Glioblastoma and calcium signaling - analysis of calcium toolbox expression

NOÉMIE ROBIL[1,2], FABIEN PETEL[1], MARIE-CLAUDE KILHOFFER[2] and JACQUES HAIECH[2]

[1] Programme Cartes d'Identité des Tumeurs (CIT), Ligue Nationale contre le Cancer, Paris, France and [2] Laboratoire d'Innovation Thérapeutique, Laboratoire d'Excellence Médalis, UMR 7200 Université de Strasbourg / CNRS, Illkirch, France

ABSTRACT The characteristics of a cellular calcium signal (calcium signature) are determined, at least partly, by the expression of a subset of genes encoding proteins involved in calcium entry, calcium uptake and calcium modulation. Our aim in the present work was to characterize the set of genes involved in calcium signal generation that are differentially expressed in normal brain tissues versus brain tumor and/or glioma stem cells. Public datasets were analyzed according to a four step methodology consisting of: 1. detecting the outliers by using principal component analysis of the whole transcriptome; 2. building a calcium toolbox composed of 260 genes involved in the generation and modulation of the calcium signal; 3. analyzing the calcium toolbox transcriptome of different human brain areas and 4. detecting genes from the calcium toolbox preferentially expressed in tumor tissues or tumor cells compared to normal brain tissues. Our approach was validated on normal brain tissue. Tumor sample analysis allowed us to disclose a set of eighteen genes characteristic of glioblastoma tissues or glioma stem cells. Interpreting the set of genes highlighted in the study led us to propose that i) the mechanism of store operated calcium entry is strongly perturbed in cancer cells and tissues, ii) the process of calcium reuptake into mitochondria is more important in cancer cells and tissues than in their normal counterparts and iii) these two mechanisms may be coupled in at least one subgroup of the glioblastoma stem cells.

KEY WORDS: *calcium signal, glioblastoma, cancer stem cell, calcium toolbox*

Introduction

Glioblastoma is rare, but one of the most deadly adult tumors. Similarly to other solid tumors, it can be considered as organ-like, being composed of differentiated cancer cells that exhibit low tumor genesis capacity and cancer stem cells that are able to resume the tumor after treatment of the patient tumor by surgery, chemotherapy and radiotherapy (Levy *et al.*, 2014). Normal cells from the host also infiltrate the tumor which therefore appears as a heterogeneous cellular mass behaving as a microsociety. In addition, the tumor cells appear to exhibit an exquisite plasticity with differentiated cells able to recover stemness properties, a characteristic that is reported as a hallmark of their tumorigenicity (Chaffer and Weinberg, 2015).

Calcium is considered as the main second messenger in eukaryotic cells. Previous and recent papers have established a generally accepted vocabulary to describe the different elements involved in setting up a specific calcium signal in a given cell and in a given physiological state (for review see (Haiech *et al.*, 2011)). The human genome contains more than 250 genes coding for such elements. This ensemble of genes constitutes the calcium toolbox. In a given cell, a subset of the calcium toolbox will be expressed (forming the cell calcium signalosome). Given signalosome proteins assemble to form macromolecular complexes, called calcisomes. Such calcisomes are going to be key elements to encode external information into a calcium signal and then to decipher this calcium signal in order to transduce the external information into cellular events (Haiech *et al.*, 2011).

In cancer cells, such cellular events may represent any of the hallmarks described by Weinberg (Hanahan and Weinberg, 2011)

Address correspondence to: Jacques Haiech. Laboratoire d'innovation thérapeutique, UMR 7200 University of Strasbourg / CNRS, 74, route du Rhin, 67401 Illkirch, France. Tel: +33-3-6885-4270. Fax: +33-3-6885-4310. E-mail: haiech@unistra.fr - web: http://www-medchem.u-strasbg.fr

for solid tumors, namely metabolism regulation, cell cycling, differentiation, apoptosis and more generally cell death, genome integrity, host immunological system evasion, cell mobility, angiogenesis …

Taking this rationale, the aim of the paper was to compare the calcium signalosomes in glioblastoma tissue and cancer stem cells isolated from glioblastoma to normal brain tissues. To take into account biological variability (the tumor and cell heterogeneity), human variability (annotations errors) and technical variability (differences in experimental protocols), we developed a workflow for this meta-analysis by appreciating the biological coherency of the results obtained.

The analysis was divided into four steps:

1. To remove the putative outliers from our set of data, the overall homogeneity of the different datasets was tested using a principal component analysis (Shlens, 2014) on the complete transcriptomic data associated with each sample,

2. A human calcium toolbox focused on the genes encoding proteins involved in calcium signal generation and modulation (calcium entry, calcium outflow and calcium binding proteins involved in the modulation of calcium fluxes) was built. We checked that the selected set of genes was able to discriminate the different groups of samples, indicating that the calcium signalosome is a specific signature for a given tissue or a specific cell.

3. Using public data showing gene expression profiles of different human brain areas, we optimized a workflow to extract a transcriptomic signature characteristic of the brain tissue and signatures specific to the different brain areas. The signatures obtained were coherent with the current knowledge of brain biology.

4. A clustering method was then used to define homogenous groups of glioblastoma, non-glioblastoma brain cancers, glioma cancer cell lines, glioblastoma and normal human embryonic stem cells. Those groups were added to the set of brain tissues and the previous analysis (step 3) was redone. By that way, common calcium signalosome signatures (genes expressed in all samples)

TABLE 1

PUBLIC DATASETS USED IN THE PRESENT ANALYSIS

Datasets	Publication	Samples used in our analysis
GSE7181	(Beier *et al.*, 2007)	6 glioblastoma stem cells
GSE23806	(Schulte *et al.*, 2011)	36 glioma cell lines, 27 glioblastoma stem cells, 12 GBM
GSE18015	(Garcia *et al.*, 2010)	Cells isolated from gliomas, 8 with CD133+ and 8 with CD133-
GSE7307		677 samples from normal and diseased human tissues. Only normal neural tissues were used (229 samples)
GSE4290	(Sun *et al.*, 2006)	157 tumor samples, including 26 astrocytomas, 50 oligodendrogliomas and 81 GBM.Samples from epilepsy patients were not used
GSE21514	(Moser and Fritzler, 2010)	2 human astrocytes samples
GSE17312	(Bernstein *et al.*, 2010)	4 embryonic stem cell samples
GSE20126	(Fong *et al.*, 2011)	4 embryonic stem cell lines
GSE34200	(Mallon *et al.*, 2013)	12 embryonic stem cell lines
GSE39762	(Aldaz *et al.*, 2013)	3 embryonic stem cell lines
GSE44841	(Aldaz *et al.*, 2013)	8 glioblastoma stem cells
GSE46016	(Rheinbay *et al.*, 2013)	10 glioblastoma stem cells
GSE46531	(Ye *et al.*, 2013)	12 glioblastoma stem cells
GSE51822	(Zorniak *et al.*, 2015)	2 glioblastoma stem cells

and some specific signatures (genes mostly present in a small subset of samples) could be highlighted.

