Genomic and Transcriptomic Characterization of the New Human Glioblastoma Cell Line AHOL1

Wallax Augusto Silva Ferreira¹; Carolina Koury Nassar Amorim¹; Rommel Rodriguez Burbano²,³,⁴; Rolando André Rios Villacis⁵; Fabio Albuquerque Marchi⁶; Tiago da Silva Medina⁶; Edivaldo Herculano Correa de Oliveira¹,⁷*.

¹ Laboratório de Cultura de Tecidos e Citogenética, SAMAM, Instituto Evandro Chagas, 67030-000, Ananindeua, PA, Brazil.

² Laboratório de Citogenética Humana, Instituto de Ciências Biológicas, Universidade Federal do Pará (UFPA), 66075-110, Belém, Pará, Brazil.

³ Núcleo de Pesquisas em Oncologia, Hospital Universitário João de Barros Barreto, 66.073-000, Belém, Pará, Brazil.

⁴ Laboratório de Biologia Molecular, Hospital Ophir Loyola, 66063-240, Belém, Pará, Brazil.

⁵ Department of Genetics and Morphology, Institute of Biological Sciences, University of Brasília-UnB, Brasília 70910-900, Brazil.

⁶ International Research Center (CIPE), A.C. Camargo Cancer Center, 01508-010, São Paulo, Brazil.

⁷ Instituto de Ciências Exatas e Naturais, Faculdade de Ciências Naturais, Universidade Federal do Pará (UFPA), 66075-110, Belém, Pará, Brazil.

*Corresponding author

Dr. Edivaldo Herculano Correa de Oliveira

Laboratório de Cultura de Tecidos e Citogenética, SAMAM, Instituto Evandro Chagas, Rodovia BR-316 km 7, s/n, Leivilândia, 67030-000, Ananindeua, Pará, Brasil.

E-mail: edivaldodeoliveira@iec.pa.gov.br
ABSTRACT

Cancer cell lines are widely used as in vitro models of tumorigenesis, facilitating fundamental discoveries in cancer biology and translational medicine. Currently, there are few options for glioblastoma (GBM) treatment and limited in vitro models with accurate genomic and transcriptomic characterization. Here, a detailed characterization of a new GBM cell line, namely AHOL1, was conducted in order to fully characterize its molecular composition based on its copy number alteration (CNA) and transcriptome profiling, followed by the validation of key elements associated with GBM tumorigenesis. Large numbers of CNAs and differentially expressed genes (DEGs) were identified. CNAs were distributed throughout the genome, including gains at Xq11.1-q28, Xp22.33-p11.1, Xq21.1-q21.33, 4p15.1-p14, 8q23.2-q23.3 and losses at Yq11.21-q12, Yp11.31-p11.2 and 12p13.31 positions. Nine druggable genes were identified, including HCRTR2, ETV1, PTPRD, PRKX, STS, RPS6KA6, ZFY, USP9Y and KDM5D. By integrating DEGs and CNAs, we identified 57 overlapping genes enriched in fourteen pathways. Altered expression of several cancer-related candidates found in the DEGs-CNA dataset was confirmed by RT-qPCR. Taken together, this first comprehensive genomic and transcriptomic landscape of AHOL1 provides unique resources for further studies and identifies several druggable targets that may be useful for therapeutics and biologic and molecular investigation of GBM.

Keywords: array-comparative genomic, gliomas, Cell culture, Cancer genomics, Cancer Transcriptomics, brain tumors, cell line, glioblastoma.
INTRODUCTION

Glioblastomas (also known as GBM or Glioblastoma Multiforme) are tumors that arise from the accumulation of somatic mutations in neural stem cells within the subventricular zone (SVZ) [1, 2]. They are the most lethal and common malignancy among all brain tumors, with incidence rate of 3.21 cases per 100,000 individuals, median survival rate of 12-18 months [3] and higher predominance in males [4]. GBMs are commonly diagnosed in elderly patients (median of 65 years) [3], increasing with age peaking at 75-84 years and declining after 85 years [4].

According to the new classification for Central Nervous System (CNS) tumors proposed by World Health Organization (WHO) in 2016, GBMs are classified as grade IV and included in diffuse astrocytic and oligodendroglial tumors [5]. Based on the mutational pattern of the isocitrate dehydrogenase (IDH) gene, they are further classified as (a) IDH-wildtype GBM (90% of cases), which frequently correspond to clinically defined primary GBM (or de novo GBM), arising predominately from supratentorial region in patients with median age of ~62 years at diagnosis and whose mean length of clinical history is 4 months; or (b) IDH-mutant-type GBM (10% of cases), which are defined as secondary GBM preferentially arising from the frontal region of younger patients (median age at diagnosis of ~44 years) and whose prognosis is usually better than those with IDH-wildtype [5-8].

Currently, standard treatment for both GBM entities encompasses surgical resection followed by radiotherapy (RT) and chemotherapy (mainly using temozolomide – TMZ) [9-12]. However, these aggressive treatments are not effective in controlling the disease [13, 14], thus evidencing a high demand for new efficacious therapies to improve outcomes of patients with GBM [15].

