UNIVERSITY OF UDINE

DOCTORATE COURSE IN
BIOMEDICAL AND BIOTECHNOLOGICAL SCIENCES
CICLE XVII

RESEARCH DOCTORATE THESIS

DEFINITION OF BIOLOGICAL RESPONSES THROUGH THE ANALYSIS OF GENE EXPRESSION PROFILES

Candidate:
Raffaella Picco

Tutor:
Prof. Claudio Brancolini
Prof. Federico Fogolari

ACADEMIC YEAR
2014/2015

A Marian, a Carla
e alle mie figlie

TABLE OF CONTENTS

1. ABSTRACT

The microarray technology has revolutionized the study of the complexity in the biological systems and of the biological responses to the environment. In order to define new scenarios and new biological pathways, the aim of this PhD thesis was the acquisition of a mastery in the management and analysis of microarray data, deepening the basic concepts of statistical analysis, to make knowledgeably choices about the test more suitable to a particular set of data. After in-depth knowledge of the different phases, the aim was to build up a generic pipeline that can be used to analyze any type of microarray data. The computational pipeline for processing raw microarray data (images) was implemented in R, using mostly Bioconductor packages. Implementation aimed to define gene expression levels, to provide experiment quality assessment and significative statistical tests.

During the first part of the PhD thesis my purpose was the determination of the gene function combining experiments of silencing with the gene expression analysis.

Caspase-2 is a member of the cystein-protease family of enzymes, which carry out important roles in apoptosis and in inflammation. Although *CASP2* is highly conserved through the evolution and was the first caspase identified, several contradictory results are found in the literature. Being expressed at high level during the neurological development and with a strong involvement in the apoptotic processes in the adult central nervous system, we decided to proceed with its silencing in glioblastoma cells, to evaluate the effect on gene expression. The comparative analysis of expression profiles of silenced cells respect to the control, highlighted the relation between *CASP2* and genes involved in the cholesterol metabolism. Previous studies have suggested for this enzyme a role in the control of intracellular level of this metabolite. Therefore, we decided to use data stored in public databases in order to extend the investigation, including all the other caspases and all the genes in some way connected to cholesterol. Computational analysis of the correlation between expression levels of these different genes families, has allowed us to: i) define correlations among cholesterol genes and caspases and ii) perform a hierarchical clustering of the different caspases in term of correlational profile. The analysis was expanded to normal brain and liver tissues and a correlation between expression levels of certain caspases and aging was found in human brain.

During the second part of the work, I performed gene expression profile analysis to define

signaling pathways and resistance mechanisms elicited after treatment of chronic B lymphocytic leukemia cells with a new class of ubiquitin-proteasome system (UPS) inhibitors.

Through the comparison of transcriptional profiles before and after treatments, many genes and relative pathways, whose expression was altered by the UPS inhibitor, were identified. These genes have allowed us to define new mechanisms of drug action. Furthermore, considering the difference in terms of responsiveness of the analyzed patients, we propose some genes as responsible for the differential cellular reactivity to the specific inhibitor.

2. INTRODUCTION

2.1 MICROARRAY

The microarray technique, developed in the 90s (Schena, 1995), enables the analysis of gene expression by monitoring at once the RNA products of thousands of genes.

Thanks to the production of high quality platforms, the creation of standardized experimental protocols, the construction of sophisticated devices for the recording of the intensity values and the development of robust computational tools for data management, this technology became a technique of everyday use in biological laboratories (Hoheisel, 2006; Trevino, 2007).

Unlike traditional investigation techniques, that make use of probes labeled with fluorescent substances directed against specific sequences of DNA anchored to a solid support, microarray technology overturns this relationship: in this case the non-labeled probes are fixed on the same support and the labeled sample is instead in the liquid phase. The hybridization of all the probes and their respective targets occurs simultaneously and this produces information on the global transcriptional state of the sample. After removal of excess solution that did not interact with the probes, the bond between the probe and target is translated into a numerical value, after recording the emitted fluorescence, which is supposed to be proportional to the amount of the respective transcript present in the sample under investigation.

The microarray analysis provides a matrix of genes (along rows) and samples (along columns). The values in the matrix are intensity measurements that are proportional to the mRNA quantity of each gene present in a cell.

MICROARRAY TECHNOLOGY

The choice of the material for the microarray construction has a fundamental importance. At the beginning, DNA arrays were based on nylon membranes. Nowadays, both porous and non-porous materials can be selected for their construction, each of them with specific advantages. The non-porous ones, such as glass, are the most useful, given the low intrinsic fluorescence and the poor dissemination.

These features allow to create small sized spots in a very precise way, permitting the use of an

extremely low sample quantity. Unfortunately the glass has a high affinity for dust and for the other contaminants present in the air and a smaller binding surface with respect to the other porous materials.

A recent technology makes use of glass beads to which the probes are linked before being spotted in a completely random way. The positions are then determined through a complex pseudo-sequencing process. The small bead size allows for greater number of spots per microarray slide and as many as 12 simultaneous sample analyses on a single platform are possible.

An additional microarray classification is made on the basis of the support dimensionality: there are 2D-microarray and 3D-microarray. In this case each probe (specific for a single gene) is linked to the microchannel walls, process that allows to increase in a considerable way the quantity of probes used (therefore with an augmentation of sensitivity) and the possibility to use the much more stable chemiluminescence (instead of the fluorescence).

Three techniques for microarray manufacturing exist.

In the first technique single strand cDNA molecules (length between 200 and 2000 bases) or pre-synthesized oligonucleotides (50-100 bases) in solution are deposited onto a solid support through the use of a robot. These are spotted microarray.

The second technique exploits the in-situ synthesis of 60-bases oligonucleotides through ink-jet printing technology (Agilent): this is a very flexible method that allows the production of personalized array for two-dye experiments. A computer controls the printing process and the sequences of all the oligonucleotides are contained in a file. This methodology makes particularly efficient the manufacturing process of custom microarray. Another positive feature is the uniformity of the spot shape and dimension.

In the third technology small oligonucleotides are directly synthesized on a silica support, through photolithographic techniques (Affymetrix) that make use of masks and light-reactive compounds with the selective activation of specific nucleotides in well-determined positions. The sequential use of different masks leads to the synthesis of all probes essential for the whole genome analysis.

The high cost of this technology does not justify the realization of custom array. The maximum oligonucleotide length allowed is 25 bases and the density reached on the spot is very high (over 700000 probes), more than 10 times higher with respect to that obtained through other techniques.

In order to represent each gene, a combination (called probeset) of 11-20 different oligonucleotides are used in the PM (Perfect Match) form and likewise in the MM (MisMatch) form, each of them interrogating a different part of the sequence of a specific gene.
In particular, the PM oligos are perfectly complementary to the transcribed mRNA, while the MM oligo are exactly like the equivalent PM except that the central-position nucleotide (the thirteenth) is substituted by its complementary (A ↔ T, C ↔ G) base.
The MM oligos are used to highlight the non-specific hybridization of the corresponding PM, which is useful to quantify weakly-expressed mRNA.

A further classification of microarray platform is based on whether one-dye or two dyes are used. The RNA extracted from the sample is labeled with a fluorescent molecule and then it incubated with the support. Different fluorescent substances can be used to hybridize simultaneously with two samples on the same platform. In this case the competitive binding between the two targets and the probes in the array is exploited. The experiment goes on with the scanning step and the recording of emitted signals using lights with different wave-length to excite the two fluorophores (red and green fluorescence: excitation at 550 nm and 640 nm, emission at 581 nm and 670 nm, respectively). The ratio of the signals reveals the relative abundance of the different mRNAs between the two samples.
In the alternative scheme (the case of the genechip Affymetrix), only a single sample is used, it commonly emits yellow fluorescence. The final result of this procedure generates monochromatic 16 bits images, the pixel intensities can range in value from 0 (black) to 65535 (white) shades of gray.

IMAGE ANALYSIS: gridding, segmentation, intensity extraction

Gridding

The array is physically scanned to produce a digital record. After the hybridization of fluorescently labeled cDNA molecules to the platform, the array is stimulated with a laser and the emitted fluorescence is measured. Several scanner settings can be varied to improve the sensitivity of the resulting image, with the brightest pixels settled just below the level of saturation in order to increase the power of the analysis in the case of less bright pixels. The first step of the analysis is the gridding that is usually a semi- or fully automated process. It consists in overlaying a rectangular grid onto the pixels in a manner that isolates each spot

within a platform, identifying spots through coordinates.
In the Affymetrix platforms, control probes located at the corners of the probe array are used to align a grid to delineate the probe cells.

Segmentation

The second step is the segmentation: every single pixel has to be classified as foreground or background. The signal observed is a combination of the foreground and the background, the specific hybridization of interest and the non-specific one, respectively.
The background is due to non-specific bindings, to contamination and to the intrinsic fluorescence emitted from other chemicals on the glass. The classical approach performs the simple subtraction of the background value from the mean or from the median of the foreground intensity. A global background can be considered (global background correction), by subtracting the same value from all the spots in the platform. The background level is calculated by considering a set of negative control spots or, alternatively, the third percentile of the intensities can be taken as global background.
Another possibility is to consider a local background. Depending on the tool used to analyze the image, the areas around each spot taken as a reference for the background assessment are different.
In the case of high-density array, it is possible to consider wider areas (subgrids) to do this.
For the foreground evaluation, there are spatial and distributional methods.
Fixed or adaptive segmentation: the first category tries to establish the spot shape, by defining a fixed circle with a constant diameter, not optimal in the most cases, or a circle with a variable diameter or another shape that considers the real appearance of each spot.
In this case, we have a Fixed circle segmentation, Adaptive circle segmentation and Adaptive shape segmentation, respectively.
The second method called histogram segmentation summarizes instead the signal coming from all pixels, not necessarily connected, that are over a treshold, commonly included between the 5^{th} and the 20^{th} percentile of the signal intensities. After foreground pixels identification, the intensity value for the spot is usually estimated as the average of all these measurements.

Intensity extraction

The extraction of reliable intensities from foreground and background pixels involves

methods to control the effects of all the possible sources of variation during the experimental procedures. It includes several substeps.

– Filtering

The aim of this procedure is to reduce the data variability and dimension by removing all the genes whose measurements are not enough accurate and by deleting genes weakly-expressed or with an almost constant expression (low variance of log ratio across samples).

– Summarization

This is the final stage in preprocessing Affymetrix GeneChip data. It is the process of combining the multiple probe intensities for each probeset to produce a unique expression value, by combining them in some manner.

– Normalization

The normalization process has the aim the removing of all the variability sources (bias) without biological origin. These contributions can be classified in this way:

i) Dye-effect: it depends on the different level of incorporation of the fluorophores used in two-dye experiments. The Cy5 molecule is more bulky and labile: the red emission is systematically less intense that the green one.

ii) Print-tip effect: it is due to the difference in term of genetic material spotted on the glass because of the microscopic differences in the tip conformations (used by the robot to print the platform).

iii) Array-effect: it can derive from the differences in the global intensities between the arrays that are linked to the different conditions of preparation, extraction, hybridization of the sample and to the scanning step.

Normalization can be carried out within each array or between arrays. The common approach to normalize the data is the global one. Averages of the overall expression for all genes within an array are set to be the same across different arrays. The amount of transcription is essentially similar across samples and over-expressed genes and under-expressed ones are roughly balanced.

Normalization procedures can be divided as follow:

– Total intensity normalization: this kind of procedure assumes the same mRNA quantity for both samples. A Constant K is searched so that the two samples have the same mean or median.

– Normalization based on regression methods: it starts from the assumption that most of the genes are expressed at the same level. There are several regression techniques:

linear regression (the coefficient of the regression line has an angular coefficient equal to 1) or local linear regression (LOWESS - “LOcally WEighted Scatter plot Smoothing”). In this case, instead of having a single correction factor, a function that compensates for bias intensity dependent is used (Cleveland, 1979).

Other methods that consider only a small part of all genes exist.

We can choose the normalization procedure that uses housekeeping genes: these are genes that are expressed at the same level ubiquitously. They are involved in the maintenance of the steady-state and in the survival. Often, however, it is difficult to define which genes actually belong to this category. Affymetrix has inserted in its platforms a set of genes that have very low variability among a big set of experiments with samples coming from different tissues and that can be considered as real housekeeping genes.

