Data in Brief

Analysis of cutaneous MRGPRD free nerve endings and C-LTMRs transcriptomes by RNA-sequencing

Ana Reynders a, Aziz Moqrich a,⁎

a Aix-Marseille-Université, CNRS, Institut de Biologie du Développement de Marseille, UMR 7288, case 907, 13288 Marseille Cedex 09, France

A R T I C L E   I N F O

Article history:
Received 15 May 2015
Accepted 24 May 2015
Available online 30 May 2015

Keywords:
Cutaneous sensory neurons
Free endings
C-LTMRs
RNA-seq
FACS

A B S T R A C T

The skin is the largest sensory organ that is densely innervated by highly specialized sensory neurons allowing the detection of a wide range of stimulations including light touch, temperature, itch and pain. Our knowledge of the sets of genes instructing the functional specialization of sensory neurons is just emerging. In a previous study, we have identified a new Gαi inhibitory interacting protein (GINIP) that marks two distinct subsets of skin-innervating sensory neurons conveying noxious and pleasant touch: the MRGPRD-expressing C-fibers specialized in noxious touch and the TH+/TAFA4+/V-GLUT3+ C-Low Threshold MechanoReceptors (C-LTMRs), part of neurons processing pleasant touch. In the recent study published by Reynders et al. (2015), we took advantage of GINIPmCherry mouse model in combination with Isolectin B4 (IB4) cell surface labeling and fluorescence activated cell sorting (FACS). We successfully purified MRGPRD+, C-LTMRs and a heterogeneous population of sensory neurons and subjected their RNA contents RNA-deep sequencing (RNA-seq). The subsequent RNA-seq experiment led to the generation of unique sets of data representative of pure transcriptome profiles of each subset. As a result of this pioneering approach, we established the combinatorial expression of the sets of genes that could dictate the functional specializations of MRGPRD+ neurons and C-LTMRs. Herein we provide details regarding the experimental design, the quality controls and statistical analysis of the data deposited at Gene Expression Omnibus under the accession number GSE64091.

© 2015 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64091.

2. Experimental design, materials and methods

2.1. Identification and FACS-sorting of MRGPRD+ neurons, C-LTMRs and double negative DRG neurons

The FACS-sorting strategy takes advantage of the GINIPmCherry mouse model in which the mCherry fluorescent protein is expressed at Ginip locus, allowing high fidelity identification of MRGPRD+ neurons through live mCherry fluorescence [4]. As shown in Fig. 1, combined analysis of GINIP and IB4 expression segregates dorsal root ganglia (DRG) neurons into 4 distinct subsets: GINIP+/IB4+ double positive (DP), GINIP+/IB4-, GINIP-/IB4+ simple positive (SP) and GINIP-/IB4- double negative (DN) neurons. We have previously shown that DP neurons corresponded to the cutaneous MRGPRD+ free nerve endings that convey noxious touch [1,4] and GINIP+/IB4- to the C-LTMRs, involved in pleasant touch [2,4,5].

http://dx.doi.org/10.1016/j.gdata.2015.05.022
2213-5960/© 2015 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Brieﬂy, GINIPmCherry DRG cell suspensions were labeled with Alexa647-conjugated IB4 and with Sytox Blue dye for exclusion of dying cells prior to FACS sorting. DRG cells suspensions from wild-type (WT) mice were used as a negative control. Given that the ﬁnal aim of this experiment is the extraction of high quality RNA from selected neuronal subsets, the ﬁne calibration of the FACS experiment is a critical step. The basal ﬂuorescence of DRG cell suspensions in the mCherry detection channel was determined by plotting the structure (SSC) parameter versus mCherry parameter of sample from WT animals (Fig. 2A). This allowed reliable and speciﬁc detection of mCherry-positive cells within the GINIPmCherry sample (Fig. 2B–C). Further analysis enabled the elimination of dead and auto-ﬂuorescent cells (Fig. 2D–E).

The FACS-gating strategy was as follows: we ﬁrst deﬁned (A) mCherry− and (B) mCherry+–Sytox Blue− living cell gates. SP and DN subsets were sorted from gate A and DP and C-LTMRs were sorted from gate B (Fig. 2E–F). Two independent FACS sorting experiments were performed to obtain two independent biological replicates.
2.2. RNA extraction and purity control

Cells were sorted directly in RLT lysis buffer from RNeasy Micro Kit (Quiagen) and total RNA was extracted according to manufacturer’s instructions. Prior to RNA-seq, the purity of the 4 sorted samples and the eventual cross contaminations were evaluated by RT-PCR. The expression patterns of Ginip, mCherry, Mrgprd, Taf4, Trpv1 and TrkA transcripts were analyzed and found accurate (data not shown). However, SP cells were found to be highly contaminated by CD31-expressing endothelial cells that also bind IB4 and thus were excluded from the RNA-seq experiment (data not shown). The quality of the RNA obtained from each sorted sample was determined with respect to RNA profiles generated by the Agilent Bioanalyzer 2000 (Agilent Inc.).