In vitro cultures of GBM cell lines have been widely used as an important model for understanding GMB heterogeneity, drug sensitivity and resistance, evaluation of new therapeutic approaches and to search for novel biomarkers [16-20]. The Human Glioblastoma Cell Culture (HGCC) biobank has assembled a panel of 53 cell lines derived from surgical samples of GBM patients [21]. However, there is a limited number of GBM cell lines deposited in HGCC or other biobanks (Broad-Novartis Cancer Cell Line Encyclopedia and American Type Culture Collection), given the heterogeneity of each molecular subtype of GBM [22-29]. Thus, new, well-characterized cell lines that resemble these different molecular subtypes of GBM are still needed to better comprehend the molecular mechanisms involved in GBM.
tumorigenesis. To this end, the goal of this study was to characterize the molecular composition based on copy number alteration (CNA) and transcriptome profile of a newly established glioblastoma cell line, as a strategy to discover potential druggable targets that might prevent GBM development and progression.

MATERIAL AND METHODS

Study approval by the Research Ethics Committee and collection of non-neoplastic samples

This study was approved by the Research Ethics Committee of the Instituto Evandro Chagas/IEC/SVS/MS (Process Number 192.336). The use of the sample to establish the AHOL1 cell line was approved by the Research Ethics Committee of the Ophir Loyola Hospital, and a written informed consent was obtained from the patient.

Ten non-neoplastic samples from nervous tissue were obtained from biopsies of patients from Ophir Loyola Hospital, Belém, Pará, Brazil. All tissue samples were immediately frozen in liquid nitrogen and stored in DNA/RNA Shield™ (Zymo Research) at -80°C until the extraction stage.

Materials

Dulbecco's modified Eagle's medium (DMEM), fetal bovine serum (FBS), Trypsin/EDTA, penicillin G and streptomycin were obtained from Gibco (Grand Island, NY, USA) and used to grow GBM cell lines in culture.

Culture of human glioblastoma cell lines

The main subject of this study was AHOL1 (Astrocytoma Ophir Loyola Hospital 1), a cell line established by our group at the Human Cytogenetics Laboratory, Federal University of Pará (UFPA), from a secondary GBM obtained from the tumor resection of a 43-year-old multiracial male patient treated in the Neurological Clinic of Ophir Loyola Hospital (Belém, PA, Brazil) with a histopathologic diagnosis of GBM (grade IV) that evolved from a grade III astrocytoma [30, 31]. As comparatives to common alterations observed in glioblastoma cell lines, we used three well stablished human glioblastoma cell lines in our experiments: U-343 MGa cell line kindly provided by the Cytogenetics and Mutagenesis Laboratory, University of São Paulo (Ribeirão
Preto, SP, Brazil), derived from a primary GBM from a Caucasian adult patient and obtained from CLS Cell Lines Service (CLS order number 300365) [32], U-87 MG purchased from American Type Culture Collection (ATCC; Manassas, VA, USA) (catalog number ATCC® HTB-14™) and 1321N1 cell obtained from European Collection of Authenticated Cell Cultures (ECACC) (catalog number 86030402).

All cell lines were cultured separately in 25-cm² culture flasks using DMEM supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin and 100 µg/mL streptomycin at 37 °C in a humidified 5% CO₂ atmosphere. The medium was changed every 2-3 days, and cells were sub-cultured when confluency reached 70-80% using 0.25% trypsin at 37°C.

**Nucleic acids extraction**

When cells reached total confluence, they were washed with PBS, detached with 0.25% Trypsin/EDTA (Gibco™), and suspended in PBS. DNA and RNA were extracted using the Wizard® Genomic DNA Purification Kit (Promega Corporation) and SV Total RNA Isolation System (Promega Corporation) respectively, according to the manufacturer’s protocol.

DNA and RNA purity and integrity were assessed on the Agilent 2200 TapeStation (Agilent Technologies, Santa Clara, CA, USA) with D1000 ScreenTape (Agilent Technologies, Santa Clara, CA, USA) and High Sensitivity RNA ScreenTape (Agilent Technologies, Santa Clara, CA, USA) respectively, as previously described [33-35]. Only samples with DNA Integrity Number (DIN) and RNA integrity (RIN) >7 were used for downstream analyses.

**Array-Based Comparative Genomic Hybridization Analysis**

**Chromosomal Imbalances Analysis**

Array-CGH (aCGH) experiments were performed on an Agilent microarray platform (Agilent Technologies Inc., Santa Clara, CA USA) with a SurePrint G3 Cancer CGH+SNP Microarray 4x180K slide (Agilent). Sample preparation, labelling, and microarray hybridization were performed according to the Agilent CGH Enzymatic Protocol version 7.5. Slides were scanned using the Agilent G2565CA scanner. Data were extracted with Feature Extraction software (v9.1 Agilent Technologies) and
analyzed with Genomic Work Bench 11.0.1, Agilent Cytogenomics 5.0 and GeneSpring GX 14.5. LogRatio > 0.25 and logRatio < -0.25 were defined as copy number gains and losses, respectively. The ideogram showing the CNAs identified in AHOL1 genome was constructed using the PhenoGram online software [36]. The CNAs information of 1087 cancer cell lines from CellMiner [37] and the Cancer Cell Line Encyclopedia (CCLE) [38], 1987 human GBMs from TCGA database [39-45], stored in the cBioPortal for Cancer Genomics (accessed in June 2019) [46], was used to explore the similarities with CNAs of AHOL1 cell line. Additionally, Candidate Cancer Gene Database (CCGD) was used to identify candidate cancer genes from CNAs of AHOL1 genome [47].

