RESEARCH ARTICLE

Analysis of SOX2-Regulated Transcriptome in Glioma Stem Cells

Arlie M. Acanda de la Rocha1,2,3, Hernando López-Bertoni4,5, Elizabeth Guruceaga1,6, Marisol González-Huarriz1,2,3, Naiara Martínez-Vélez1,2,3, Enric Xipell1,2,3, Juan Fueyo7, Candelaria Gomez-Manzano7, Marta M. Alonso1,2,3*

1 The Health Research Institute of Navarra (IDISNA), Pamplona, Spain, 2 Program in Solid Tumors and Biomarkers, Foundation for the Applied Medical Research, Pamplona, Spain, 3 Department of Pediatrics, University Hospital of Navarra, Pamplona, Spain, 4 Hugo W Moser Research Institute at Kennedy Krieger, Baltimore, Maryland, United States of America, 5 Department of Neurology, The Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America, 6 Bioinformatics Unit, Center for Applied Medical Research, Pamplona, Spain, 7 Brain Tumor Center, University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America

* mmalonso@unav.es

Abstract

Introduction

Glioblastoma is the most malignant brain tumor in adults and is associated with poor survival despite multimodal treatments. Glioma stem-like cells (GSCs) are cells functionally defined by their self-renewal potential and the ability to reconstitute the original tumor upon orthotopic implantation. They have been postulated to be the culprit of glioma chemo- and radio-resistance ultimately leading to relapse. Understanding the molecular circuits governing the GSC compartment is essential. SOX2, a critical transcription regulator of embryonic and neural stem cell function, is deregulated in GSCs however; the precise molecular pathways regulated by this gene in GSCs remain poorly understood.

Results

We performed a genome-wide analysis of SOX2-regulated transcripts in GSCs, using a microarray. We identified a total of 2048 differentially expressed coding transcripts and 261 non-coding transcripts. Cell adhesion and cell-cell signaling are among the most enriched terms using Gene Ontology (GO) classification. The pathways altered after SOX2 down-modulation includes multiple cellular processes such as amino-acid metabolism and intercellular signaling cascades. We also defined and classified the set of non-coding transcripts differentially expressed regulated by SOX2 in GSCs, and validated two of them.

Conclusions

We present a comprehensive analysis of the transcriptome controlled by SOX2 in GSCs, gaining insights in the understanding of the potential roles of SOX2 in glioblastoma.
Introduction

Glioblastoma is the most common and deadly primary brain tumors in adults and despite multiple treatments, the survival of glioma patients remains poor, with a median time between 12–15 months [1–3].

Glioblastoma contains a subpopulation of tumor propagating stem-like cells, known as glioma stem-like cells (GSCs) [4], which display the ability to self-renew, to differentiate into distinct lineages and to efficiently initiate and propagate tumors in xenografts models that recapitulate the phenotypic characteristics of the initial tumor from which they were derived [5,6]. Moreover, GSCs have been shown to increase resistance to radio- and chemotherapy [7,8], explaining in part the poor overall survival despite multiple treatments.

SOX2, a member of the SRY gene family, is a key transcription factor in the regulation of stemness properties, and it is essential in early embryonic development [9]. SOX2 has been reported to be deregulated in several human cancers [10–12] including glioblastoma where it is over expressed due to several mechanisms such as amplification and promoter hypomethylation [13]. SOX2 is enriched in human-derived GSCs where it sustains stemness, migration, invasion and maintenance of tumorigenicity [13,14]. Although SOX2 response program in a glioblastoma cell line has been analyzed [15], to the best of our knowledge, an exhaustive analysis of SOX2-regulated molecular circuitries in GSCs has not been performed. Deciphering the molecular circuitries controlled by SOX2 in GSCs will provide insights about glioma development, biology and possible novel molecular therapies.

Besides coding genes, long non-coding RNAs (lncRNAs) are an emerging class of RNAs with no functional protein-coding ability that consists of more than 200 nucleotides [16]. Recent discoveries have proven that they play important roles regulating gene expression and function. These non-coding RNAs actively participate in many pathological processes in human malignancies [17–19] including cancer where a number of lncRNAs have been shown to act as oncogenes or tumor suppressors [20]. Recently, different groups published a signature of lncRNAs with aberrant expression in glioblastoma [21] and a set of prognostic lncRNAs that could have clinical implications in the sub-classification of this disease [22]; though the functional effect of lncRNAs in glioblastoma is not well understood.

Given that SOX2 is predominantly expressed in the GSCs compartment, which plays prominent roles in driving the growth, treatment resistance and recurrence of glioblastoma, the elucidation of the transcriptome and the molecular pathways involved in the generation and maintenance of this recalcitrant population is critical to understand the molecular underpinnings of glioblastoma malignancy. The aim of this work was to characterize the transcriptome regulated by SOX2 in GSCs. We set out to describe not only the coding genes but also the lncRNAs, which have been shown to play predominant roles in cancer. In this study we present a comprehensive analysis of the transcriptome controlled by SOX2 in GSCs, gaining insights in the understanding of the potential roles of SOX2 in glioblastoma.

Materials and Methods

Cell Lines, culture and transfection

Neurosphere cultures (GSC11 and GSC23), a kind gift of Dr. Lang at UT MD Anderson Cancer Center, were established from acute cell dissociation of human glioblastoma surgical specimens and maintained in Dulbecco's modified Eagle's medium/nutrient mixture F12 supplemented with B27 (Invitrogen, Carlsbad, CA), epidermal growth factor, and basic fibroblast growth factor (20 ng/mL each; Sigma-Aldrich, St Louis, MO).
To inhibit SOX2 expression, transient transfection assays were performed using two commercially available, specific siRNA against human SOX2 (si-SOX2, s13295 and s13294, Ambion) and a non-targeting control siRNA (si-scramble) (Ambion) in four independent experiments. The siRNA transfections were performed according to the manufacturer's instructions using Lipofectamine 2000 (Invitrogen). The cells were then cultured for 72 h after transfections and subjected to different analysis.