2.3. RNA-Sequencing and quality control

For each sample, 50 ng of high quality total RNAs were amplified using the Amino Allyl Message Amp II amplification kit (Life Technologies) and 100 ng of resulting amplified RNAs were used for libraries building with TrueSeq Sample preparation (low throughput protocol) kit from Illumina. Libraries were sequenced in 100-bp paired-end cycles using the Illumina HiSeq2000 system with SBS technology. Image analysis and base calling were performed using the HiSeq Control Software and Real-Time Analysis component provided by Illumina. Quality of the sequencing was assessed using FastQC and the Illumina software SAV (Sequence Analysis Viewer); a representative example is shown in Fig. 2.

For each sample, the median and the mean distribution of the quality scores was above 30 and data acquisition was not affected by base loss (Fig. 3A–B and data not shown). Altogether, these criteria attest for the good quality of the RNA-seq experiment.

2.4. Data analysis

Base calling yielded between 40 and 50 million reads (Table 1). RNA-seq reads were aligned to Mus musculus genome (UCSC mm10) using the eland_rna module from CASAVA (refFlat.txt file downloaded from UCSC on October 10, 2012). As CASAVA software is not designed for pair-end analysis, only the first pair was used to generate the counts. Reads mapping to contaminants including 28S, 18S and 5S ribosomal RNAs, mitochondrial chromosome, PhiX genome (Illumina control) and Illumina adaptors as well as reads mapping to multiple splicing positions and with no match, were removed. Gene counting was carried with the counting module from CASAVA. Table 1 gives an overview of the number of counts generated and shows those that mapped to exons that were used for the creation of the Raw Data file. Genes generating less than 10 counts were excluded from the subsequent statistical analysis.

The first normalization method used was the Reads Per Kilobase per Million of mapped reads (RPKM) [6]. Analysis of RPKM values generated for genes expected to be enriched in each defined subset, showed substantial and accurate enrichment in Mrgprd and Trpv3 in DP subset Taf4 and Th in C-LTMRs and Cgrp and Sp in DP subset, providing a proof of concept of the RNA-seq data presented in our study (Fig. 4A).

Because our aim was to determine the sets of differentially expressed genes (DEGs) in one subset as compared to the two others and that the RPKM normalization method was shown to be inappropriate for such an analysis [3], we chose the Trimmed Mean of M-values (TMM) normalization factors provided by edgeR package (p < 0.01, fold change > 2). Smear-plot representation of pairwise comparison of TMM-normalized datasets showed that the log fold change (FC) was centered on zero, attesting that the libraries were correctly normalized (see Fig. 4B). 486 DEGs were declared enriched in DP as compared to the other subsets, 549 in C-LTMRs and 2916 in DN.

2.5. Data validation

Only DEGs that generated at least 1000 reads in both biological replicates were kept for downstream data validation. As a result, data validation was carried on 156, 184 and 784 DP-, C-LTMRs- and DN-DEG, respectively. For this, we performed an extensive in situ hybridization (ISH) screen on DRG cryosections followed by either a) GINIP or IB4 immunostaining for DP and C-LTMRs datasets or b) GINIP and TrkA immunostaining for DN dataset.

Table 1

|                | DP Sort no. 1 | DP Sort no. 2 | C-LTMRs Sort no. 1 | C-LTMRs Sort no. 2 | DN Sort no. 1 | DN Sort no. 2 |
|----------------|---------------|---------------|-------------------|-------------------|---------------|---------------|
| Total clusters | 55,860,206    | 50,236,510    | 45,620,214        | 43,638,958        | 51,801,750    | 43,045,079    |
| Usable         | 42,253,049    | 38,469,481    | 34,361,200        | 32,425,037        | 39,156,811    | 32,418,056    |
| Splice usable  | 9,459,613     | 8,67,128      | 7,992,789         | 7,956,586         | 9,451,395     | 7,789,216     |
| Genome usable  | 32,793,436    | 29,832,353    | 26,607,223        | 25,223,801        | 29,971,384    | 25,234,178    |
| Exon counts    | 37,491,412    | 34,400,856    | 30,327,985        | 28,750,186        | 34,988,317    | 29,045,318    |