Gene Expression Microarray

Gene expression profiling analysis was performed using the Agilent Oligo Microarray Kit 8×60K according to the Agilent One-Color Microarray-based Gene Expression Analysis Protocol (Agilent Technologies, Santa Clara, CA, USA). Data were extracted with Feature Extraction software (v9.1 Agilent Technologies) and analyzed with GeneSpring software GX 14.5 (Agilent Technologies). Raw data were normalized by robust multiarray average (RMA) quantile normalization analysis algorithm with the GeneSpring GX 14.5 software (Agilent Technologies, Santa Clara, CA, US) to generate CEL intensity files. We performed the quality control following diagnostic plots: principal component analysis (PCA), boxplots, Pearson’s correlation, and MvA plots.

Significantly differentially expressed genes (DEGs) were identified by using the mixed model analysis of variance [11] with a false discovery rate (FDR) cut-off < 0.01 and absolute fold-change values ≥2. Gene ontology enrichment analysis was performed using DAVID (Database for Annotation, Visualization and Integration Discovery) Bioinformatic tools. Pathway enrichment analysis was performed by using the KEGG database and PANTHER (http://www.pantherdb.org/).

Integrative analysis of CNAs and DEGs

To identify the significant genes that exhibited CNA and gene expression alterations, we integrated the significant CNAs and DEGs using GeneSpring software GX 14.5 (Agilent Technologies) as described elsewhere. [48].
Search for drugs targeting CNAs

The Drug-Gene Interaction Database (DGIdb) [49, 50] was used to search potential drugs targeting CNAs. This database provides gene druggability information from databases (such as therapeutic target database [51]; DrugBank [52]; pharmacogenomics knowledge database [53]), papers and web resources [50].

Reverse transcription qPCR (RT-qPCR)

For the cDNA synthesis, we used the GoScript™ Reverse Transcription System (Promega Corporation) following the manufacturer's instructions. Real time PCR (qPCR) was performed as described by Ferreira et al. [54], using GoTaq® Probe qPCR Master Mix (Promega Corporation). All reactions were carried out in triplicate in 96-well PCR plates, using CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). Data analysis was performed using the Bio-Rad CFX Manager™ 3.1 software (Bio-Rad). Following the MIQE guidelines [55], the expression levels were normalized using TBP and GAPDH in non-neoplastic samples. The relative gene expression was calculated using the $2^{-\Delta\Delta C_T}$ formula ($p <0.05$) [56].

The expression of the genes ANOS1, ETV1, XPNPEP2 and PCDH11Y was quantitated using Taqman® gene expression assays (Applied Biosystems, Foster City, CA, USA) (Table 1).

Table 1. Targets and housekeeping genes used in this study.

<table>
<thead>
<tr>
<th>Official Gene Symbol*</th>
<th>Official Full Name*</th>
<th>Assay ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>Hs02786624_g1</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-box binding protein</td>
<td>Hs00427620_m1</td>
</tr>
<tr>
<td>ANOS1</td>
<td>Anosmin 1</td>
<td>Hs01085107_m1</td>
</tr>
<tr>
<td>ETV1</td>
<td>ETS variant 1</td>
<td>Hs00951951_m1</td>
</tr>
<tr>
<td>XPNPEP2</td>
<td>X-prolyl aminopeptidase 2</td>
<td>Hs00950918_m1</td>
</tr>
<tr>
<td>PCDH11Y</td>
<td>Protocadherin 11 Y-linked</td>
<td>Hs06651077_g1</td>
</tr>
</tbody>
</table>

*Official symbols and names of the genes were based on HUGO gene nomenclature committee (HGNC).

RESULTS

CNA profiling of AHOL1 cell line

A global view of the AHOL1 CNAs composition was generated using the a-CGH results. A total of 19 CNAs (17 gains and 2 losses) were identified, ranging from...
0.28 Mb to 93 Mb. A full list of the CNAs and their corresponding chromosome localization, cytobands, type of alteration, p value and genes are provided in Table S1. Copy number gains were located at chromosomes 4, 6, 7, 8, 9, 10, 11, 14, 17 and 19, while losses at chromosome 15 (Figure 1; Table S1). Whole chromosome gains and losses were observed at chromosomes X and Y, respectively.