The results of this meta-analysis are discussed in the paper. They constitute a starting point to understand the differences between calcium signals in various tissues or cells. Moreover, this analysis is also a starting point to decipher the calcium signaling pathways prevailing in normal and tumor cells.

Results and Discussion

Outliers and description of different groups

To perform our meta-analysis, several datasets from the GEO database (see Table 1) were selected.

To begin, we decided to use the following groups:
- Normal brain tissue
- Brain cancer: all brain cancers except the glioblastoma tumors
- Glioblastoma tumors (GBM)
- Glioblastoma cell lines, meaning the cell lines derived from GBM tumors
- Glioblastoma cancer stem cells (gCSC), meaning cancer stem cells isolated from GBM tumor biopsies
- Primary human astrocytes from ATTC (HA)
- Embryonic stem cell lines (ESC).

Principal Component Analysis (PCA) aims at transforming a number of possibly correlated variables into a smaller number of variables called principal components. In our case, the variables are either all the genes of the human genome (more than 20000 genes) or the genes of the calcium toolbox (260 genes, Supplementary materials table 1). One sample is therefore a point in a space of dimension 20 000 in one case or 260 in the other case. The aim of a Principal Component Analysis is to find a plane in this space where the representation of the samples conserves the overall shape seen in the Multidimensional space. Therefore, an examination of the reduced dimension dataset allows the user to spot data outliers, far more easily than would have been possible without performing the principal component analysis.
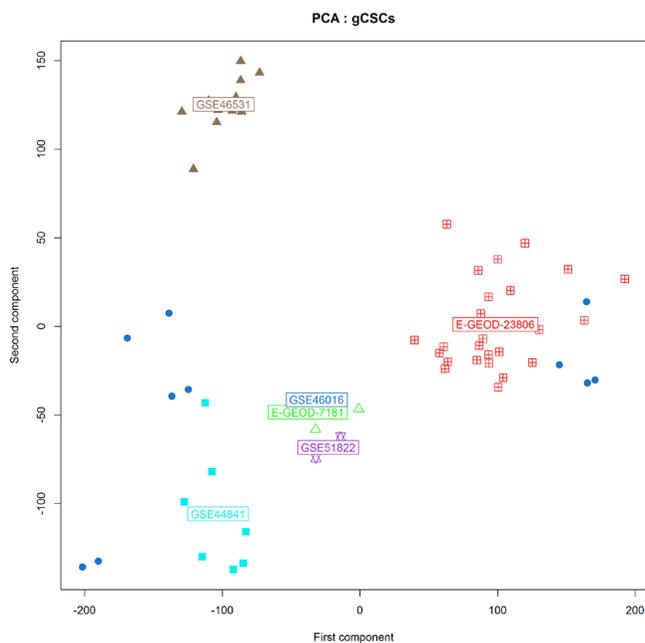

Fig. 1. Principal Component Analysis (PCA) of glioblastoma cancer stem cells (gCSC) from the different datasets presented in Table 1.
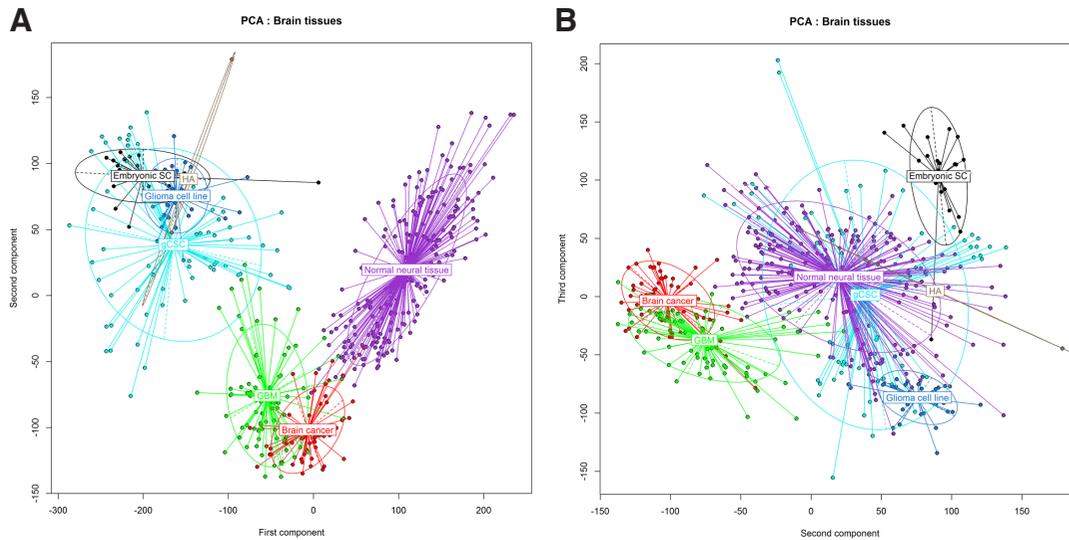
A

B

Fig. 2. Principal Component Analysis (PCA) of different brain tissue samples and embryonic stem cells (ESC).
(A) *First and second components and* **(B)** *second and third components.*

The principal component analysis will look for a new set of orthogonal axes. This new set of axes is ranked based on the percentage of sample variance held by each axis. The first axis is the first principal component and so on. In general, the plane used to project the samples is composed of the first and second principal component. However, if the first component held more than 50% of the variance of the sample, it may be interesting to use the plane defined by the second and the third principal components.

The coherency of the sample annotations was determined by a principal component analysis (PCA) on the transcriptomic data. This allowed to visualize the different groups and to detect the putative outliers. Outliers were defined as samples of a specific group closer to the barycenter of another group.

Performing this grouping, 18 brain cancer samples among 76 were found to be closer either to the GBM group (15 samples) or to normal brain tissue group (3 samples). Out of the 93 GBM samples, 13 were closer to the brain cancer group and 3 to normal brain tissue group. When brain cancer or GBM samples were found to be closer to normal brain tissues, we may hypothesize that the tumor biopsies were strongly contaminated with normal brain tissue. It was not possible to get information to confirm or infirm the proposed public annotations. If true, that would estimate the error rate of brain tumor annotations to be around 15 to 20 %. As a consequence, outliers were removed from our analysis without trying any re-annotation.

No outliers were found among the cancer cell lines and the ESC groups. Concerning the gCSC, 16 were detected as outliers. Among them, 11 belong to the same dataset (GSE18015 in Table 1). The entire dataset (16 samples) was therefore removed. Among the 5 other samples found as outliers, 4 were closer to the cancer cell line group and one to the GBM group.

Leaving apart the dataset GSE18015 that comprised 16 samples, the annotation error rate was around 6% pointing to a strong homogeneity in the gCSC group. Those cell lines were isolated in different laboratories using slightly different protocols. The cell lines isolated at the department of Medicine of the University of Salamanca (GSE18015) appeared different from all the other cell lines, but no technical reason could be found to explain the observation.

To better visualize the homogeneity of the group, we plotted a Principal Component Analysis of gCSC alone (Fig. 1) after outlier removal. Samples from the different datasets were mixed together in the PCA, except those from GSE46531 that clustered apart from the others.