Over the years, several algorithms have been developed to normalize the microarray data.

Currently over 50 methods are described and compared at http:// affycomp.biostat.jhsph.edu/.

The most popular ones for affymetrix arrays are RMA e MAS5.

With Microarray Analysis Suite (MAS) 5.0, proposed by Affymetrix (Statistical Algorithms Description Document (2002 Affymetrix Inc.), the chip is divided in 16 zones, for each of them the 2% of the lower intensities are used to calculate the local background. This value is subtracted from all the intensities in the region, taking into account the distance of the spot from the centroid of the grid. In this way, the algorithm corrects both PM and MM values, before using both for the assessment of the final result.

Because approximately 30% of the MM intensities are higher than the respective PM, in this cases an ideal mismatch is used.

MAS5 normalizes each array independently and sequentially, after the summarization process. The results are not log transformed.

Robust Multi-array Average (RMA) proposed by Irizarray in 2003 (Irizarry, 2003) uses a multi-chip model. It imposes the same empirical distribution of intensities in each array. RMA does not use the mismatch probes, because their intensities are often higher than the match probes, making them unreliable as indicators of non-specific binding. RMA values are in log2 units, so they are not directly comparable with MAS5 normalized data. Nevertheless, normalized data are log-transformed at the end. The advantage of using this transformation is that the resulting data reflect the up-regulation and down-regulation values in a symmetrical scale. Without log conversion, ratios of down-regulated genes have their values between 1 and 0, while up-regulated ratios can have large values.

GRAPHICAL PRESENTATION OF THE DATA

Graphic displays can help assess the success of the experiment, guide the choice of analysis tools, and highlight specific problems. The first and most obvious diagnostic graphic is the well-known image in which the scanned microarray output images of the Cy3 and Cy5 channels are false-colored green and red, respectively, with yellow representing an equal balance of the two. Co-registration and overlay of the two channels offers a quick visualization of the experiment, revealing information on color balance, uniformity of hybridization, spot uniformity, background, and artifacts such as dust or scratches.

Boxplots: comparison of intensity distribution among the arrays of a dataset. The box is delimited by the 3rd and the 1st quartile values. The line in the middle of the box is the median value. The horizontal lines connected to the box via the dash line represent the higher and the lower values, without considering the outliers. Median is more robust to outliers than the mean.

An alternative way of representing the data is provided by histograms, which give information on the distribution of intensity values.

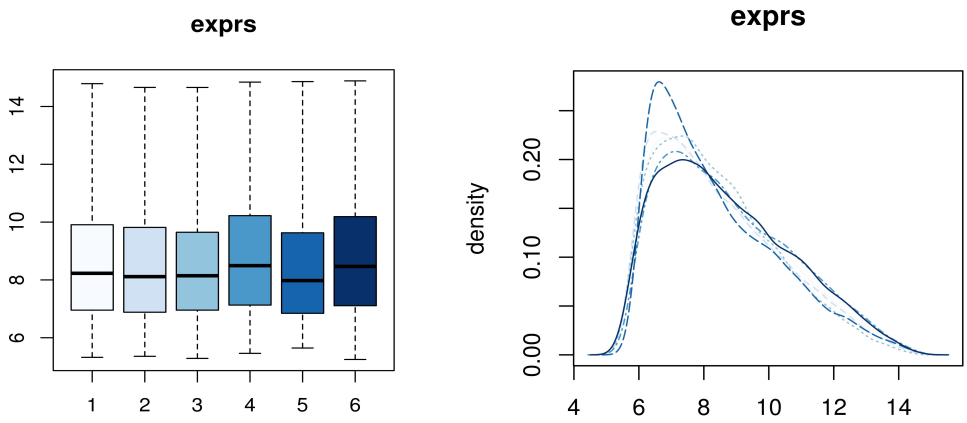


Figure 1 Boxplots and smoothed histograms of unprocessed log scale probe intensities for 6 arrays from the same data set.

QUALITY ASSESSMENT OF AFFYMETRIX GENECHIP

There are several useful tools implemented to assess the quality of microarray data. A typical first step is to look at pseudo-images created starting from the raw probe-level data: in this way it is possible to identify artifacts on the array. RNA Degradation Plot gives a good indication of the quality of the sample that has been hybridized to the array. Because this kind of degradation starts at the 5’ end of the molecule and progresses to the 3’ end it can be easily

measured using oligonucleotide arrays, where each PM probe is numbered sequentially from the 5' end of the targeted mRNA transcript. When RNA degradation is advanced, PM probe intensity at the 3' end of a probeset should be higher when compared with the 5' end one. In the graph, the lines have to be shifted from the original data for a clearer view, but, if there are no problems, the slopes remain unchanged because it is platform specific.

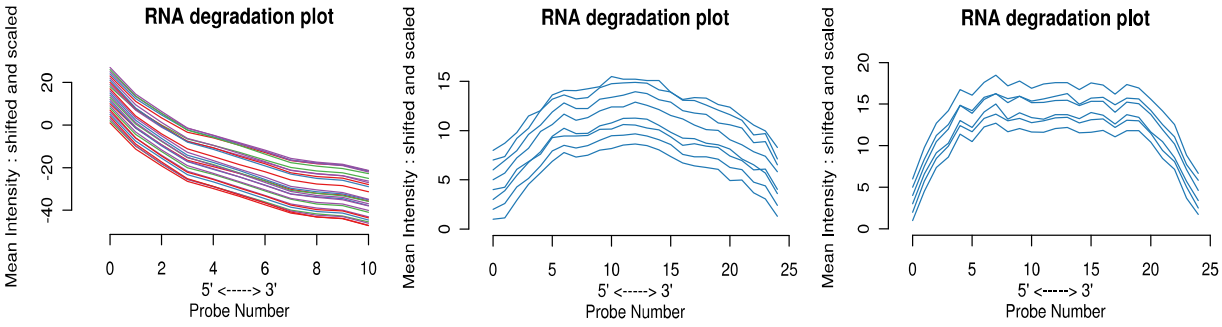


Figure 2 Each plot represents an microarray experiment realized on a specific platform. Plotted on the Y axis is mean intensity by probeset position. Intensities have been shifted from original data for a clearer view, but slope is unchanged.

Another quality assessment method uses the probe level model (Bolstad, 2004). According to this model, the intensity value of each probe depends on the actual expression of the corresponding gene in the array, the sequence of the probe and the error of measurement. This model uses the robust linear regression that is able to highlight whether the errors in the measurements are not uniformly distributed or whether several outliers are present. Based on this model, several quality assessment tools were constructed. Different pseudo-images can be created pointing out subtle artifacts that might otherwise be missed completely.

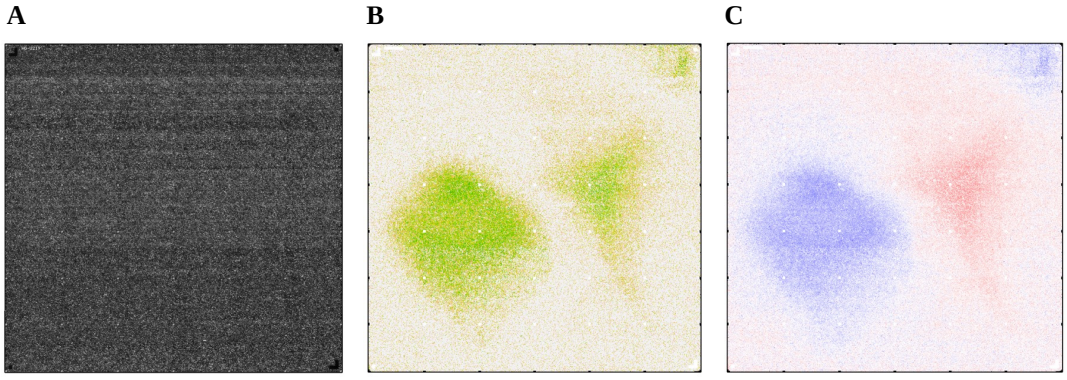


Figure 3 Chip pseudo-images based on PLM fit make visible subtle artifacts. Images of A) the raw probe intensities (log transformed) B) weights and C) residuals.

Two different plots are constructed based on the probe level model. The Relative Log Expression (RLE) plot shows for each array and for each gene across all arrays, the deviation of gene expression level from the median gene expression level. The majority of genes does not change the mRNA quantities. Hence, the boxplot with the difference calculated between the estimated expression value derived from the regression and the median across arrays should be centered around zero with a small spread. An array with quality problems may show significantly different values compared to the majority of arrays, resulting in a RLE box with greater spread or a median that deviates from 0.

The Normalized Unscaled Standard Error (NUSE) plot portrays the chip-wise distribution of standard error estimates, obtained for each gene on each array. To account for the fact that variability differs considerably between each gene, the error estimates are standardized so that the median standard error across arrays is 1 for each gene.

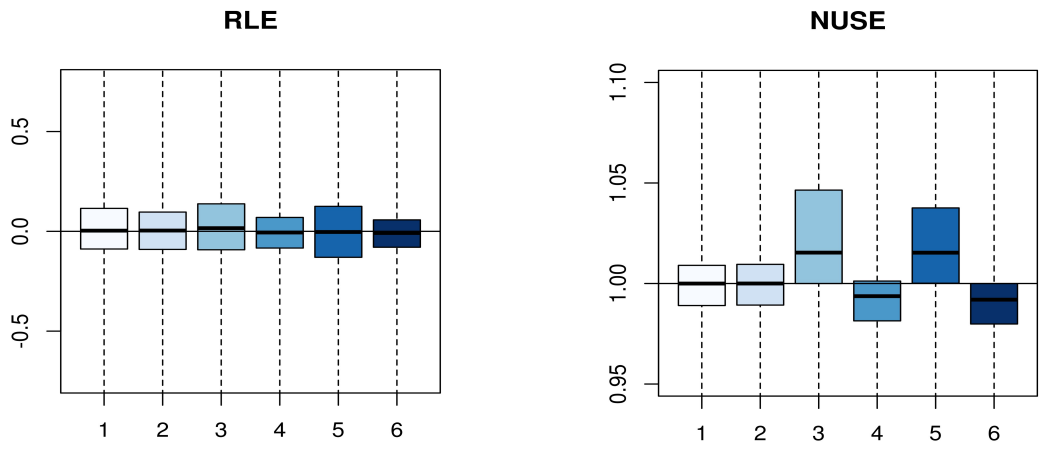


Figure 4 RLE plot and NUSE plot for the same data set. The array with the spatial effect deviates considerably from the other arrays.

DIFFERENTIALLY EXPRESSED GENES

The most common and basic question in DNA microarray experiments is whether genes appear to be down-regulated (the expression has decreased) or up-regulated (the expression has increased) between two or more groups of samples.

Assessing the differential expression of genes allows a deeper understanding about the difference between two groups of samples when the biological functions of those genes are well-known. Additionally the results can suggest possible functions for genes.

A gene is considered differentially expressed if its expression value between the compared conditions are significantly different. The relative gene expression value also known as "fold

change" (FC) is expressed as the logarithm of the ratio between the intensity value of the gene in an experimental condition and that in the reference condition. A threshold has to be fixed in order to distinguish up-regulated or down-regulated genes.

It is important to keep in mind that the absolute difference of expression simply quantifies the degree of intensity variation between the two groups, whereas it is essential to understand if the observed difference is significant. One of the most used methods to determine the significance of differentially expressed genes between two experimental conditions is the t-test. It compares the averages of the two groups of values (normalized by the standard deviation) in order to assess if their difference is significant, or whether it is only due to chance.

The bigger is the value of the t-statistics, the more significant is the difference between the expression in the two experimental conditions. t assumes high values when the averages are very different or when the variances of the two groups are low, indicating a very low dispersion (spread) of the data. The latter conditions may be due to chance in case of insufficient sampling.

The t-test is useful when the experimental conditions to be compared are only two. When the goal of a study is to compare several experimental conditions it is necessary to perform an analysis of variance (ANOVA) among the samples (Pavlidis, 2003). Anova is one of the most used methods to identify differentially expressed genes in complex experimental designs.