The highest number of gains was found at chr X at q11.1-q28 (93,148.679 kb – with 3016 segments), p22.33 - p11.1 (55,799.123 kb – with 1745 segments) and q21.1 - q21.33 (13,545,889 kb) positions. The second highest number of gains was documented at chromosome 4 (cytoband p15.1 - p14) corresponding to 7,800 kb, followed by chromosome 8 (cytoband q23.2 - q23.3) corresponding to 5,104 kb.

Additionally, we identified that the highest frequency of losses occurred on the Y chromosome at q11.21 - q12 (45,327.039 kb) and p11.31 - p11.2 (7,119.122 kb) positions (Figure 1).

**Figure 1.** Representative CNAs (ideogram) showing gains (blue), and losses (red) in the AHOL1 genome.

We further explored whether AHOL1 CNAs were recurrent in cancer cell lines from CellMiner and Cancer Cell Line Encyclopedia (CCLE) databases (N=1087 cell lines) (≥ 10% of frequency). Indeed, in silico analysis revealed that thirty-one genes were covered by CNAs in most cancer cell lines (Table S2). In addition, by performing Ingenuity Pathway Analysis (IPA), it was revealed a total of 240 cancer-related genes,
15 of them exclusively related with brain cancer (Table S3). Also, we analyzed whether CNAs detected in AHOL1 were frequently found across primary GBM tumors from TCGA database. Interestingly, the vast majority of genes covered by CNAs were commonly altered across several primary GBM tumors.

We also conducted an analysis to detect the main altered pathways affected by CNAs in the AHOL1 genome. Amplifications affected 60 pathways, such as putrescine degradation III, melatonin degradation II and leucine degradation pathways, while losses have no impact in any pathway (Table S4).

**AHOL1 is genomically similar to other human GBM cell lines**

Considering that the AHOL1 cell line was established from a GBM patient, we expected it to share common CNAs with commercial GBM cell lines (U87MG, U343 and 1321N1). As shown in Figure 2, our results indicate the existence of a common genomic signature between AHOL1 and commercial GBM cell lines (1321N1, U343 and U87), thus confirming its GBM identity.

**Figure 2.** Heatmap of all CNAs present in each cell line (AHOL1, 1321N1, U343 and U87).

**Drugs targeting CNAs**

Potential drugs targeting CNAs were investigated using the DGIdb database. Nine druggable genes were identified, including *HCRTR2* (hypocretin receptor 2), *ETV1* (ETS variant 1), *PTPRD* (protein tyrosine phosphatase, receptor type D), *PRKX* (protein kinase X-linked), *STS* (steroid sulfatase), *RPS6KA6* (ribosomal protein S6 kinase A6), *ZFY* (zinc finger protein Y-linked), *USP9Y* (ubiquitin specific peptidase 9 Y-linked) and *KDM5D* (lysine demethylase 5D) (Table 2).
Table 2. Druggable genes of AHOL1 cell line. Information of each gene regarding their chromosomal location, cytoband, size (kb), type of alteration found (Gain or Loss), drugs and the type of interaction of each drug is provided below.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Cytoband</th>
<th>Size (Kb)</th>
<th>Variant type</th>
<th>Gene</th>
<th>Drug</th>
<th>Interaction Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>p12.1</td>
<td>1,752</td>
<td>Gain</td>
<td>HCRTR2</td>
<td>Suvorexant SB-649868</td>
<td>Antagonist</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lemborexant</td>
<td>Antagonist</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Almorexant</td>
<td>Antagonist</td>
</tr>
<tr>
<td>7</td>
<td>p21.3 - p21.2</td>
<td>3,393</td>
<td>Gain</td>
<td>ETV1</td>
<td>Trametinib</td>
<td>n/a</td>
</tr>
<tr>
<td>10</td>
<td>q21.1</td>
<td>3,864</td>
<td>Gain</td>
<td>PTPRD</td>
<td>Cucurbitacin</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Teprotumumab</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cixutumumab</td>
<td>n/a</td>
</tr>
<tr>
<td>X</td>
<td>p22.33 - p11.1</td>
<td>55,799</td>
<td>Gain</td>
<td>PRKX</td>
<td>GSK-690693</td>
<td>Inhibitor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
</tr>
<tr>
<td>X</td>
<td>q21.1 - q21.33</td>
<td>13,545</td>
<td>Gain</td>
<td>RPS6KA6</td>
<td>AT-9283 Chembl573107</td>
<td>Inhibitor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chembl573107</td>
<td>Inhibitor</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chembl573107</td>
<td>Inhibitor</td>
</tr>
<tr>
<td>Y</td>
<td>p11.31 - p11.2</td>
<td>7,119</td>
<td>Loss</td>
<td>ZFY</td>
<td>Chembl383208</td>
<td>n/a</td>
</tr>
<tr>
<td>Y</td>
<td>q11.21 - q12</td>
<td>45,327</td>
<td>Loss</td>
<td>USP9Y</td>
<td>Testosterone</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ascorbate</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Transcriptome characterization of AHOL1 cell line

In total, we identified 1,837 differentially expressed genes (DEGs). Among these, 713 genes were up-regulated, whereas 1,124 genes were down-regulated (FC ≥2 and p < 0.05) (Table 3). A full list of differentially expressed genes and their corresponding fold changes in expression and p values is provided in Table S5. Ret Finger Protein-like 4A-like 1 (RFPL4AL1) was the most up-regulated (FC: 55.87), and lincRNA lnc-CHIC1-2:1 the most down-regulated mRNA (FC: -65.82) (Table 3).