RNA extraction and Real Time PCR analysis

Total cellular RNA was isolated from the cultured cells using a Trizol reagent (Ambion) according to the manufacturers' protocols. For IncRNAs analysis, total RNA was subjected to DNase I treatment to digest the DNA. RNA quantity and quality were measured by NanoDrop ND-1000. The RNA samples were then reversely transcribed into cDNA using the Taqman miRNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions. Real Time PCR (RT-PCR) was performed using the Sybr Green Fast Master Mix (Applied Biosystems) in the ABI 7700 sequence detection system (Applied Biosystems, Foster City, CA). The quality of the products was controlled by the melting curve. Transcript levels were normalized against human GAPDH. Transcripts expression levels relative to GAPDH were calculated using the ddCt method. Primers for IncRNA detection and quantification were designed at Universal ProbeLibrary Assay Design Center (http://www.roche-applied-science.com/). All primer sequences are listed below (Tables 1 and 2):

Immunoblotting Assay

For the western blot assay, cells were lysed in RIPA buffer (Triton and PBS) for 30 min on ice. Samples containing identical amounts of protein (30 μg) were resolved in a 12% polyacrylamide gel, transferred to polyvinylidene membranes, and blocked in 5% nonfat milk in phosphate-buffered saline/Tween-20. Membranes were incubated with the following antibodies:

Table 1. Primers of lncRNAs for qRT-PCR.

<table>
<thead>
<tr>
<th>IncRNA position</th>
<th>TCONS</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr19:28281401−28284848</td>
<td>TCONS_00027256</td>
<td>GCCCAAAGTTGATTTTCGTG</td>
<td>CGAGGTC7AACCAGGGTGCT</td>
</tr>
<tr>
<td>chr11:121899032−121899389</td>
<td>TCONS_00020142</td>
<td>GCCGAGCCTTCATGAAAAAT</td>
<td>GTGCATGCTCCATGACCA</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0163155.t001

Table 2. Primers of genes for qRT-PCR.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>AGCCACACGCTCGAGCACAC</td>
<td>GCCAAATACGACAAATCC</td>
</tr>
<tr>
<td>SOX2</td>
<td>AGCTCGACGACCTAGCAA</td>
<td>CCGGGAGAATACATGCTGAT</td>
</tr>
<tr>
<td>PLP1</td>
<td>ACCATGAGCTGACGCTTAG</td>
<td>TGGGAGAAAGCATGACTAGC</td>
</tr>
<tr>
<td>COL2A1</td>
<td>TGGACGCCATAGGAGGTTTT</td>
<td>TGGGAGGATTCATGACTTC</td>
</tr>
<tr>
<td>ATP8B1</td>
<td>AGCAGAATGGGAGAGAATTC</td>
<td>GATCGTTGTCATGACAGGTC</td>
</tr>
<tr>
<td>PPP1R1B</td>
<td>CAGACTGGAAGACACCACCC</td>
<td>GCTGTGTTCTACATGAAATTC</td>
</tr>
<tr>
<td>CMTM5</td>
<td>GGAGGACACATCAGGCTAGAT</td>
<td>CGGAGGATGGAGGACAGAT</td>
</tr>
<tr>
<td>GALNT14</td>
<td>CAGTCGGGGATTCATGACACG</td>
<td>CGGATCGAGATGCTGAGG</td>
</tr>
<tr>
<td>F11R</td>
<td>GTGCTTTGTCGCGTCTTTCTT</td>
<td>GTCACCCGCGCTCTGATAGGA</td>
</tr>
<tr>
<td>SYT4</td>
<td>ATGGGATACCCTACCAACAAT</td>
<td>TCCGGAGAGAGAAATTAGAATT</td>
</tr>
<tr>
<td>SLC18A1</td>
<td>GTGGTGTGATCTGCTCTTGTG</td>
<td>CGGAGGTAGGAGAGAGAT</td>
</tr>
<tr>
<td>ITLN2</td>
<td>GCCAGGCCAAAGGAGACATA</td>
<td>CAGGGGCGCTTTCTCCAA</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0163155.t002
SOX2 (Cell Signaling, Danvers, MA) and α-Tubulin (Sigma-Aldrich) using 1:1000 dilution. The membranes were developed according to the protocol for enhanced chemiluminescence from Perkin Elmer.

Microarray expression analysis

Total RNA was isolated from scrambled and SOX2-siRNA GSC11 cells using Trizol extraction. RNA was purified by the QIAGEN RNAeasy mini kit (QIAGEN) according to the manufacturer’s protocol. One-color Cy3 RNA labeling, array hybridization to Agilent SurePrint G3 8 × 60 K Human Gene Expression Arrays (Agilent Technologies), data collection, and analysis were performed at the Bioinformatics Unit (Fundación para la Investigación Médica Aplicada, CIMA, Pamplona, Spain). Normalization of microarray data was performed using quantile algorithm. After quality assessment a filtering process was carried out to eliminate low expression probe sets. Applying the criterion of an expression value greater than 64 in 2 samples of at least one of the experimental conditions, 40986 probe sets were selected for statistical analysis. LIMMA (Linear Models for Microarray Data) [23] was used to find out the probe sets that showed significant differential expression between experimental conditions. Genes were selected as significant using a B statistic cut off B > 0. Data processing and statistical analyses were performed with R and Bioconductor [24]. The microarray data from this study are publicly available at Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE79302.