A multiple-testing correction such as FDR (false discovery rate) can be performed at the end to control the expected proportion of false positives among the declared significant results.

CLUSTERING

The values of gene expression can be used to classify conditions or gene expression profiles. The clustering procedure is complex. There are many different techniques to deal with it and also in only a few cases a unique solution is found. Nevertheless, any reasonable solution can help to simplify data analysis. Depending on the way in which the data are clustered we can distinguish between hierarchical and non-hierarchical clustering and supervised and unsupervised clustering (Eisen, 1998; Chipman, 2003; Gollub, 2006). The hierarchical techniques allow to establish an ordered relation among clusters, while, the non-hierarchical procedures, or at least most of these, allocate the expression profiles into a predefined number of clusters, without providing information on mutual ordering relations among them.

An unsupervised clustering is performed when no previous information about the samples is used to classify them and therefore the clustering is a consequence of data alone.
In the case of supervised clustering, instead, previous knowledge about the data is used to create different groups in order to make a class comparison with the extraction of a list of differentially expressed genes. The usage in grouping genes is based on the premise that co-expression is a result of co-regulation (Do, 2008). Genes with an identical profile of expression may represent a co-ordinate response to a stimulus. In order to quantify this relationship, and therefore compare the expression profile of two different genes, we need to use an appropriate metric. This metric can be changed in relation to the features, which are more important to be analyzed. Among the several metrics that can be defined in the context of gene expression profile analysis, the two most widely used are the Euclidean distance and Pearson correlation coefficient. The first metric is used to select genes whit similar magnitude of expression; the second is generally employed to search for genes expressed at different level but with the same overall profile (Dopazo, 2001).
Having defined a metric, one of the most used hierarchical clustering algorithm for microarray is UPGMA (Unweighted Pair Group Method with Arithmetic mean). In this method at each step two gene expression profiles and/or groups of gene expression profiles, which average distance is the shortest are clustered together, until no gene expression profile is left. UPGMA indeed is just a particular case of the so-called Linkage method. In this method the scheme is similar to that described for UPGMA with the difference that instead of choosing the average distance as the distance between two clusters (Average Linkage) we can choose the minimum distance of each member of a cluster from each member of the other (Single Linkage) or the maximum distance among them (Complete Linkage).

PCA

Clustering techniques can be used in combination with other exploratory methods that allow to visualize the complexity of the data in a two or three-dimensional space. PCA is able to select a small number of variables that can explain the most variability of the data by removing the background noise from the dataset. At the end of this procedure, some principal components are identified: weighted combinations of the original variables that have maximum variance and that are uncorrelated with one another (Alter, 2000; Raychaudhuri, 2000; Wall, 2001).

PUBLIC DATABASES OF GENE EXPRESSION PROFILES

Given the huge amount of data produced and the enormous utility derived from the opportunity to compare data coming from different experiments, several public repositories have been created to store and retrieve datasets of gene expression profile experiments. Moreover, to make the different data comparable, in 2001, a standard called MIAME (Minimal Information About a Microarray Experiment) has been developed in order to correctly and homogenously depositing microarray data (Brazma, 2001). The required information allow the right interpretation of the results and the possibility of reproducing the experiments.

The main public databases are Arrayexpress at the EBI – European Bioinformatics Institute (http://www.ebi.ac.uk/arrayexpress/) and GEO at the NCBI – National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/). In addition to these databases that contain data with different origin, there are others that store data related to specific sectors, such as caArray at the NCI – National Cancer Institute and Stanford Microarray Database.

BIOLOGICAL INTERPRETATION OF THE RESULTS

Gene Set Enrichment Analysis (GSEA) is a computational method for interpreting gene expression data that determines whether an a priori defined set of genes shows statistically significance, concordant differences between two biological states (Subramanian, 2005). All analysis can be performed using the GSEA software available at http://www.broadinstitute.org/gsea/index.jsp

DAVID is an integrated biological knowledge-base and analytic tool (available at http://david.abcc.ncifcrf.gov) that allows to extract biological meaning from large gene lists derived from high-throughput genomic experiments (Sherman, 2007). It uses more than 40 publicly available functional annotation sources and a novel agglomeration algorithm in order to define for each gene some identifiers. This analysis permits to associate not simply a biological enriched themes but also a set of gene clusters, a gene functional groups involved in well-defined functional pathways.

2.2 CASPASES

Caspases (cysteinyl aspartate proteinases) are a family of aspartate-specific cysteine proteases whose catalytic activity depends on the presence of a cysteine residue inside the highly conserved active site (pentamer QACRG). Caspases act on their substrate in a selective way: they recognize at least 4 contiguous amino acids in their substrate and they cleave the peptide bond after the 1st residue, which is usually an aspartic acid (Alnemri, 1996). Cys side chain is used as a nucleophile to accomplish peptide bond hydrolysis in several protease families.

Caspases are involved in the launch and in the execution of the apoptotic process, one of the most well-studied forms of programmed cell death. This process is genetically controlled and evolutionary conserved because of its fundamental role for the embryogenesis and the maintaining of the tissue homeostasis in the adult organism. Moreover, caspases are involved in different processes such as inflammatory responses, proliferation and differentiation (Launay, 2005; Lamkanfi, 2007). They were discovered initially in *Caenorhabditis elegans*, the organism where for the first time the apoptotic phenomenon was observed and characterized during the normal development. Until now, 11 genes, coding for 11 different caspases, are known in humans (Li, 2008). They are widely expressed with the exception of caspase-14 that is limited to keratinocytes (Denecker, 2008).

Caspases are synthesized like zymogens constituted by a small and a larger subunit and their activation necessitates the proteolytic removal of N-terminal prodomain. Caspases can be classified in two different ways: or in relation to the main function, as pro-apoptotic and pro-inflammation enzymes or on the bases of the prodomain dimension, as initiatior and effector caspases. Initiator caspases contain a long prodomain that mediates both homotypic interaction forming dimers or oligomers or heterotypic interaction with additional proteins that can exhibit both repressive and activation functions. In the initiator caspases the prodomains are structurally defined domains characterized by 6 alpha-helices. Two principally sub-types of pro-domain exist: the CARD (CAspase Recruitment Domain) and the DED (Death Effector Domain) both involved in interacting with the upstream adaptor molecules (Figure 5).

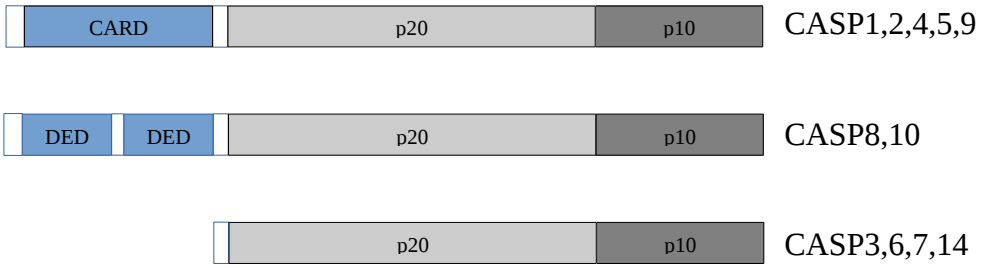


Figure 5 Domain organization of caspases.

Effector caspases present a short prodomain. They perform the downstream execution steps of apoptosis by cleaving multiple cellular substrates and are typically processed and activated by upstream caspases (Jin, 2005). Over 400 different substrates are known, information about these can be found in a public databases (Lüthi, 2007).

Caspases are not all equally active from the proteolytic point of view. The initiators seem to be more specific respect to the effectors, they are not able to cleave many substrates in addition to the precursors and the downstream caspases. Differently, effector caspases and caspase-3 in particular are responsible for most of the proteolytic processes occurring during apoptosis (Thornberry, 1998; Slee, 1999; Barnhart, 2003).

CASPASE-2

Caspase-2, initially described as Nedd-2/Ich-1 has been identified as a protein related to the *C. elegans* cell death protein CED-3 and mammalian interleukin-1β-converting enzyme (caspase-1). Despite the high degree of conservation, the contribution of caspase-2 to cell death is still mysterious and the mechanism controlling its regulation and function is much less understood when compared to other regulative caspases (Vakifahmetoglu-Norberg, 2010; Krumschnabel, 2009).

Caspase-2 was initially identified as a neuronal expressed gene that is down-regulated during brain development. Its mRNA level is much higher in developing CNS than in fully differentiated adult brain (Kumar, 1992). It presents a CARD domain followed by the large (p19) and the small subunits (p12), separated by a linker region. The two subunits are generated after proteolytic processing at Asp residues. The proenzyme can undergo to auto-processing and both subunits contribute to the formation of the catalytic site.

Two distinct caspase-2 mRNA species derived from alternative splicing encode two proteins, caspase-2L and caspase-2S. Overexpression of caspase-2L induces cell death, whereas

overexpression of caspase-2S can antagonize cell death. Caspase-2L is the dominant isoform that is expressed in most tissues (Wang, 1994).

CASP2 shows sequence homology with CASP9 and CASP1, but a specificity of cleavage similar to CASP3 and CASP7. Importantly its preference goes toward a pentapeptide (VDVAD/G) instead of the classical tetrapeptide (DEVD/G). In summary this protein shares features both with initiator and effector caspases.

Caspase-2 presents two signals for nuclear localization in the prodomain region. Mutation of a specific lysine residue within these NLS is sufficient to prevent the nuclear accumulation and to induce its accumulation into the cytoplasm (Paroni, 2002). There are evidences that in untreated cells, the only caspase present in the nucleus is CASP2, while after the activation of the apoptotic program both CASP2 and CASP3 can be found in the nuclear fraction (Zhivotovsky, 1999). CASP2 has been also localized in the mitocondria, in the endoplasmic reticulum and in Golgi-like structures, although the relative data are controversial.

CASP2 activation is primed by its dimerization. Subsequently, removal of the CARD domain and cleavages at the linker region between the large and the small subunits can be observed, to generate the tetrameric form. Caspase-2 is activated by proximity-induced oligomerization and trans-cleavage in vitro, and ectopic overexpression is sufficient for its activation in cells. In fact at high concentrations, self-association and subsequent oligomerization through the CARD domain occurs (Baliga, 2004).

Through its card domain, CASP2 interacts with several proteins. In response to DNA damage, the PIDDosome has been proposed to control CASP2 activation. The PIDDosome is a multiproteins complex containing the PIDD, encoded by a TP53–induced gene and CRADD/RAIDD (Tinel, 2004; Park, 2007). The latter contains one CARD and one DD domain, which respectively associate with the CARD domain of CASP2 and the DD domain of PIDD, respectively. PIDD may act as a switch molecule controlling the balance between cell survival and cell death in response to DNA damage. By increasing its expression or inhibiting its activity, it is possible to promote or to attenuate p53-mediated apoptosis, respectively (Lin, 2000; Tinel, 2004; Baptiste-Okoh, 2008).

Both apoptosis and CASP2 processing have been studied in PIDD-deficient mice. PIDDosome is not essential for DNA damage-induced cell death, but it is able to modulate CASP2 activity in processes leading to cell survival rather than cell death. Moreover, in cell types where CASP2 localization is mainly nuclear the PIDDosome is needed for its activation (Vakifahmetoglu, 2008; Shi, 2009). In other cells where the protein remains in the cytosol the

activation is independent from this complex (Vakifahmetoglu, 2006; Olsson, 2009; Manzl, 2009). These results suggest that there are other protein complexes involved in CASP2 activation (Manzl, 2012) (several evidences published) or in some cases, it is possible that this process is due to an autoproteolytic cleavage without the requirement of any adaptor protein for their oligomerization (Baliga, 2003). Furthermore in PIDD-null cells, CASP2 can be cleaved by CASP3 or CASP7 (Manzl, 2009), while in CASP9 deficient cells apoptotic stimuli abolish proteolytic processing of CASP2 (Paroni, 2001; Samraj, 2007).

In summary it remains unclear whether CASP2 really requires an activation platform or it may dimerize, auto-process and autoactivate in the absence of such a platform. It is also possible that different mechanisms of activation may be involved in response to different cellular stresses.

