Genomics 100 (2012) 363–369

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

RNA-seq transcriptome analysis of male and female zebra finch cell lines

Christopher N. Balakrishnan [a,b,*], Ya-Chi Lin [b], Sarah E. London [b,c], David F. Clayton [b,1]

[a] Department of Biology, East Carolina University, Greenville, NC 27858, USA
[b] Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[c] Department of Psychology, University of Chicago, Chicago, IL, USA

ARTICLE INFO

Article history:
Received 8 March 2012
Accepted 5 August 2012
Available online 21 August 2012

Keywords:
Zebra finch
RNA-seq
Song learning
Gene expression
Illumina
Dosage compensation
Bird
Sex chromosome

ABSTRACT

The derivation of stably cultured cell lines has been critical to the advance of molecular biology. We profiled gene expression in the first two generally available cell lines derived from the zebra finch. Using Illumina RNA-seq, we generated ~93 million reads and mapped the majority to the recently assembled zebra finch genome. Expression of most Ensembl-annotated genes was detected, but over half of the mapped reads aligned outside annotated genes. The male-derived G266 line expressed Z-linked genes at a higher level than did the female-derived ZFTMA line, indicating persistence in culture of the distinctive lack of avian sex chromosome dosage compensation. Although these cell lines were not derived from neural tissue, many neurobiologically relevant genes were expressed, although typically at lower levels than in a reference sample from auditory forebrain. These cell lines recapitulate fundamental songbird biology and will be useful for future studies of songbird gene regulation and function.

1. Introduction

Songbirds are intensively studied across a diversity of fields ranging from ecology to neuroscience. Songbirds are arguably the best animal models for the study of learned vocal communication [1] and have yielded important insights into the mechanisms of evolutionary adaptation [2,3] and sexual differentiation of brain and behavior [4–7]. Genomic tools for songbirds have now come of age with publication of the complete genome sequence [8] and development of high throughput gene expression assays [9–11] for the zebra finch, the most common songbird in laboratory research. These tools have now been used to identify genes associated with perception of song [12–14], singing behavior [8,11], seasonally regulated courtship and territorial behavior [15,16] and sex-specific brain development [17].

A critical task now is to develop functional tests of specific genes identified through songbird genomics. To this end, recent studies *in vivo* have used pharmacological manipulations [18] and RNA interference [19] to affect zebra finch behavior, and transgenic zebra finches have also now been produced [20]. However, whole-animal manipulations are laborious and expensive and many basic aspects of functional characterization could be carried out more efficiently in cell lines (e.g., assaying consequences of specific gene knockdown on gene expression networks, probing gene dosage compensation mechanisms, or testing microRNA–mRNA interactions). Recently, cultured cell lines from zebra finches been established [21]. Although some experimental objectives may be accomplished using cell lines from other organisms, it remains possible and even likely that transcriptional control networks (e.g., for dosage compensation) and specific molecular interactions are sufficiently different to warrant specific study in cells and tissues from the zebra finch.

In this report we contribute to the characterization of the two tumor-derived cell lines of Itoh and Arnold [21]. One of the lines was derived from a male, which in birds are the homogametic sex (ZZ), and the other from a female bird (ZW). Both tumors were removed from non-neural tissues (although the exact cellular origin of neither line is known [21]). We used Illumina mRNA sequencing (RNA-seq) to generate gene expression profiles of these two cell lines, and analyzed the data specifically to evaluate potential utility of the cell lines for study of sex differences and for genes of neurobiological interest. RNA-seq expression profiling is still a relatively new and evolving methodology [22–25], and our study is one of the first applications of this method in songbird research [see also 8,26,27], or for *de novo* characterization of cell lines from any species.

2. Results

2.1. RNA-seq and read mapping

One lane of sequencing of the zebra finch cell lines on the Illumina HiSeq2000 platform yielded 92,609,701 reads. These were distributed

* Corresponding author at: Department of Biology, East Carolina University, Greenville, NC 27858, USA.
E-mail address: balakrishnanc@ecu.edu (C.N. Balakrishnan).
[1] Present address: School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK.

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2012.08.002

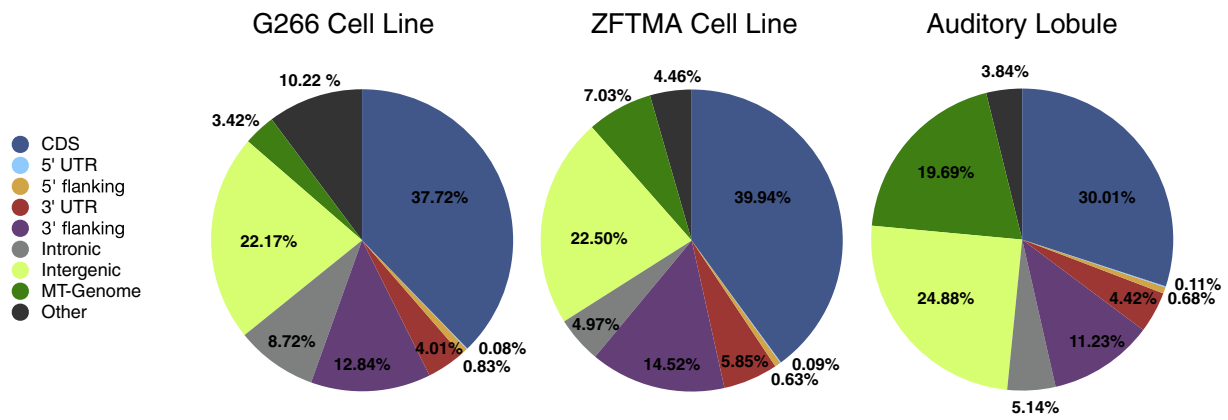

Fig. 1. Distribution of uniquely mapping RNA-seq reads among genomic compartments. Flanking regions are defined as being within 1 kb up or downstream of the Ensembl cDNA. Reads outside of annotated gene models are termed "intergenic" although it is possible likely of these are actually novel genes. The "other" category includes reads of ambiguous mapping location (reads spanning two compartments) as well as reads mapping to annotated telomeres.

