Data in Brief

Transcriptome analysis of the mouse E14.5 (TS23) developing humerus and differential expression in muscle-less mutant embryos lacking mechanical stimulation

Rebecca A. Rolfe a,b,* Elaine M. Kenny c, Paul Cormican c, Paula Murphy a,b

a Department of Zoology, School of Natural Sciences, Trinity College Dublin, Dublin, Ireland
b Trinity Centre for Bioengineering, School of Engineering, Trinity College Dublin, Dublin, Ireland
c TrinSeq, Institute of Molecular Medicine, Ireland

**Abstract**

Mechanical stimulation is important for the correct formation of the skeleton. Splotch-delayed mutant embryos (Pax3Spd+/−) that develop with no limb muscle and therefore no limb movement experience an altered mechanical environment resulting in specific defects in ossification and joint formation, particularly in the forearm. To test the hypothesis that mechanical stimuli influence the regulation of genes important in skeletal development we generated a transcriptome profile of the developing humerus at Theiler stage 23 (TS23), and then identified differentially expressed genes in muscle-less mutant embryos compared to control littermates. Here we describe the experimental methods and analysis of the resulting data, publically available in the ArrayExpress database under E-MTAB-1745 (Transcriptome of control humerus), E-MTAB-1744 (Microarray; differential expression) and E-MTAB-1746 (RNA-sequencing; differential expression). Our data provide a resource for exploring the transcriptome that underlies skeletal development at TS23 in the mouse humerus. The interpretation and description of this data can be found in a recent publication in BMC Genomics [1]. This is a resource for exploring the molecular mechanisms that are involved in skeletal development and mechanotransduction.

© 2014 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/3.0/).
RNA was extracted (SV Total RNA Isolation System: Promega, UK). RNA quantity and integrity were assessed on a 2100 Bioanalyser (Agilent Technologies). RNA samples with RNA integrity number (RIN) values greater than 8 were used for microarray and RNA-sequencing analysis. Details of the samples used for array hybridization and RNA-sequencing as described in ArrayExpress are shown in Table 1.

**Microarray study design**

Four independent pooled sets of biological replicate samples were used for microarray (n = 4 biological replicates) analysis, as shown in Table 1. All microarrays were processed at IMGM® Laboratories (Martinsried, Germany). The protocol involved the reverse transcription of 100 ng of total RNA per sample into cDNA and then the conversion into labelled cRNA by in vitro transcription (Low input quick-amp labelling kit one-colour, Agilent Technologies) incorporating cyanine-3-CTP. All cRNA was quality control checked to assess yield and activity prior to hybridisation. The cRNA values ranged from 4.53 μg to 1.70 μg and completely fulfilled the quality criteria defined by Agilent. The cRNA was cleaned up and quantified and 1.65 μg of each Cyanine-3-labelled cRNA sample was fragmented and prepared for one-colour based hybridisation (Gene Expression Hybridisation Kit, Agilent Technologies). The cRNA samples were hybridised at 65 °C for 17 h on separate Mouse GE v2 Microarrays (4x44K format) (G4846A, Agilent Technologies). The cRNA values ranged from 4.53 μg to 1.70 μg and completely fulfilled the quality criteria defined by Agilent. The cRNA was cleaned up and quantified and 1.65 μg of each Cyanine-3-labelled cRNA sample was fragmented and prepared for one-colour based hybridisation (Gene Expression Hybridisation Kit, Agilent Technologies). The cRNA samples were hybridised at 65 °C for 17 h on separate Mouse GE v2 Microarrays (4x44K format) (G4846A, Agilent Technologies), which contain 39,485 coding and non-coding sequences of the mouse genome. Microarrays were then washed with increased stringency using Gene Expression Wash Buffers (Agilent Technologies) followed by drying with Acetonitrile (Sigma). Fluorescent signal intensities were detected with Scan Control A.8.4.1 software (Agilent Technologies) in the Agilent DNA microarray scanner and extracted from the images using Feature Extraction 10.7.3.1 software (Agilent Technologies). A workflow of the individual protocol followed for microarray preparation can be seen in Table 2.