Functional group analysis

In this study we applied Gene Ontology (GO) analysis to find the primary function of the differential expression of mRNAs regulated by SOX2, using online software DAVID (Database for Annotation, Visualization and Integrated Discovery, http://david.abcc.ncifcrf.gov/). GO analysis can organized genes into hierarchical categories (Gene Ontology Consortium). To find out the significant pathway of the differential genes participating we performed gene regulatory network analysis using Ingenuity Pathway Analysis (IPA) software (http://www.ingenuity.com), which can integrate gene-expression data with other molecular databases to facilitate the development of new and more complete pathway maps. Fisher’s exact test was used to select the significant GO categories. The threshold of significance was defined by P value with a cutoff set in 0.05.

Statistical analysis

Experimental data are represented as the mean ± SD of three biologic replicates and were compared using Student’s t-test. Significant P-values are indicated with asterisks as follows: *P < 0.05, **P < 0.01.

Results

Transcripts regulated by SOX2

Since SOX2 is a key driver in the maintenance of the GSCs phenotype and therefore in the perpetuation of this devastating tumor we down-regulated the expression of this gene in the GSC11 cell line in four independent experiments, using a SOX2 specific siRNA (Fig 1A). The efficiency of SOX2 knockdown was assessed by real-time PCR and western blot (Fig 1B) and was also confirmed in our array results. Microarray data identified a total of 2048 differentially expressed coding transcripts and 261 non-coding transcripts (B value >0) (Fig 1A and S1 Table).
SOX2 Transcriptome in GSCs

A

Sox2 down-regulation

3 independent experiments

GSC-11 cells mRNA Array Agilent SurePrint G3 Human gene expression

B

Transcripts regulated by Sox2

Sox2

Tubulin

siScramble siSox2

Sox2 relative expression

si-Scrbl si-Sox2

mRNA Sox2 relative expression
SOX2 controls a wide spectrum of protein-coding genes and pathways in GSCs

To further narrow the coding transcripts data a cut-off 1 logarithmic fold difference between SOX2 knockdown and scrambled GSC11 cells was set, identifying 35 up-regulated and 100 down-regulated genes, which suggest that SOX2 act primarily as a transcriptional activator. In Table 3 we showed the top-10 up or down-regulated protein coding-genes, and select the top 5 candidates of each group for further validation by qRT-PCR using the GSC11 and GSC23 cell lines. We confirmed the observed microarray expression changes in 5 out of 5 down-regulated coding-genes (Fig 2B) and in one out of 5 up-regulated coding-genes in GSC11 cells (Fig 2A). Regarding GSC-23 cells, we down-modulate SOX2 expression using two different siRNAs against human SOX2, and we validate the expression of 5 out of 5 up-regulated coding-genes (Fig 2A) and of 4 out of 5 down-regulated coding genes (Fig 2B), partially validating the microarray results.

To understand the significance of differential gene expression, bioinformatics analysis related to Gene Ontology (GO) Classification and pathway analysis were performed. GO classifications using the DAVID web tool and pathway analysis using Ingenuity Pathway Analysis (IPA) was performed. For these analysis, gene lists were classified based upon decreased (logFC < -1) or increased (logFC > 1) expression relative to control and analyzed altogether as a single list.

Table 3. The top 10 up- and down-regulated genes in Sox2-downmodulated GSC-11 cells, organized by logFC.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>logFC</th>
<th>P.Value</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLP1</td>
<td>2.443</td>
<td>4.73E-05</td>
<td>2.488</td>
</tr>
<tr>
<td>COL2A1</td>
<td>2.122</td>
<td>2.18E-05</td>
<td>3.281</td>
</tr>
<tr>
<td>ATP8B1</td>
<td>1.954</td>
<td>1.50E-06</td>
<td>5.943</td>
</tr>
<tr>
<td>PPP1R1B</td>
<td>1.925</td>
<td>7.27E-06</td>
<td>4.393</td>
</tr>
<tr>
<td>CMTM5</td>
<td>1.702</td>
<td>1.12E-04</td>
<td>1.586</td>
</tr>
<tr>
<td>ELMO1</td>
<td>1.66</td>
<td>5.97E-05</td>
<td>2.246</td>
</tr>
<tr>
<td>ITIH5L</td>
<td>1.555</td>
<td>6.67E-05</td>
<td>2.133</td>
</tr>
<tr>
<td>IGFBP5</td>
<td>1.522</td>
<td>1.82E-06</td>
<td>5.759</td>
</tr>
<tr>
<td>SCARNA9</td>
<td>1.498</td>
<td>7.10E-05</td>
<td>2.067</td>
</tr>
<tr>
<td>SCARNA17</td>
<td>1.435</td>
<td>4.46E-06</td>
<td>4.879</td>
</tr>
<tr>
<td>GALNT14</td>
<td>-3.006</td>
<td>0.000299</td>
<td>0.569</td>
</tr>
<tr>
<td>F11R</td>
<td>-2.824</td>
<td>1.13E-07</td>
<td>8.308</td>
</tr>
<tr>
<td>SYT4</td>
<td>-2.664</td>
<td>2.61E-06</td>
<td>5.405</td>
</tr>
<tr>
<td>SLC18A1</td>
<td>-2.335</td>
<td>4.01E-06</td>
<td>4.984</td>
</tr>
<tr>
<td>ITLN2</td>
<td>-2.329</td>
<td>3.57E-07</td>
<td>7.288</td>
</tr>
<tr>
<td>RASEF</td>
<td>-2.294</td>
<td>9.38E-07</td>
<td>6.39</td>
</tr>
<tr>
<td>GADD45G</td>
<td>-2.231</td>
<td>3.55E-08</td>
<td>9.277</td>
</tr>
<tr>
<td>CYP26A1</td>
<td>-2.056</td>
<td>7.66E-10</td>
<td>11.938</td>
</tr>
<tr>
<td>KRTAP21-1</td>
<td>-2.028</td>
<td>0.00021</td>
<td>0.94</td>
</tr>
<tr>
<td>PNLIPRP2</td>
<td>-1.992</td>
<td>0.000495</td>
<td>0.042</td>
</tr>
</tbody>
</table>

**SOX2 Transcr**
Enrichment analysis of GO categories including biological process (BP), molecular function (MF), and cellular component (CC) were obtained using DAVID web tool (Fig 3). We observed the highest enrichment in the categories related to “cell adhesion”, “biological adhesion”, “cell-cell signaling”, “extracellular region” and “calcium ion binding”.