Table 3. Characteristics of the top 20 differentially expressed mRNAs in AHOL1 cell line sorted by fold change (Fold change ≥2 and p < 0.05).
<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Chromosome</th>
<th>Description</th>
<th>Fold-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFPL4AL1</td>
<td>chr19</td>
<td>Ret Finger Protein-like 4A-like 1</td>
<td>55.87</td>
</tr>
<tr>
<td>lnc-WDR5-2</td>
<td>chr9</td>
<td>lncRNA</td>
<td>27.04</td>
</tr>
<tr>
<td>ERICH1-AS1</td>
<td>chr8</td>
<td>lncRNA</td>
<td>26.70</td>
</tr>
<tr>
<td>LRRN4</td>
<td>chr7</td>
<td>Leucine Rich Repeat Neuronal 4</td>
<td>21.44</td>
</tr>
<tr>
<td>CMTR1</td>
<td>chr6</td>
<td>Cap Methyltransferase 1</td>
<td>21.38</td>
</tr>
<tr>
<td>XLOC_12_012743</td>
<td>chr6</td>
<td>lncRNA</td>
<td>11.53</td>
</tr>
<tr>
<td>LINC01297</td>
<td>chr14</td>
<td>lncRNA</td>
<td>10.86</td>
</tr>
<tr>
<td>KDM4E</td>
<td>chr11</td>
<td>Lysine demethylase 4E</td>
<td>9.08</td>
</tr>
<tr>
<td>lnc-RBPJ-1:1</td>
<td>chr4</td>
<td>lncRNA</td>
<td>8.59</td>
</tr>
<tr>
<td>lnc-FBXO25-5:4</td>
<td>chr8</td>
<td>lncRNA</td>
<td>6.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Chromosome</th>
<th>Description</th>
<th>Fold-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnc-CHIC1-2:1</td>
<td>chrX</td>
<td>lncRNA</td>
<td>-65.82</td>
</tr>
<tr>
<td>SHANK3</td>
<td>chr22</td>
<td>SH3 and multiple ankyrin repeat domains 3</td>
<td>-36.71</td>
</tr>
<tr>
<td>SNORD114-3</td>
<td>chr14</td>
<td>Small nucleolar RNA, C/D box 114-3</td>
<td>-18.68</td>
</tr>
<tr>
<td>LOC403323</td>
<td>chr9</td>
<td>Uncharacterized LOC403323</td>
<td>-12.19</td>
</tr>
<tr>
<td>ZFP57</td>
<td>chr6</td>
<td>ZFP57 zinc finger protein</td>
<td>-10.78</td>
</tr>
<tr>
<td>CYP11A1</td>
<td>chr15</td>
<td>Cytochrome P450 family 11 subfamily A member 1</td>
<td>-9.88</td>
</tr>
<tr>
<td>DSP</td>
<td>chr6</td>
<td>Desmoplakin</td>
<td>-9.16</td>
</tr>
<tr>
<td>PLVAP</td>
<td>chr19</td>
<td>Plasmalemma vesicle associated protein</td>
<td>-8.89</td>
</tr>
<tr>
<td>C4BPA</td>
<td>chr1</td>
<td>Complement component 4 binding protein alpha</td>
<td>-8.72</td>
</tr>
<tr>
<td>lnc-ZNF100-2</td>
<td>chr19</td>
<td>lncRNA</td>
<td>-7.92</td>
</tr>
</tbody>
</table>

To better understand the biological processes associated with DEGs, Gene Ontology (GO) analysis was conducted. The majority of DEGs was distributed into four GO categories: biological process, molecular function, cellular component and protein class (Figure 3). A full list of GO terms is provided in Table S6.

In the biological process category, the most enriched terms were related to biological regulation and metabolic process (Figure 3A). Binding and catalytic activity mostly accounted for terms related to the molecular function category (Figure 3B). Within the cellular component category, the GO term with the highest level of significance was cell, followed by membrane and organelle (Figure 3C). Finally, in the protein class category, the terms hydrolase and transcription factor exhibited the highest significance (Figure 3D).
Figure 3. Gene Ontology (GO) functional annotations for DEGs of AHOL1 cell line. Bar graphs show four GO categories: (A) Biological processes, (B) Molecular functions, (C) Cellular components and (D) Protein classes. The X-axis represents the number of DEGs and the Y-axis shows the GO terms of each category (Fold change ≥2 and p<0.05).

Pathway analysis was performed to investigate the biological significance of these DEGs. Thirteen pathways were significantly affected in the AHOL1 cell line, such as mevalonate, IL-1, glycogenolysis and mRNA capping pathways in cancer (Table 4).

