Supplementary Information

DCX-EMAP is a candidate gene for mechanoreception in *Drosophila* sensory cilia.
Susanne Bechstedt, Jörg T. Albert, David P. Kreil, Thomas Müller-Reichert, Martin C. Göpfert, Jonathon Howard

Contents

Supplementary Methods 2
GeneChip Experiment 2
Experimental Description and Analysis 2
Microarray validation 3
Quantitative real-time PCR 3
qRT-PCR - experimental procedure 3
In-situ hybridization 4

List of Figures

Supplementary Figure S1 Tissue Preparation . . . . . . . . . . . . . . . . . . . . . . . . 5
Supplementary Figure S2 GeneChip - qRT-PCR Comparison . . . . . . . . . . . . . . . 5
Supplementary Figure S3 EMAP in situ hybridization in zebrafish. . . . . . . . . . . . . . 6

List of Tables

Supplementary Table S1 Results of qRT-PCR and Comparison to Array M-value. . . . . 6
Supplementary Table S2 qRT-PCR Primer . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

References 7
Supplementary Methods

GeneChip Experiment

Experimental Description and Analysis

To measure the relative expression of genes in mechanoreceptor-rich tissue compared to mechanoreceptor-poor tissue, we extracted RNA separately from the two proximal segments and the one distal segment of 50 halteres. RNA was also extracted from thoracic ganglion to serve as a neuronal model tissue. Because the yield was low, the RNA was amplified.

Exploratory plots confirmed the need for inter-chip normalization, making a subtraction of the gcma heuristic estimate for optical instrument background unnecessary, which was hence skipped to reduce noise. Defaults were used for all other gcma parameters. Chip signal distributions showed no systemic differences and random fluctuations between chips could be equalized using quantile-quantile normalization [1]. After normalization, robust summaries of probe set signals were obtained for each gene transcript through an iterative weighted least squares fit of a linear probe level model [2] using the affyPLM library. The fitPLM function automatically identified unreliable chip areas and correspondingly downweighted outlier probes. No areas were dominated by outliers, indicating good hybridization quality for all chips.

The normalized data on log$_2$ scale were then fitted gene by gene with a linear model with a separate effect for each tissue class (Pedicel, Capitellum, Neuronal), using the lmFit function of the Bioconductor package limma [3]. The analysis also includes q-values as indicators of significance of contrasts after correction for multiple testing with strong control of the False Discovery Rate under arbitrary dependency structures [4]. In addition to the direct comparisons of P versus C and P versus N, a combined contrast of P versus (C and N) was computed. A contrast difference M is an uncalibrated log$_2$-ratio: M = 0 means no change; values M = 1 and M = -1 mean a two-fold increase or decrease in array signal. Being uncalibrated, this does not directly translate to sample mRNA fold-change [5], but this does not affect the accuracy of a screen for differences. For the statistical tests of significance, individual gene variances have been moderated using an Empirical Bayes approach that draws strength from transferring variance characteristics from the set of all genes to the test for each individual gene [3]. The prior was set self-consistently for the calculation of the log odds ratio B of differential expression: a log odds ratio of zero for a transcript means there is no evidence to indicate whether it is differentially expressed for the contrast considered; a value of B = 1 corresponds to a posterior probability of 73% for differential expression.

Table S3 compiles the M and B values for the three contrasts of interest (P vs. C, P vs. N and P vs. C & N). The Table is ranked by the log odds ratio for differential expression in the biologically most relevant contrast of pedicel versus capitellum $B_{PC}$. Considering that the significance estimates are more uncertain than the effect strengths M themselves, moderate significance thresholds combined with thresholds on effect strength have been shown to provide more robust results than stricter thresholds on significance alone [6]. We therefore threshold both $B_{PC} > 1$ and show only strongly up-regulated transcripts with $M_{PC} > 1$. To focus on targets that are not also enriched in neuronal tissue, we further require $B_{PN} < 1$. 


In order to assign a statistical significance to the relative gene-expression levels, we computed B-values for the pairs of tissues using the measured variances within and between samples (reference, and see Supplementary Table S3, legend). For $M_{PC} > 1$, $B_{PC}$ for a particular gene is the loge of the ratio of the probability that this gene is expressed more highly in pedicel relative to capitellum over the probability that it is is not. We estimated that 1% of the genes are differentially expressed between the pedicel and the capitellum. These are potential mechanotransduction genes. Because we are not interested in neuronal genes, we excluded from our list those genes for which $M_{PN} < 1$ and $B_{PN} < 1$ (i.e. we took only those genes that had two-fold higher expression in pedicel over thoracic ganglion and for which we had 73% probability of being overexpressed in pedicel). In addition we excluded genes that had $M_{PC} < 1$ (i.e. we kept genes that had two-fold higher expression in pedicel than capitellum). We ranked our list of genes by $B_{PC}$ to create a candidate gene list (Supplementary Data contains the top 625 genes).