Different mechanisms act to regulate CASP2 activity. The first is the production of PIDD-fragment whit different dimensions, where the shorter is produced when cells undergo apoptosis, while the longer when repair mechanisms are needed after DNA damage (Janssens, 2005; Tinel, 2007).

Differential expression levels have been found for the two splicing variants 2L e 2S: the latter has a protective role against cell death probably through the formation of inactive caspase-2L/S heterodimers at the expense of pro-apoptotic, active caspase-2L homo-dimers (Droin, 2001).

CASP2 can be inactivated through phosphorylation of several residues and it can undergo N-terminal acetylation or SUMOylation (important for its subcellular localization) (Shirakura, 2005).

Function: *Casp2*-deficient mice don't help to evaluate the requirement for this enzyme in the apoptosis or in other physiological process. Mice carrying a null mutation for caspase-2 develop normally without severe phenotypic abnormalities (Bergeron, 1998). It is likely that other caspases compensate for its deficiency. Hence, this study did not contributed significantly to clarify the biological function of this protease.

Several studies have proposed several interaction partners for CASP2, suggesting the involvement of this protease in different signaling pathway. Here, a brief summary of the CASP2 possible functions as hypothesized in the literature (see Krumschnabel, 2009; Fava, 2012 and references cited therein).

Caspase-2 was proposed to be involved in DNA damage response (Zhivotovsky, 2005) but the exact role is debated and controversial. Some results suggest its engagement as initiator,

before the release of the cytochrome c and the apoptosome formation (Guo, 2002), others contradict this idea by indicating a role as downstream caspase in genotoxic cell death (Paroni, 2001; Paroni, 2002; O'Reilly, 2002; He, 2004; Bonzon, 2006).

Caspase-2 activation after DNA damage seems to be under the control of ataxia telangiectasia mutated (ATM) kinase, that is able to induce its activation through the phosphorylation of PIDD that leads to the PIDDosome formation (Ando, 2012). CASP2 is involved also in ER stress: under this condition IRE1α promotes a rapid degradation of microRNAs that target its mRNA, causing a fast protein synthesis, contributing to the induction of apoptosis (Upton, 2012).

Several non-apoptotic functions have been proposed to be influenced by caspases. An increase of CASP2 activity during the cell cycle leads to a delay in the G2 phase progression into M phase in an ATM-dependent manner (Shi, 2009). Caspase-2 is a sensor of the metabolic state of the cell: the phosphorylation of Ser135 by CaMKII kinase and the opposite process catalyzed by PP1 are both under the metabolic control (Nutt, 2005). Treatment of Xenopus oocytes with glucose-6-P activates the kinase that inhibits CASP2. When phosphorylated CASP2 is bound by 14-3-3ζ proteins and in this way maintained in an active state. Nutrient depletion promotes CASP2 dephosphorylation thanks to the action of PP1 and the release from the 14-3-3 (Nutt, 2009). Glucose-6-P acts on the interaction ability of 14-3-3ζ proteins, controlling in a still undefined manner its acetylation state (Andersen, 2011).

CASP2 is involved in aging: loss of CASP2 induces a premature aging due to an increase of oxidative stress (Shalini, 2012). It is also considered a tumor suppressor (reviewed in Puccini, 2013a and Olsson, 2014) since its absence provides some advantages in terms of oncogene-induced transformation (Ho, 2009; Parsons, 2013; Puccini, 2013b) even though in a tissue and context-specific manner (Dorstyn, 2014).

Sequence analysis of a 5-kb region upstream of the first non-coding exon of caspase-2 revealed the presence of 6 different putative binding site for the sterol regulatory element binding proteins SREBPs (Logette, 2005). Experiments of progressive deletions and mutations in these putative sites coupled with a ChIP assay, confirmed the binding of these TFs to the promoter region of CASP2. Using lipid-lowering drugs, such as the statins, which are able to trigger cholesterol depletion and a compensatory feedback mechanism that induces SREBP maturation, the transcription of their target genes as well as the level of CASP2 mRNA increase. Overexpression of the SREBPs result in a large increase of lipid levels; the simultaneous silencing of CASP2 reduces this augmentation by 50%.

CASPASES in the Central Nervous System (CNS)

Several studies have confirmed the engagement of caspases in the apoptotic processes that occur in CNS (Hyman, 2012). They are involved in the normal neuronal development and in programmed cell death events that accompanies normal cerebral processes or the appearance of acute and chronic neurodegenerative diseases. The global and full activation of caspases leads to apoptosis, whereas maintaining limited and localized activations it is possible to control normal physiology in living neurons (Hyman, 2012).

Apoptosis strongly characterized brain homeostasis. Approximately 30–70% of newly generated neurons in the adult brain undergo programmed cell death within 1 month of their initial production. Neurons undergo apoptosis also in case of trophic factor deprivation. Several caspases are involved in this process, both in developing and mature brain (Haviv, 1997; Deshmukh, 1997; Li, 1998).

Caspases have also non-apoptotic roles in CNS. They are able to regulate important neuronal functions, such as axon pruning and synapse elimination, processes that are the bases of neuronal plasticity, essential to refine mature neuronal circuits (Williams, 2006; Li, 2012).

Caspases are involved in normal synaptic physiology. The long-term depression (LTD), in which synapses become less sensitive to stimuli with a significant change in the dendrites' size at the spine level, can be blocked after inhibition of either caspase-3 and caspase-9 (Li, 2010).

Not surprisingly in chronic neurological diseases, like Alzheimer's disease and other forms of dementia caspases can play a role. They are activated after an acute neurological injury such as ischaemia in a region-specific and in a time-dependent manner. Generally at early times, in the necrotic core, caspase-8 and caspase-1 activation can be observed whereas at later point, in the surrounding region the involvement of caspase-9 was reported (Benchoua, 2001).

With regard to caspase-2, it is expressed at high level during brain development but it has also a strong involvement in the apoptotic processes of the adult CNS. Caspase-2 has been shown to be required for neuronal death induced by several stimuli, including NGF (nerve growth factor) deprivation and Aβ (β-amyloid) (Troy, 2000).

Both NGF deprivation and Aβ treatment of neurons induce caspase-2 activation and this activation depends on RAIDD expression, but is independent from PIDD expression (Ribe, 2012).

2.3 CHRONIC LYMPHOCYTIC LEUKEMIA

In the second part of my PhD thesis I used the microarray technology to define the genetic elements of responsible for the responsiveness of B-cell chronic lymphoid leukemia (B-CLL) to a new inhibitor of the ubiquitin proteasome system (Aleo, 2006; Fontanini, 2009; Foti, 2009).

Leukemias are malignant diseases characterized by a tumor proliferation of lymphoid or myeloid tissues, with the development of lymphoproliferative or myeloproliferative disorders, respectively. In both cases, the tissues present a decrease in differentiation and alterations in the mechanisms controlling programmed cell death. It can occur in an acute form, if the proliferation of tumor tissues affects cells unable to differentiate, or in a chronic form, when proliferation concerns cells able to differentiate, although not completely.

The chronic lymphoid leukemia (CLL) is a chronic lymphoproliferative disorder characterized by the proliferation and subsequent accumulation of mature lymphocytes inactive from the immunological point of view and not able to divide (Bertilaccio, 2010; Klein, 2010; Zenz, 2010) .

Accumulation of these lymphocytes affects the primary lymphoid organs (bone marrow) and secondary (lymph nodes and spleen). Compared to the other type of leukemia, CLL usually grows slowly and many patients may not have any symptom for years. It is generally harder to cure than acute leukemias. In some cases, the patient has several infections in the period before the diagnosis (Dameshek, 1967; Twomey, 1973).

There are two different subtypes of CLL. A subtype grows very slowly and for this reason it may take a long time before the patient needs treatment. On the opposite the second subtype grows faster and is a more serious disease (Rosenwald, 2001).

CLL is the most common form of leukemia in Western countries. It is a disease typical of the adult life with a median age at diagnosis of about 65 years, twice as common in males. 95% of CLL cases is derived from B cells, and only the remaining 5% by T lymphocytes (Seligmann, 1973).

The most widely used staging systems for CLL are two: the Rai staging system (Rai, 1975) and the Binet classification (Binet, 1981).

RAI SYSTEM			
Rai stage	**Modified Rai stage (risk)**	**Clinical characteristics**	**Median survival (yr)**
0	Low	Lymphocytosis in peripheral blood and bone marrow only	> 10
I	Intermediate	Lymphocytosis and enlarged lymph nodes	6
II		Lymphocytosis and enlarged spleen and/or liver	
III	High	Lymphocytosis and anemia (hemoglobin < 11 g/dL)	2
IV		Lymphocytosis and thrombocytopenia (platelets < 100×10^9/L)	

BINET SYSTEM		
Binet stage	**Clinical characteristics**	**Median survival (yr)**
A	Hemoglobin level ≥ 10 g/dL, platelet count ≥ 100×10^9/L, and < 3 areas involved	> 7
B	Hemoglobin level ≥ 10 g/dL, platelet count ≥ 100×10^9/L, and ≥ 3 areas involved	< 5
C	Hemoglobin level < 10 g/dL, platelet count < 100×10^9/L, or both (independent of areas involved)	< 2

Table 1 Rai and Binet staging systems for classification of CLL.

These two staging systems, although they are the most widely used in clinical practice and they are able to predict the clinical course of the disease in terms of survival from diagnosis, have limitations and not consider a number of biological parameters introduced recently that have now a fundamental role in the prognostic stratification of the disease and sometimes in the diversification of the therapeutic treatment (Döhner, 2000; Montillo, 2005; Hallek, 2008; Zenz; 2011).

The most important prognostic factors in CLL, in addition to the clinical stage, are markers of tumour load (the lymphocyte count, the lymphocyte doubling time, the serum LDH level or the bone marrow (BM) infiltration pattern), cellular protein expression (ZAP70 and CD38) and genetic parameters including immunoglobulin heavy chain variable gene segment (IGHV) mutational status, genomic aberrations and individual gene mutations (Kröber, 2002).

B-cell response to antigen is mediated through the B cell receptor (BCR) in normal and malignant B cells. CLL patients can be classified based on the assessment of B cell receptor mutational status (Hamblin, 1999; Hamblin, 2002).

While CLL with unmutated IGHV follows an unfavorable course with rapid progression and earlier death, CLL with mutated IGHV often shows slow progression and long survival.

Approximately 80% of CLL cases exhibit aberrations in a few recurrently affected chromosomal regions (Stilgenbauer, 2000). The aberrations linked to short survival have a

higher incidence in the case of IGHV unmutated status, even if both mutated and unmutated present genomic alterations (Stilgenbauer, 2002).

Principal genetic alterations in CLL

Deletion 13q14 occurs in more than half of the cases, it correlates with the best survival expectancy. In contrast to other recurrent aberrations, 13q14 deletions may be hetero- or (less commonly) homozygous. Studies of serial samples suggest that heterozygous deletion of 13q14 is an early event, whereas deletion of the second copy of this region occurs at a later stage (Shanafelt, 2006; Stilgenbauer, 2007).

Deletions of 11q22-q23: it affects patients (about 25%), which have a more rapid disease progression. These deletions often lead to a hemizygous loss of the ataxia telangiectasia (ATM) gene in almost all cases. No homozygous 11q deletions have been described. The ATM is a protein kinase involved in the response to DNA damage and DNA double-strand breaks. ATM disruption in CLL may occur by mutation, deletion, or a combination of both events: mutations of the ATM gene have been shown to occur in 12% of patients with CLL and 30% of patients with 11q deletion (Austen, 2005; Austen, 2007).

Trisomy 12 is a frequent aberration in CLL (10–20%), resulting in an intermediate outcome.

The molecular genetic defects associated with these risk categories are unknown. For trisomies in general, it is often assumed that there is a gene dosage effect of a candidate oncogene on the additional chromosome, but such a gene has yet to be identified on chromosome 12 (Winkler, 2005; Porpaczy, 2009).

Deletions 17p. It is often associated with unmutated IgHV and it confers the highest risk (poor prognosis and refractory disease).