**RNA-sequencing alignment**

A total of six independent DNase-treated RNA (3 μg) samples of the control and mutant preparations, as indicated in Table 1, were used to prepare RNA-sequencing libraries with the TruSeq RNA Sample Prep kit for RNA-sequencing. Libraries were prepared according to the Illumina protocol (Illumina Part #15008136 Rev. A) and sequenced on an Illumina GAII sequencer (Trinity Genome Sequencing Laboratory, Institute of Molecular Medicine, Trinity College Dublin, Ireland). The workflow to prepare the libraries consisted of purifcation and fragmentation of the mRNA, first strand cDNA synthesis, second strand cDNA synthesis, repair of fragmented ends into blunt ends, adenylation 3’ ends, ligation of adapters, PCR amplification and library validation. The six constructed cDNA libraries represent triplicate biological replicates for each group. 40 bp single end reads were obtained from an Illumina GAII in FASTQ format, with one sample per sequencing lane (n = 6). The Tophat aligner (http://tophat.cbcb.umd.edu/) was used to align the reads to the mouse reference genome (mm9). After alignment the read counts for each gene were extracted using htseq-count (http://www-huber.embl.de/users/anders/HTSeq/) based on an mm9 Refseq gff file. A workflow of the individual protocols followed for RNA-sequencing preparation can be seen in Table 2.
Table 3

| Genotype   | Probability |
|------------|-------------|
| 11100022R   | 11100059R   | 15000072R   | 15000123R   | 15000150R   | 15000192R   | 20100160R   | 20101009R   | 24000061R   | 31000074R   | 34500104R   | 35000107R   | 36500123R   |
| 15000202R   | 15000240R   | 15000282R   | 15000320R   | 15000360R   | 15000400R   | 31000040R   | 31000080R   | 31000120R   | 31000014R   | 31000054R   | 31000094R   | 31000134R   |
| 17001252H   | 18100145R   | 20100211R   | 20100251R   | 20100291R   | 20100331R   | 20100371R   | 20100411R   | 20100451R   | 20100491R   | 20100531R   | 20100571R   | 20100611R   |
| 22100404K   | 23100169R   | 23100219R   | 23100269R   | 23100319R   | 23100369R   | 23100419R   | 23100469R   | 23100519R   | 23100569R   | 23100619R   | 23100669R   | 23100719R   |
| 29000058R   | 29000099R   | 29000140R   | 29000190R   | 29000240R   | 29000290R   | 29000340R   | 29000390R   | 29000440R   | 29000490R   | 29000540R   | 29000590R   | 29000640R   |
| 34831419R   | 49225021R   | 49305118R   | 49305168R   | 49324318R   | 49334300R   | 53904123R   | 53904173R   | 62000417R   | 63500453R   | 64500512R   | 64500562R   | 64500612R   |
| 64500662R   | 64500712R   | 64500762R   | 64500812R   | 64500862R   | 64500912R   | 64500962R   | 64500012R   | 64500062R   | 64500112R   | 64500162R   | 64500212R   | 64500262R   |

Legend:
- **R.A. Rolfe et al. / Genomics Data 2 (2014) 32–36**

**680 Down-regulated DE genes**

**432 Up-Regulated DE genes**
Whole transcriptome profile of developing humerus at TS23

The three constructed libraries of the TS23 control developing humerus samples representing distinct biological replicates were sequenced using Illumina GAII and aligned to the mm9 mouse reference genome. Following alignment, an average read count value was computed from the three replicates and the total number of transcripts with greater than/equal to 5 reads were defined as expressed.

Differential expression quantification

Microarray

The software tools Feature Extraction 10.7.3.1, GeneSpring GX 11.5.1 (both Agilent Technologies) and Spotfire Decision Site 9.1.2 (TIBCO) were used for quality control and statistical data analysis. Quantile normalisation was applied to each data set in order to impose the same distribution of probe signal intensities for each array [4], thus adjusting them to a uniform level that can allow for comparable downstream analysis. Welch’s approximate t-test (“unpaired unequal variance”, parametric) was applied to compare the control and mutant groups. A corrected p-value was calculated based on the algorithm of Benjamini and Hochberg [5], based on control of the false discovery rate (FDR). In each comparison, two groups were compared in a pairwise manner. The extent and direction of differential expression between the groups were determined by calculating a fold change value. A fold change of $\geq 2$ and FDR-adjusted p-value of $\leq 0.05$ were used as criteria to indicate differential expression (DE) between the two groups. For example a fold change value of 2 indicates a two-fold repression or down-regulation. The fold change threshold signifies that probes are induced by a factor of 2 (equal to 100% increase in averaged normalised signal) or repressed by the same factor (equal to a decrease of the initial signal by 50%).

