F1000Research

DATA ARTICLE

Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs [v1; ref status: indexed, http://f1000r.es/w2]

Kyle S Hubbard, Ian M Gut, Megan E Lyman, Patrick M McNutt

United States Army, Medical Research Institute of Chemical Defense, MD, 21010, USA

Abstract

Using paired-end RNA sequencing, we have quantified the deep transcriptional changes that occur during differentiation of murine embryonic stem cells into a highly enriched population of glutamatergic cortical neurons. These data provide a detailed and nuanced account of longitudinal changes in the transcriptome during neurogenesis and neuronal maturation, starting from mouse embryonic stem cells and progressing through neuroepithelial stem cell induction, radial glial cell formation, neurogenesis, neuronal maturation and cortical patterning. Understanding the transcriptional mechanisms underlying the differentiation of stem cells into mature, glutamatergic neurons of cortical identity has myriad applications, including the elucidation of mechanisms of cortical patterning; identification of neurogenic processes; modeling of disease states; detailing of the host cell response to neurotoxic stimuli; and determination of potential therapeutic targets. In future work we anticipate correlating changes in longitudinal gene expression to other cell parameters, including neuronal function as well as characterizations of the proteome and metabolome. In this data article, we describe the methods used to produce the data and present the raw sequence read data in FASTQ files, sequencing run statistics and a summary flatfile of raw counts for 22,164 genes across 31 samples, representing 3-5 biological replicates at each timepoint. We propose that this data will be a valuable contribution to diverse research efforts in bioinformatics, stem cell research and developmental neuroscience studies.

Article Status Summary

Referee Responses

Referees	1	2
v1 published 07 Feb 2013	✓ report 1	✓ report

1   **Cliff Ragsdale**, University of Chicago USA, **Caroline Albertin**, University of Chicago USA

2   **Joyce van de Leemput**, Neural Stem Cell Institute USA

Latest Comments

No Comments Yet

Corresponding author: Patrick M McNutt (patrick.mcnutt@us.army.mil)

Competing Interests: No competing interests were disclosed.

Objectives

Transcriptional profiling by RNA sequencing (RNAseq) enables the sensitive and accurate characterization of the transcriptome[1–4]. Following enumeration of reads, various methods are available to normalize counts, estimate probability distributions and identify differential gene expression between biological conditions[3,5,6]. RNAseq has the added advantages of being extremely high-throughput and relatively inexpensive, with a high signal-to-noise ratio and a dynamic range encompassing 4–5 orders of magnitude. The combination of throughput and sensitivity enables the detection of rare transcripts from nanograms of RNA.

We have described a method to produce large quantities of highly enriched, electrically active glutamatergic neurons (ESNs) from suspension-adapted mouse embryonic stem cells (ESCs)[7,8]. This technique is a modification of the 4/4 method, and is based on the spatiotemporal changes in morphogens that occur during neural induction and patterning[9–11]. In this method, neuroepithelial stem cells (NESCs) are derived from ESCs by the withdrawal of leukemia inhibitory factor (LIF), then induced to undergo neurogenesis and neural patterning by supplementation with all-*trans* retinoic acid (RA) in the presence of fetal bovine serum[12]. We have modified the 4/4 method to include feeder cell-free, suspension culture of ESNs; differentiation under rotary conditions to normalize the intra-aggregate environment; and neuronal induction and neural induction and patterning using 6 μM RA. These refinements have resulted in a facile and economical method to generate large quantities of highly enriched glutamatergic neurons[13]. Using immunocytochemistry, we have shown that ESNs are composed mostly of glutamatergic neurons (~95%), with about 5% GABAergic neurons, and no evidence of dopaminergic, serotonergic, cholinergic or glycinergic subtypes[7]. Expression profiling using RNA sequencing has confirmed that derived cultures are primarily glutamatergic, and identified the abundant expression of a wide range of cortical markers, including reelin, Pax6, Otx1, Ctip2 and Cux1/2[13–19].