While the deletion frequently encompasses most of short arm of chromosome 17, the minimally deleted region always involves 17p13, the locus of the TP53 gene.

Most cases (over 80%) with 17p deletion show loss of one copy and mutation of the remaining copy. The sole TP53 mutation occurs in approximately 5% of patients (Zenz, 2008; Dicker, 2009; Rossi, 2009).

Patients affected by CLL are treated with chemotherapy drugs or with monoclonal antibodies against surface antigens present in the B-lymphocytes. Taking into account the great heterogeneity, the prognosis is extremely variable, with a survival time between 2 and 20 years (Rodríguez-Vicente, 2013). Unfortunately, despite by combining different drug treatments it is possible to prolong the time free of disease, the CLL tends to relapse. The only

curative strategy is the allogeneic transplantation, used in very few cases (above all with young patients), given the very high mortality rate (Dreger, 2007; Gribben, 2009).

3. MATERIALS E METHODS

PIPELINE

Packages installation

R is a free software environment for statistical computing and graphics. It has several advantages: an active user community, a constant update, an easy installation and high quality standards.

It is possible to download R at http://cran.r-project.org/. It runs on all common operating systems.

The Bioconductor project supply additional libraries to manage biological data (Gentleman, 2004).

To quickly install a set of the most common packages

```
source("http://www.bioconductor.org/biocLite.R")
biocLite()
```

To install R libreries

```
install.packages("name_package_R”)
```

In the Affymetrix system, the raw image data are stored in so-called DAT files. Currently, Bioconductor software does not handle the image data, and the starting point for most analyses are the CEL files. These are the result of processing the DAT files using Affymetrix software to produce estimated probe intensity values. Each CEL file contains data on the intensity at each probe on the GeneChip, as well as some other quantities. To import CEL file data into the Bioconductor system, we use the ReadAffy function contained in the library *affy*. So first it is necessary to load the library *affy*, and then we issue the command ReadAffy:

```
library(affy)
Data <- ReadAffy()
```

which reads all the CEL files in the current working directory and return the probe-level data in an object of class AffyBatch. The function *list.celfiles()* can be used to show the CEL files that are located in the directory.

Affymetrix provides probe information in what are referred to as CDF files. These files denote which probe belongs to which probeset and whether the probe is a PM or MM.

The *affy* package automatically load a CDF package, when the package does not exist on the users system, it tries to download and install it.

For custom arrays, *makecdfenv* library can be used to convert the CDF file into an R package that has to be installed using the *R CMD INSTALL* command (Dai, 2005; Ferrari, 2007).
In this case, in order to use custom CDF packages, within the AffyBatch object, the value of the slot *cdfName* has to be replaced with the custom CDF name.

```
Data<-ReadAffy(cdfname ="custom_CDF_name")
```

The *pm* and *mm* accessory functions return the PM and MM probe intensities.
The function *probeNames* returns a vector of probeset IDs where the i^{th} entry corresponds to the i^{th} row of the PM matrix.The functions s*ampleNames* and *geneNames* access the names of the samples and Affymetrix probeset IDs stored in the AffyBatch object, respectively.

```
names_CEL<- sampleNames(Data)
names_gene<- geneNames(Data)
```

In order to modify and simplify sample names *sub* function can be used.

```
sampleNames(Data)<- sub("old_pattern" ,"new_pattern",
sampleNames(Data))
```

Preprocessing

Affymetrix expression arrays usually involves three steps: background adjustment, normalization, and summarization.

```
eset <- rma(Data)
eset <- mas5(Data)
```

The *expresso* function provides quite general facilities for computing expression summary values. In particular it allows most background adjustment, normalization, and summarization methods to be combined.

```
eset <- expresso()
```

In order to record normalize data, a dataframe has to be created. The *exprs* function creates an expression matrix in which columns represent the arrays and rows the probesets.

```
values<- exprs(eset)
norm<-data.frame(probeid=row.names(values),values)
write.table(norm, "datix.txt",sep="\t")
```

Quality Assessment of Affymetrix GeneChip

Various graphical tools that can be used to facilitate the decision of whether to remove an array from further analysis Exploratory data analysis has been the tool of choice for detection of problematic arrays. The *image* function can be used to create chip images of the raw

intensities.

To record the image in a png file, the object has to be create before the image creation and than it has to be *closed.*

```
png(" ")
image(name_CEL_i)
dev.off()
```

To generate an unique image with all the arrays together, the *par* function can be use.

```
par (mfrow=c(n_row,n_column))
{for (q in 1:length(names_CEL)image(Data[,i])}
```

Looking at the distribution of probe intensities across all arrays at once can sometimes demonstrate that one array is not like the others. The *boxplot* gives a simple summary of the distribution of probes. In the *affy* package the boxplot method is available for the AffyBatch class.

Looking at histograms of the probe-level data can also reveal problems. The hist function creates smoothed histograms (or density estimators) of the intensities for each array.

```
boxplot(Data)
hist(Data)
```

The *RColorBrewer* package is used to select a nice color palette.

install.packages("RcolorBrewer")

#Show all the colour schemes available

```
display.brewer.all()
scale<-colorRampPalette(brewer.pal(9,"Blues"))
(length(names_CEL))
scale<-
colorRampPalette(colors=c("color_1","color_2",...,"color_i"))
(length(names_CEL))
```

To apply the color scale, input the name of your custom scale in the argument for col.

RNA degradation

3'/5' ratios for several control genes can be used for quality assessment purposes.

We use the *AffyRNAdeg* function for this evaluation.

```
RNAdeg <- AffyRNAdeg(Data)
plotAffyRNAdeg(RNAdeg)
```

A summary of the slopes can be viewed in the following way

```
summaryAffyRNAdeg(RNAdeg)
```

Different chip types have different characteristic slopes because of the differences in probeset architecture.

Probe Level Model (PLM)

Numerous useful quality assessment tools can be derived from the output of the PLM fitting procedures.

Two different packages allow to do this evaluation.

```
library(oligo)
celFiles = list.celfiles ()
raw = read.celfiles (celFiles)
plm<-fitProbeLevelModel(raw)
show(plm)

library("affyPLM")
plm<-fitPLM(Data)
show(plm)
```

This library allows to obtain three different pseudo-images that highlight also subtle artifacts that could otherwise be missed.

```
image(plm,type = "weights")
image(plm,type = "resids")
image(plm,type = "sign.resids")
```

By using the PLM output, it is possible to use two other graphical tools.

In the RLE plot, boxes have small spread and they are centered near 0, problematic arrays deviate considerably from the others.

```
rle<-RLE(plm, las=2,cex.axis=0.6)
```

The NUSE plot shows boxes centered near 1, in the case of arrays with no spatial effect.

```
nuse<-NUSE(plm, las=2,cex.axis=0.6)
```

arrayQualityMetrics is a Bioconductor package that provides an exposition of diagnostic plots for one or two colour microarray data (Kauffmann, 2009). It uses powerful and comprehensive tools and generates an automatic report, in an *html* format.

```
library(arrayQualityMetrics)
arrayQualityMetrics(expressionset = Data,outdir = "datiraw")
```

By inserting *do.logtransform = TRUE* as option, the data are log converted, before doing the assessment.

Filtering

The aim of the filtering step is to discard the probesets with very low expression across the samples (and that provide no biological information) in order to reduce noise in data and to avoid wrong interpretations of the final results. In differential analysis for example, the number of hypothesis to test is therefore reduced (which is very useful in the multiple testing context).

A useful library to do this is EMA (Easy Microarray data Analysis). The function used is *expFilter.*

```
filter<-expFilter(eset,threshold = 3.5,p=0.01,graph = TRUE)
```

In this way, it keeps only probes with at least 1 sample with an expression level higher than 3.5.

With the functions *dim* and *summary* it is possible to obtain information about the result of this step.

An alternative approach provides for the use of *genefilter* package.

```
f1<- kOverA(1, 3.5)
ffun<- filterfun(f1)
filter<- genefilter(exprs(eset), ffun)
```

The result is the same.

Here f1 is a function that implies “expression measure above 3.5 in at least 1 sample” criterion, the function ffun is the filtering function that by using genefilter is applied to the data.

Annotating a platform

Bioconductor provides an extensive catalog of R packages that encodes annotation resources. From the web site is possible to download a library containing all the informations about the probesets in the platform. This .db package has to be installed and loaded to be opened. So fields of interest can be chosen, after the visualization with the *ls* function.

```
library(annotate)
library("platform_name.db")
ls("platform_name.db")
```

A dataframe with all the probe informations has to be prepared and recorded in a .csv file.

```
ID <- featureNames(eset)
Symbol <- getSYMBOL(ID, "platform_name.db")
Name <- as.character(lookUp(ID, "platform_name.db",
"GENENAME"))
Ensembl <- as.character(lookUp(ID, "platform_name.db",
"ENSEMBL"))
df <- data.frame(ID=ID,Symbol=Symbol, Name=Name,
Ensembl=Ensembl)
write.table(df,"Annotation.csv",sep="\t")
```

In the case of Exon GeneChip, we can adopt a different strategy. The CEL file has to be read in an alternative way in order to create a different object.

It allows the user to obtain summaries at the probeset level and also at the transcript level. For Exon arrays, there are three possible options for transcript level summarization: core, full and extended.

```
library(oligoData)
exonCELs <- list.celfiles()
affyExonFS <- read.celfiles(exonCELs)
data(affyExonFS)
exonPS <- rma(affyExonFS, target = "probeset")
exonCore <- rma(affyExonFS, target = "core")
```

Biological annotation can be obtained with the *getNetAffx* function. It will return an AnnotatedDataFrame object, which contains all the information about the platform features.

```
featureData(exonPS) <- getNetAffx(exonPS, "probeset")
featureData(exonCore) <- getNetAffx(exonCore, "transcript")
```

In order to select only few fields, we visualize all the content available through the index, obtained in the following way

```
varLabels(featureData(exonCore))
```

To estract all the gene symbols

```
x<-pData(featureData(exonCore))[1:length(exonCore),
"geneassignment"]
write.table(x,"geneassignmentCore.txt",sep="\t")
```

For Illumina platform the annotations can be directly downloaded from the web site as .bgx

file. The *illuminaio* package is then needed to convert this in csv format.

```
library(illuminaio)
A<-readBGX("file.bgx")
```

To extract the list of gene symbols

```
A1<-A[[1]]
write.table(A1,"file.csv",sep="\t")
```

Differentially Expressed Genes (DEG) selection

The choice to use *maanova* package instead of the classical aov analysis, was motivated by the fact that it is a powerful tool to manage both small and large scale microarray data sets, also arising from sophisticate designed experiments.

The first step is the preparation of design file with a well-defined fields. One has to check order and names of samples to set up design matrix.

One must have an Array column (in the same order respect to the names_CEL object) and other three columns, containg Strain, Sample and Dye. Dye is reserved for the staining dye, and Sample is reserved to identify biological replicates. Usually we have to assign each biological subject a unique sample number.

```
design.table <- data.frame(Array=row.names(pData(Data)),
Strain = c(" ",...), Sample = c(1,...), Dye = c(1,...))
```

This object can be recorded as a table to be used several times.

read.madata function allows to connect the normalized data to the structure of the whole experiment. It reads microarray data from TAB delimited text file or matrix type R object (in this case the *values* object).

```
pdata <- read.madata(values, designfile=designfile)
```

Model fitting

To fit model, one needs to specify the formula that includes factors affecting the gene expression among ones specified in designfile. The *fitmaanova* function adapts the ANOVA model to the microarray data. This is a time-consuming step.

```
fit.full.mix <- fitmaanova(pdata, formula = ~Strain+Sample,
random = ~Sample)
```

or, in the cases of absence of experimental replicates, a more semplified formula

```
fit.full.mix <-fitmaanova(pdata,formula = ~Sample, random =
~1)
```

To test how significantly a given term or terms affect to the gene expression level, R/maanova calculates F-test or t-test. The classical ANOVA F-test is a generalization of the t-test that allows for the comparison of more than two samples. The F-test is designed to detect any pattern of differential expression among several conditions by comparing the variation among replicated samples within and between conditions.

