Developmental Biology ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Developmental Biology

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

Resource

A gene expression atlas of early craniofacial development

Eric W. Brunskill [a,1], Andrew S. Potter [a,1], Andrew Distasio [a], Phillip Dexheimer [b], Andrew Plassard [b], Bruce J. Aronow [b], S. Steven Potter [a,*]

[a] Cincinnati Children's Medical Center, Division of Developmental Biology, 3333 Burnet Avenue, Cincinnati, OH 45229, USA
[b] Cincinnati Children's Medical Center, Division of Bioinformatics, 3333 Burnet Avenue, Cincinnati, OH 45229, USA

ARTICLE INFO

Article history:
Received 31 December 2013
Received in revised form
19 April 2014
Accepted 21 April 2014

Keywords:
Mammalian craniofacial development
Gene expression atlas
RNA-seq
Microarrays

ABSTRACT

We present a gene expression atlas of early mouse craniofacial development. Laser capture microdissection (LCM) was used to isolate cells from the principal critical microregions, whose development, differentiation and signaling interactions are responsible for the construction of the mammalian face. At E8.5, as migrating neural crest cells begin to exit the neural fold/epidermal ectoderm boundary, we examined the cranial mesenchyme, composed of mixed neural crest and paraxial mesoderm cells, as well as cells from adjacent neuroepithelium. At E9.5 cells from the cranial mesenchyme, overlying olfactory placode/epidermal ectoderm, and underlying neuroepithelium, as well as the emerging mandibular and maxillary arches were sampled. At E10.5, as the facial prominences form, cells from the medial and lateral prominences, the olfactory pit, multiple discrete regions of underlying neuroepithelium, the mandibular and maxillary arches, including both their mesenchymal and ectodermal components, as well as Rathke's pouch, were similarly sampled and profiled using both microarray and RNA-seq technologies. Further, we performed single cell studies to better define the gene expression states of the early E8.5 pioneer neural crest cells and paraxial mesoderm. Taken together, and analyzable by a variety of biological network approaches, these data provide a complementing and cross validating resource capable of fueling discovery of novel compartment specific markers and signatures whose combinatorial interactions of transcription factors and growth factors/receptors are responsible for providing the master genetic blueprint for craniofacial development.

Introduction

An atlas of gene expression profiles can provide a valuable resource for the research community. A prime example is the Allen Brain Atlas, which was initiated in 2003 to create a comprehensive expression dataset to advance fundamental discovery into brain function (Lein et al., 2007). Hundreds of thousands of *in situ* hybridizations were carried out to define gene expression patterns in the developing and adult mouse brain, human brain, and mouse spinal cord.

The GenitoUrinary Molecular Anatomy Project (GUDMAP.ORG) provides another example of a gene expression atlas (Harding et al., 2011; McMahon et al., 2008). A few thousand *in situ* hybridizations were carried out. In addition, however, the diverse compartments of the kidney were gene expression profiled using a combination of laser capture microdissection (LCM) and microarrays (Brunskill et al., 2008) as well as RNA-seq (Brunskill et al., 2011a, 2011b, 2011c; Brunskill and Potter, 2010). The results define the changing waves of gene expression as the kidney progenitor cells progress through the different stages of nephrogenesis.

The FACEBASE Consortium was established by NIH to provide a resource for the craniofacial research community (Hochheiser et al., 2011). One purpose of this consortium is to generate a gene expression atlas of mouse craniofacial development. In this report we describe the results of an extensive LCM/microarray/RNA-seq analysis of the gene expression patterns of early mouse craniofacial development, during E8.5, E9.5 and E10.5. At each developmental stage the multiple craniofacial compartments were isolated by LCM and gene expression patterns characterized by microarray and RNA-seq. The results define the gene expression blueprint of early craniofacial development. All growth factor, receptor and transcription factor domains of expression are defined. Novel compartment specific molecular markers are identified. In addition the RNA-seq data defines RNA splice patterns and provides a comprehensive catalog of non-coding transcripts, including those derived from enhancers. In summary, this is an extensive gene expression compendium meant to augment craniofacial research.

* Corresponding author.
E-mail address: Steve.potter@cchmc.org (S.S. Potter).
[1] These two authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ydbio.2014.04.016
0012-1606/

Materials and methods

LCM protocols

In brief, embryos were rapidly collected from CD1 outbred mice (Charles River) with the day of vaginal plug designated E0.5. Embryos were flash frozen in O.C.T. (Sakura Finetek) with liquid nitrogen cooled isopentan. Cryostat sections were made and processed, and LCM was carried out with an Arcturus Veritas machine, with membrane slides and using a combination of UV cutting and IR capture lasers as previously described (Potter and Brunskill, 2012). For a typical sample approximately 10–30 LCM collected tissue sections were pooled for analysis. Detailed protocols, with representative LCM images, are available at https://www.facebase.org/node/154.

RNA purifications and amplifications

RNA was purified using the ZR RNA MicroPrep kit (Zymo). Nugen RiboSpia Ovation Pico WTA System V2 was used for target amplification for RNA-seq. For microarrays we used the SCAMP method previously described (Brunskill et al., 2011c). RNA-seq was carried out using 50 b single end reads on the Illumina Hi-Seq 2500 machine according to Illumina protocols, with read depths of >40 million. For microarrays we examined a minimum of three biological replicates, and commonly four to six. Over one hundred microarrays in total were used. Exact replicate numbers are shown in heatmaps and at FaceBase.Org. For RNA-seq single samples were analyzed.