Although aspects of corticogenesis have been replicated *in vitro* by the directed differentiation of pluripotent stem cells, spatiotemporal changes in gene expression responsible for regionalization and apical-to-basal patterning of the cerebral cortex are not well understood[20–22]. The embryological origin of the cerebral cortex is the telencephalon, which is the most anterior structure of the mammalian neural tube[23]. Cortical glutamatergic neurons are generated by proliferation and laminar patterning of the dorsal telencephalon, whereas inhibitory GABAergic neurons derive from the ventral telencephalon and migrate tangentially into cortical layers[24,25]. Our ability to derive predominantly glutamatergic neurons that express markers of the telencephalon and all six cortical layers is consistent with an *in vitro* model of the developing telencephalon and cortical layer formation[23]. We propose to apply this model to identify the temporal changes in gene and isoform expression associated with neural patterning and corticogenesis. If successful, we intend to use ESNs to functionally interrogate the roles of individual genes in executing the transcriptional programs underlying the formation of the cerebral cortex.

To evaluate the transcriptional changes associated with differentiation of cortical glutamatergic neurons, we conducted a longitudinal expression profile of the deep transcriptome during neurogenesis from days *in vitro* (DIV) -8 to 28, where DIV 0 corresponds to the end of differentiation. High quality RNA was isolated from ESCs (DIV -8; n=4); neuroepithelial stem cells (DIV -4; n=3); radial glia (DIV 0; n=3); developmental stage (DS) I/II neurons (DIV 1; n=4); DS III/IV neurons (DIV 7; n=5); and maturing DS IV/V neurons at DIV 16 (n=4); DIV 21 (n=4); and DIV 28 (n=4). The summary data, including quality scores, are presented in Data File 1, and the raw transcript read counts for each biological replicate are presented in Data File 2. The FASTQ files generated for each biological replicate are available at the Sequence Read Archive (SRA), a freely accessible database provided by NCBI (http://www.ncbi.nlm.nih.gov/sra/), under the accession number PRJNA185305.

In addition to providing the basis for a study to characterize transcriptional processes involved in corticogenesis and neuronal maturation, we can foresee several other applications for this data. First, the identification of genes and isoforms that exhibit differential expression in synchronization with known markers is expected to provide novel insight into molecular mechanisms of neurogenesis and neural patterning. Second, a few algorithms are available for statistical determination of differential gene expression within longitudinal data sets. The depth and quality of these data suggests it may be well-suited for the development and validation of such methods. Third, we envision this dataset facilitating inter-specific comparisons of transcription during neurogenesis and corticogenesis, providing insight into transcriptional mechanisms common to and unique in the evolution of the mammalian cerebral cortex. For these reasons, we are making this data publically available for research efforts in bioinformatics, stem cell research and developmental neuroscience studies.

Materials and methods

Suspension adaptation and continuous culture of mouse ESCs

ESCs were adapted to feeder cell-free, suspension culture and maintained as previously described[7,13]. In brief, aliquots of R1 ESCs (ATCC, Manassas, VA) were thawed and maintained at 37°C at 5% CO_2 in 90% relative humidity in 10 cm bacterial plates in ESM (Knockout DMEM supplemented with 100 μM β-mercaptoethanol, 0.1mM nonessential amino acids, 2.0 mM L-glutamine, 5000 units/mL penicillin/streptomycin, 1000 units/mL recombinant mouse LIF [all Life Technologies, Carlsbad, CA] and 15% ES qualified fetal calf serum [ATCC, Manassas, VA])[26]. Cells were passaged once aggregates first became clearly visible to the naked eye (4–8 days) and maintained for at least 5 passages prior to differentiation. For passaging, aggregates were allowed to settle by gravity, washed once with 0.5 mL PBS and dissociated for 3 min at 37°C with 0.5 ml of TrypLE Express (Life Technologies). Dissociation was terminated by addition of 0.5 mL ESM followed by gently trituration with a P1000 pipette to achieve a single-cell suspension. ESNs were counted manually using a hemocytometer and ~1.5x10⁶ mESCs were transferred to 10 mL ESM in a fresh 10 cm bacterial dish.

Neuronal differentiation

ESCs were differentiated into neurons between 5–30 passages after adaptation to suspension culture. A modified 4/4 protocol was used for neuron differentiation[7,27]. Following routine sub-passaging,

3.5x10⁶ mouse ESCs were transferred to 30 mL differentiation medium (ESM modified to contain 10% ESC-qualified fetal calf serum and without LIF) in a 10 cm ultra-low attachment suspension culture dish (Corning, Lowell, MA). This was designated as DIV -8. Differentiating aggregates were maintained on a rotary shaker at 45 rpm at 37°C, 5% CO2 and 90% relative humidity. Complete media changes were conducted every 48 h, and media was supplemented with 6 μM retinoic acid (Sigma-Aldrich) at DIV -4 and DIV -2.