In the Principal Component Analysis obtained using all the samples except the outliers, the first component accounted for more than 50 % of the information, indicating that most of the data variance was retained in this component. Therefore to better visualize the different groups, results of the PCA were also presented with the second and third components (Fig. 2A and 2B) (Shlens, 2014).

After outlier removal, the different sets of samples were considered homogenous enough. Therefore, the sensibility of the used analytical
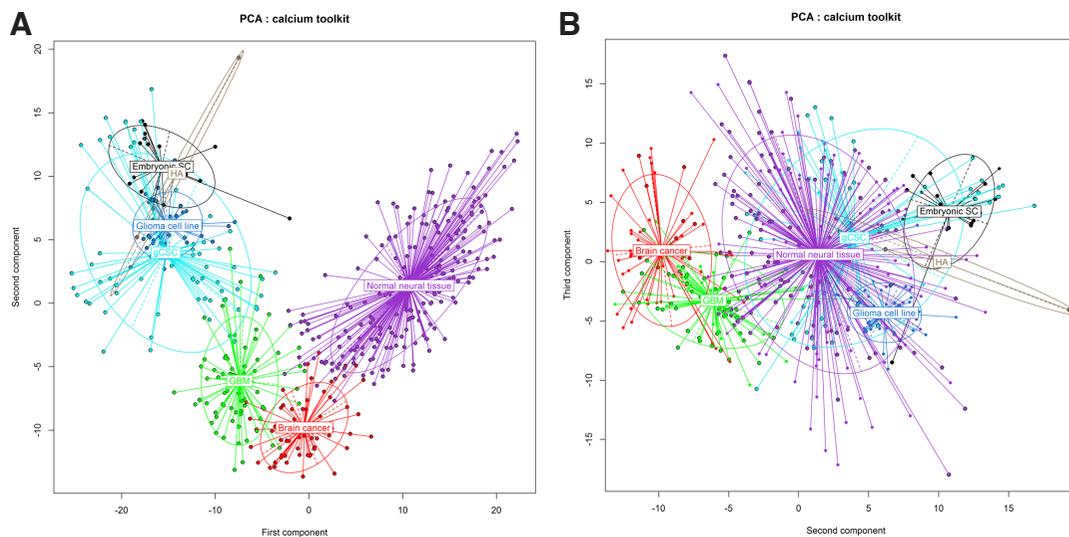
A

B

Fig. 3. Principal Component Analysis (PCA) of different brain tissue samples and embryonic stem cells (ESC), except outliers computed from the calcium toolbox genes. (A) *First and second components and* **(B)** *second and third components.*

tools to wrong annotations was not checked.

Comparison of the similarity of brain tissue sample groupings obtained using calcium toolbox gene expression and global gene expression

Our working hypothesis considered that the calcium signature was characteristic of a given cell or a given tissue. A calcium signature relies on the calcium signalosome, which is the subset of genes expressed from the calcium toolbox (see Supplementary Table 1).

PCA analysis was therefore restricted to the calcium toolbox genes and compared to that obtained in Fig. 2B using all the genes. As shown in Fig. 3 the grouping looked similar although it lost in structure and clarity. Restricting the analysis to the calcium toolbox genes is thus accompanied by some loss of information. However focusing on a specific signaling pathway will facilitate the understanding of the results from a biological point of view.

To move the analysis further, a clustering was performed with all the samples, using either all the genes or only the genes from the calcium toolbox.

A consensus clustering, based on filtered probe sets was performed. Only probe sets with a mean expression value in all samples comprised between 6 and 12 were kept (the probe set expression ranges between 0 and 14). After probe sets filtering, our "all genes" dataset was composed of 26702 probe sets. A consensus clustering analysis was performed considering from 2 to 10 clusters in order to find the most stable partition. The most stable partition was obtained using 5 clusters. A heatmap presenting the clustering is shown in Fig. 4A. The samples divided into 3 major clusters and two small ones. The 3 major clusters are correlated with annotations: one for normal brain tissues, one for GBMs and other brain cancers and the last one gathering all the samples

corresponding to cell lines, i.e. HA, glioma cell lines, gCSC and ESC. Concerning the two other clusters, one corresponded to one GBM sample and the second to the three "pituitary gland" tissues from normal brain tissues.

The same analysis was performed using exclusively the calcium toolbox genes. In this case, the number of probe sets amounted to 260 and the best classification appeared to be consistent when using 6 clusters. Those divided into four main clusters and two smaller ones. We kept the cell line cluster and most of the brain cancer cluster, obtained with the whole set of probes (Fig. 4a). A major difference concerned normal brain tissues. With all genes, they are clustered together whereas with the calcium genes, they are divided into 4 clusters (Fig. 4B).

Calcium toolbox genes does not group samples in the same way the whole set of genes and therefore, give us a different point of view that will be easier to interpret as we will link the classification with specific expression of genes involved with known function in neuronal tissues.

Is it possible to find a calcium toolbox signature specific to a given set of brain tissue samples?

To start our calcium signature analysis, we focused on normal brain tissues. A public dataset containing 229 samples of normal brain tissues with 38 different annotations corresponding to different areas of the brain and the medulla was used (Supplementary materials, table 2, 1st column). As previously indicated, outliers were defined as samples of a specific group closer to the barycenter of a group with a different annotation.

In these annotations, 7 contained only one sample, namely cerebellar hemisphere, cerebellar vermis, fetal brain, globus pallidus, pons, prefrontal cortex and thalamus lateral nuclei and 1 contained two samples, namely frontal cortex. These samples
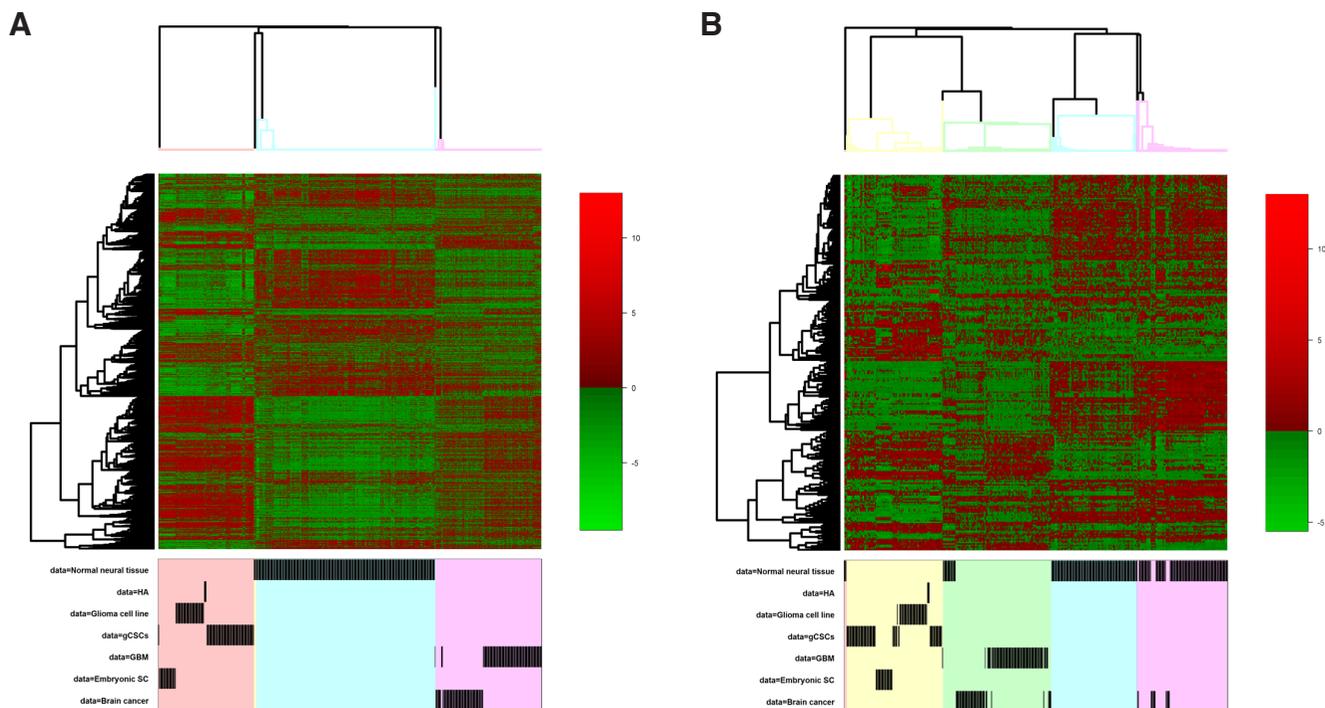