In order to obtain all possible pairwise comparison we use the *PairContrast* function.

```
C = PairContrast(n)
```

where *n* is the number of levels in test term.

If one is interested in specific pairwise comparisons, the contrast matrix has to be created. This is a matrix in which the number of columns corresponds to the number of levels and the rows to the comparisons that are going to be made. This is an example

```
C = matrix(c(-1,1,0,0,-1,1,-1,0,1), ncol=3, byrow=T)
```

At this point, we use the *matest* function to perform the statistical evaluation. R/maanova provides permutation method to calculate the significance of each test statistics.

```
ftest.pair <- matest(pdata, fit.full.mix, Contrast = C,
term="Strain", n.perm=100)
```

The final step is the data export through the *summarytable* function that provide summary of matest, such as pvalue for the classical F statistic (F1), for F statistic based on the James-Stein shrinkage estimates of the error variance (Fs) and fold-change.

```
summarytable(ftest.pair,outfile='summarytable.csv')
```

To select the DEGs both fold-change and pvalue have to be taken in account.

Paired samples

The computation of the significance in the case of natural pairing of experimental units requires the use of paired t-test. The variable under examination is the difference between couple of values.

The starting point is a different table in which the two groups are clustered together.

```
even_indexes<-seq(2,length(names_CEL),2)
odd_indexes<-seq(1,length(names_CEL),2)
ct=A[even_indexes]
tr=A[odd_indexes]
B<-cbind(tr,ct)
pvalue<-apply(B, 1,function(x)
{ t.test(x[1:length(even_indexes)], x[(length(odd_indexes)+1):
(length(even_indexes)*2)],paired=TRUE)$p.value } )
```

The p-value correction for multiple testing must be finally applied:

```
pvalue_adj<-p.adjust(pvalue, method ="method_chosen")
```

Annotation insertion

The creation of the summarizing table containing normalized data, annotations, anova results is done in two ways: in the case of files with the same order we can use *cbind* function, otherwise we are obliged to use *merge*, selecting a common field as probeID.

Correlations and heatmap

In order to calculate the correlation between couple of genes, two different functions can be used.

R provides the *cor* function that allows to select the method with no correction for multiple testing. In the *psych* library, with *corr.test* it is possible to choose both the method and the adjustment.

The first step is the table reading and transposition

```
A<-read.table("Table.csv",header=TRUE,row.names = 1)
A<-t(A)
```

the computation of global correlation

```
B<-cor(A,method="method_chosen”)
```

For a specific gene against all the others, we have to create two matrices

```
library(psych)
single_gene<-A[,gene_position]
single_gene_matrix<-as.matrix(single_gene)
all_genes<-A[,1:length(names_gene)]
B<-corr.test(single_gene_matrix, all_genes,
method="method_chosen”, adjust="adjust_chosen”)
```

To extract the correlation values and the respective significance from the list just created

```
C=B[[1]]
D=B[[4]]
```

These objects have to be transposed before being inserted in a dataframe for the final recording

```
correlazione=t(C)
pvalue_adj=t(D)
gene_symbol=row.names(correlazione)
results<-
```

```
data.frame(gene_symbol,correlazione,pvalue_adj,stringsAsFactor
s=F)
write.table(results,"corr_risp.csv",sep="\t")
```

To simplify the interpretation of this large amount of data identifying genes with similar correlation values, we have used *hclust* and *agnes* functions

```
A<-t(A)
d <- dist(A,method="method_chosen")
fit <- hclust(d, method="method_chosen")
plot(fit)

library(cluster)
agn<-agnes(A,method="method_chosen",metric = "metric_chosen")
plot(agn)
```

Reducing the dimension of the data, it is possible to use the heatmap for the visualization. Three different functions exist.

```
library(gplots)
A_heatmap <- heatmap(A_matrix)
A_heatmap <- heatmap.2(A_matrix)

library(pheatmap)
A_matrix <- data.matrix(A)
A_heatmap <-pheatmap(A_matrix)
```

It is important to choose correctly the range of color to maximize the contrast. The classical scale for microarray is the red/green, but more options exist.

In order to make different images comparable, an absolute scale is created. Correlation values are between 1 and -1. The choice of the resolution value is arbitrary.

```
resolution=400
dosata=greenred(resolution)
a=ceiling((min(A_matrix)+1)/2*resolution)
b=ceiling((max(A_matrix)+1)/2*resolution)
A_heatmap <-pheatmap(A_matrix,col=dosata[a:b])
```

Partial Least Square (PLS) regression

In the case that samples are characterized by different features important for instance for their classification, PLS can be performed. Among all the genes it is possible to select a subset that can be used to predict a specific characteristic based on the expression level only.

```
library(pls)
fitplsr <- plsr(B ~ A)
pred <- predict(fitplsr, ncomp = 2)
```

New responses or scores are predicted using a fitted model and a new matrix of observations.

```
cor<-cor(y,pred)
plot(y,pred)

library(mixOmics)
plsr<-pls(A,B)
```

AWK

AWK is a programming language to analyze and to elaborate text files by organizing data in a table format.

It requires the definition of a condition and an action to be performed on the input data.

For instance, to extract the differentially expressed genes from the summarytable, we can insert in the awk command after "if" the selection criteria that has to be applied on the columns containing fold-change and Pvalue (the column number has to be preceded by the simbol $) and print $0 as action to extract the entire row when the criteria are satisfied.

An example

```
awk 'BEGIN { FS = "\t"} ; {if (($n_i< -1.5 &&  $n_j<0.05) || ($n_i> 1.5 &&  $n_j<0.05)){print $0} }' input.csv>output.csv
```

With *BEGIN* we specify the action to do before starting the input data reading. In this case, we indicate with *FS* the separation field of the input file. The boolean operators "or" (||), "and" (&&), and "not" (!) are used to define the selection criteria.

Shell scripts

To make multiple selections in an automatic way, the Unix shell is used. Simple but very useful scripts are created to increase the speed of each step and to guarantee the correctness of the results.

Basic linux commands (cat, ls, grep, comm, sort, and so on) are used inside a *while* and *for*

loops to have a repetitive execution of a list of these.

The syntax is:

```
while/for [ condition ]; do list of commands; done
```

Graphics

One of the strengths of the R language is the graphics. Simple, exploratory, high quality graphics are easy to produce.

R provides different possibilities of exporting graphs in png, pdf, jpeg format with R base package and emf format using the *devEMF* library.

Graphical functions are

plot(), boxplot(), hist(), barplot(), stripchart(), dotchart(). Two or more graphs can be overlayed using *add=TRUE* as option.

We can insert title, texts , legend, extra-axis, points, marginal texts and lines, changing several graphical parameters, into a pre-existent graphic with appropriate functions to obtain the best result.

Not only 2D plots can be created but also three-dimensional ones (*scatterplot3d* package).

```
library(scatterplot3d)
scatterplot3d(x,y,z)
plot(x,y)
```

The *grid* package allows to control the appearance and arrangement of graphical output. A well-defined graphic region is created through the *viewport* and *pushViewport* function.

Grid.rect and *grid.text* are used to insert a rectangle and a text in the image. The first function needs x and y coordinates, width, and height for specifying the locations and sizes of the rectangles to draw. The other one requires the text to draw as its first argument.

To set the margins of a graph in R, the *par()* function with the mar argument can be called.

```
par(mar=c(bottom,left,top,right))
```

With *grid.arrange* function in the gridExtra package it is possible to create multiple plots when more than one graphic has to be flanked in a single figure. This provides somewhat similar functionality to *par(mfrow=c())*.

A special graph is the Venn diagram. This allow to present logical relations between different sets. In the case of comparisons of different gene lists, the Venn diagram highlights the proportion of elements in common. Two libraries permit to do this.

```
library(gplots)
A<-read.table("tableA.csv")
```

```
B<-read.table("tableB.csv")
list<-list(A,B)
venn(list)

library(VennDiagram)
common <- intersect(A[,1],B[,1])
draw.pairwise.venn(length(A[,1]),length(B[,1]),length(common))
```

4. RESULTS and DISCUSSION

4.1 RESULTS OF THE FIRST PART

The manuscript published in the peer-reviewed journal PlosONE enclosed with this thesis and entitled “Transcriptomic Analysis Unveils Correlations between Regulative Apoptotic Caspases and Genes of Cholesterol” contains the results of my thesis work and all the significant figures and the biological conclusions in a much extensive form. In this paragraph, its subject is briefly summarized.

Although caspase-2 is the subject of many studies, its role in physiology and disease in not well-defined because of controversial results. Several reviews discuss the involvement of caspase-2 in different cellular processes (Krumschnabel, 2009; Vakifahmetoglu-Norberg, 2010; Fava, 2012).

To identify genes and pathways under its influence, its expression was silenced in a glioblastoma cell line, U87MG, considering the suggested roles of caspase-2 in brain development and homeostasis.

Glioblastomas (GBM) are among the most lethal tumors, highly recalcitrant to radiotherapies and chemotherapies.

After silencing only a small number of genes resulted differentially expressed: choosing a 1.5 fold cut-off, 24 genes were significantly down-regulated and 17 were significantly up-regulated.

Despite the limited transcriptional alterations, the results were in any case interesting. Among the genes up-regulated after caspase-2 silencing that could mark compensatory responses, *LIPA*, *CYP51A1* and *CYP1B1* are linked with cholesterol metabolism, and two others are in some way connected with it.

Taking into account that previous studies showed that caspase-2 is involved in the regulation of the levels of this metabolite, we decided to further investigate the relationship between this cystein-protease and cholesterol genes.

Since complex regulative networks influence cholesterol homeostasis in vivo, cell culture models are not the best tool to investigate correlations of gene expression. We decided to expand our study by retrieving all data available in public repositories about gene expression

profile studies in glioblastoma and human brain. Data coming from 13 datasets including 327 microarrays were used to assess the relation between *CASP2* and *LIPA* in glioblatoma. A weak but significant inverse correlation was confirmed.

To clarify whether the expression of other caspases found correlations with LIPA, we expanded this analysis to all caspases, as well as to elements of the molecular platform involved in caspase-2 activation (CRADD/RAIDD and PIDD). Furthermore, gene expression profile data from 12 datasets including 293 microarrays of normal brain were recovered for comparison. We found for *CASP2/LIPA* a good and significant inverse correlation in cerebellum and a positive one in hippocampus. This result could depend on the difference of cholesterol genes expression between different brain areas (Segatto, 2012).

Therefore, we decided to investigate with a more comprehensive approach the correlations between caspase-2 and the expression of genes involved in cholesterol homeostasis. In order to clarify whether or not these correlations are limited to caspase-2 we extended this study to all caspases.

Genes involved in cholesterol homeostasis, including steroidogenesis were extracted from Gene Ontology (GO) and integrated from literature data. In total we selected 256 genes, of which 166 were grouped into 5 categories: *biosynthesis*, *adsorption/import*, *export*, *steroid and bile acid synthesis* and *transcriptional regulators*.

To understand whether our discoveries were limited to CNS or could be observed also in other tissues, we decided to compare variations in the expression levels of caspases and of cholesterol genes in human liver, an essential organ for cholesterol homeostasis. Gene expression profiles from 5 datasets including 106 microarrays of normal human liver were interrogated.

Since changes in the expression of cholesterol genes are linked to several physiological/pathological conditions. To understand the significance of the describe correlations, we decided to interrogate more precisely the public available datasets form human brain used in our studies. We restricted our inquiries to one dataset (GSE17612) in which, information about the specific brain area, the anterior prefrontal cortex (Brodmann area 10) and subjects age were available.

4.2 RESULTS OF THE SECOND PART

In the last part of my PhD, I worked on microarray data with the purpose to build up a methodological approach to define pathways activated after drug treatment through gene expression profile analysis. In the specific case as a model, in order to develop a pipeline able to extract information concerning the cellular response to a specific compound, I used data coming form chronic lymphocytic leukemia (CLL) cells exposed to two different pharmacological treatments.