On DIV 0, aggregates were dissociated with TrypLE Express for 5 min at 37°C. Trypsinization was halted with 5 mL of 1% soybean trypsin inhibitor (Life Technologies), the aggregates were gently dissociated by trituration with a 10 mL pipet, and the cell suspension was filtered through a 40 μm cell strainer (Thermo Scientific, Waltham, MA). Cells were pelleted for 5 min at 300 x g, washed in N2 medium (Neurobasal-A medium with 1 x N2 vitamins, 2 mM glutamine and antibiotics [Life Technologies]) and counted manually using a hemocytometer. Neuronal progenitors were plated at 1.5x10⁶ cells/cm² in poly-D coated dishes. Complete washes with N2 medium were conducted at 4 h and 24 h to remove residual serum and non-adherent cells. At DIV 2, N2 was replaced with B27 medium (Neurobasal-A supplemented with antibiotics, 2 mM glutamine and 1 x B27 vitamins [Life Technologies]). Subsequently, ESNs underwent full medium changes with B27 on DIV 4, 8 and 12. On DIV 8 the media was supplemented with 30 μM 5-fluoro-2'-deoxyuridine and 70 μM Uridine (Sigma-Aldrich) to select against any remaining glia. Following DIV 12, ESNs were left undisturbed until RNA harvest. Cultures remained healthy and viable until at least DIV 28 under these conditions.

Expression profiling

Neurons were harvested from 6 cm dishes at DIV -8, -4, 0, 1, 7, 16, 21 and 28 (n=3 to 5) and RNA was isolated by QIAcube (Qiagen, Valencia, CA) using the RNeasy mini kit protocol (Qiagen) and submitted to Expression Analysis, Inc. (Durham, NC) for library preparation and sequencing. Typical recovery was ~3 μg per 6 cm dish, with an RNA integrity number exceeding 8 and a 260:280 of 1.8 – 2.2. PolyA+ RNA was purified from 500 ng of total RNA using polyA selection, chemically fragmented and reverse transcribed using random hexamers. This was followed by second strand synthesis and end repair. Libraries were prepared for paired-end sequencing using the TruSeq™ RNA sample prep kits (Illumina, San Diego, CA) per manufacturer's instructions, and library size and integrity were determined using the Agilent Bioanalyzer 2100 (Santa Clara, CA). Libraries were bound to flow cell surfaces using the Standard Cluster Generation Kit v5 (Illumina). Flow cells were transferred to the Illumina HiSeq 2000 and run using TruSeq SBS Kits (Illumina). Paired-end sequencing data were generated over 2x50 sequencing cycles. Sequence information and quality scores are available at SRA in FASTQ format, with the forward and reverse reads appended with an "F" or "R" (e.g., DIV7.1F and DIV7.1R). Raw sequence reads were aligned using the University of California, Santa Cruz's mouse knownGENE track, and transcriptome abundance estimation was performed on completed alignments using RSEM (RNA-Seq by Expectation Maximization)[28]. A summary flatfile containing the total raw reads for each gene in each biological replicate is presented in Data File 2. Genes were excluded if no reads occurred in any biological sample.

Data File 1: summary data

http://dx.doi.org/10.6084/m9.figshare.154969

Data File 2: raw transcript read counts

http://dx.doi.org/10.6084/m9.figshare.154970

References

1. Bullard JH, Purdom E, Hansen KD, *et al.*: **Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments.** *BMC Bioinformatics.* 2010; **11**: 94.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Griffith M, Griffith OL, Mwenifumbo J, *et al.*: **Alternative expression analysis by RNA sequencing.** *Nat Methods.* 2010; **7**(10): 843–847.
    PubMed Abstract | **Publisher Full Text**