**Microarray validation**

**Quantitative real-time PCR**

Real-time PCR is a common tool to validate microarray data. Here we show qRT-PCR evidence for 12 different genes using 15 primer pairs. We compared capitellum and pedicel and calculated qRT-PCR fold changes, where a positive fold change indicates enrichment in the pedicel (Table Supplementary Table S1). As displayed in Figure Supplementary Figure S2 and Table Supplementary Table S1 qRT-PCR confirms the enrichment of the candidate genes in the mechanoreceptor-rich tissue (pedicel). For DCX-EMAP we designed three different primers. M-values are uncalibrated and therefore only a measure of array signal. They do not directly translate to sample mRNA fold-change (Irizarry et al., 2006). Hence, we do not expect a slope of 1 in Figure Supplementary Figure S2, but a linear relationship. Indeed we obtain a slope of $1.93 \pm 0.22$. From the qRT-PCR experiments we conclude that the microarray experiment delivered reliable data for relative (not absolute) expression level comparison between the tissues of interest.

**qRT-PCR- experimental procedure**

Tissue collection, RNA extraction and reverse transcription was performed as described in the microarray procedure. The cDNA obtained by this procedure was then used with the 2x QPCR Mastermix + Sybr Green (Abgene) according to the manufacturer’s protocol. Gene specific primers used for the reactions are given in Table Supplementary Table S2. The MX3000P cycler (Stratagene) was used for performing the qRT-PCR reaction. Data analysis and presentation was performed with IgorPro 5 (Wavemetrics Inc.’www.wavemetrics.com).
**In-situ hybridization**

A question arising about DCX-EMAP is, whether its function in sensory systems is conserved. Especially its role as a potential Usher syndrome gene [7] and therefore its possible function in vertebrate hair cells is intriguing. We where therefore asking whether EMAP is expressed in sensory cells in vertebrates. As a model organism we chose *Danio rerio*. Therefore, a 320 bp fragment of zebrafish EMAP was cloned and in situ hybridizations with a DIG-labeled RNA probe of that sequence was performed. Total RNA was extracted from wild type zebrafish (72 hpf old) with the RNeasy kit (Qiagen). Reverse transcription was performed using Superscript II (Invitrogen) and gene specific primers: drEMAPinsituFW2 TCACTTTGTGTGGCGGTAGACG drEMAPinsituBW2 CCTTTAGCCCAGATGTAGATGTTCC according to manufacturer's protocol.

Digoxygenin (DIG)-labeled probe was prepared from the drEMAP fragment in pGEM-T easy by using the DIG-RNA labeling mix (Roche, Cat.-Nr.: 1277073) and SP6 promotor (GATTAGGTGACACTATAG) and T7 promotor (TAATACGACTCACTATAGGG) primer according to the manufacturer's protocols.

In situ hybridization with drEMAP riboprobes was performed as described [8]. The antisense probe clearly stains dots that represent the lateral line organ neuromasts as well as extensive structures in the zebrafish head (Figure Supplementary Figure S3). Neuromasts are clusters of hair cells that are used to detect hydrodynamic flow on the fish’s surface or vibrations and movement in water. We therefore conclude that zebrafish EMAP is expressed in neuromasts in zebrafish and its expression in sensory systems is conserved from invertebrates to vertebrates. In higher vertebrates (mammals like mice and men) up to six different EMAP homologs have been identified so far. We speculate that some of them might have function in sensory systems and some will have acquired new functions in different tissues.
Figure Supplementary Figure S1: Tissue Preparation
After dissection of the haltere and the thoracic ganglion from the fly’s thorax it was cut with a 20-gauge needle precisely at the joint between the pedicel and the capitellum. The tissues were separated, pooled and total RNA was extracted.

Figure Supplementary Figure S2: GeneChip - qRT-PCR Comparison
Plot of the $\log_2$ of the fold change as obtained by quantitative real-time PCR and the M-value obtained by the microarray experiment. The red line represents the line fit ($y = a + bx; a = 0$) with coefficient values $\pm$ one standard deviation: $b = 1.93 \pm 0.22$. Results for the three different primer pairs for DCX-EMAP are marked by a blue circle.
Figure Supplementary Figure S3: EMAP in situ hybridization in zebrafish.

**A**: Schematic drawing of a zebrafish larvae. Adapted from Whitfield et al. [9]. Neuromast (red dots) are distributed along the lateral line and the head region. **B**: Zebrafish in situ showing 72 hpf old larvae. The lateral line neuromasts are clearly stained. The arrow points towards neuromast P6 in A and B.