Given the high frequency of relapse and the difficulties in treating this disease, many efforts have been done to find novel effective and save treatments for chronic lymphocytic leukemia. In our work, CLL cells from patients affected by this disease, under treatment at the Hematology Department of the University of Padova, were treated with two different molecules both impacting on the ubiquitin-proteasome system (UPS). Due to the importance of the proteasome in the regulation of different cellular functions, targeting the ubiquitin-proteasome system has been used as a strategy for developing new treatments against cancer (Huang, 2009; Adams, 2004).

Several chemotherapeutic drugs induce cell death through the activation of apoptotic programs. By considering the fact that chemo-resistance is a consequence of mutational accumulations in genes involved in apoptosis, our group in the past focused the attention in the identification of new compounds able of inducing cell death through alternative mechanisms (Henderson, 2005). Fifty-seven small molecules were tested in a screening (Aleo, 2006) and at the end two were selected and only one underwent further investigations (Foti, 2009). Both isolated compounds were UPS inhibitors, which belong to the class of non-selective isopeptidase inhibitors. This family of compounds is, able to act against several isopeptidases, offering advantages in terms of effectiveness on different tumors.

UPS used in our work were bortezomib and a derivative of G5 named 2cPE.

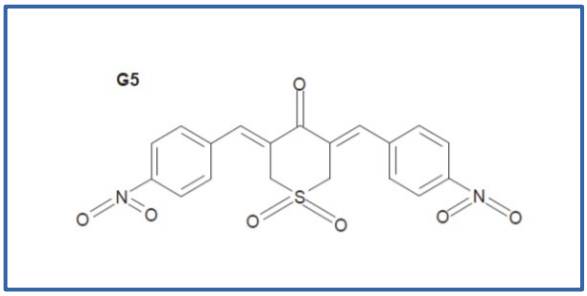


Figure 1. Chemical structure of G5

Bortezomib is a novel, slow-reversible, and highly specific proteasome inhibitor, which has been approved for the treatment of multiple myeloma and mantle cell lymphoma (Pérez-Galán, 2006).

G5 (Figure 1) was modified in order to optimize its use in vivo. Out from this approach a new molecule was generated named 2c that was next conjugated to polyethylene glycol (PEG), a common strategy to improve drug solubility and bioavailability (see section 5.2 of this thesis containing the recently published manuscript). The new compound was named 2cPE (Cersosimo, 2015).

Drug treatment and evaluation of cell viability

Initially, we compared the ability of the two UPS inhibitors to trigger cell death of CLL cells obtained from different patients. The percentages of cell survival after treatments for both drugs and the principal characteristics of the patients are reported in the table 1. Overall cells from different patients show certain variability in the cell death response and in particular to 2cPE treatment. These experiments were performed at the University of Padova.

	% of cell alive after 24 hours		prognostic markers			standard classification	
patient code	2cPE	BORT	SHM	FISH	CD38	RAI	BINET
LLC122	85	41	unmut	11q-	POS	0	A
LLC195	98	14	unmut	11q-; 13q-	POS	2	B
LLC270	83	35	unmut	11q-	POS	2	B
LLC300	12	11	unmut	13q-	NEG	2	
LLC305	25	23	mut	normale	NEG	0	A
LLC351	19	8	unmut	17q-; 13q-	POS	1	A
LLC366	55	33	nv	13q-; 12+; p53 mut	POS	0	A
LLC37	92	62	mut	13q-	NEG	2	
LLC43	32	18	mut	12+	POS	1	A
LLC4	18	19	mut	nd	NEG	1	

Table 1. Summary of the genetic profiles of the patients, which cells were object of the study. In columns 2 and 3 are reported the percentage of cell survival after treatment with 2cPE and bortezomib. Patients were categorized based on the standard classification and a set of prognostic factors. SHM =immunoglobulin receptor heavy chain (IGHV) hypermutation status; FISH = more frequent genetic aberrations; CD38 expression.

Single patient evaluation

Cells from the available patients were evaluated for the responsiveness to both UPS inhibitors, but since the principal aim of this study was the characterization of the response to 2cPE treatment, only a subgroup of cells incubated with bortezomib underwent the microarray analysis.

The RNA was extracted and a microarray analysis was performed. I processed all the CEL files together, by using standard tools available within the *affy* package. I used a UniGene ID

centered CDF (Chip Description file) in order to have only one intensity value per gene (Dai, 2005).

The cells obtained from different patients exhibited a different response in particular to 2cPE. In order to unveil the possible mechanisms of resistance to treatment each patient was considered separately. In this manner it was possible to appreciate which was the magnitude of the response in terms of alteration of gene expression.

	2cPE		Bortezomib	
	FC 1.5	FC 2.0	FC 1.5	FC 2.0
LLC122	109	17	211	25
LLC195	18	4	282	50
LLC270	49	12	162	23
LLC300	535	92	-	-
LLC305	2525	783	-	-
LLC351	632	169	-	-
LLC366	195	34	179	12
LLC37	258	71	283	71
LLC43	120	26	111	13
LLC4	1647	457	-	-

Table 2. Patient response in terms of number of genes, which expression was changed after treatments, as indicated.

Table 2 shows that in some cases, responsiveness has a direct correlation with the number of genes, which expression is altered after treatment, whereas in one case, the alteration is remarkable but the drug is not able to induce cell death. Hence, two mechanisms of resistance can be hypothesized. In the first case (patients LLC122, LLC195, LLC270) the drug is not able to trigger the initial response that is normal converted in the activation of specific genetic programs (gene expression changes). In the second case (patient LLC37) a block in the apoptotic responsiveness should be present at a certain level, after the engagement of the transcriptional response.

Extraction of genes differentially responsive to drug treatments

After the initial, single patient analysis, next step was to identify a list of genes which expression is influenced by the two compounds in a significative manner among all patients. The idea was to define common and specific patterns of response to the two drugs.

In order to obtain these information, a set of differentially expressed genes (DEG) was extracted, after an accurate no conventional quality assessment. Since the data of treated

samples are linked to their respective controls, we decided to use a *paired t-test* instead of the standard statistical method previously chosen. Therefore, the analysis was conducted by taking into account that, every single alteration in gene expression has to be referred to the corresponding level before treatment. In both cases, for each compound used, a list of genes was obtained (cut off: fold change 1.5; p value 0.05).

Figure 2 shows the result of such analysis.

For both down-regulated and up-regulated genes, there is only a partial overlapping of modulated genes (common pattern). Instead, the vast majority of these genes were modulated independently by the two compounds.

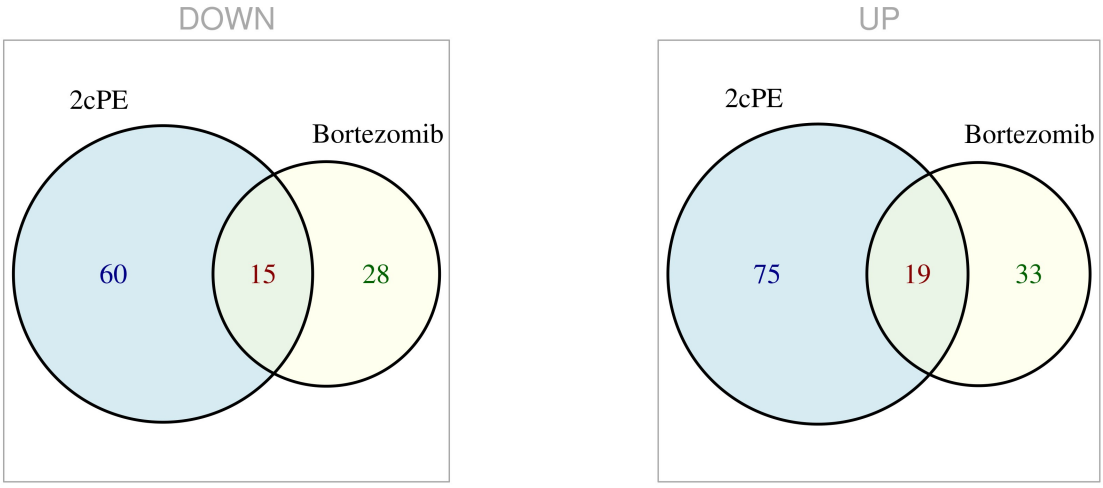


Figure 2. Venn diagram of down and up-regulated genes after 2cPE and Bortezomib treatments

Gene symbol	Bortezomib		2cPE	
	pvalue	FC	pvalue	FC
ADA	0.0477	-1.6127	0.0318	-1.6760
AKAP5	0.0116	-1.6217	0.0244	-1.6535
BTLA	0.0052	-1.5515	0.0492	-1.6378
DUSP4	0.0213	-1.9980	0.0059	-2.1808
ENC1	0.0146	-1.5824	0.0170	-1.6999
FAM111B	0.0094	-1.6207	0.0200	-1.7306
GAPT	0.0079	-1.6868	0.0283	-1.7190
IL2RA	0.0189	-1.6484	0.0228	-1.7765
LILRB2	0.0318	-1.6981	0.0296	-1.6349
MS4A7	0.0400	-1.9534	0.0291	-1.6023
PVRIG	0.0041	-2.0139	0.0262	-2.3021
PYCARD	0.0223	-1.5048	0.0224	-1.6480
RGCC	0.0256	-1.5511	0.0190	-1.5693
RNASE6	0.0142	-1.6640	0.0205	-1.9681
SOWAHD	0.0158	-1.5023	0.0286	-1.7865

Table 3. List of common down-regulated genes (Fold change > 1.5).

Gene symbol	Bortezomib		2cPE	
	pvalue	FC	pvalue	FC
BAG3	0.0086	2.1355	0.0262	4.2462
DNAJB1	0.0068	1.7294	0.0226	3.8534
DNAJB4	0.0042	1.6538	0.0186	4.4083
E2F6	0.0068	1.8537	0.0194	1.8695
GADD45A	0.0059	1.6606	0.0302	2.0034
GCLM	0.0043	1.9413	0.0142	5.0008
HSPA1A	0.0104	3.1499	0.0155	4.6622
HSPA1B	0.0102	3.2281	0.0183	6.9492
HSPA6	0.0209	1.8104	0.0315	7.6563
HSPA8	0.0159	1.6203	0.0179	1.9407
HSPH1	0.0074	1.6869	0.0203	3.2081
KEAP1	0.0064	1.5879	0.0114	1.5506
LRP8	0.0095	1.6167	0.0231	1.5270
MAFG	0.0051	1.6000	0.0188	2.1370
MIR22HG	0.0027	2.0390	0.0203	2.7014
PSMC2	0.0038	2.1857	0.0262	1.5162
PSMD14	0.0027	2.1162	0.0318	1.5110
SRXN1	0.0142	1.7259	0.0116	2.8933
ZFAND2A	0.0106	1.7024	0.0373	2.2129

Table 3. List of common up-regulated genes (Fold change > 1.5).

DAVID analysis

In order to understand pathways and cellular functions modified by the two UPS, I performed an over-representation analysis on each DEG list using DAVID (Sherman, 2007). Results are summarized in table 4 and table 5. It is evident that the two UPS inhibitors exert profoundly different effects on the cellular functions, possibly indicating different intracellular targets. In the case of 2cPE several genes linked protein misfolding, cellular stress and cell survival resulted differentially expressed. qRT-PCR experiments on selected genes (SQSTM1, HMOX1, GCLM, LCK) of the identified pathways confirmed the regulation operated by 2cPE (data not shown). On the contrary, after Bortezomib treatment several genes coding for proteasome components showed an altered expression, as previously observed (Mitsiades, 2002).

Term	Count	%	PValue	Benjamini	FDR
GO:0006986~response to unfolded protein	13	8.333	1.66013E-012	2.67113E-009	2.78174E-009
GO:0010033~response to organic substance	25	16.026	3.10830E-008	1.25031E-005	5.20837E-005
GO:0006457~protein folding	13	8.333	7.73350E-008	1.77759E-005	0.000129585
GO:0042981~regulation of apoptosis	25	16.026	2.38466E-007	4.79604E-005	0.0003995807
GO:0043066~negative regulation of apoptosis	16	10.256	8.25416E-007	0.0001106684	0.001383084

Table 4. Results of DAVID analysis of differentially expressed genes after 2cPE treatment.