Fig. 4. Heatmap of normal brain tissues and embryonic stem cells. *Clustering of samples was done using consensus clustering algorithm.* **(A)** *Using all genes and* **(B)** *using the calcium toolbox genes.*

Fig. 5. Network of normal brain tissues, analyzed according to calcium toolbox genes overexpressed in these tissues.

were suppressed, except the first and second ones (cerebellar hemisphere and cerebellar vermis) which are close to cerebellum and which were re-annotated. On the remaining 222 samples, 55 were found to be closer to the barycenter of another annotation group than to their own. We hypothesized that brain tissues are not as far apart from each other and decided to keep the sample annotations when the distance to their own annotation barycenter was less than 100 and the relative difference between the distances to their own barycenter and that to the closest barycenter of another annotation less than 12 %. These thresholds were set arbitrarily from our data. Indeed out of a total of 229 analyzed samples, 24 with a distance to their own barycenter above 100 were located closer to another barycenter. Those samples were considered at outliers. All of them exhibited a relative difference between the distance of their own barycenter and that to the closest barycenter of at least 12%. We therefore used this threshold of 12 % as a minimum to consider a sample as an outlier even when the distance to their own annotation barycenter was less than 100.

Concerning the "cerebral cortex" annotation, 6 out of 9 samples were defined as outliers. These 9 samples were removed. Thus, a total of 35 samples were suppressed. Our final dataset was made up with 194 samples, annotated in 29 different groups of tissues (see Supplementary Table 2). This pointed to a noise in this dataset of c.a. 16%. This annotation noise in biological datasets was considered to be acceptable.

In order to explore the differential expression of the calcium toolbox genes among the different normal brain tissue samples, we developed an algorithm which allowed us to find genes over-expressed in one tissue comparatively to the others (see Materials and Methods section). Results are available in Supplementary Table 3 (sheets 1 to 3 genes coding channels, pumps and calcium binding proteins).

Four genes involved in calcium entry, namely *CACFD1,*

CACNG4, *ORAI2* and *TPCN1* were expressed in all the brain tissues and 3 other genes, *CACNA1A (*also known as *Cav2.1*), *CACNA2D1* and *CACNG6* in more than 80 % of the brain tissues. This indicates that few calcium toolbox genes are involved in calcium entry (less than 10 %). *CACFD1* codes for a membrane protein with calcium channel activity and was reported to be involved in neuronal exocytosis/ endocytosis (Yao *et al.*, 2009). *TPCN1* appears to be one of the major calcium channels found in lysosomal and endosomal membranes (Neely Kayala *et al.*, 2012). *CACNA2D1,* present in all brain tissues except one is a subunit of the voltage-dependent channels, Cav2.1. Cav2.1 has also be reported to be involved in exocytosis (Weiss, 2010). *CACNG4* and *CACNG6* known as modulators of the AMPA receptors (ionotropic transmembrane receptors for glutamate that mediate fast synaptic transmission in the central nervous system) were expressed in all tissues but two for CACNG6. Finally, *ORAI2* but not *ORAI1* nor *ORAI3* was expressed in all brain tissues, pointing out the importance of a specific store operated calcium channel mechanism (Heo *et al.*, 2015, Kito *et al.*, 2015) either in microglia or in brain endothelial capillary cells.

Another set of calcium toolbox genes expressed in all brain tissues coded for pumps or exchangers. In this set, one gene (Serca2) is involved in the regulation of endoplasmic vesicles (Baba-Aissa *et al.*, 1998), two genes (*PMCA2* and *PMCA4)* code for pumps localized in plasma membrane (Baba-Aissa *et al.*, 1998), a set of genes (*MICU1, MICU2, SLC25A23)* are involved in mitochondrial calcium homeostasis (Hoffman *et al.*, 2014) and another set correspond to calcium exchangers (*SLC24A2* (Li *et al.*, 2002) and *SLC25A12* (Rueda *et al.*, 2014)). 32% of the genes were expressed in at least 25 tissues, pointing out the importance of calcium reuptake in the different calcium stores of the different normal brain tissues.

Finally, 17 genes coding for calcium binding proteins (*CAB39, CABP4*, *CALCOCO1*, *CALCOCO2*, *CALR*, *CHP1*, *EFCAB14*, *HPCAL1*, *MYL12B*, *MYL6*, *MYL6B*, *RASEF*, *RCN2*, *S100A13*, *S100A16*, *S100A6*, *SRI/sorcin*) were expressed in all brain tissues and 6 genes (*CETN2*, *EFHD1*, *S100B*, *S100A1*, *VSNL1*, *NCS1*) were expressed in more than 80 % of the brain tissues. That represents 15 % of the genes coding for calcium binding proteins that were selected in our study.

The next step was to detect genes specifically overexpressed in a tissue or a subset of neural tissues. Signatures specific to a given brain area could be obtained (see Supplementary Table 3, sheets 4 to 6 for genes coding channels, pumps and calcium binding proteins).

For calcium entry, the specificity resided in the expression of alpha and beta subunits of voltage dependent calcium channels, inositol phosphate dependent receptor, ryanodine receptor and in a subset of TRP channels.

For calcium pumps and exchangers, tissue specificity was mainly given by the difference in expression of the sodium/calcium exchangers.

Concerning the calcium toolbox genes coding for calcium binding proteins, signature for a given brain tissue or a small subset

of brain tissues was given by a set of 39 overexpressed genes as listed in Supplementary Table 3, sheet 6.