Table 4. Pathway analysis of DEGs from AHOL1 cell line. Pathways were selected according to the P value (Fold change ≥2 and p<0.05).

<table>
<thead>
<tr>
<th>Pathways</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trans, trans-farnesyl diposphate biosynthesis</td>
<td>0.0262</td>
</tr>
<tr>
<td>Superpathway of geranylgeranyldiphosphate biosynthesis I (via mevalonate)</td>
<td>0.0391</td>
</tr>
<tr>
<td>Mevalonate pathway</td>
<td>0.0391</td>
</tr>
<tr>
<td>IL1</td>
<td>0.0397</td>
</tr>
<tr>
<td>4-hydroxy-2-nonenal detoxification</td>
<td>0.0414</td>
</tr>
<tr>
<td>Glycogenolysis</td>
<td>0.0311</td>
</tr>
<tr>
<td>mRNA capping</td>
<td>0.0417</td>
</tr>
<tr>
<td>Pathways in cancer</td>
<td>0.0091</td>
</tr>
<tr>
<td>C20 prostanoid biosynthesis</td>
<td>0.0169</td>
</tr>
</tbody>
</table>
Plasmalogen degradation 0.0229
Rapoport-Luebering glycolytic shunt 0.0229
Pyridoxal 5'-phosphate salvage 0.0229
Choline degradation 0.0229

**AHOL1 cell line shares several DEGs with other commercial GBM cell lines**

To determine the number of genes shared between AHOL1 and the commercial GBM cell lines, Venn diagrams were created. Our results showed that AHOL1 shared several transcripts with the commercial GBM cell lines.

The analysis of transcriptomes highlighted that AHOL1 cell line has several changes common to the different GBM commercial cell lines. All cell lines had shared 756 up-regulated and 281 down-regulated genes (Figure 4).

![Venn diagram showing the number of genes up-regulated (a) and downregulated (b) in all GBM cell lines.](image)

**Integrative Analysis of CNA and Gene Expression Profiling**

To explore and compare how CNAs affect AHOL1 transcriptional program, we performed an integrative analysis of CNAs with DEGs. Most of DEGs showed no positive correlation with CNAs. Only fifty-six genes displayed significant CNA-DEGs correlation in AHOL1 cell line. A full list of these overlapping genes is found in Table S7. Fourteen genes showed a positive association, while 42 genes presented an inverse association (Table 5).

**Table 5.** Overlapping genes identified after integration of CNAs and gene expression data of AHOL1 cell line.

<table>
<thead>
<tr>
<th>CNA</th>
<th>Gene expression</th>
<th>Overlapping genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>Up-regulated</td>
<td>13</td>
</tr>
<tr>
<td>Gain</td>
<td>Down-regulated</td>
<td>32</td>
</tr>
<tr>
<td>Loss</td>
<td>Up-regulated</td>
<td>11</td>
</tr>
<tr>
<td>------</td>
<td>--------------</td>
<td>----</td>
</tr>
<tr>
<td>Loss</td>
<td>Down-regulated</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>57</td>
</tr>
</tbody>
</table>

To further understand the biological function of these 57 overlapping genes, they were submitted to functional annotation and classification analysis. Most of these genes were enriched for metabolic processes, biological regulation and cellular processes in the biological category (Figure 5A). For the molecular function category, the GO terms with the highest levels were catalytic activity, binding and transcription activity (Figure 5B). Within the cellular component category, cell was the highest term, followed by membrane and organelle terms (Figure 5C). Of note, in the protein class category, the terms nucleic acid binding and transcription factor exhibited the highest frequencies (Figure 5D).

Next, we performed separately GO functional annotation of each group shown in Table 7. As shown in Table S8, most genes covered by gains whose expression was up-regulated mainly affected the nucleus (Cellular Component), provoking changes in the DNA-binding transcription factor (TF) activity (Molecular Function) of the helix-turn-helix and zinc finger TFs (Protein Class), which consequently had an effect on the cell development and cell differentiation (Biological process).
Genes covered by gains and whose expression was down-regulated mostly affected the membranes (Cellular Component), especially disturbing the hydrolase activity (Molecular Function), modifying the cell communication, signal transduction, cellular response to stimulus and cellular metabolic process (Biological Process).

Furthermore, genes covered by losses and whose expression was up-regulated essentially affected the cytoplasm (Cellular Component), altering the RNA binding (Molecular Function) of the ribosomal proteins and translation factors (Protein Class), influencing the translational elongation and termination (Biological Process). Finally, *PCDH11Y* loss and down-regulation affected plasma membrane, by altering the cell adhesion and inducing changes in the calcium ion binding (Table S8).

**Figure 5.** Functional classification of integrated DEGs-CNAs data from AHOL1 cell line using Gene ontology (GO) analysis. Bar graphs show four independent GO information categories: (A) Biological processes, (B) Molecular functions, (C) Cellular components and (D) Protein classes. The Y-axis represents the number of genes of integrated DEGs-CNAs data and the X-axis shows the GO terms for each category (p<0.05).
We also conducted a global analysis with all 57 genes from the DEGs-CNAs dataset, in order to detect the main altered pathways from AHOL1. Signaling pathways via Receptor-type tyrosine protein phosphatases, Cadherin, Wnt, FGFR1 and Akt, as well as metabolism of proteins were statistically significant.