We used IPA analysis to undercover the canonical pathways regulated by SOX2 in GSCs. Our results showed 13 pathways significantly altered (Table 4). Most of them related with
A

GO-Biological Process

- cell adhesion
- biological adhesion
- growth
- bone development
- transmission of nerve impulse
- synaptic transmission
- skeletal system development
- regulation of specific transcription from RNA polymerase II promoter
- cell-cell signaling
- cellular amino acid derivative metabolic process

Number of Molecules

B

GO-Cellular Component

- extracellular region
- extracellular region part
- extracellular matrix
- proteinaceous extracellular matrix
- extracellular space
- cell projection
- dynein complex
- cytoskeleton
- receptor complex
- basolateral plasma membrane

Number of Molecules

C

GO-Molecular Function

- calcium ion binding
- carboxylic acid binding
- endopeptidase inhibitor activity
- peptidase inhibitor activity
- manganese ion binding
- metalloendopeptidase activity
- carbohydrate binding
- metallopeptidase activity
- amine binding
- enzyme inhibitor activity

Number of Molecules
amino-acid metabolism pathways, such as “histamine biosynthesis”, where histamine is an important regulator of numerous physiological processes including neurotransmission in the central nervous system (CNS) [25]; “L-cysteine degradation process” where cystathionine γ-lyase (CTH) activity has been related with glioblastoma treatment [26] and “serotonin receptor signaling pathway” being serotonin an important neurotransmitter in the CNS during neuronal development [27]. Other enriched pathways were “hematopoiesis from multipotent stem cells”, where KITLG has been reported to regulate neoplastic processes such as growth and invasion [28]; apoptosis [29] and cell adhesion [30]. Role of JAK2 in Hormone-like Cytokine Signaling stood out because GHR (growth hormone receptor) and IRS1 (insulin receptor substrate 1) has been linked with glioma progression [31,32]. A well-characterized pathway frequently altered in tumors is the NOTCH signaling cascade, which was also enriched in our analysis. The NOTCH pathway is a conserved intercellular signaling route that has been implicated in different developmental processes. Interestingly, NOTCH pathway is deregulated in human glioblastoma and plays a key role in maintaining the growth, the undifferentiated state of glioma cells and tumorigenesis [33–35]. The integrated analysis of SOX2 enriched canonical pathways revealed the link between this transcription factor and multiple cellular processes such as amino-acid metabolism and intercellular signaling cascades, like NOTCH pathway.

The IPA analysis also showed the most relevant biological functions and diseases in our data set. The most significant bio-functions altered following SOX2 down-modulation are shown in Table 5. The set of SOX2-associated genes were assigned mainly to the following networks: “cancer”, “organismal injury and abnormalities”, “cellular movement”, “tissue morphology”, “cellular development” and “hematopoiesis”. Interestingly, most of these networks involved very well-known functions of SOX2 such as morphology determination [36], development [37] and cellular proliferation and migration in glioma [13,13]. Fig 4 shows the most relevant selection of bio-function categories: disease and disorders, molecular and cellular functions and physiological system development and function, obtained by using IPA software organized by p-value.

### Table 4. List of top-13 canonical pathways identified by IPA software.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>-log(p-value)</th>
<th>Ratio</th>
<th>Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine Betaine Degradation</td>
<td>2.63E+00</td>
<td>2.50E-01</td>
<td>DMGDH, PIPOX</td>
</tr>
<tr>
<td>Hepatic Stellate Cell Activation</td>
<td>2.38E+00</td>
<td>4.42E-02</td>
<td>LY96, COL2A1, COL22A1, L1FBP5, COL28A1</td>
</tr>
<tr>
<td>Histamine Biosynthesis</td>
<td>2.03E+00</td>
<td>1.00E+00</td>
<td>HDC</td>
</tr>
<tr>
<td>L-cysteine Degradation II</td>
<td>2.03E+00</td>
<td>1.00E+00</td>
<td>CTH</td>
</tr>
<tr>
<td>Triacylglycerol Degradation</td>
<td>2.02E+00</td>
<td>1.25E-01</td>
<td>PNLP2P, CES1</td>
</tr>
<tr>
<td>Retinol Biosynthesis</td>
<td>2.02E+00</td>
<td>1.25E-01</td>
<td>PNLP2P, CES1</td>
</tr>
<tr>
<td>Hematopoiesis from Multipotent Stem Cells</td>
<td>1.73E+00</td>
<td>5.00E-01</td>
<td>KITLG</td>
</tr>
<tr>
<td>Cysteine Biosynthesis/Homocysteine Degradation</td>
<td>1.73E+00</td>
<td>5.00E-01</td>
<td>CTH</td>
</tr>
<tr>
<td>Role of JAK2 in Hormone-like Cytokine Signaling</td>
<td>1.61E+00</td>
<td>7.69E-02</td>
<td>GHR, IRS1</td>
</tr>
<tr>
<td>Serotonin Receptor Signaling</td>
<td>1.58E+00</td>
<td>7.41E-02</td>
<td>SLC18A1, HTR1D</td>
</tr>
<tr>
<td>Phenylethylamine Degradation</td>
<td>1.56E+00</td>
<td>3.33E-01</td>
<td>AOC3</td>
</tr>
<tr>
<td>Notch Signaling</td>
<td>1.35E+00</td>
<td>5.56E-02</td>
<td>HES5, HEY1</td>
</tr>
<tr>
<td>Lysine Degradation V</td>
<td>1.34E+00</td>
<td>2.00E-01</td>
<td>PIPOX</td>
</tr>
</tbody>
</table>

This selection is organized by the negative logarithm of p-values (Fisher Test), calculated by IPA ([-Log (0.05) = 1.3]).

doi:10.1371/journal.pone.0163155.t004
These results established a signature of protein coding-genes regulated by SOX2 in GSCs with biological functions relevant to glioblastoma growth and maintenance of its malignant phenotype. The tight overlap between the existing literature and our enrichment analysis highlights the robustness of our results and predicts that this approach will be an excellent discovery platform to identify novel SOX2 targets.