When using genes overexpressed in a subset of brain tissues, a graph could be built where tissues are connected when they share a given number of genes overexpressed in a tissue when compared to at least 14 others tissues. Each edge of the graph represents genes overexpressed in common between the two nodes (Berthold *et al.*, 2007, Shannon *et al.*, 2003).

Spinal cord and nerve ganglia clearly gathered in one group and 15 brain tissues in another group (Fig. 5). This overall analysis allowed to propose a calcium genomic signature and to link this signature to functional cellular traits.

Calcium signature for brain tumors and glioblastoma stem cells

The same rationale was used to establish a calcium toolbox expression signature for brain tumors. To do so, our heterogeneous set of tumor samples was first classified into homogenous classes.

We started from the previous clustering presented in section 2 using the calcium toolbox genes and tried to refine it. In section 2, five clusters were defined: 3 with normal brain tissues, one with brain cancer and glioblastoma (GBM) samples (named "cancer cluster") and one named "cell line cluster " with glioblastoma cell lines, glioblastoma stem cells (gCSCs), embryonic stem cells (ESC) and human astrocytes (HA). The last two clusters were refined as previously using a consensus clustering on the calcium toolbox genes filtered probe sets.

The "cell line cluster" was divided into 7 sub-clusters with 4 major ones (Fig. 6A) composed of cell lines and HA for the first one, gCSCs for the second and third one, ESC for the last one. Each of these 4 sub-clusters contained respectively 38, 40, 12 and 21 samples

The "cancer cluster" was divided into 15 sub-clusters (Fig. 6B)

with 4 major clusters and 11 composed of one to four samples. These latter clusters were too small and were not used in our analysis. Considering the major clusters, one contained all pituitary gland samples, the second the majority of brain cancers and a few GBM, the third and fourth GBM and a few brain cancers. Since we cannot verify sample annotations, we decide to keep only samples with a logical coherency between annotation and the overall characteristic of the cluster. Hence, 3 sub-clusters were defined for the "cancer cluster" samples namely, brain cancer, GBM1, GBM2, containing respectively 35, 19 and 41 samples.

Seven new groups were thus added to our previous analysis of normal brain tissues using the calcium toolbox genes, namely brain cancer, GBM1, GBM2 (corresponding to the 3 sub clusters retained from the "cancer cluster") and gCSC1, gCSC2, cell lines, ESC (corresponding to the 4 sub-clusters retained in the "cell line cluster").

Comparative gene expression between groups was then performed. We looked for genes overexpressed in one group with respect to the others, using each group as control. Due to noise in microarrays and to increase analysis robustness, we focused on genes overexpressed for every group. Genes highly expressed in brain tumors or in cancer cells but not in normal brain tissues were then selected (see Supplementary Table 3 sheets 7 to 9 for genes expressed in brain cancer and normal tissues and sheets 10 to 12 for genes specific to brain cancer and normal tissues).

This disclosed 6 genes involved in calcium entry (*STIM2*, *ORAI1*, *TRPA1*, *TRPP2/PKD2*, *TRPM7* and *ITPR3*), 2 genes involved in calcium reuptake (*MCU* and *SLC8A3*) and 10 genes coding for calcium binding proteins putatively involved in the modulation of calcium fluxes (*STC1*, *S100A11*, *S100A2*, *S100A4*, *S100A8*, *EFCAB7*, *EFCAB11*, *CETN3*, *RCN1* and *CALU*).

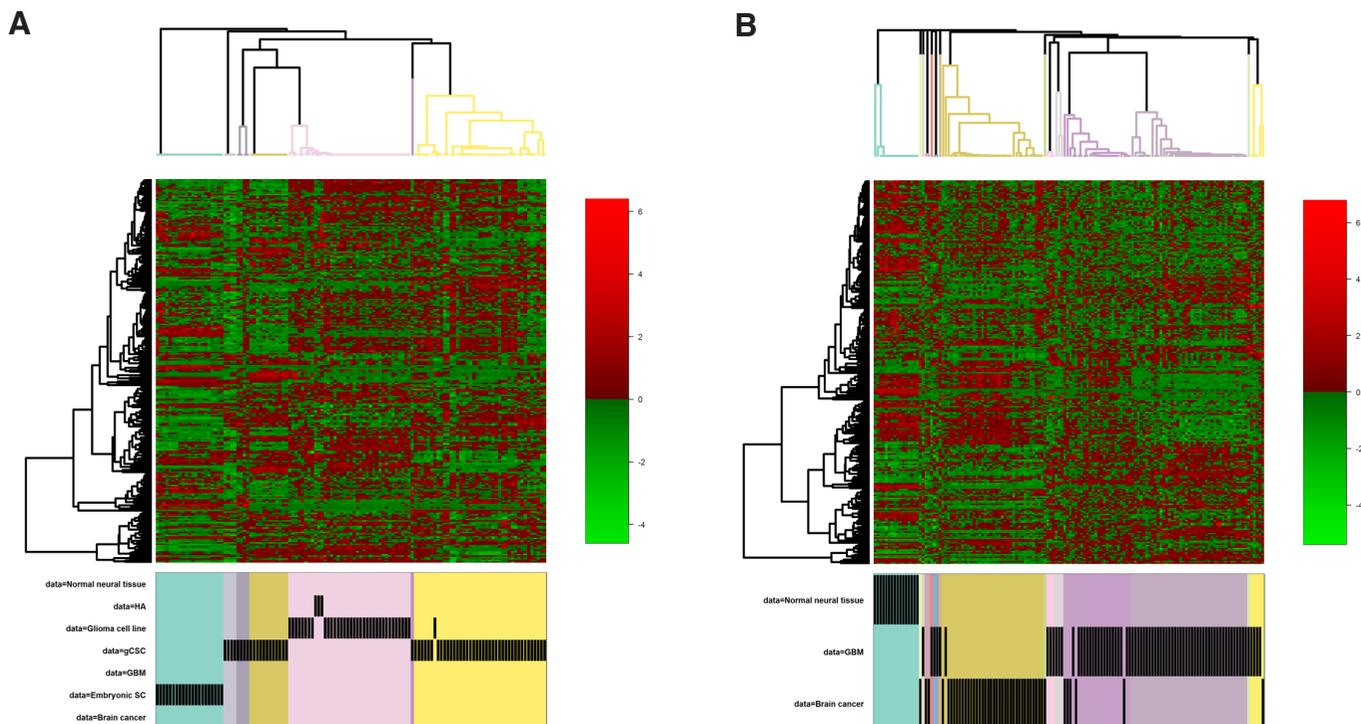*TRPM7* has recently been described as a modulator of glioma

Fig. 6. Refined clustering of samples according to the expression of the calcium toolbox genes. (A) *"Cell lines cluster"* and **(B)** *"cancer cluster".*

stem cell proliferation and metastasis (Liu *et al.*, 2014). We found this gene more specifically expressed in the group of normal ESCs and in one of the two groups of gCSC. This seems to confirm the heterogeneity of the gCSC isolated in different laboratories and suggests that *TPM7* may be used as a marker to classify different populations of those cells.

