Archival Report

Biological Psychiatry

A Comprehensive Analysis of Cell Type–Specific Nuclear RNA From Neurons and Glia of the Brain

Adarsh S. Reddy, David O'Brien, Nilambari Pisat, Claire T. Weichselbaum, Kristina Sakers, Miriam Lisci, Jasbir S. Dalal, and Joseph D. Dougherty

ABSTRACT

BACKGROUND: Studies in psychiatric genetics have identified >100 loci associated with disease risk, yet many of these loci are distant from protein coding genes. Recent characterization of the transcriptional landscape of cell lines and whole tissues has suggested widespread transcription in both coding and noncoding regions of the genome, including differential expression from loci that produce regulatory noncoding RNAs that function within the nucleus; however, the nuclear transcriptome of specific cell types in the brain has not been previously investigated.

METHODS: We defined the nuclear transcriptional landscape of the three major cellular divisions of the nervous system using flow sorting of genetically labeled nuclei from bacTRAP mouse lines. Next, we characterized the unique expression of coding, noncoding, and intergenic RNAs in the mature mouse brain with RNA-Seq and validation with independent methods.

RESULTS: We found diverse expression across the cell types of all classes of RNAs, including long noncoding RNAs, several of which were confirmed as highly enriched in the nuclei of specific cell types using anatomic methods. We also discovered several examples of cell type–specific expression of tandem gene fusions, and we report the first cell type–specific expression of circular RNAs—a neuron-specific and nuclear-enriched RNA arising from the gene *Hnrnpu*.

CONCLUSIONS: These data provide an important resource for studies evaluating the function of various noncoding RNAs in the brain, including noncoding RNAs that may play a role in psychiatric disease.

Keywords: Circular RNA, *Hnrnpu*, lincRNA, *Mirg1*, ncRNA, Nuclear

http://dx.doi.org/10.1016/j.biopsych.2016.02.021

The central nervous system (CNS) has remarkable cellular diversity with hundreds of distinct cell types, each with unique morphology, connectivity, and function. However, all this diversity must arise from essentially identical copies of the genome in every cell, and more recent work has highlighted the substantial range of epigenetic and transcriptional variation across cultured cells and whole tissues in the body (1). Likewise, gene expression plays a key role in defining the cellular diversity of the brain, and previous work demonstrated significant differences in transcript abundance for protein-coding genes (2–4). It is now recognized that transcription also occurs beyond protein-coding genes (5). Furthermore, the nucleus has a clearly distinct RNA profile relative to the cytoplasm, suggesting there may be many additional RNA molecules with roles particularly within this subcellular compartment (5,6). However, previous work focused largely on cell lines, and the CNS remained relatively unstudied.

In the last decade, there has also been a substantial expansion of the known roles for noncoding RNAs (ncRNAs), and many new species have been defined, including the long intergenic noncoding RNAs (lincRNAs) (7–10). Work in embryonic stem cells suggests that many lincRNAs localize to the nucleus and regulate cell fate decisions: 38% bind known chromatin remodeling complexes (11), and experimental overexpression or suppression of ~100 lincRNAs caused changes in gene expression and cell fate in vitro (12). Some lincRNAs have been identified as important for neural stem cell differentiation in primary cultures (13). In addition, more recent studies have highlighted the prevalence of a new class of circular RNAs (ciRNAs) (14,15), at least one of which has a role in microRNA regulation (16,17).

The major cell types of the nervous system—neurons, astrocytes, and oligodendrocytes—have distinct messenger RNA profiles in vivo (2,3). These profiles have subserved a wide range of informative secondary analyses (18–20). However, little is known about the nuclear transcriptome of these cells. Also, given a nuclear role for some lincRNAs, and generally low expression levels, analysis of nuclear transcriptomes might provide better assessment of ncRNAs. In this study, we describe a powerful method to study cell type–specific nuclear RNA in vivo. We demonstrate a diversity of nuclear transcripts, replicating previous analyses of

 ISSN: 0006-3223

protein-coding genes and greatly extending them to include a thorough description of nuclear ncRNA, ciRNA, and other RNA species. We provide an overview of the data, describe differences in protein coding and lincRNA transcription across these three cell types, and identify novel tandem gene fusions and ciRNAs detectable in each.

METHODS AND MATERIALS

Mice

All procedures in mice were approved by the Washington University Animal Studies Committee. Lines included B6.FVB-Tg(Snap25-EGFP/Rpl10a)^{JD362Htz}/J, B6.FVB-Tg(Aldh1L1-EGFP/Rpl10a)^{JD130Htz}/J, and B6.FVB-Tg(Cnp-EGFP/Rpl10a)^{JD368Htz}/J.

