Marine Genomics 25 (2016) 89–94

Contents lists available at ScienceDirect

Marine Genomics

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

Quantitative developmental transcriptomes of the Mediterranean sea urchin *Paracentrotus lividus*

Tsvia Gildor [a,1], Assaf Malik [b,1], Noa Sher [b,1], Linor Avraham [a], Smadar Ben-Tabou de-Leon [a,*]

[a] Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa 31905, Israel
[b] Bionformatics Core Unit, University of Haifa, Haifa 31905, Israel

ARTICLE INFO

Article history:
Received 27 October 2015
Received in revised form 26 November 2015
Accepted 26 November 2015
Available online 3 December 2015

Keywords:
Quantitative developmental transcriptomes
Sea urchin
Echinoderms
Embryogenesis

ABSTRACT

Embryonic development progresses through the timely activation of thousands of differentially activated genes. Quantitative developmental transcriptomes provide the means to relate global patterns of differentially expressed genes to the emerging body plans they generate. The sea urchin is one of the classic model systems for embryogenesis and the models of its developmental gene regulatory networks are of the most comprehensive of their kind. Thus, the sea urchin embryo is an excellent system for studies of its global developmental transcriptional profiles. Here we produced quantitative developmental transcriptomes of the sea urchin *Paracentrotus lividus* (*P. lividus*) at seven developmental stages from the fertilized egg to prism stage. We generated de-novo reference transcriptome and identified 29,817 genes that are expressed at this time period. We annotated and quantified gene expression at the different developmental stages and confirmed the reliability of the expression profiles by QPCR measurement of a subset of genes. The progression of embryo development is reflected in the observed global expression patterns and in our principle component analysis. Our study illuminates the rich patterns of gene expression that participate in sea urchin embryogenesis and provide an essential resource for further studies of the dynamic expression of *P. lividus* genes.

1. Introduction

Embryonic development is controlled by networks of regulatory genes that drive the differential activation of genes in different embryonic territories through developmental time (Davidson, 2010; De-Leon and Davidson, 2007; Peter and Davidson, 2011). Understanding how this control system produces embryonic structures is a key to the mechanistic understanding of developmental processes (De Robertis, 2008; Shubin et al., 2009). The availability of developmental quantitative transcriptomes had greatly improved the ability to study the relation between developmental gene expression programs and the morphology of embryonic structures they generate (e.g., Brown et al., 2014; Gong et al., 2013; Lappalainen et al., 2013; Nodine and Bartel, 2012; Tu et al., 2014, 2012). Studying the changing landscape of gene expression through embryogenesis can shed light on the relationship between the increasing complexity of embryonic morphology and the complexity of expressed transcripts. Analyses of developmental transcriptomes, specifically in organisms where both embryogenesis and gene regulatory networks are well studied, can significantly contribute to the understanding of transcriptional regulation in development.

The sea urchin embryo presents a relatively simple developmental program that had been studied in detail for more than a century (Harvey, 1909; Hörstadius, 1939; Moore, 1929; Osborn, 1893). The current models of the gene regulatory networks that control sea urchin embryogenesis are among the most comprehensive of their kind (Andrikou et al., 2015; Annunziata et al., 2014; Ben-Tabou de-Leon et al., 2013; Garfield et al., 2013; Li et al., 2013; Luo and Su, 2012; Oliveri et al., 2008; Peter and Davidson, 2011; Saudemont et al., 2011; Tu et al., 2014). These models are based on experimental analyses in a few main species. One of the well-studied species is *Paracentrotus lividus* (*P. lividus*) that inhabits the eastern Atlantic Ocean and the Mediterranean Sea (e.g. Bessodes et al., 2012; Duboc et al., 2005; Gildor and Ben-Tabou de-Leon, 2015; Haillot et al., 2015; Lapraz et al., 2015; Molina et al., 2013; Röttinger et al., 2008; Saudemont et al., 2010). Despite the intensive studies in this species that greatly contributed to our understanding of developmental gene regulatory networks and their evolution the quantification of this species developmental transcriptomes is still lacking.