STIM2, *ORAI1* and *TRPP2* may probably be involved in the store operated calcium entry (SOCE) mechanism although the role of *TRPP2* remains elusive (Saul *et al.*, 2014). However, our results suggest that GBM and gCSC may exhibit a different molecular mechanism to control the SOCE mechanism.

As far as calcium reuptake is concerned, the two main genes highly expressed in gCSC, ESC and glioblastoma cancer cell lines and to a lower level in normal brain tissues are *MCU* coding for the mitochondrial calcium uniporter and *SLC8A3* coding for a sodium/calcium exchanger present both in the plasma membrane and in mitochondria (Scorziello *et al.*, 2013). Taking into account the expression of the inositol triphosphate receptor *ITPR3* and the probable change in the CRAC channel in cancer tissues and cancer cells, we hypothesize that calcium reuptake by mitochondria upon calcium entry via CRAC channels may play an important role in tumor tissues or tumor cells.

The role of the calcium binding proteins overexpressed in the tumor tissues and/or the tumor cells will not be discussed. We may just notice that the specific set of S100 proteins found to be overexpressed in brain tumors and cancer cells, namely *S100A2*, *S100A4*, *S100A8* and *S100A11*, have been reported as markers of tumorigenicity (Rand *et al.*, 2008). Moreover, the plasma concentration of some of those calcium binding proteins may be used as serum biomarkers with potential utility in brain cancer classification and GBM prognosis (Sreekanthreddy *et al.*, 2010).

Conclusion

The present paper proposes a general workflow composed of four steps to extract potential genes coding for proteins involved in the generation and modulation of the calcium signal in normal and cancer tissues or cells.

In this methodology, the first step aimed at decreasing the noise and heterogeneity of the public datasets used. The noise level was estimated to range between 10 to 20 %. Outliers were identified and removed stringently. The following steps dealt with a subset of genes involved in the generation and modulation of the intracellular calcium signal triggered by an external stimulus. By using a well-defined set of genes, the biological coherency of our analytical workflow could be checked more easily and a better setting of the different thresholds used for the analysis could be obtained. Moreover, although in general these types of analysis lead to correlation between a set of genes and a specific phenotype defined by a given annotation, by focusing on the calcium toolbox, it became possible to propose a biological hypothesis that is testable.

Following this approach and focusing on the calcium toolbox involved in calcium fluxes (entry and exit from cytoplasm) we propose that:

1. Major modifications appear in the regulation of the SOCE mechanism in glioblastoma and glioblastoma stem cells, compared to normal brain tissues.

2. In some types of glioblastoma cancer stem cells, mitochondria may play an important role in the calcium uptake mechanism. This role may be coupled with the modifications of the SOCE molecular mechanism.

As we did not have access to transcriptomic data of glioblastoma stem cells grown under different conditions, the next step will be to investigate if the modifications of the calcium signal management are induced by tumorigenicity or if they determine the fate of the cell.

Materials and Methods

Transcriptomic data

Data from 14 independent public gene expression data sets (all using Affymetrix HG-U133 Plus 2.0 arrays) and 2 samples hybridized in our laboratory (one Human Astrocyte cell line andU87) were taken for the analysis. Public samples used are detailed in Table 1.

Calcium toolbox definition

We considered that the characteristics of the calcium signature generated in a given cell upon an external stimulus were mainly due to the expression of a subset of genes from the calcium toolbox and to the localization of the expressed proteins.

The calcium toolbox is composed of the genes involved in calcium entry (all kinds of calcium channels), calcium removal from the cytoplasm (calcium pumps and calcium exchangers) and calcium detection (calcium binding proteins belonging to the EF-hand family) (Haiech *et al.*, 2004).

For the calcium toolbox analysis, 260 genes including few genes annotated as pseudogenes were selected (see Supplementary Table 1). These included 75 calcium channels, 25 calcium pumps or calcium exchangers, 160 calcium binding proteins belonging to the EF-hand domain family (Haiech *et al.*, 2011). Such a toolbox is by far probably not exhaustive but it facilitates biological interpretation. Moreover, it would be easy to modify the calcium toolbox content and redo the overall analysis.

Data analysis

Data normalization

All data were normalized with R software using the Robust Multi-array Average method from the justRMA function in affy package.

Outliers analysis

Samples were annotated in 7 distinct groups: normal brain tissue, brain cancer, glioblastoma (GBM), glioblastoma stem cells (gCSC), human astrocytes (HA), embryonic stem cells (ESC), glioma cell lines. The barycenter for each group and the distance from each sample to each barycenter were computed. Samples closer to another barycenter than to their own were defined as outliers and were removed.

Selection of normal neural tissue retained for the analysis

Normal neural tissues were annotated in 38 groups for the outlier analysis using barycenters and distances from barycenters. We suppressed groups containing one sample, but kept the annotation if the sample was almost at the same distance from another barycenter as from its own, suppressed the samples if the distance was different by more than 20% and re-annotated cerebellar hemisphere and cerebellar vermis (alone in their group) into cerebellum (closest barycenter).

Method to detect genes overexpressed in tissue or cell sub-populations

The algorithm used aimed at identifying and sorting genes strongly expressed in a subpopulation of a given dataset in comparison with a control dataset. Input data were the results of mRNA expression microarray analysis. For every probe set, the algorithm (ALGO) computed the maximum of expression in control samples (M) and selected all case samples with an expression superior to this maximum plus a threshold. A score was computed for the probe set reflecting the proportion of case samples selected and the delta of expression between case samples and M.

Algorithm:

Let the gene expression values of the n_1 CTRL control samples for a gene be given by

$$\{x_{1i}\}_{i \in \{1,2,\dots,n_1\}}$$

and the gene expression values of the n_2 CASE case samples for a gene be given by

$$\{x_{2i}\}_{i \in \{1,2,\dots,n_2\}}$$

The maximal gene expression value of a gene among the control samples was first defined

$$\hat{x}_1 = \max\{X_{1i}\}$$
$$t = 0.5$$
$$pop_1 = \{x_{2i} > \hat{x}_1 + t\}$$
$$n_{1p} = size\{pop_1\}$$
$$\Delta_1 = median\{pop_1\} - \hat{x}_1$$
$$Score_1 = 2^{\Delta_1}\frac{n_{1p}}{n_2}$$

Calcium genes specific to normal brain tissue subgroups

We defined the calcium signalosome as the subset of calcium toolbox genes expressed in a given cell or tissue and the calcisome the macromolecular protein complex built from the proteins coded by the calcium signalosome.

Analysis was performed in order to know if calcium signalosomes allowed discriminating tissues from normal brain areas. For this purpose, we used the precedent algorithm (ALGO) in a recursive way with the normal brain dataset containing 29 different tissue annotations. This algorithm works by finding genes overexpressed in subgroups; it was therefore used with one annotation as "control" and all the others as "case". Its output is a list of probe sets with a score and a list of samples that overexpressed it. If the list of samples is empty, the score is zero. After each run of ALGO, we kept the information "probe set overexpressed" if it is overexpressed for all samples in one annotation.