3. Marioni JC, Mason CE, Mane SM, *et al.*: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.** *Genome Res.* 2008; **18**(9): 1509–1517.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Hebenstreit D, Fang M, Gu M, *et al.*: **RNA sequencing reveals two major classes of gene expression levels in metazoan cells.** *Mol Syst Biol.* 2011; **7**: 497.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Sultan M, Schulz MH, Richard H, *et al.*: **A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome.** *Science.* 2008; **321**(5891): 956–960.
    PubMed Abstract | **Publisher Full Text**

6. Robinson MD, Smyth GK: **Moderated statistical tests for assessing differences in tag abundance.** *Bioinformatics.* 2007; **23**(21): 2881–2887.
    PubMed Abstract | **Publisher Full Text**

7. McNutt P, Celver J, Hamilton T, *et al.*: **Embryonic stem cell-derived neurons are a novel, highly sensitive tissue culture platform for botulinum research.** *Biochem Biophys Res Commun.* 2011; **405**(1): 85–90.
    PubMed Abstract | **Publisher Full Text**

8. Hubbard KS, Gut IM, Lyman ME, *et al.*: **High yield derivation of enriched glutamatergic neurons from suspension-cultured mouse ESCs for neurotoxicology research.** *BMC Neurosci.* 2012; **13**: 127.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Bain G, Kitchens D, Yao M, *et al.*: **Embryonic stem cells express neuronal properties *in vitro*.** *Dev Biol.* 1995; **168**(2): 342–357.
    PubMed Abstract | **Publisher Full Text**

10. Bibel M, Richter J, Lacroix E: **Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells.** *Nat Protoc.* 2007; **2**(5): 1034–1043.
    PubMed Abstract | **Publisher Full Text**

11. Glaser T, Brustle O: **Retinoic acid induction of ES-cell-derived neurons: the radial glia connection.** *Trends Neurosci.* 2005; **28**(8): 397–400.
    PubMed Abstract | **Publisher Full Text**

12. Levine AJ, Brivanlou AH: **Proposal of a model of mammalian neural induction.** *Dev Biol.* 2007; **308**(2): 247–256.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Hubbard KS, Gut IM, Lyman ME, *et al.*: **High yield derivation of enriched glutamatergic neurons from suspension-cultured mouse ESCs for neurotoxicology research.** *BMC Neurosci.* 2012; **13**: 127.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Germain N, Banda E, Grabel L: **Embryonic stem cell neurogenesis and neural specification.** *J Cell Biochem.* 2010; **111**(3): 535–542.
    PubMed Abstract | **Publisher Full Text**

15. Weimann JM, Zhang YA, Levin ME, *et al.*: **Cortical neurons require Otx1 for the refinement of exuberant axonal projections to subcortical targets**. *Neuron.* 1999; **24**(4): 819–831.
    PubMed Abstract | **Publisher Full Text**

16. Chen B, Schaevitz LR, McConnell SK: **Fezl regulates the differentiation and axon targeting of layer 5 subcortical projection neurons in cerebral cortex.** *Proc Natl Acad Sci U S A.* 2005; **102**(47): 17184–17189.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Meyer G, Goffinet AM: **Prenatal development of reelin-immunoreactive neurons in the human neocortex.** *J Comp Neurol.* 1998; **397**(1): 29–40.
    PubMed Abstract | **Publisher Full Text**

18. Georgala PA, Manuel M, Price DJ: **The generation of superficial cortical layers is regulated by levels of the transcription factor Pax6.** *Cereb Cortex.* 2011; **21**(1): 81–94.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Li N, Zhao CT, Wang Y, *et al.*: **The transcription factor Cux1 regulates dendritic morphology of cortical pyramidal neurons.** *PLoS One.* 2010; **5**(5): e10596.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Gaspard N, Bouschet T, Hourez R, *et al.*: **An intrinsic mechanism of corticogenesis from embryonic stem cells.** *Nature.* 2008; **455**(7211): 351–357.
    PubMed Abstract | **Publisher Full Text**

21. Hansen DV, Rubenstein JL, Kriegstein AR: **Deriving excitatory neurons of the neocortex from pluripotent stem cells.** *Neuron.* 2011; **70**(4): 645–660.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Jing Y, Machon O, Hampl A, *et al.*: ***In vitro* differentiation of mouse embryonic stem cells into neurons of the dorsal forebrain.** *Cell Mol Neurobiol.* 2011; **31**(5): 715–727.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Sansom SN, Livesey FJ: **Gradients in the brain: the control of the development of form and function in the cerebral cortex.** *Cold Spring Harb Perspect Biol.* 2009; **1**(2): a002519.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