Nuclear Capture and RNA Purification

Nuclear sorting was conducted as described (21). We processed four animals from each line (6–10 weeks, two of each sex) as independent replicates. Gates were set for green fluorescence protein using nuclei from wild-type mice. Nuclei were sorted with MoFlo (Beckman Coulter, Brea, CA) directly into TRIzol LS (Thermo Fisher Scientific, Waltham, MA) for purification, followed by RNeasy MinElute (Qiagen, Valencia, CA) columns. Quality and quantity were confirmed with Bioanalyzer (Agilent Technologies, Santa Clara, CA) and quantitative polymerase chain reaction (PCR).

Library Preparation

We generated double-stranded complementary DNA using Ovation RNA-Seq System V2 (NuGEN Technologies, Inc., San Carlos, CA), starting from 5 ng of RNA. Standard sequencing libraries (Illumina, Inc., San Diego, CA) were generated from 1–2 μg of cDNA, sheared to ~200 bp.

Sequencing Analysis

Reads were trimmed with Trimmomatic (22). Ribosomal RNA reads were removed by mapping to rRNA sequences using Bowtie 2 (23). Remaining reads were mapped to Ensembl version 75 using STAR (24), counted with BEDTools (25), and tested by edgeR (26).

ciRNA

A custom genome was created by extracting every gene ± 1 kb from repeat masked Ensembl version 75 mouse genome. Each gene sequence was duplicated in tandem and treated as a separate chromosome. Cleaned reads were mapped using STAR with the following additional parameters: mapping 90% of read length, canonical splice junctions only, and no maximum intron size. An additional filtering removed reads that mapped twice to a "chromosome" (unspliced), lacking a splice junction spanning the duplicate gene copies, or where splicing was linear.

Gene Ontologies

The messenger RNAs (mRNAs) with false discovery rate < 0.05 and fold change >10 were ranked by counts per million (CPM), and the 25 most highly expressed from each cell type were analyzed with BiNGO v2.44 in Cytoscape v2.8 for biological processes overrepresented at $p < 1e\text{-}4$ with Benjamini-Hochberg correction. Results using longer lists were very similar.

Quantitative PCR Validation

Reverse transcription was performed using qScript cDNA SuperMix (Quanta BioSciences, Gaithersburg, MD). cDNA was amplified with a ViiA 7 Real-Time PCR system (Thermo Fisher Scientific) using a comparative CT protocol and Qantas 2X Master Mix (Quanta BioSciences). For ciRNA, DNase1-treated total RNA was incubated with or without RNaseR (Epicentre, Madison, WI) and purified by phenol-chloroform extraction. Reverse transcription was performed using either random hexamers or oligo(dT) and Superscript III (Thermo Fisher Scientific).

In Situ Hybridization

Probe templates were amplified by PCR from cDNA. Complementary RNA probes were generated using digoxigenin RNA labeling mix (Roche, Indianapolis, IN). Brains were cryosectioned after 4% paraformaldehyde perfusion, postfixed in 4% paraformaldehyde, and prehybridized for 1 hour. Probes were hybridized overnight at 65°C and washed for stringency, then detected with sheep anti-digoxigenin and developed using Cyanine 3 Tyramide Signal Amplification kit (PerkinElmer, Waltham, MA), followed by anti-Gfp (GFP-1020; Aves Labs, Inc., Tigard, OR) and mouse anti-CNP1 (MAB326; EMD Millipore, Billerica, MA), detected with appropriate Alexa Fluor dye–labeled secondaries. Slides were imaged using an UltraVIEW VoX spinning disk confocal microscope (PerkinElmer).

RESULTS

Isolation of Nuclear RNA From Neurons, Astrocytes, and Oligodendrocytes

We previously generated and validated bacTRAP lines expressing Gfp-ribosomal fusion proteins to capture mRNA from neurons (Snap25-EGFP/Rpl10a), astrocytes (Aldh1L1-EGFP/Rpl10a), and oligodendrocytes (Cnp-EGFP/Rpl10a) in vivo (2,27), as outlined in Figure 1A. Because ribosomes are assembled in the nucleolus, bacTRAP lines have a non-diffusible Gfp signal within the nucleus (Figure 1B, C), facilitating its purification with fluorescent-activated cell sorting (21). In the present study, we adapted these bacTRAP lines to study nuclear-localized coding (mRNA) and noncoding (ncRNA) transcripts in vivo in a cell type–specific manner.

We first tested this approach with astrocytes. RNA Bioanalyzer confirmed that RNA from the sorted nuclei was intact, and quantitative PCR confirmed the presence of both ncRNA and mRNA (Figure 1D, E). As expected, ncRNAs with known roles in the nucleus were clearly enriched in nuclei compared with total brain homogenate (Figure 1D), and mRNAs known to be translated in astrocytes were already enriched in their sorted nuclei (Figure 1E). We then repeated this procedure with four animals from each line and confirmed sort quality with fluorescence microscopy (not shown), RNA integrity with RNA Bioanalyzer, and specificity via quantitative PCR for mRNAs known to be enriched in each cell type (Figure S1 in Supplement 1). Finally, we generated RNA-Seq libraries from Gfp-positive nuclei, presorted nuclei, and cytoplasmic controls.