Let the different annotations be given by $\{annot_i\}_{i=1,2,\dots,29}$
The samples be given by $\{samples_i\}_{i=1..n_s}$
And the final output with the samples overexpressed for one probe set vs one condition be given by "output"

For i in 1:29:

RES=ALGO(CASE=$\{annot_t\}_{t \neq i}$, CTRL=$annot_i$)

For $probeset_j$ in RES:

For $annot_k$ in $\{annot_k\}_{k \neq i}$:

If $\{samples\}_{annot=annot_k} \subset RES_{probeset_j}$

$(probeset_j, annot, annot_i) \in output$

Once all probe sets overexpressed for one annotation versus another one were obtained, we looked for genes specific for one annotation.

First we aggregated probe sets for one gene and used the gene annotation instead of probe set.

Then we selected, for each annotation, genes that were overexpressed at least against 14 annotations (half of them).

Let the final result with genes specific to one annotation be "specific"

For i in 1:29

For $gene_j$ in $output(gene_j, annot_i,)$

If length($output(gene_j, annot_i,)$) >14

$(annot_i, gene_j) \in specific$

References

ALDAZ, B., SAGARDOY, A., NOGUEIRA, L., GURUCEAGA, E., GRANDE, L., HUSE, J.T., AZNAR, M.A., DIEZ-VALLE, R., TEJADA-SOLIS, S., ALONSO, M.M. *et al.,* (2013). Involvement of miRNAs in the differentiation of human glioblastoma multiforme stem-like cells. *PLoS One* 8: e77098.

BABA-AISSA, F., RAEYMAEKERS, L., WUYTACK, F., DODE, L. and CASTEELS, R. (1998). Distribution and isoform diversity of the organellar Ca2+ pumps in the brain. *Mol Chem Neuropathol* 33: 199-208.

BEIER, D., HAU, P., PROESCHOLDT, M., LOHMEIER, A., WISCHHUSEN, J., OEFNER, P.J., AIGNER, L., BRAWANSKI, A., BOGDAHN, U. and BEIER, C.P. (2007). CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles. *Cancer Res* 67: 4010-4015.

BERNSTEIN, B.E., STAMATOYANNOPOULOS, J.A., COSTELLO, J.F., REN, B., MILOSAVLJEVIC, A., MEISSNER, A., KELLIS, M., MARRA, M.A., BEAUDET, A.L., ECKER, J.R. *et al.,* (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 28: 1045-1048.

BERTHOLD, M.R., CEBRON, N., DILL, F., GABRIEL, T.R., KOTTER, T., MEINL, T., OHL, P., SIEB, C., THIEL, K. and WISWEDEL, B. (2007). KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization.*

CHAFFER, C.L. and WEINBERG, R.A. (2015). How does multistep tumorigenesis really proceed? *Cancer Discov* 5: 22-24.

FONG, C.Y., CHAK, L.L., BISWAS, A., TAN, J.H., GAUTHAMAN, K., CHAN, W.K. and BONGSO, A. (2011). Human Wharton's jelly stem cells have unique transcriptome profiles compared to human embryonic stem cells and other mesenchymal stem cells. *Stem Cell Rev* 7: 1-16.

GARCIA, J.L., PEREZ-CARO, M., GOMEZ-MORETA, J.A., GONZALEZ, F., ORTIZ, J., BLANCO, O., SANCHO, M., HERNANDEZ-RIVAS, J.M., GONZALEZ-SARMIENTO, R. and SANCHEZ-MARTIN, M. (2010). Molecular analysis of ex-vivo CD133+ GBM cells revealed a common invasive and angiogenic profile but different proliferative signatures among high grade gliomas. *BMC Cancer* 10: 454.

HAIECH, J., AUDRAN, E., FEVE, M., RANJEVA, R. and KILHOFFER, M.C. (2011). Revisiting intracellular calcium signaling semantics. *Biochimie* 93: 2029-2037.

HAIECH, J., MOULHAYE, S.B. and KILHOFFER, M.C. (2004). The EF-Handome: combining comparative genomic study using FamDBtool, a new bioinformatics tool, and the network of expertise of the European Calcium Society. *Biochim Biophys Acta* 1742: 179-83.

HANAHAN, D. and WEINBERG, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144: 646-674.

HEO, D.K., LIM, H.M., NAM, J.H., LEE, M.G. and KIM, J.Y. (2015). Regulation of phagocytosis and cytokine secretion by store-operated calcium entry in primary isolated murine microglia. *Cell Signal* 27: 177-186.

HOFFMAN, N.E., CHANDRAMOORTHY, H.C., SHANMUGHAPRIYA, S., ZHANG, X.Q., VALLEM, S., DOONAN, P.J., MALLIANKARAMAN, K., GUO, S., RAJAN, S., ELROD, J.W. *et al.,* (2014). SLC25A23 augments mitochondrial Ca(2)(+) uptake, interacts with MCU, and induces oxidative stress-mediated cell death. *Mol Biol Cell* 25: 936-47.

KITO, H., YAMAMURA, H., SUZUKI, Y., YAMAMURA, H., OHYA, S., ASAI, K. and IMAIZUMI, Y. (2015). Regulation of store-operated Ca(2)(+) entry activity by cell cycle dependent up-regulation of Orai2 in brain capillary endothelial cells. *Biochem Biophys Res Commun* 459: 457-462.

LEVY, S., CHAPET, S. and MAZERON, J.J. (2014). [Management of gliomas]. *Cancer Radiother* 18: 461-467.

LI, X.F., KRAEV, A.S. and LYTTON, J. (2002). Molecular cloning of a fourth member of the potassium-dependent sodium-calcium exchanger gene family, NCKX4. *J Biol Chem* 277: 48410-48417.

LIU, M., INOUE, K., LENG, T., GUO, S. and XIONG, Z.G. (2014). TRPM7 channels regulate glioma stem cell through STAT3 and Notch signaling pathways. *Cell Signal* 26: 2773-2781.

MALLON, B.S., CHENOWETH, J.G., JOHNSON, K.R., HAMILTON, R.S., TESAR, P.J., YAVATKAR, A.S., TYSON, L.J., PARK, K., CHEN, K.G., FANN, Y.C. *et al.,* (2013). StemCellDB: the human pluripotent stem cell database at the National Institutes of Health. *Stem Cell Res* 10: 57-66.

MOSER, J.J. and FRITZLER, M.J. (2010). The microRNA and messenger RNA profile of the RNA-induced silencing complex in human primary astrocyte and astrocytoma cells. *PLoS One* 5: e13445.

NEELY KAYALA, K.M., DICKINSON, G.D., MINASSIAN, A., WALLS, K.C., GREEN, K.N. and LAFERLA, F.M. (2012). Presenilin-null cells have altered two-pore calcium channel expression and lysosomal calcium: implications for lysosomal function. *Brain Res* 1489: 8-16.

RAND, V., PREBBLE, E., RIDLEY, L., HOWARD, M., WEI, W., BRUNDLER, M.A., FEE, B.E., RIGGINS, G.J., COYLE, B., GRUNDY, R.G. *et al.,* (2008). Investigation of chromosome 1q reveals differential expression of members of the S100 family in clinical subgroups of intracranial paediatric ependymoma. *Br J Cancer* 99: 1136-1143.