Single cell

Single, neural-crest cells were isolated from the cranial mesenchyme using *WNT1CRE-Rosa26GFP* reporter mice. The cranial mesenchyme and overlying epidermal ectoderm were separated from the neural fold, transferred to an eppendorf tube, digested with 0.05% trypsin for 5 min and the digestion stopped with ice-cold 1% FBS/PBS. Under a fluorescent microscope GFP-positive cells, neural crest cells, representing neural crest, and GFP-negative paraxial mesoderm and epidermal ectoderm cells, were isolated using pulled glass pipettes. Each cell was serially transferred through several Petri dishes to confirm that only a single-cell was isolated. The cell was then transferred to an eppendorf tube containing lysis buffer and quick-frozen on dry-ice for later SCAMP amplification and microarray analysis.

Data analysis

Microarray and RNA-seq data were analyzed by a combination of Bowtie, Tophat, and GeneSpring versions 7.3.1 and 12.6 software. A standard workflow for RNA-seq analysis included removal of probesets that did not map uniquely to the mm9 ENSEMBL genome or ENSEMBL-annotated genes. For the microarray samples, probesets were selected whose RMA-estimated expression level was at least 6.2 in at least one sample. Differentially expressed genes were identified using both unpaired *t*-tests for pairwise comparisons and one way ANOVA for multiple compartment comparisons (FDR <0.05), with subsequent fold-change of 2–5 fold depending on the comparisons so as to optimize the identification of expression signatures for each compartment. RNA-seq BAM files generated using the Bowtie–Tophat2 pipeline were analyzed for the expression of known and unknown genes/transcripts using both Cufflinks2 and GeneSpring 12.6. Workflows included filtering to remove duplicate reads, and those with post-aligned read metrics mapping quality below 40, as well as the exclusion of samples that exhibited strong outlier 3′/5′ read distribution ratios. Transcript/isoform and gene summarized expression tables were filtered to identify entities whose expression was at least 3 FPKM (Cufflinks) or 3 nRPKM (GeneSpring) in at least one sample. Differentially expressed gene signatures were identified using the Audic Claverie tests ($P<0.05$) and the Welch *t*-tests (FDR <0.05) followed by a two to five fold change requirements. Gene Ontology and other enrichment and biological network analysis were carried out with GeneSpring, ToppGene (http://toppgene.cchmc.org), and ToppCluster (http://toppcluster.cchmc.org/) yielding both similar and complementary results.

Biological replicate microarray Pearson correlation coefficients were generally in the range of 0.93–0.99. For example, at E10.5, we observed for the lateral nasal prominence (0.944, 0.937, 0.962), medial nasal prominence (0.950, 0.950, 0.943), olfactory pit (0.966, 0.956, 0.955), mandibular arch (0.943, 0.947, 0.978), and maxillary arch (0.976, 0.992, 0.972).

Data is available on the FaceBase.Org website and in GEO under superset GSE55967, which includes individual data series GSE55964, GSE55965, and GSE55966 for ST1.0 single cell samples, LCM microarray, and LCM RNA-seq samples, respectively.

Generating a global Facebase datamine

In order to establish integrated, comprehensive, mineable, and community-useful tables and maps of gene expression patterns across the series of cell types, regions and developmental stages that were profiled in this project, we followed a strategy that allowed us to generate two large normalized data matrix files (one being RNA-seq, the other Affymetrix GeneChip ST1.0) of all the measured values in each sample in the study. Prior to "baselining", these normalized "raw" values correspond to the estimated expression level of each gene, transcript, or probeset as measured using either technology. The RNA-seq values are FPKM values, and the microarray are based on the RMA normalization approach. These absolute expression value tables were then baselined to a global median expression value.

The RMA microarray data algorithm provides a log 2 based relative intensity value for each probeset for each microarray. Our baseline referencing approach for these RMA values was to convert these into relative ratios per sample using the transform of Intensity$=2\widehat\widehatRMA$, and then to define the ratio of each gene/probeset's expression relative to its median expression value across all samples in the dataset. Thus, the relative expression profile for a given gene is relative to this baseline. The data submitted to GEO described above includes the global normalized value data matrices that we used in these analyses. For the FPKM-based RNA-seq data, we carried out a similar strategy, but used a denominator for relative expression that was the ratio of FPKM/sample *versus* the global median of FPKM$+1$ so as to define the lowest level of log 2FPKM$=0$. Thus, the final normalized and baselined expression values represent a ratio of expression in each sample relative to the median of that gene or transcript across the sample series.

To define compartment-specific genelists, we subjected the two expression tables to a "shredding protocol" in which a series of expression signatures per sample type were generated using relative expression rank cutoffs of 100, 250, 500 and 1000 transcripts/compartment. Additional sub-signatures from these parent lists were then generated by subjecting each genelist to *K*-means clustering using 5 Pearson-correlation-based *K*-groups, thus providing genesets that are both highly ranked for a given compartment, and similarly expressed across the other samples in the datasets when they belong to the same *K*-means cluster (note that different *K*-means clusters have different numbers of genes/cluster). For each list of mouse genes, Toppgene only lists those that have an NCBI Homologene-mapped human gene ortholog.

Please cite this article as: Brunskill, E.W., et al., A gene expression atlas of early craniofacial development. Dev. Biol. (2014), http://dx.doi.org/10.1016/j.ydbio.2014.04.016