SOX2–regulated non-coding RNAs in GSC

Reprogramming transcription factors, including SOX2, have been shown to regulate both coding and non-coding RNAs [38]. LncRNAs are emerging as key regulators of biological processes and disease [39] therefore, seems reasonable to hypothesize that SOX2 will regulate this class of genes as well. The strength of our data-sets allowed us to identify potential non-coding transcripts differentially expressed (B value > 0) regulated by SOX2 in GSCs. After biotype distribution analysis we identify protein coding RNAs (44% for up-regulated and 41% for down-regulated), while the rest were classified as different types of non-coding transcripts. Out of the total number of transcripts differentially expressed we identify 80 upregulated and 181 down-regulated and we classify them as intergenic RNAs, antisense, processed transcripts, transcripts derived from pseudogenes and unassigned transcripts (Fig 5). The transcripts classified as “others” correspond to transcripts derived from miRNAs, rRNAs, sense-overlapping and sense intronic transcripts. The lncRNA annotation was performed with the Bioconductor package ChIPpeakAnno [40] and using Gencode v19 as reference [41]. The gene type corresponding to the gene that overlaps with the lincRNA locus was assigned to each lincRNA. Table 6 shows the top 25 non-coding transcripts regulated by SOX2 in GSCs.

The top four differentially expressed lncRNAs (Table 7) that presented chromatin marks and high abundance in brain were selected using GRCh37/hg19 assembly in UCSC Genome Browser, and their expression was validated using qRT-PCR in GSC11 cells, comparing SOX2-siRNA versus scrambled control. Our data indicated that the expression of chr19:28,281,401–28,284,848 (TCONS_00027256) was significantly down-regulated (p value = 0.018), while chr11:121899032–121899389 (TCONS_00020142) was significantly up-regulated (p value = 0.042) after SOX2 inhibition (Fig 6A). The results were consistent with the microarray data (Fig 6B).
Fig 4. 10-Top Bio Functions categories altered following SOX2 inhibition. The categories listed are Physiological System Development and Function, Molecular and cellular Functions and Disease and Disorders, identified using IPA software. Bars represent the number of genes in the specified category, organized by p-value.

doi:10.1371/journal.pone.0163155.g004
Altogether, these results identified and confirmed the non-coding transcript profile controlled by SOX2 in GSCs. Characterizing the functional relevance of these lncRNAs will undoubtedly impact our understanding of glioblastoma biology.

**Fig 5. SOX2 regulated non-coding transcripts.** (A) A total of 261 transcripts were found differentially expressed (B > 0), which were distributed in 80 upregulated and 181 downregulated transcripts. (B) Biotype distribution of the differentially expressed transcripts following SOX2 downmodulation in GSC11 cells.

doi:10.1371/journal.pone.0163155.g005
### Table 6. List of the top 25 non-coding transcripts regulated by Sox2, organized by B value.

