasets, where available. Hyperlinks are provided to access more detailed gene-level data in each dataset separately. The gene list is provided in alphabetical order. Note, as different microarray chips were used for the two studies, instances where a probeset for a particular gene was available on the microarray for one analysis, but missing for the other, are indicated by a dash. In such instances, hyperlinks in the left column will only be available to connect to that gene in one dataset.

Summary of Whole Prominence Dataset

Legend:

Color Scale: Gene data are presented per time point in a given prominence relative to the average expression level across all replicates and all transcript ids representing the gene. Colored boxes indicate gene expression level relative to the mean:

Red =over-expressed relative to mean, **Blue** =under-expressed relative to mean, **White** =expressed at same level as mean.

Abbreviations: **b**=E10.5, **c**=E11.0, **d**=E11.5, **e**=E12.0, **f**=E12.5
Ecto/Mesen=Link to equivalent Ectoderm/Mesenchyme data, if available, n/a otherwise

Gene in Whole Prominence Data	Mandibular	Maxillary	Frontonasal	Equivalent Ecto/Mesen Data	Description
	bcdef	bcdef	bcdef		
0610005C13Rik				n/a	RIKEN cDNA 0610005C13 gene
0610006L08Rik				n/a	RIKEN cDNA 0610006L08 gene
0610007N19Rik				n/a	RIKEN cDNA 0610007N19
0610007P14Rik				see other dataset	RIKEN cDNA 0610007P14 gene
0610009B22Rik				see other dataset	RIKEN cDNA 0610009B22 gene
0610009D07Rik				see other dataset	RIKEN cDNA 0610009D07 gene
0610009I18Rik				see other dataset	RIKEN cDNA 0610009I18 gene

Fig. 2. Whole prominence summary index: An overview of the expression profiles in the whole prominence dataset is provided for each named gene, with hyperlinks to the equivalent data in the ectoderm/mesenchyme dataset. If the probesets/gene are not available in the other dataset, this is indicated by n/a.

Summary of Ectoderm/Mesenchyme Dataset

Legend:

Color Scale: Gene data are presented per time point in a given prominence and tissue type (Ectoderm or Mesenchyme) relative to the average expression level across all replicates and all transcript ids representing the gene in a given tissue type (Ectoderm or Mesenchyme, respectively). Colored boxes indicate gene expression level relative to the mean:

Red =over-expressed relative to mean, **Blue** =under-expressed relative to mean, **White** =expressed at same level as mean.

Abbreviations: **MdP**=mandibular prominence, **MxP**=maxillary prominence, **FNP**=frontonasal prominence
b=E10.5, **d**=E11.5, **f**=E12.5

WholeProm=Link to equivalent Whole Prominence data, if available, n/a otherwise

Gene in Ecto/Mesen data	Ectoderm			Mesenchyme			Equivalent WholeProm Data	Description
	MdP	MxP	FNP	MdP	MxP	FNP		
	bdf	df	df	bdf	df	bdf		
0610007P14Rik							see other dataset	RIKEN cDNA 0610007P14 gene
0610009B22Rik							see other dataset	RIKEN cDNA 0610009B22 gene
0610009D07Rik							see other dataset	RIKEN cDNA 0610009D07 gene
0610009L18Rik							see other dataset	RIKEN cDNA 0610009L18 gene
0610009O20Rik							see other dataset	RIKEN cDNA 0610009O20 gene

Fig. 3. Ectoderm/Mesenchyme summary index: An overview of the expression profiles in the ectoderm/mesenchyme dataset is provided for each named gene, with hyperlinks to the equivalent data in the whole prominence dataset. If the probesets/gene are not available in the other dataset, this is indicated by n/a.

there is a specific index for the whole prominence or ectoderm/mesenchyme dataset (Figs. 2 and 3, respectively). There is also a folder "HTMLS" that connects directly to the list of genes and a folder "JPEGS" which has a list of images used to populate the html pages. Within the whole prominence or ectoderm/mesenchyme datasets, a gene-specific webpage visualizes the expression and detection values for each gene as heatmaps and line graphs (raw and \log_2 scale) (Fig. 4). Each gene-specific webpage also lists annotations from the Mammalian Phenotype [4], Kyoto Encyclopedia of Genes and Genomes (KEGG) [5], InterPro [6] and Gene Ontology (GO) databases [7]. Terms relevant to craniofacial biology are highlighted in red (Fig. 5).