Here we generated a *P. lividus* reference transcriptome from RNA-seq data collected at seven developmental stages. We annotated and quantified the expression levels of 29,817 *P. lividus* genes at these developmental time points and verified the expression kinetics by QPCR for a subset of genes. The change in gene expression patterns and in transcript number between the maternal stages and the early zygotic stages in *P. lividus* reflect large scale transcriptional changes of

* Corresponding author.
E-mail address: sben-tab@univ.haifa.ac.il (S. Ben-Tabou de-Leon).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.margen.2015.11.013
1874-7787/

developmental, housekeeping and homeostasis genes. Thus our studies portray *P. lividus* transcript composition and expression dynamics through embryo development and provide an important resource for further studies of *P. lividus* expression dynamics.

2. Materials and methods

2.1. Sea urchin embryo cultures and RNA extraction

Adult sea urchins were supplied from a mariculture facility of the Israel Oceanographic and Limnological Research in Eilat. Sea urchin eggs and sperm were obtained by injecting adult sea urchin with 0.5 M KCl. All developmental stages were from the same parental animals. Embryos were cultured at 18 °C in 0.2 μ filtered sea water from the coasts of Haifa. Embryos growing in these conditions show slower developmental rates compared to those grown in artificial sea water (Gildor and Ben-Tabou de-Leon, 2015), yet their morphology is normal and their survival rates are similar. Total RNA was extracted using Qiagen mini RNeasy kit from embryos at indicated time points.

2.2. Reference transcriptome assembly and annotations

2.2.1. RNA-seq preparation

Samples were PolyA selected and prepared for Illumina sequencing using NEB's Ultra Directional RNA Library Prep Kit for Illumina (NEB#7520) according to the manufacturer's protocols, using 400 ng RNA per developmental time point. All libraries were verified by Tapestation and quantitated by Qubit. Equimolar concentrations were loaded at 8 pM and were sequenced with a 100 bp PE run on an Illumina HiSeq 2500 using one lane of a v3 flow cell. On average, 45 million (M) 100-bp paired-end reads were obtained for each sample, and in total, the seven samples yielded 314 M reads.

2.2.2. Transcriptome assembly and annotation

Paired End reads (PE) were cleaned from adapters, and low-quality regions, using Trimmomatic 0.3 (Bolger et al., 2014). The quality of reads was further inspected using Fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). PE reads were assembled using Trinity (version r20140717) PE *de-novo* assembly (Grabherr et al., 2011; Haas et al., 2013), to give 188,147 contigs (the Trinity equivalent of transcripts, N50 of 1498 bp), which were grouped into 137,315 Trinity-genes groups.

To assess the completeness of the transcriptome assembly, the CEGMA (Core Eukaryotic Genes Mapping Approach) software was applied (Parra et al., 2007). The analysis indicated that 80% of the 248 widely conserved core genes we tested were completely assembled and an additional 2% were partially assembled, resulting in a total of 82% representation of the core genes in the transcriptome. To examine the contiguity (Martin and Wang, 2011) of the Trinity assembly, the number of full-length transcripts was assessed according to transcript sequence representation of homology to known proteins, as previously described (Haas et al., 2013). The *P. lividus* transcripts were searched against the *Strongylocentrotus purpuratus* protein database (in http://www.echinobase.org) using blastx with an e-value cut-off of 10^{-5}. The analysis indicated that ~70% of the matching target proteins (10,451 out of 15,015 proteins) are covered by at least 70% of their length by the best matching query contig. The large proportion of fully reconstructed transcripts supports the high quality and contiguity of the assembled transcriptome.