<table>
<thead>
<tr>
<th>Probe</th>
<th>GeneName</th>
<th>Classification</th>
<th>logFC</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_19_P00320471</td>
<td>linRNA:chr9:2535671–2536375_R</td>
<td>antisense</td>
<td>-1.28</td>
<td>11.75</td>
</tr>
<tr>
<td>A_19_P00315804</td>
<td>linRNA:chr9:2530903–2539456_R</td>
<td>antisense</td>
<td>-1.15</td>
<td>10.77</td>
</tr>
<tr>
<td>A_19_P00320469</td>
<td>linRNA:chr9:2535671–2536375_R</td>
<td>antisense</td>
<td>-1.27</td>
<td>9.70</td>
</tr>
<tr>
<td>A_33_P3397743</td>
<td>LOC100128088</td>
<td>pseudogene</td>
<td>-1.83</td>
<td>7.42</td>
</tr>
<tr>
<td>A_19_P00321203</td>
<td>linRNA:chr6:72126155–72129954_R</td>
<td>lincRNA</td>
<td>-0.92</td>
<td>7.33</td>
</tr>
<tr>
<td>A_19_P00322118</td>
<td>linRNA:chr2:39745746–39826668_F</td>
<td>antisense</td>
<td>-0.86</td>
<td>6.90</td>
</tr>
<tr>
<td>A_23_P3552</td>
<td>LOC730092</td>
<td>pseudogene</td>
<td>-0.65</td>
<td>6.50</td>
</tr>
<tr>
<td>A_19_P00322220</td>
<td>linRNA:chr20:3705062–3706887_R</td>
<td>processed_transcript</td>
<td>-0.69</td>
<td>6.28</td>
</tr>
<tr>
<td>A_33_P3392460</td>
<td>LOC100128077</td>
<td>processed_transcript</td>
<td>-1.47</td>
<td>6.18</td>
</tr>
<tr>
<td>A_19_P00322149</td>
<td>linRNA:chr6:72126142–72129923_R</td>
<td>lincRNA</td>
<td>-0.91</td>
<td>6.16</td>
</tr>
<tr>
<td>A_19_P00317793</td>
<td>linRNA:chr20:3705062–37068916_R</td>
<td>processed_transcript</td>
<td>-0.68</td>
<td>6.10</td>
</tr>
<tr>
<td>A_19_P0080846</td>
<td>linRNA:chr21:17992729–18010729_F</td>
<td>lincRNA</td>
<td>-0.64</td>
<td>6.09</td>
</tr>
<tr>
<td>A_19_P00318304</td>
<td>linRNA:chr20:37050986–37063998_R</td>
<td>processed_transcript</td>
<td>-0.67</td>
<td>5.94</td>
</tr>
<tr>
<td>A_19_P00316341</td>
<td>linRNA:chr7:13060080–130606702_F</td>
<td>lincRNA</td>
<td>-0.86</td>
<td>5.93</td>
</tr>
<tr>
<td>A_19_P00316985</td>
<td>linRNA:chr6:72126162–72129969_R</td>
<td>lincRNA</td>
<td>-0.91</td>
<td>5.93</td>
</tr>
<tr>
<td>A_19_P00322967</td>
<td>linRNA:chr20:37050934–37057222_R</td>
<td>processed_transcript</td>
<td>-0.69</td>
<td>5.54</td>
</tr>
<tr>
<td>A_19_P00802908</td>
<td>linRNA:chr2:3579550–3585150_R</td>
<td>lincRNA</td>
<td>-0.58</td>
<td>5.15</td>
</tr>
<tr>
<td>A_24_P756289</td>
<td>SOX2OT</td>
<td>other</td>
<td>-0.86</td>
<td>5.08</td>
</tr>
<tr>
<td>A_33_P3613516</td>
<td>LOC254057</td>
<td>antisense</td>
<td>-1.10</td>
<td>4.94</td>
</tr>
<tr>
<td>A_19_P00318114</td>
<td>linRNA:chr2:3579840–3584422_R</td>
<td>lincRNA</td>
<td>-0.73</td>
<td>4.52</td>
</tr>
<tr>
<td>A_33_P3287710</td>
<td>chr10:79,686,570–79,689,583</td>
<td>unassigned</td>
<td>-0.67</td>
<td>4.46</td>
</tr>
<tr>
<td>A_33_P3405043</td>
<td>LOC100133264</td>
<td>unassigned</td>
<td>-0.72</td>
<td>4.43</td>
</tr>
<tr>
<td>A_32_P88349</td>
<td>LOC730256</td>
<td>pseudogene</td>
<td>-0.48</td>
<td>4.36</td>
</tr>
<tr>
<td>A_33_P3705884</td>
<td>chr19:28,281,401–28,284,848</td>
<td>lincRNA</td>
<td>-0.89</td>
<td>4.27</td>
</tr>
<tr>
<td>A_19_P00321044</td>
<td>linRNA:chr16:50682543–50683160_F</td>
<td>lincRNA</td>
<td>1.04</td>
<td>6.09</td>
</tr>
<tr>
<td>A_19_P00315647</td>
<td>linRNA:chr11:121899032–121899389_R</td>
<td>other</td>
<td>0.65</td>
<td>5.63</td>
</tr>
<tr>
<td>A_32_P63013</td>
<td>LOC283174</td>
<td>unassigned</td>
<td>1.32</td>
<td>5.08</td>
</tr>
<tr>
<td>A_32_P47157</td>
<td>LOC92973</td>
<td>unassigned</td>
<td>0.74</td>
<td>4.95</td>
</tr>
<tr>
<td>A_19_P00317484</td>
<td>linRNA:chr3:112315643–112316945_R</td>
<td>lincRNA</td>
<td>0.55</td>
<td>4.11</td>
</tr>
<tr>
<td>A_19_P00809440</td>
<td>linRNA:chr11:133765815–133774297_R</td>
<td>other</td>
<td>1.23</td>
<td>4.06</td>
</tr>
<tr>
<td>A_33_P3798382</td>
<td>chr10:65,224,989–65,226,322</td>
<td>antisense</td>
<td>0.57</td>
<td>3.98</td>
</tr>
<tr>
<td>A_19_P00321420</td>
<td>linRNA:chr11:133766329–133767054_R</td>
<td>unassigned</td>
<td>1.34</td>
<td>3.74</td>
</tr>
<tr>
<td>A_19_P00332120</td>
<td>linRNA:chr13:156455706–156471081_R</td>
<td>lincRNA</td>
<td>0.59</td>
<td>3.57</td>
</tr>
<tr>
<td>A_19_P00320101</td>
<td>linRNA:chr11:133767609–133771496_R</td>
<td>other</td>
<td>1.06</td>
<td>3.51</td>
</tr>
<tr>
<td>A_19_P00812924</td>
<td>linRNA:chr11:121895965–121904065_R</td>
<td>other</td>
<td>0.53</td>
<td>3.46</td>
</tr>
<tr>
<td>A_19_P00326763</td>
<td>linRNA:chr6:12308735–12318605_R</td>
<td>lincRNA</td>
<td>0.48</td>
<td>2.73</td>
</tr>
<tr>
<td>A_33_P3753757</td>
<td>LOC158402</td>
<td>other</td>
<td>0.53</td>
<td>2.46</td>
</tr>
<tr>
<td>A_33_P3393679</td>
<td>LOC645323</td>
<td>lincRNA</td>
<td>0.47</td>
<td>2.24</td>
</tr>
<tr>
<td>A_19_P00315649</td>
<td>linRNA:chr11:121899032–121899389_R</td>
<td>other</td>
<td>0.55</td>
<td>2.23</td>
</tr>
<tr>
<td>A_19_P00809838</td>
<td>linRNA:chrX:100247844–100257469_R</td>
<td>unassigned</td>
<td>0.79</td>
<td>2.20</td>
</tr>
<tr>
<td>A_19_P00331576</td>
<td>linRNA:chr3:114043485–114052926_F</td>
<td>unassigned</td>
<td>0.36</td>
<td>2.00</td>
</tr>
<tr>
<td>A_33_P3259557</td>
<td>LOC440104</td>
<td>pseudogene</td>
<td>0.45</td>
<td>1.67</td>
</tr>
<tr>
<td>A_19_P00319347</td>
<td>linRNA:chr2:168149680–168414843_F</td>
<td>lincRNA</td>
<td>0.77</td>
<td>1.67</td>
</tr>
<tr>
<td>A_19_P00320212</td>
<td>linRNA:chr1:114795825–114797203_R</td>
<td>other</td>
<td>0.41</td>
<td>1.63</td>
</tr>
<tr>
<td>A_24_P393703</td>
<td>LOC440104</td>
<td>pseudogene</td>
<td>0.40</td>
<td>1.50</td>
</tr>
<tr>
<td>A_19_P00318878</td>
<td>linRNA:chr1:247350513–247352101_R</td>
<td>lincRNA</td>
<td>0.38</td>
<td>1.44</td>
</tr>
</tbody>
</table>