Fig. 3. Quality control of RNA-seq experiment. A — Shows the quality scores (QS) distribution across all bases. The red line represents the median QS. B — Shows the QS distribution across all sequences. The peak depicts the mean quality per sequence (Phred Score). Only one representative sample is illustrated herein.
immunostaining for DN. The criteria for validation were the following: a) a gene declared enriched in DP should be detected by ISH as highly expressed in GINIP+/IB4+ neurons but excluded from GINIP+/IB4− neurons; a) a gene declared enriched in C-LTMRs should exhibit the opposite expression pattern and c) a gene declared enriched in DN should be excluded from GINIP+ neurons.

We thus monitored the expression of 48 DP-, 68-C-LTMRs and 13-DN-enriched transcripts thereby confirming the validity of RNA-seq data generated in this study (Fig. 5A–B and data not shown). Moreover, we generated two open-access ISH libraries (DP ISH library: http://www.ibdm.univ-mrs.fr/equipes/reynard/DP-ISHlibrary.pptx and C-LTMRs ISH library: http://www.ibdm.univ-mrs.fr/equipes/reynard/CLTMRs-ISHlibrary.pptx) in which we provide the digital prevalence and the expression patterns of these genes in DRG neurons.

3. Discussion

We described here a unique dataset encompassing the combinatorial gene expression enriched in MRGPRD+, C-LTMRs and DN DRG sensory neurons. We highlight here that our published RNA-seq data are of good quality and concordant with our validation strategy. These data provide a wealth of information regarding the presence and the prevalence of transcripts in MRGPRD+, C-LTMRs and DN sensory neurons, they highlight the differential gene expression in two functionally opposite classes of cutaneous afferents and they allow open-access monitoring the cellular distribution of over 100s of genes [7].

Acknowledgments

We are grateful to the members of the laboratory for scientific discussions, to M. Barad, A. Zouine and M. Malissen from CIML flow cytometry facility for FACS assistance. We thank S. Alonso for comments on the manuscript, M. Deage for the pilot experiment on Tafa4 DRG neurons as well as IBDM imaging and animal facilities for assistance. This work has been funded by ERC-Starting grant paineurons 260435 to A.M.

References

[1] D.J. Cavanaugh, H. Lee, L. Lo, S.D. Shields, M.J. Zylka, A.I. Basbaum, D.J. Anderson, Distinct subsets of unmyelinated primary sensory fibers mediate behavioral responses to noxious thermal and mechanical stimuli. Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 9075–9080.
[2] M.C. Delfini, A. Mantilleri, S. Gaillard, J. Hao, A. Reynders, P. Malapert, S. Alonso, A. Francois, C. Barrere, R. Seal, M. Landry, A. Eschallier, A. Alloui, E. Bourinet, P. Delmas, Y. Le Feuvre, A. Moqrich, TAF4A, a chemokine-like protein, modulates injury-induced mechanical and chemical pain hypersensitivity in mice. Cell Rep. 5 (2013) 378–388.
[3] M.A. Dillies, A. Rau, J. Aubert, C. Henniquet-Antier, M. Jeanmougin, N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, C. Guernic, B. Jagla, L. Jouveau, D. Lalo, C. Le Gall, B. Schaeffer, S. Le Crom, M. Gueld, J. Jaffrezic, A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief. Bioinform. 14 (2013) 671–683.
[4] S. Gaillard, L. Lo Re, A. Mantilleri, R. Hegg, L. Urien, P. Malapert, S. Alonso, M. Deage, C. Kambrun, M. Landry, S.A. Low, A. Alloui, L. Lambolez, G. Scherrer, Y. Le Feuvre, E. Bourinet, A. Moqrich, GINIP, a Galphai-interacting protein, functions as a key modulator of peripheral GABAB receptor-mediated analgesia. Neuron 84 (2014) 123–136.
[5] L. Li, M. Rutlin, V.E. Abraira, C. Cassidy, L. Kus, S. Gong, M.P. Jankowski, W. Luo, N. Heintz, H.R. Koerber, C.J. Woodbury, D.D. Ginty, The functional organization of cutaneous low-threshold mechanoreceptive neurons. Cell 147 (2011) 1615–1627.
[6] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5 (2008) 621–628.
[7] A. Reynders, A. Mantilleri, P. Malapert, S. Rialle, S. Nidelet, S. Laffray, E. Bourinet, A. Moqrich, Transcriptional Profiling of Cutaneous MRGPRD Free Nerve Endings and C-LTMRs. Cell Rep, 2015.