24. Butler AB: **The evolution of the dorsal pallium in the telencephalon of amniotes: cladistic analysis and a new hypothesis.** *Brain Res Rev.* 1994; **19**(1): 66–101.
    PubMed Abstract | **Publisher Full Text**

25. Flames N, Marin O: **Developmental mechanisms underlying the generation of cortical interneuron diversity.** *Neuron.* 2005; **46**(3): 377–381.
    PubMed Abstract | **Publisher Full Text**

26. Nagy A, Rossant J, Nagy R, *et al.*: **Derivation of completely cell culture-derived mice from early-passage embryonic stem cells.** *Proc Natl Acad Sci U S A.* 1993; **90**(18): 8424–8428.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Bibel M, Richter J, Schrenk K, *et al.*: **Differentiation of mouse embryonic stem cells into a defined neuronal lineage.** *Nat Neurosci.* 2004; **7**(9): 1003–1009.
    PubMed Abstract | **Publisher Full Text**

28. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics.* 2011; **12**: 323.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Current Referee Status:

Referee Responses for Version 1

Joyce van de Leemput
Neural Stem Cell Institute, Rensselaer, NY, USA

Approved: 14 May 2013

Referee Report: 14 May 2013
The depth and temporal nature of the dataset presented in this paper will be beneficial to any researcher interested in cortical development in general, and potentially lead to many new insights and avenues to pursue.

A point of note, in my experience differences in passage number of the cells used for differentiation can affect gene expression levels throughout. The authors state "ESCs were differentiated into neurons between 5-30 passages after adaptation to suspension culture.", I wonder if that is why the DIV21 samples cluster in between the DIV16 and DIV28 when performing a PCA analysis on the transcript read counts (obtained from Data File 2)?

Related question, how raw are the transcript read counts in Data File 2, as I thought raw counts would have to be integers whereas the counts given have decimal points?

Finally, with regard to the previous Ref Report (Ragsdale and Albertin; 12 March 2013), have you considered comparative analysis using the Allen Brain Atlas/ Mouse Brain expression data for the thalamic and cortical areas and see which region your samples resemble most?

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Cliff Ragsdale[1], **Caroline Albertin**[2]

[1] Department of Neurobiology, University of Chicago, Chicago, IL, USA

[2] Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL, USA

Approved: 12 March 2013

Referee Report: 12 March 2013
There are a growing number of protocols for differentiating stem cells into particular neural cell types. This paper demonstrates the great potential of RNAseq technologies for assessing the identities of such differentiated cells in culture. The authors' goal is an *in vitro* population of 'glutamatergic cortical neurons'. Although many of the genes catalogued show the anticipated profiles across the differentiation process (*Otx2* abundance decreases with DIV while *Kcnh5* reads increase), the dataset also demonstrates that this

culture protocol may not be the best for 'glutamatergic cortical neuron' study as transcripts for the predominant cortical vesicular glutamate transporter gene, *Vglut1/Slc17a7*, are barely detected in the differentiated cell populations.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

1 Comment

Author Response

Patrick McNutt, Research Division, United States Army Medical Research Institute of Chemical Defense (USAMRICD), USA
Posted: 19 Mar 2013

Despite the presence of a number of transcripts suggestive of telencephalogenesis, the singular derivation of vglut2+/vglut1- neurons has bothered us as well, and was one of the motivating factors underlying this experiment. While vglut2 expression has been reported in the neocortex, transcripts have generally been described to appear postnatally (e.g., see allen brain institute expression data for slc17a6 in the developing mouse brain) and vglut1 expression is much stronger than vglut2 in cortical glutamatergic neurons. We are still wading our way through the differential gene expression data, but preliminary, analysis suggests that the RNAseq data may be more consistent with a thalamic origin than a neocortical origin. Not only is thalamogenesis consistent with a vglut2+/vglut1- phenotype, but it is also less well described than neocortical development. We are currently working toward a clear differential diagnosis based on gene expression (should one exist) in an attempt to more clearly elucidate the structural origin of neurons derived in our model.
Competing Interests: There are no competing interests to disclose.