(Continued)
Discussion

Our work provides a comprehensive view of the genome wide SOX2 regulated transcripts in GSCs, illustrating a complex scenario where SOX2 is the central player regulating different molecules and pathways in glioblastoma.

In this study, we used state-of-the-art microarray technology to query the SOX2 coding and non-coding RNA transcriptome in GSCs. It is interesting to note that among the down-regulated genes following SOX2 knockdown, F11R has been shown to be overexpressed in glioblastoma cells [42,43]. F11R is necessary and sufficient for GSC maintenance and self-renewal and of clinical significance is associated with increased malignancy and poor patient prognosis [42,43]. On the other hand, we found several interesting over-expressed candidates controlled by SOX2; for example, PPP1R1B is a well-known striatal projection neuron signature marker [44]. The fact that its expression increases following SOX2 inhibition is in line with its role in neuronal differentiation.

In a previous work where the SOX2 response program in a glioblastoma cell line was analyzed [15], authors identified 489 genes whose expression were altered in response to SOX2 knockdown, using several genomic technologies. Interestingly several of these genes are also differentially expressed in our array data, such as NGFR, CEBPA, BCL2, BNIP3, EBF4, ALCAM, protocadherins and solute carrier family members. Overall these results highlight the strength of our array data and the link between SOX2 and GSCs biology. The work of Fang and colleagues exhibits some similarities with our study, such as the analysis of SOX2 regulated-coding genes. However, we focused in the molecular circuitries controlled by SOX2 in GSCs, meanwhile they performed their study in an established glioma cell line. Additionally, our study addressed the SOX2-regulated IncRNAs in GSCs. Altogether both works provide clues regarding SOX2 functions in glioblastoma.

Gene-set enrichment analysis shows SOX2 is involved in regulating "cell adhesion", "biological adhesion", "cell-cell signaling", and "calcium ion binding" pathways, uncovering its key function as a driver of the glioma stem-like phenotype [45–48].

We also analyzed the canonical pathways regulated by SOX2 in GSCs. Pathways related with amino-acid metabolism were among the most deregulated, illustrating that SOX2 expression is critical for maintaining metabolic homeostasis in the GSC population, and plays important role in different tumor microenvironment conditions, such as hypoxic stress conditions.

**Table 6. List of the top-four IncRNAsregulated by SOX2, organized by p-value.**

<table>
<thead>
<tr>
<th>IncRNA</th>
<th>logFC</th>
<th>P.Value</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>lincRNA:chr6:72126155–72129954</td>
<td>-0.9188744</td>
<td>3.4E-07</td>
<td>7,3327381</td>
</tr>
<tr>
<td>lincRNA:chr6:29701971–29740296</td>
<td>-0.78724372</td>
<td>3.9E-07</td>
<td>7,2007586</td>
</tr>
<tr>
<td>lincRNA:chr19:28,281,401–28,284,848</td>
<td>-0.89247792</td>
<td>8.2E-06</td>
<td>4,265306</td>
</tr>
<tr>
<td>lincRNA:chr11:121899032–121899389</td>
<td>0.64633267</td>
<td>2.07E-06</td>
<td>5,6334606</td>
</tr>
</tbody>
</table>

This selection was evaluated by the presence of histone modifications and high abundance in brain according to UCSF genome browser tool.
Other enriched pathway altered in our analysis was the NOTCH pathway, where Hes5 and Hey1 had the most significantly down-modulated expression. Hes5 is a marker of neural multipotent progenitors with stem cell properties [50] where it sustains progenitors proliferative state inhibiting their differentiation into neurons [51]. On the other hand Hey1 has been related to a subset of molecules directly associated with hypoxia in glioblastoma tumors [52];
and might be used as a marker to distinguish glioblastoma patients with a relative good prognosis (negative Hey1 expression) [53]. Furthermore, Hey 1 is up-regulated in glioma samples correlating with tumor grade, and functionally its down-regulation results in a proliferation reduction [54], suggesting a role in the progression of glioblastoma. Taking all this into account, the canonical pathways more significantly altered after SOX2 inhibition are those related with intracellular signaling cascades and amino-acid metabolism pathways associated with tumor propagation.

Consistent with what is known about SOX2 biological function, our data-set is enriched with genes involved in morphology determination, development and cellular proliferation and migration. Interestingly, “proliferation of tumor cells” (S2 Table), is one of the most repeated subcategories for which IPA analysis assigned an activation Z score close to -2, predicting it’s inhibition, which is in line with a putative role of SOX2 in cell proliferation.