2. Experimental design, materials and methods

2.1. Data processing

Details of data collection and capture were described previously in [1–3]. Here, for the whole prominence datasets, the raw image data from the array scans were processed within the Affymetrix GeneChip Operating Software (Affymetrix, Santa Clara, CA) using the option in which a scaling factor was applied to bring the average intensity for all probes on the array to the same target intensity value (TGT) of 500 and exported as Signal.PivotData.dat and Detection.PivotData.dat text files. Transcript identifiers were annotated by gene name, gene description and database identifiers using the Affymetrix-provided annotation files MOE430A.na33.annot.csv and MOE430B.na33.annot.csv. Only data for transcript identifiers associated to a named gene are used for display in the html pages (39103 transcript identifiers corresponding to 22107 distinct gene symbols). All data from our analysis are MIAME compliant and are available via GEO (<http://www.ncbi.nlm.nih.gov/geo>) with the accession number GSE7759.

For separated ectoderm and mesenchyme datasets, the raw image data from the array scans were processed using the apt-probeset-summarize command from the Affymetrix Power Tools (APT) software suite, with options "-a rma -a pm-only,dabg" to perform 1) Robust Multi-array Average (RMA) with background correction, quantile normalization, and median polish summarization in order to get

Ectoderm/Mesenchyme

Transcriptional Profiling in the Ectoderm and Mesenchyme of Facial Prominences during Development

Corresponding Author: [Trevor Williams](#), Ph.D, University of Colorado Department of Craniofacial Biology

Myh3 MGI:1339709 [\[Image Legend\]](#) [\[Database Annotations\]](#) [\[Go to ENTREZ\]](#) [\[Go to MGI\]](#)

Description: myosin, heavy polypeptide 3, skeletal muscle, embryonic
Synonyms: Myhse1Myhs-e1MyHC-emb

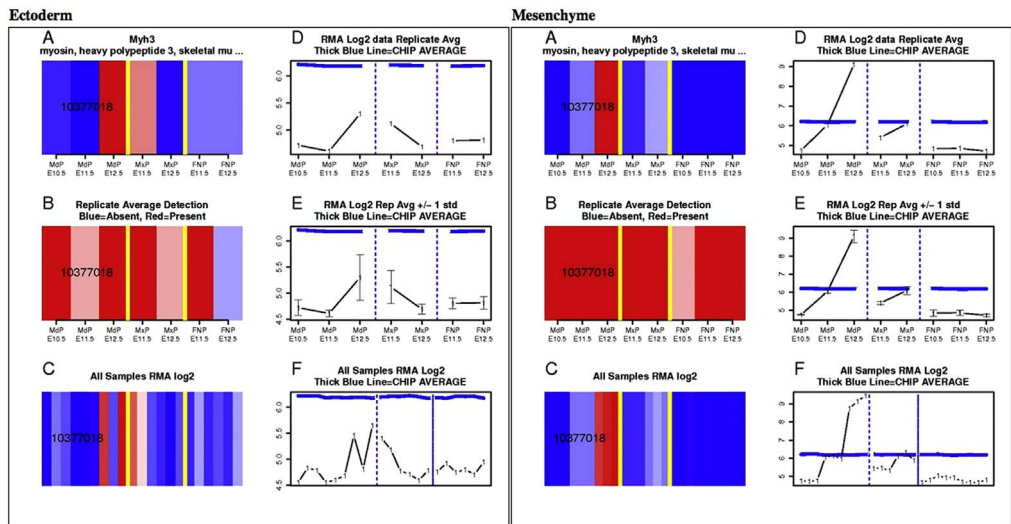


Fig. 4. Example profile visualization: Expression signal and detection profiles of all probe sets representing the gene on the microarray are visualized as heatmaps and line graphs (raw and log-scale for whole prominence; log-scale for ectoderm and mesenchyme). Explanatory legend, database annotations, and database references are also available on the html page. Active links also provide access to the legend and the annotations as well as further information on the mouse gene at NCBI and MGI. The gene shown, *Myh3*, has only one probeset in these ectodermal and mesenchymal profiles. Genes with additional probesets will have data from each shown in the equivalent heatmap and as additional line(s) on the line graph.

Database Annotations (Pheno, KEGG, InterPro, GO)

[\[Database References\]](#)

Note: Terms highlighted in Red for the Mammalian Phenotype Ontology and Gene Ontology indicate cellular, molecular, developmental and pathological processes linked with craniofacial biology, including but not limited to: the regulation of gene expression by transcription and post-transcriptional events, cell proliferation and cellular morphogenesis, cell differentiation, the extracellular matrix and cell-cell signaling as well as the processes related to the formation of bone, cartilage, vasculature, glands, placodes, cranial neural crest and the cranial neural tube that affect facial shape and patterning.