RNA-sequencing

Differential expression (DE) at a gene level in our two groups (control and mutant) was evaluated using DESeq version 1.4.1, implemented in R 2.14.1. DESeq uses a negative binomial distribution to model genic read counts following normalisation based on size factors and variance. As with the microarray data the extent and direction of differential expression was determined by calculating a fold change value. The read count for an expressed gene from the control sample ($n = 3$) was directly compared to the read count from the mutant ($n = 3$) in order to determine the fold change value. Values $\geq 2$ were taken as an indication of induction or up-regulation of expression, while values $\leq 0.5$ were taken to indicate repression or down-regulation of expression. The p-values presented are adjusted for multiple-testing with the procedure of Benjamini and Hochberg to control the type I error rate, and a cut off of $p \leq 0.05$ was used as a threshold to define differential expression (Table 3).

Discussion

The data described here catalogue the transcriptome of a developing skeletal rudiment (humerus; TS23), as well as reveal differentially expressed genes between humeri developing normally and humeri with reduced mechanical stimulation (Table 3). These datasets have recently been analysed and interpreted to reveal aspects of the molecular mechanisms involved in mechanotransduction during skeletal development [1]. Here we describe the publically available datasets in order to facilitate future analyses and integration with other appropriate studies. The normal transcriptome (E-MTAB-1745) provides a valuable resource to help understand the processes that are occurring during skeletogenesis at this stage of development when ossification and joint specification is occurring. These data can be used in combination with other genome wide databases and tools, for example EurExpress (www.eurexpress.org) which examines the spatial expression patterns of 18,000 coding genes and over 400 microRNAs in the whole mouse embryo at the same developmental stage examined here, TS23 [6]. Combining these resources gives information on the quantitative and spatial expression of individual genes providing the basis to explore regulatory networks active during the development of skeletal rudiments. Differential expression was revealed by both Microarray (374) and RNAseq (1037) approaches with RNAseq showing greater sensitivity in this respect [1]. These datasets can now be mined and integrated with the results of other experiments examining the impact of mechanical stimuli on developing tissues (e.g.[7]) and cells (e.g. [8]), focusing perhaps on specific pathways or genes. This holds the potential to reveal highly informed and supported hypotheses to drive functional studies, for example, asking if similar pathways and component genes are affected by mechanical stimuli in different biological contexts.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

[1] R.A. Rolfe et al., Identification of mechanosensitive genes during skeletal development: alteration of genes associated with cytoskeletal rearrangement and cell signaling pathways. BMC Genomics 15 (1) (2014) 48.

[2] K. Theiler., The House Mouse. Springer, Berlin, Heidelberg, New York, 1972.

[3] C.R. Keller-Peck, R.J. Mullen, Altered cell proliferation in the spinal cord of mouse neu- ronal tube mutants curly tail and Pax3 splotch-delayed. Brain Res. Dev. Brain Res. 102 (2) (1997) 177–188.

[4] B.M. Bolstad, et al., A comparison of normalization methods for high density oligonu- cleotide array data based on variance and bias. Bioinformatics 19 (2) (2003) 185–193.
[5] Y. Benjamini, Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B. 57 (1) (1995) 289–300.

[6] G. Diez-Roux, et al., A high-resolution anatomical atlas of the transcriptome in the mouse embryo. PLoS Biol. 9 (1) (2011) e1000582.

[7] C.M. Reijnders, et al., Increased expression of matrix extracellular phosphoglycoprotein (MEPE) in cortical bone of the rat tibia after mechanical loading: identification by oligonucleotide microarray. PLoS One 8 (11) (2013) e79672.

[8] R.K. Sironen, et al., cDNA array reveals mechanosensitive genes in chondrocytic cells under hydrostatic pressure. Biochim. Biophys. Acta 1591 (1–3) (2002) 45–54.