nearly evenly between the two cell lines, male G266 (45,268,521 reads) and tetraploid female ZFTMA (47,340,550 reads). Using Tophat [28] we were able to map over half (G266: 58.8%; ZFTMA: 53.7%) of the reads unambiguously to a single location in the genome. Allowing for multireads (reads mapping up to 20 places in the genome), we were able to map 69.8% and 71.2% of G266 and ZFTMA reads, respectively. Of the uniquely mapped reads, only approximately 40% mapped to genome regions covered by current Ensembl coding regions (Fig. 1, Supplementary Table 1). The remaining reads mapped to regions currently annotated as introns, intergenic regions or UTRs.

To provide a point of reference for cell line gene expression profiles we also sequenced RNA from the auditory forebrain ("auditory lobule" or "AL" [29]) of female zebra finches. The auditory lobule is composed of three forebrain subregions, Field L, the caudomedial nidopallium (NCM) and the caudomedial mesopallium (CMM). We focused on the auditory lobule as it is a focal point for recent research on gene expression in the zebra finch brain [8,12,18,30]. Three lanes of sequencing of zebra finch auditory forebrain samples on an Illumina Genome Analyzer produced 69,836,901 reads of which 68.1% were mapped uniquely (these results were generated using a different library preparation and sequencing platform, see Materials and methods).

Across all samples (cell lines and auditory lobule), less than 6% of reads mapped to known UTR regions, likely reflecting the incomplete state of zebra finch gene annotations. The large proportion of reads mapped to regions flanking Ensembl gene models (~11–15%) suggests that these areas are in fact transcribed regions that have yet to be formally annotated. In particular, a relatively large proportion of reads mapped to the region within 1 kb of the 3′ end of Ensembl models (Fig. 1).

Read mapping from cell lines versus auditory lobule also showed some distinctive differences. Despite a higher overall mapping rate (68.1% versus <60% for cell lines), a lower proportion of the auditory forebrain reads (only 30.0%) mapped to Ensembl coding regions (Fig. 1). We also note that a much higher proportion of auditory lobule reads mapped to the mitochondrial genome (19.69%) than the samples from the cell lines (G266 = 3.42%, ZFTMA = 7.03%; Fig. 1).

2.2. Functional annotation of genes present in cell lines

We detected 13,333 Ensembl-annotated genes with at least one read in each cell line (Fig. 2), and we carried out further functional analyses of this set of "cell line expressed" genes. Statistical over and under-representation of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway representation was assessed using Fisher's Exact Test and p-values were adjusted for multiple testing. Canonical signaling pathways are well represented among this set (Table 1). Only one KEGG pathway was significantly underrepresented among the 132 detected pathways: Calcium signaling pathway (gga04020, $p = 0.00018$). One pathway, Ribosome (gga03010) was significantly enriched ($p = 0.043$). Genes associated with 44 Gene Ontology (GO) terms were underrepresented ($p < 0.05$) and another 45 terms were over-represented ($p < 0.05$) in the cell lines (Supplementary Table 2). Categories that were enriched often involved cellular components (e.g., cytoplasm, mitochondrion, endoplasmic reticulum) whereas categories that were underrepresented included a number of signaling processes (olfactory receptor activity, G-protein coupled receptor activity) and immune components (immune response, MHC Class II protein complex, MHC Class I protein complex). We also specifically examined the list of expressed genes for GO terms related to neurobiology (121 GO categories containing "neuro" or "synap"). Of these, only three terms were significantly underrepresented among expressed genes, indicating a relative lack of neuronal post-synaptic proteins and receptors (Table 2).

2.3. Gene expression differences between cell lines

We identified 98 genes that were differentially expressed between the two cell lines (FDR $p < 0.01$ Fig. 3). This gene list was significantly

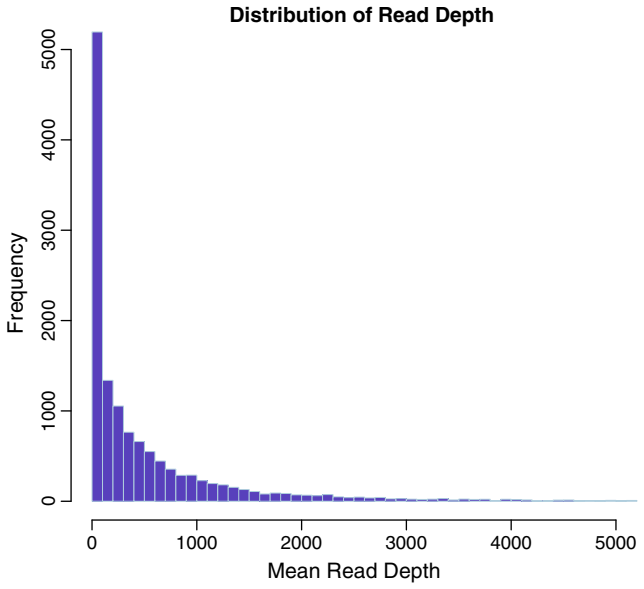

Fig. 2. Distribution of normalized read count across 13,333 genes that were represented by at least one read in each of the two cell lines.