One of the most exciting aspects of our study involved expanding our knowledge of the SOX2 transcriptome into the realm of lncRNAs. In this study we showed and classified the lncRNA landscape regulated by SOX2 in GSCs. Our microarray results showed a strong correlation with published reports, demonstrating the strength of our approach and providing confidence that we can use this data-set for de novo discovery of novel SOX2 targets, including lncRNAs. One previous study determined the differentially expressed lncRNAs between glioblastoma and brain tissues, showing 654 IncRNAs upregulated and 654 down-regulated [55]. To our knowledge, this is the first study that evaluates the differentially expression of lncRNAs in GSCs controlled by SOX2.

Among the transcripts regulated by SOX2 we found that SOX2OT was down-regulated in our data set, even though we did not validate it. SOX2OT is a lncRNA which harbors SOX2 gene in its intronic region and is transcribed in the same orientation as SOX2 [56]. Several studies have demonstrated a role of SOX2OT in the regulation of SOX2 gene in human stem cells [57,58] although little is known about the exact role of this non-coding RNA. SOX2OT has been associated with carcinogenesis and, for example in breast cancer is involved in the induction and/or maintenance of SOX2 expression [59], in esophageal squamous cell carcinoma has been shown to play a role in tumor initiation and/or progression as well as in regulation of the pluripotent state of stem cells [58], and proliferation in lung cancer cells [60]. Askarian-Amiri et al demonstrated that SOX2OT has a positive effect on SOX2 expression [59]. Published data suggest the mediation of IncRNA SOX2OT in pluripotency and tumorigenesis events, probably through regulation of SOX2 expression. These data together with our own results suggest a possible role of SOX2OT in the malignant phenotype of glioblastoma, however further functional and mechanistic studies will be necessary to elucidate the precise role of SOX2OT and other lncRNA candidates in the tumorigenicity of glioblastoma.

Although advance in managing and treating glioblastoma have been made, tumor recurrence and treatment resistance remains the major cause of glioblastoma mortality. Understanding how factors such as SOX2 drive the glioblastoma tumor phenotype will aid in the development of new therapeutic approaches based on targeting GSCs. Our study integrates for the first time the coding and non-coding transcriptome controlled by SOX2 in GSCs, gaining new insights about the molecular circuitries governing glioblastoma biology.

**Conclusion**

In conclusion we have performed a comprehensive analysis of differential expression of coding and non-coding transcripts controlled by SOX2 in GSCs. We performed gene set enrichment analysis to find the most relevant pathways and biological functions altered in our data set. This integrated analysis allows for a better understanding of the SOX2 transcriptome in GSCs.
Supporting Information

S1 Table. Differentially expressed coding transcripts from microarray data. (XLS)

S2 Table. List of disease and biofunctions data from Ingenuity analysis. (XLS)

Acknowledgments

We are grateful to Dr. Yolanda Sánchez and Dr. John Laterra for critical reading of the manuscript.

Author Contributions

Conceptualization: JF CGM MMA AMAR.

Data curation: EG HLB.

Formal analysis: AMAR HLB MGH NMV EX.

Funding acquisition: AMAR MMA.

Investigation: AMAR HLB.

Methodology: AMAR HLB MGH.

Project administration: MMA.

Software: EG.

Supervision: HLB JF CGM MMA.

Validation: AMAR MGH NMV EX.

Writing – original draft: AMAR MMA.

Writing – review & editing: AMAR HLB EG MGH NMV EX JF CGM MMA.

References


sistance by preferential activation of the DNA damage response. Nature. 2006; 444: 756–760. PMID: 
17051156

8. Auffinger B, Spencer D, Pytel P, Ahmed AU, Lesniak MS. The role of glioma stem cells in chemother-
doi: 10.1586/14737175.2015.1051968 PMID: 26027432


recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. PLOS ONE. 2010; 5: 
e8960. doi: 10.1371/journal.pone.0008960 PMID: 20126410

1242. doi: 10.1038/ng.465 PMID: 19801978

499–507. doi: 10.1038/ng.127 PMID: 18443585

netic modifications of Sox2 contribute to the invasive phenotype of malignant gliomas. PLOS ONE. 
2011; 6: e26740. doi: 10.1371/journal.pone.0026740 PMID: 21991387

14. Fang X, Yoon JG, Li L, Yu W, Shao J, Hua D, et al. The SOX2 response program in glioblastoma mul-
tiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis. BMC Genomics. 


c2mb25386k PMID: 18948646

17. Jiang X, Yoon JG, Li L, Yu W, Shao J, Hua D, et al. The SOX2 response program in glioblastoma multi-
forme: an integrated ChIP-seq, expression microarray, and microRNA analysis. BMC Genomics. 

155–159. doi: 10.1038/nrg2521 PMID: 19188922

48: R45–53. doi: 10.1530/JME-12-0008 PMID: 22393162

3390/jms14034655 PMID: 23443164

coding RNA transcriptomes. PLOS ONE. 2011; 6: e25915. doi: 10.1371/journal.pone.0025915 PMID: 
21991387

48: R45–53. doi: 10.1530/JME-12-0008 PMID: 22393162

23. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in micro-

24. Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, Bioinformatics and computational biology solu-
NY: Springer; 2005.

709: 95–107. PMID: 21618891

3892/or.2015.3712 PMID: 25585997

Suppl 1: 19–21. PMID: 8732440

28. Yang S, Li WS, Dong F, Sun HM, Wu B, Tan J, et al. KITLG is a novel target of miR-34c that is associ-
ated with the inhibition of growth and invasion in colorectal cancer cells. J Cell Mol Med. 2014; 18: 
32. Minchenko DO, Kharkova AP, Huben'eva OV, Minchenko OH. Insulin receptor, IRS1, IRS2, INSIG1, INSIG2, RRAD, and BAIAP2 gene expressions in glioma U87 cells with ERN1 loss of function: effect of hypoxia and glutamine or glucose deprivation. Endocrin Regul. 2013; 47: 15–26. PMID: 23363253