**Gene expression Validation of the Microarray Results**

To validate the DEGs-CNAs dataset generated by microarray, we selected four genes related to cancer. Three of them (ANOS1, ETV1 and XPNPEP2) were up-regulated via copy number gain and one gene (PCDH11Y) was down-regulated via copy number loss. Transcription levels of selected genes are shown in Figure 6.

Gene expression showed that ANOS1, ETV1 and XPNPEP2 were up-regulated in AHOL1 relative to normal brain tissue fragments (2-fold; >4 fold; 2.7-fold respectively), which is in line with the microarray data (Figure 6 and Table S5). Of note, PCDH11Y was 2.3-fold down-regulated in AHOL1 relative to normal brain tissues, in agreement with the microarray data (Figure 6D; Table S5).
Figure 6. Validation of the microarray data by qRT-PCR. qRT-PCR was performed for *ANOS1 (A)*, *ETV1 (B)*, *XPNPEP2 (C)* and *PCDH11Y (D)* genes. Data were expressed as fold-change in mRNA expression and compared with a tissue fragment from normal brain as a control. Each bar represents mean ± SD values for each gene with three technical replicates. Statistical significance was determined using student's t test and denoted by asterisks: * indicates p < 0.05; ** indicates p < 0.01; *** indicates p < 0.001.

DISCUSSION

Recent data provided by The Cancer Genome Atlas (TCGA) consortium have shown great genetic and epigenetic diversity among GBM tumors [5, 23-28]. The elucidation of the biological mechanisms and the complexity behind this diversity is a central challenge for achieving precision medicine in GBM subtypes. For this reason, the availability of a large number of molecularly well-characterized GBM cell lines may have a high impact on understanding the complex biology of this tumor, thus contributing to identification of new therapeutic targets.

As an attempt to better understand the complexity of the GBM biology and the phenomena that pervade the aspects of its intricate mechanisms, we provide the first comprehensive molecular profile of the new human glioblastoma cell line AHOL1. This cell line clearly exhibited recurrent CNAs consistent with those observed in primary GBM tumors from TCGA database, CellMiner and Cancer Cell Line Encyclopedia, as well as the commercial GBM cell lines (1321N1, U343 and U87) used in this study and other reports [57-67]. Among these, the most prominent CNAs were documented at X (gain of entire chromosome) and Y-chromosomes (loss of entire chromosome), in line with previous cytogenetics and molecular reports for GBM tumors [65, 68-70]. Although many functional aspects of CNAs at chr X and chr Y have not yet been extensively studied in cancer, a preliminary study suggested that high expression of *SPIN4* and *ASB12* was correlated with amplification at Xq11 in GBM [71], thus affecting several important biological processes, such as mitosis, Wnt signaling pathway, H3K4me3 and post-translational protein modifications [72]. Recent reports also support that the X chromosome gain plays a prominent role in the neoplastic
transformation of breast cancer [73-78], chronic neutrophilic leukemia [79], non-Hodgkin lymphoma [80, 81], papillary renal cell carcinoma [82], renal cell carcinoma [83] and prostate cancer [84].

Massive amounts of cancer genomics data generated from next-generation sequencing have motivated investigators to develop novel computational approaches for the identification of new druggable genes, which can be used as therapeutic targets in the precision cancer medicine [85]. Our AHOL1 genomic data showed that nine genes (HCRTR2; ETV1; PTPRD; PRKX; STS; RPS6KA6; ZFY; USP9Y and KDM5D) are known as potential druggable anticancer targets, and some of them could be druggable vulnerabilities in some subtypes of GBM [86]. Together, these studies have highlighted the great potential of AHOL1 cell line for in vitro studies.

The association between CNAs and gene expression profiling suggested that the identified CNAs could contribute to the expression of some but not all genes. In some cases, expression changes were inconsistent with the CNAs. This might be influenced by other factors that contribute to gene expression variation, such as epigenetics changes, non-coding RNA regulation, gene mutation and altered expression of TFs [87]. This apparent discrepancy between copy number status and gene expression has also been observed in other cancers [88-93].

In this study, pathway enrichment of CNA-driven DEGs indicated significant changes in six pathways, where the majority of genes were prominently involved in the signaling by tyrosine phosphorylation. Tyrosine phosphorylation plays an important role in regulating cellular function and is a central feature in signaling cascades involved in oncogenesis [94, 95]. This process is coordinately controlled by protein tyrosine phosphatase (PTPs) and protein tyrosine kinases (PTKs) [95], which are altered in a variety of human cancers [96, 97], including GBMs [98-106]. Among the tyrosine phosphatases, the tumor suppressor PTPRD is one of those often inactivated by deletion (>50% of cases), whose loss of expression promotes gliomagenesis through aberrant STAT3 activation [107] and is related with poor prognosis in GBM patients [98]. Our observations, therefore, suggested that this gene was amplified and down-regulated in AHOL1. Perhaps, this gene might be epigenetically silenced in AHOL1, once GBMs that do not harbor loss of PTPRD have this gene inactivated by its promoter hypermethylation [98, 108].