We compared the obtained *P. lividus* contigs to protein predictions of the spbase *S. purpuratus* genome-based RNA-seq assembly (http://www.echinobase.org/Echinobase/Spbase) (Tu et al., 2012), using blastx. For each query contig, the top blastx *S. purpuratus* hit was selected (if available), after filtering out hits with e-value > 10^{-5}. In total, 44,160 *P. lividus* contigs, belonging to 27,639 Trinity-genes, were annotated using *S. purpuratus* data. Similarly, 29,834 *P. lividus* contigs, belonging to 18,016 Trinity-genes were annotated using mouse Ensembl data.

2.2.3. Gene-level expression quantitation and differential expression analysis

P. lividus PE reads were mapped to all available contigs, via RSEM 1.2.6 (Li and Dewey 2011), using the above information of genes/isoforms hierarchy. Overall, 29,817 Trinity-genes with CPM > 3 (PE read Counts Per Million mapped reads) in at least one sample before TMM normalization, were quantitated by EdgeR using R3.0.2 (Robinson et al., 2010). In EdgeR, TMM normalization was first conducted ((Robinson and Oshlack, 2010)), and then, values of CPM and Fragments Per Kilobase of transcript per Million mapped reads (FPKM) were obtained. EdgeR was further used for differential expression (DE) analysis. The current DE analysis aims to detect DE patterns specific to the tested culture, and yet, most likely similar DE patterns are shown globally in *P. lividus*. Hierarchical clustering and principal component analysis were calculated from the z-scores of the genes' CPM values. Spearman correlations were calculated using genes' $\log_{10}$FPKM values. Specifically, a hierarchical clustering on PCA was implemented by HCPC function of the R package FactoMineR (http://factominer.free.fr/classical-methods/hierarchical-clustering-on-principal-components.html). Of all 29,817 Trinity genes, 11,432 and 13,911 genes were mapped to mouse and *S. purpuratus* reference proteins, respectively. Most of these reference proteins were associated with Gene Ontology terms (11,375, and 10,571 proteins, for mouse and *S. purpuratus* respectively.).

2.2.4. Gene ontology enrichment analysis

Using the above *P. lividus* vs. *S. purpuratus* annotation data (on EdgeR-quantitated genes), *S. purpuratus* Gene Ontology (GO) terms were obtained for ~10,500 *P. lividus* genes. GO::Parser module in Perl was used in order to retrieve the GO parent terms from the go.obo database (geneontology.org). The above data, in addition to *P. lividus* gene effective-lengths (included in RSEM output), was fed into the GO enrichment program GOseq in R (Young et al., 2010). In GoSeq, correction for length selection bias was conducted using Wallenius approximation.

2.2.5. Data availability and accession numbers

Illumina short read sequences generated in this study have been submitted to the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) under accession numbers SRX749733, SRX766119, SRX766139, SRX766113, SRX766116, SRX766135, SRX766138. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GCZS00000000. The version described in this paper is the first version, GCZS00000000.1.

2.3. QPCR validation

Four independent biological repeats from different pairs of parents were used to verify RNA-seq results in *P. lividus*. Embryos were cultured at 18 °C in 0.2 μ filtered sea water from the coast of Haifa and collected at similar time points to the RNA-seq experiment. Total RNA was extracted using Qiagen mini RNeasy kit from embryos at indicated time points. 1 μg of total RNA from each time point of each independent biological replicate was used to generate cDNA using AB kit and subsequently used for QPCR.

Based on our transcriptome sequences, we designed QPCR primers for each gene using Primer3 web site (http://primer3.ut.ee/). The size of the amplicons was 120–150 bp long. Primer sequences are available in Supporting Table S1. QPCR reactions were executed in 384-well plates using 384CFX-real time machine (BioRad). Each reaction was run in technical triplicate. Every reaction contained 5 μl SYBR Green mix from BioRad including 2.5 μM forward and reverse gene specific primers and 2.5 μl of cDNA (diluted 1:50 for each assay). Thermal