RHEINBAY, E., SUVA, M.L., GILLESPIE, S.M., WAKIMOTO, H., PATEL, A.P., SHAHID, M., OKSUZ, O., RABKIN, S.D., MARTUZA, R.L., RIVERA, M.N. *et al.,* (2013). An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. *Cell Rep* 3: 1567-1579.

RUEDA, C.B., LLORENTE-FOLCH, I., AMIGO, I., CONTRERAS, L., GONZALEZ-SANCHEZ, P., MARTINEZ-VALERO, P., JUARISTI, I., PARDO, B., DEL ARCO, A. and SATRUSTEGUI, J. (2014). Ca(2+) regulation of mitochondrial function in neurons. *Biochim Biophys Acta* 1837: 1617-1624.

SAUL, S., STANISZ, H., BACKES, C.S., SCHWARZ, E.C. and HOTH, M. (2014). How ORAI and TRP channels interfere with each other: interaction models and examples from the immune system and the skin. *Eur J Pharmacol* 739: 49-59.

SCHULTE, A., GUNTHER, H.S., PHILLIPS, H.S., KEMMING, D., MARTENS, T., KHARBANDA, S., SORIANO, R.H., MODRUSAN, Z., ZAPF, S., WESTPHAL, M. *et al.,* (2011). A distinct subset of glioma cell lines with stem cell-like properties reflects the transcriptional phenotype of glioblastomas and overexpresses CXCR4 as therapeutic target. *Glia* 59: 590-602.

SCORZIELLO, A., SAVOIA, C., SISALLI, M.J., ADORNETTO, A., SECONDO, A., BOSCIA, F., ESPOSITO, A., POLISHCHUK, E.V., POLISHCHUK, R.S., MOLINARO, P. *et al.,* (2013). NCX3 regulates mitochondrial Ca(2+) handling through the AKAP121-anchored signaling complex and prevents hypoxia-induced neuronal death. *J Cell Sci* 126: 5566-5577.

SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N.S., WANG, J.T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. and IDEKER, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-504.

SHLENS, J. (2014). A Tutorial on Principal Component Analysis. *arXiv*.

SREEKANTHREDDY, P., SRINIVASAN, H., KUMAR, D.M., NIJAGUNA, M.B., SRIDEVI, S., VRINDA, M., ARIVAZHAGAN, A., BALASUBRAMANIAM, A., HEGDE, A.S., CHANDRAMOULI, B.A. *et al.,* (2010). Identification of potential serum biomarkers of glioblastoma: serum osteopontin levels correlate with poor prognosis. *Cancer Epidemiol Biomarkers Prev* 19: 1409-1422.

SUN, L., HUI, A.M., SU, Q., VORTMEYER, A., KOTLIAROV, Y., PASTORINO, S., PASSANITI, A., MENON, J., WALLING, J., BAILEY, R. *et al.,* (2006). Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* 9: 287-300.

WEISS, N. (2010). Control of depolarization-evoked presynaptic neurotransmitter release by Cav2.1 calcium channel: old story, new insights. *Channels (Austin)* 4: 431-433.

YAO, C.K., LIN, Y.Q., LY, C.V., OHYAMA, T., HAUETER, C.M., MOISEENKOVA-BELL, V.Y., WENSEL, T.G. and BELLEN, H.J. (2009). A synaptic vesicle-associated Ca2+ channel promotes endocytosis and couples exocytosis to endocytosis. *Cell* 138: 947-960.

YE, F., ZHANG, Y., LIU, Y., YAMADA, K., TSO, J.L., MENJIVAR, J.C., TIAN, J.Y., YONG, W.H., SCHAUE, D., MISCHEL, P.S. *et al.,* (2013). Protective properties of radio-chemoresistant glioblastoma stem cell clones are associated with metabolic adaptation to reduced glucose dependence. *PLoS One* 8: e80397.

ZORNIAK, M., CLARK, P.A. and KUO, J.S. (2015). Myelin-forming cell-specific cadherin-19 is a marker for minimally infiltrative glioblastoma stem-like cells. *J Neurosurg* 122: 69-77.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

Egg activation in physiologically polyspermic newt eggs: involvement of IP3 receptor, PLCγ, and microtubules in calcium wave induction
Tomoyo Ueno, Takehiro Ohgami, Yuichirou Harada, Shuichi Ueno and Yasuhiro Iwao
Int. J. Dev. Biol. (2014) 58: 315-323

Visualization, characterization and modulation of calcium signaling during the development of slow muscle cells in intact zebrafish embryos
Chris Y. Cheung, Sarah E. Webb, Donald R. Love and Andrew L. Miller
Int. J. Dev. Biol. (2011) 55: 153-174

Could modifications of signalling pathways activated after ICSI induce a potential risk of epigenetic defects?
Brigitte Ciapa and Christophe Arnoult
Int. J. Dev. Biol. (2011) 55: 143-152

Oscillatory Ca2+ dynamics and cell cycle resumption at fertilization in mammals: a modelling approach
Geneviève Dupont, Elke Heytens and Luc Leybaert
Int. J. Dev. Biol. (2010) 54: 655-665
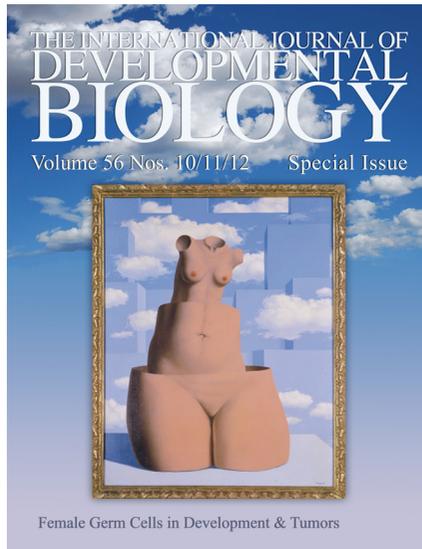
The role of ion fluxes in polarized cell growth and morphogenesis: the pollen tube as an experimental paradigm
Erwan Michard, Filipa Alves and José A. Feijó
Int. J. Dev. Biol. (2009) 53: 1609-1622

Sperm-activating peptides in the regulation of ion fluxes, signal transduction and motility
Alberto Darszon, Adán Guerrero, Blanca E. Galindo, Takuya Nishigaki and Christopher D. Wood
Int. J. Dev. Biol. (2008) 52: 595-606

The dynamics of calcium oscillations that activate mammalian eggs
Karl Swann and Yuansong Yu
Int. J. Dev. Biol. (2008) 52: 585-594

The choice between epidermal and neural fate: a matter of calcium.
Marc Moreau and Catherine Leclerc
Int. J. Dev. Biol. (2004) 48: 75-84

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
Volume 56 Nos. 10/11/12   Special Issue
Female Germ Cells in Development & Tumors

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
Volume 57 Nos. 2/3/4   Special Issue
Male Germ Cells in Development & Tumors

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
Volume 52 Nos. 5/6   2008
Fertilization

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
Volume 54 Nos. 2/3   Special Issue
Placenta