Assuming that CNA-DEG integrated data have been shown to be an efficient approach to identify genes covered by altered CNAs that directly change their
expression levels, not all genes of our integrated dataset were cancer-related [89]. In order to explore only genes related to cancer, we selected four genes, including three positively correlated genes with gain (ANOS1, ETV1 and XPNPEP2) and one negatively correlated gene with loss (PCDH11Y).

ANOS1, also known as KAL1, encodes anosmin-1, an extracellular matrix (ECM)-associated protein that plays essential roles in neural cell adhesion and axonal migration. The upregulation of this gene in the AHOL1 cell line is consistent with the results from Choy et al. [109] for GBM and low-grade astrocytic tumors, as they found that anosmin-1 enhanced cell motility and proliferation in GBM cell lines. The overexpression of ANOS1 is also related to development and metastasis of colorectal cancer and its expression is closely related to the overall survival rate of patients [110, 111].

We also identified that the up-regulation of ETV1, an important oncogene involved in transcriptional regulatory processes in GBMs [112], was due to a chromosome 7 gain in AHOL1 genome. It has been shown that chromosome 7 gain is a key event in G-CIMP-negative GBMs [23, 113], and high expression of ETV1 may be a potential downstream effector of chromosome 7 gain [112]. Previous study also revealed that ETV1 is fused with DGKB and may act as oncogenic drivers in pediatric high-grade gliomas [114].

XPNPEP2 is a membrane-bound aminopeptidase P member of the 'pita bread fold' family, which catalyzes the removal of the penultimate prolyl residue from N-termini of peptides. Although it is known that this aminopeptidase activates growth factors, hormones, coagulants, toxins, cytokines and neurotransmitters [115], the role of XPNPEP2 in cancer is still unknown. It is known that XPNPEP2 is overexpressed in cervical cancer, promoting cell invasion and migration without affecting cell proliferation and apoptosis [116]. Our gene expression results corroborate these findings and point out that XPNPEP2 can be up-regulated in GBM.

Protocadherins constitute the largest subfamily of cadherins in the genome. Protocadherins genes are predominantly expressed in the nervous system, acting in crucial functions associated with formation, maintenance, and integrity of the neural circuit [117, 118]. Recently, they have been in the spotlight for their roles in cancer [119]. Protocadherin-11Y (also named as PCDH11Y) is a proto-oncogene candidate exclusively found in man [120] and its transcription occurs mainly in the brain [118]. This gene is up-regulated in prostate carcinoma [121], inducing neuroendocrine
transdifferentiation through activation of the wnt signaling [122]. To the best of our knowledge, this is the first study showing that PCDH11Y is overexpressed in GBM, contradicting the results found elsewhere [121].

In summary, we described for the first time the genome and transcriptome of the new human cell line AHOL1, established from a GBM patient. Here, we revealed that this cell line harbors a genomic alteration spectrum similar to what is observed in commercial GBM cell lines (U87MG, U343 and 1321N1) and GBMs from the TCGA database, and some of these CNAs can be targeted by drugs, suggesting that this new cell line is a suitable model system for understanding the molecular characteristics of human GBM tumorigenesis.

Author Contributions


Funding

This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Grant Number: 455078/2014-9) and Evandro Chagas Institute.

Acknowledgments

Authors are grateful to Instituto Evandro Chagas (Ananindeua, PA, Brazil), Pró-Reitoria de Pesquisa e Pós-Graduação (PROPESP/UFPA, Belém, PA) and CNPq for financial and technical support; to the Hospital Ophir Loyola (Belém, PA) and the patient’s family for the consent to participate in this project; to Dr. Rommel Rodriguez Burbano for providing the AHOL1 cell line used in this study; to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for Wallax Augusto Silva Ferreira’s Phd fellowship.
**Conflict of interest**

The authors declare that they have no conflict of interest.

**Abbreviations**

AHOL1: Astrocytoma Ophir Loyola Hospital 1  
CNAs: Copy number alterations  
CNS: Central Nervous System  
DEGs: Differentially expressed genes  
GBM: Glioblastoma  
IDH: isocitrate dehydrogenase  
SVZ: subventricular zone  
WHO: World Health Organization
References


34. Kong, N.; Ng, W.; Cai, L.; Leonardo, A.; Kelly, L.; Weimer, B., Integrating the DNA integrity number (DIN) to assess genomic DNA (gDNA) quality control using the Agilent 2200 TapeStation system. *Agilent Technologies Application Note. doi* 2014, 10.

35. Padmanaban, A.; Salowsky, R.; Cher, C., RNA quality control using the agilent 2200 TapeStation system—assessment of the RIN e quality metric. *Agilent technologies application notes* 2012.


devoted to the understanding of human disease through the clinical application of molecular biology 2001, 6, (2), 93-108.