Mammalian Phenotypes [search MGI]	KEGG Pathways [search KEGG]	InterPro domains [search InterPro]
No annotations found.	KEGG:04530= Tight Junction KEGG:05416= Viral myocarditis	IPR000048=IQ motif, EF-hand binding site IPR01609=Myosin head, motor domain IPR002928=Myosin tail IPR04009=Myosin, N-terminal, SH3-like IPR00899=Myosin S1 fragment, N-terminal IPR015650=Myosin class 1/2/3/4/7/8/13/15 IPR027401=Myosin-like IQ motif-containing domain IPR027417=P-loop containing nucleoside triphosphate hydrolase
GO Process [search GO]	GO Function [search GO]	GO Component [search GO]
No annotations found.	GO:000166= nucleotide binding GO:0003774= motor activity GO:0003779= actin binding GO:0005516= calcium ion binding GO:0005524= ATP binding	GO:0005737= cytoplasm GO:0016459= myosin complex GO:0032962= myosin filament GO:0043292= contractile fiber

Fig. 5. Example Functional Annotation: Database annotations are given for each gene, where terms relevant for craniofacial developmental processes are highlighted in red. The example shown is for *Myh3* (see Fig. 4). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

log₂ expression signal values, and 2) Detection Above Background (DABG) with unmodified perfect match (PM) intensity values to calculate *p*-values for the probability of reliable signal detection. The DABG values were used to make 'Absent/Present' Detection calls, such that an expression value is called Present if the DABG *p*-value ≤ 0.05, and called Absent otherwise. Transcript identifiers were annotated

by gene name, gene description and database identifiers using the Affymetrix-provided annotation file MoGene-1_0-st-v1.na31.mm9.transcript.csv. Only data for transcript identifiers associated to a named gene are used for display in the html pages (25468 transcript identifiers corresponding to 23160 distinct gene symbols). All data from our analysis are MIAME compliant and are available via GEO (<http://www.ncbi.nlm.nih.gov/geo>) with the accession number GSE62214.

2.2. Data visualization and annotation

Data for each gene were visualized using custom scripts in the R language [8]. Functional annotations for each gene were taken from the following databases:

- 1) **Mammalian Phenotypes:** downloaded 30 Dec 2013 from <ftp://ftp.informatics.jax.org/pub/reports/index.html#pheno>;
see files ftp://ftp.informatics.jax.org/pub/reports/VOC_MammalianPhenotype.rpt and ftp://ftp.informatics.jax.org/pub/reports/MGI_PhenoGenoMP.rpt
- 2) **KEGG Pathways:** downloaded Feb 2011 from <http://www.kegg.jp/kegg/download/>)
- 3) **InterPro Domains:** downloaded 30 Dec 2013 from ftp://ftp.informatics.jax.org/pub/reports/MGI_InterProDomains.rpt
- 4) **Gene Ontology:** downloaded 30 Dec 2013 from <ftp://ftp.informatics.jax.org/pub/reports/index.html#go>;
see files ftp://ftp.informatics.jax.org/pub/reports/gene_association.mgi and ftp://ftp.informatics.jax.org/pub/reports/go_terms.mgi

Acknowledgements

Funding was provided by NIDCR via DE012728 and DE024429 (TW). We thank Joan Hooper for excellent input in the inception and design of this open access presentation.

Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2017.05.003>.

Appendix A. Supplementary material

Supplementary data associated with this article can be found at <http://dx.doi.org/10.17632/5y2hgyz5ng1>.

References

- [1] W. Feng, et al., Spatial and temporal analysis of gene expression during growth and fusion of the mouse facial prominences, *PLoS One* (2009) 4 (p. e8066).
- [2] J.E. Hooper et al., Systems biology of facial development: contributions of ectoderm and mesenchyme. *Developmental Biology*, 2017 (in press).
- [3] H. Li, T. Williams, Separation of mouse embryonic facial ectoderm and mesenchyme, *J. Vis. Exp.* 74 (2013) 50248.
- [4] J.A. Blake, et al., Mouse genome database (MGD)-2017: community knowledge resource for the laboratory mouse, *Nucleic Acids Res.* 45 (D1) (2017) D723–D729.
- [5] M. Kanehisa, et al., KEGG: new perspectives on genomes, pathways, diseases and drugs, *Nucleic Acids Res.* 45 (D1) (2017) D353–D361.
- [6] R.D. Finn, et al., InterPro in 2017-beyond protein family and domain annotations, *Nucleic Acids Res.* 45 (D1) (2017) D190–D199.
- [7] M. Ashburner, et al., Gene ontology: tool for the unification of biology. The gene ontology consortium, *Nat. Genet.* 25 (2000) 25–29.
- [8] R. Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2015.