Table Supplementary Table S1: Results of qRT-PCR and Comparison to Array M-value.

| Gene ID  | M   | log₂(fold change) | STDEV |
|----------|-----|-------------------|-------|
| CG31284  | 4.2 | 1.5               | 0.67  |
| CG12164  | 6.3 | 1.7               | 0.47  |
| nompC    | 2.2 | 2.6               | n.d.  |
| CG13222  | 7.5 | 3.3               | 0.28  |
| BBS2     | 2.5 | 1.4               | n.d.  |
| CG14607  | 8.8 | 3.3               | n.d.  |
| CG9813   | 6.3 | 2.5               | 0.03  |
| Patsas   | 4   | 2.2               | n.d.  |
| CG14692  | 3.9 | 2.8               | 0.53  |
| CG15714a | 4.3 | 1.3               | 0.71  |
| CG15714b | 4.3 | 3.2               | 0.25  |
| CG4648   | 1.6 | 2.2               | n.d.  |
| CG13466a | 2.6 | 0.6               | 0.35  |
| CG13466b | 2.6 | 1.9               | 0.04  |
| CG13466c | 2.6 | 0.3               | 1.11  |
Plate Supplementary Table S2: qRT-PCR Primer

| Primer | Sequence          | Primer | Sequence          |
|--------|------------------|--------|------------------|
| CG13222fq | ACTTGGGGTATGGTGAGGGAGTTG | EMAPq-F1 | ACGTCAAGTGGCTGACCAACAAC |
| CG13222bq | TTGAGCAAGCCCTTTATGAGG | EMAPq-B1 | GGAATACACCTTCGACTCGTGAAAC |
| CG15714fq | TAGAGCGGTGCTATGATGGCAAC | EMAPq-F2 | CGCAAGAACAAGCTCATCGGACCC |
| CG15714bq | TTTTGGAGGCGTTTTTGTGGG | EMAPq-B2 | AAATGCTACCGTTTTGGGAGCC |
| CG9813fq | ATGTTTGCTCAATCCCCCTCGGC | EMAPq-F3 | TGCTATTGCGTGATGATCGAGGG |
| CG9813bq | TTGGCTAGGGGCGTGTATTGG | EMAPq-B3 | CGCATTAGGGGCGTTTTTGTGGG |
| CG14607fq | TTGGCTCTGGAGTGGCGGATTAG | CG14692f | TCTGCTCAAAGCGATTCCGTGG |
| CG14607bq | AACACCGGCTAGGGCAAAACAAG | CG14692b | CTCCGAAATGGTCTCGTGCAAAAC |
| CG31284fq | GCTATTCAACGGCTACCTTCCGAC | Patsasf | TCTGCTCAAAGCGATTCCGTGG |
| CG31284bq | TTTCTCGCTGGGATTCAGTGGC | Patsasb | CTCTGCTCAAAGCGATTCCGTGG |
| CG12164fq | GGTGCAGACTGGAAAGGGAAATG | CG15714f2 | AAAAAGGCTTCCAAAAAGTGGCC |
| CG12164bq | CGTGGTGCTTCTCAACTTCTGGG | CG15714b2 | CGTCCGAAGTCTCGTGCAAAAC |
| ompCfq | CACCGATGAAGAAAGGAGACCTG | rp49fq | AGATCGTGAAGAACGACCAACAG |
| ompCbq | ATCCCGAGCGCGATATCGGAC | rp49bq | CACCGAAGTCTTCTGAATCCG |

References

[1] B Bolstad, R Irizarry, M Astrand, and T Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185–93, Jan 2003.

[2] BM Bolstad. *Low Level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization*. Phd thesis, University of California, Berkeley, 2004.

[3] Gordon K Smyth. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, 3:Article3, Jan 2004.

[4] Y Benjamini and D Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*, 29:1165–1188, Jan 2001.

[5] Rafael A Irizarry, Zhijin Wu, and Harris A Jaffee. Comparison of affymetrix genechip expression measures. *Bioinformatics*, 22(7):789–94, Apr 2006.

[6] Leming Shi, Weida Tong, Hong Fang, Uwe Scherf, Jing Han, Raj K Puri, Felix W Frueh, Federico M Goodsaid, Lei Guo, Zhenqiang Su, Tao Han, James C Fuscoe, Z Alex Xu, Tucker A Patterson, Huixiao Hong, Qian Xie, Roger G Perkins, James J Chen, and Daniel A Casiano. Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics*, 6 Suppl 2:S12, Jul 2005.

[7] J. Kaplan, S. Gerber, D. Bonneau, J.M. Rozet, O. Delrieu, M.L. Briard, H. Dollfus, I. Ghazi, J.L. Dufier, and J. Frézal. A gene for Usher syndrome type I (USH1A) maps to chromosome 14q. *Genomics*, 14(4):979–87, 1992.

[8] A.C. Oates and R.K. Ho. Hairy/E(spl)-related (Her) genes are central components of the segmentation oscillator and display redundancy with the Delta/Notch signaling...
pathway in the formation of anterior segmental boundaries in the zebrafish. Development, 129(12):2929–46, 2002.

[9] T.T. Whitfield, M. Granato, F.J. van Eeden, U. Schach, M. Brand, M. Furutani-Seiki, P. Haffter, M. Hammerschmidt, C.P. Heisenberg, Y.J. Jiang, D.A. Kane, R.N. Kelsh, M.C. Mullins, J. Odenthal, and C. Nüsslein-Volhard. Mutations affecting development of the zebrafish inner ear and lateral line. Development, 123((null)):241–54, 1996.