Term	Count	%	PValue	Benjamini	FDR
GO:0000502~proteasome complex	14	15.730	4.31329E-018	5.26221E-016	4.96617E-015
GO:0051340~regulation of ligase activity	15	16.854	6.41469E-018	2.75832E-015	9.94011E-015
GO:0010498~proteasomal protein catabolic process	15	16.854	1.90722E-016	1.90958E-014	3.44169E-013
GO:0043085~positive regulation of catalytic activity	19	21.348	1.93518E-010	8.75925E-009	2.99873E-007
GO:0007049~cell cycle	17	19.101	2.85459E-006	7.01389E-005	0.0044233329

Table 5. Results of DAVID analysis of differentially expressed genes after Bortezomib treatment.

Hierarchical clustering

As subsequent analysis we decided to use the list of genes differentially expressed in response to the different inhibitors to investigate their use as markers of the responsiveness to the compounds and of the apoptotic susceptibility. To illustrate the data in a simple manner we selected the hierarchical clustering.

The initial intensities were converted into differences of expression because they are much more informative respect to the absolute values.

The lists of DGE were fused and used to extract the values of single patient differences, in an automatic way. An unsupervised clustering analysis was performed using bioconductor packages, and the results are presented as heatmap, a coloring code graph that simplify their interpretation.

Interestingly, the cluster analysis evidenced that in the case of 2cPE treatment a clear correlation between the magnitude of the expressional responses and induction of cell death is present. With the exclusion of patient LLC37 all patients which cells are responsive to 2cPE induced apoptosis cluster together. Similarly, patients, which cells are resistant to 2cPE treatment, evidence limited transcriptional changes and cluster together. As discussed above cells form LLC37 patient probably contain some alterations in the pro-death signaling acting after the initial transcriptional response elicited by 2cPE.

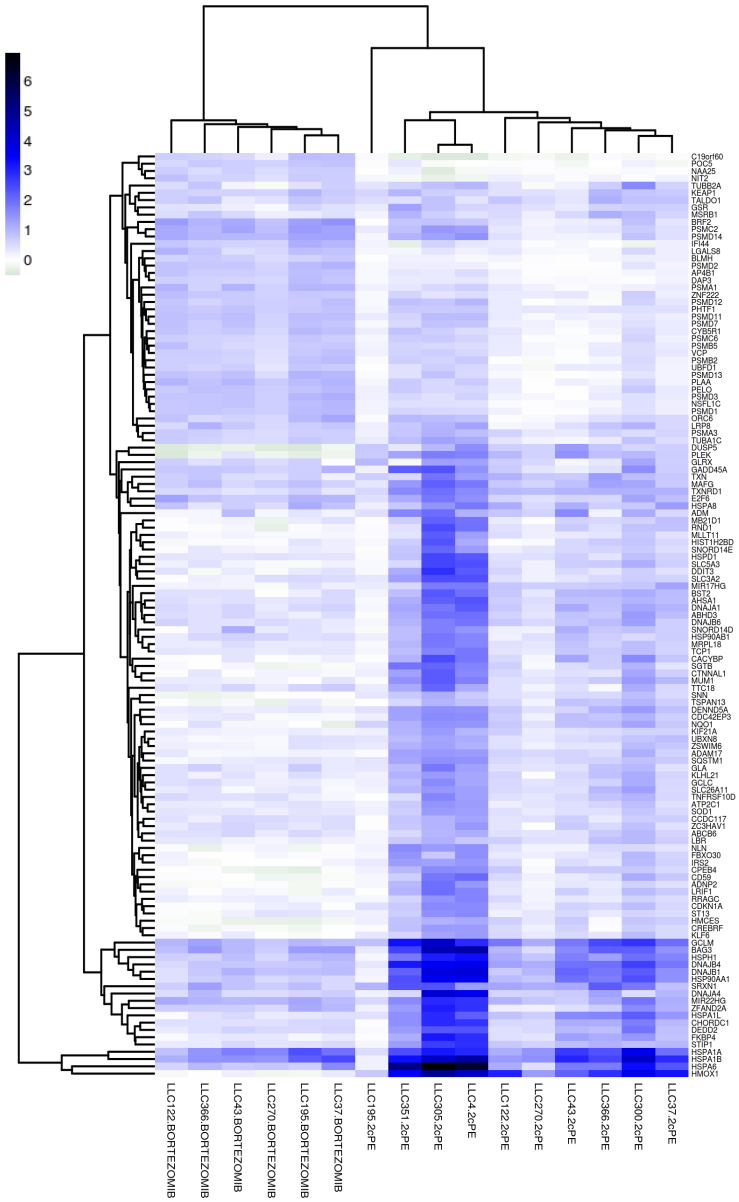


Figure 3. Differences of gene expression regarding the up-regulated genes (FC > 1.5) for 2cPE and bortezomib. The dendrograms are based on hierarchical clustering using the average linkage method.

Correlation between 2cPE treatment responses and single patient fold change

To further characterize the transcriptional response to 2cPE treatment and to identify genes, which could be involved in transducing its anti-proliferative signaling, we thought to calculate the spearman correlation between the treatment response in terms of percentage of cell survival and the single patient fold change of the differentially modulated genes (Benjamini-Hockberg was chosen as multiple test correction).

This study was done only for 2cPE given the low number of patients treated with bortezomib.

The gene that shows the best correlation, the top of the rank, in terms of responsiveness to 2cPE is SH3BP5 (r -0.9879 pvalue_adj 0.0017).

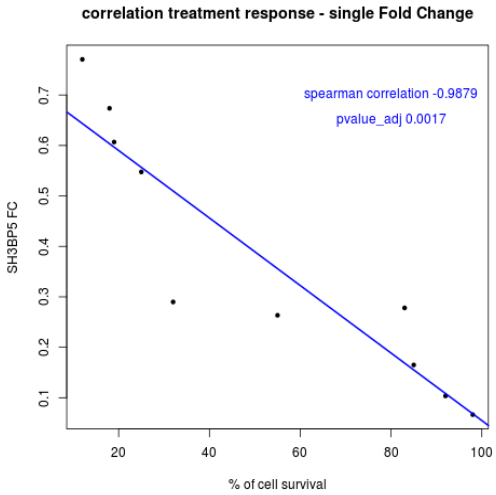


Figure 4. Plot of differences of mRNA levels for SH3BP5 versus patient responsiveness in term of % of cell survival after the treatment. Linear regression and Spearman correlation (Benjamini and Hochberg correction) are reported.

This result is particularly interesting because this gene codifies for a natural inhibitor of an important pathway involved in the regulation of CLL proliferation. B-cell receptor (BCR) signaling pathway has a central role in the CLL pathogenesis and it represents a potential therapeutic target (Burger, 2013; Woyach, 2012). SH3BP5 encodes for a negative regulatory protein that inhibits the auto and trans-phosphorylation activity of Btk, a critical component of this signaling pathway (Yamadori, 1999; Mohamed, 2009).

To further discover genes which differential expression could mark the responsiveness to 2cPE, a PLS regression (Abdi, 2013; Boulesteix, 2007) was performed over the global profile. This kind of exploratory technique allows to select variable with predictive capability. Among the measures obtained through this analysis some are more important for the determination of the best predictors (Mehmood, 2012).
The significativity of the regression will be independently assessed in the next future.
Several DEG were discovered also through this approach.

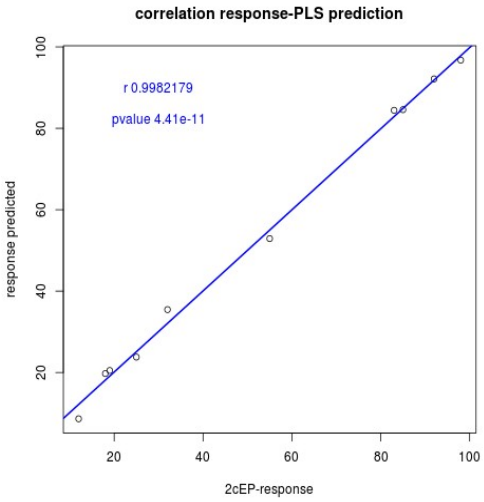


Figure 5. Plot of responsiveness in terms of % of cell survival versus the response prediction using the new factors called latent variables which have the best predictive power, identified after pls analysis.

After literature searching, we understood that several of these prognostic genes, up-regulated by 2cPE treatment are involved in the NRF2 response. This result was confirmed also by the GSEA analysis selecting the C3 (motif gene set), downloaded form the Molecular Signature Database.

NRF2 is a transcription factor activated in case of oxidative stress as a response to reactive oxygen species accumulation.

In normal conditions NRF2 is constantly degraded through the proteasome. When the levels of ROS increase, it dissociates from the regulatory protein KEAP1 (FC 1.5506) with consequent stabilization, translocation into the nucleus and transcription of the target genes (Itoh, 1999). Overall the bioinformatics analysis indicates that 2cPE triggers some sort of oxidative stress, possibly by dumping the anti-oxidant defenses (work in progress).

Gene symbol	Gene name	pvalue_adj	FC
HMOX1	heme oxygenase (decycling) 1	0.0142	7.9015
TXN	thioredoxin	0.0142	1.9739
NQO1	NAD(P)H dehydrogenase, quinone 1	0.0116	1.9085
TXNRD1	thioredoxin reductase 1	0.0114	2.2343
PRDX1	peroxiredoxin 1	0.0186	1.4725
GCLC	glutamate-cysteine ligase, catalytic subunit	0.0203	1.6711
GCLM	glutamate-cysteine ligase, modifier subunit	0.0142	5.0008
GSR	glutathione reductase	0.0170	1.5147
SOD1	superoxide dismutase 1, soluble	0.0235	1.5399
KEAP1	kelch-like ECH-associated protein 1	0.0114	1.5506
MAFG	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog G	0.0188	2.1370
CHOP	C/EBP homologous protein	0.0469	1.8440
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	0.0236	1.6544
GADD45A	growth arrest and DNA-damage-inducible, alpha	0.0302	2.0034
E2F6	E2F transcription factor 6	0.0194	1.8695
SQSTM1	sequestosome 1	0.0188	1.7508
BAG3	BCL2-associated athanogene 3	0.0262	4.2462
DNAJA1	DnaJ (Hsp40) homolog, subfamily A, member 1	0.0205	2.0938
DNAJA4	DnaJ (Hsp40) homolog, subfamily A, member 4	0.0502	2.4430
DNAJB1	DnaJ (Hsp40) homolog, subfamily B, member 1	0.0226	3.8534
DNAJB4	DnaJ (Hsp40) homolog, subfamily B, member 4	0.0186	4.4083
DNAJB6	DnaJ (Hsp40) homolog, subfamily B, member 6	0.0203	1.9334
HSP90AA1	heat shock protein 90kDa alpha (cytosolic), class A member 1	0.0220	3.7442
HSP90AB1	heat shock protein 90kDa alpha (cytosolic), class B member 1	0.0258	1.7337
HSPA1A	heat shock 70kDa protein 1A	0.0155	4.6622
HSPA1B	heat shock 70kDa protein 1B	0.0183	6.9492
HSPA1L	heat shock 70kDa protein 1-like	0.0176	2.8786
HSPA6	heat shock 70kDa protein 6	0.0315	7.6563
HSPA8	heat shock 70kDa protein 8	0.0179	1.9407
HSPD1	heat shock 60kDa protein 1 (chaperonin)	0.0378	1.8462
HSPH1	heat shock 105kDa/110kDa protein 1	0.0203	3.2081

Table 6. A subset of genes up-regulated by 2cPE

Alterations in the expression of proteasomal components

Being both the compounds ubiquitin-proteasome system inhibitors, we were interested to know how they altered the expressions of genes encoding for proteasomal components. In fact, it is well known that bortezomib, triggers compensatory response marked by the up-regulation of several proteasomal components (Mitsiades, 2002). I used Gene Ontology to extract a list of genes involved in proteasome formation assembly and homeostasis. Next I extracted the values of global (average) fold change for this gene family.

Results are presented in Figure 6. As expected in bortezomib treated cells the mRNA levels of several genes encoding for protesomal subunits were up-regulated. Surprisingly this adaptation, although present also in 2cPE treated cells was less noticeable.

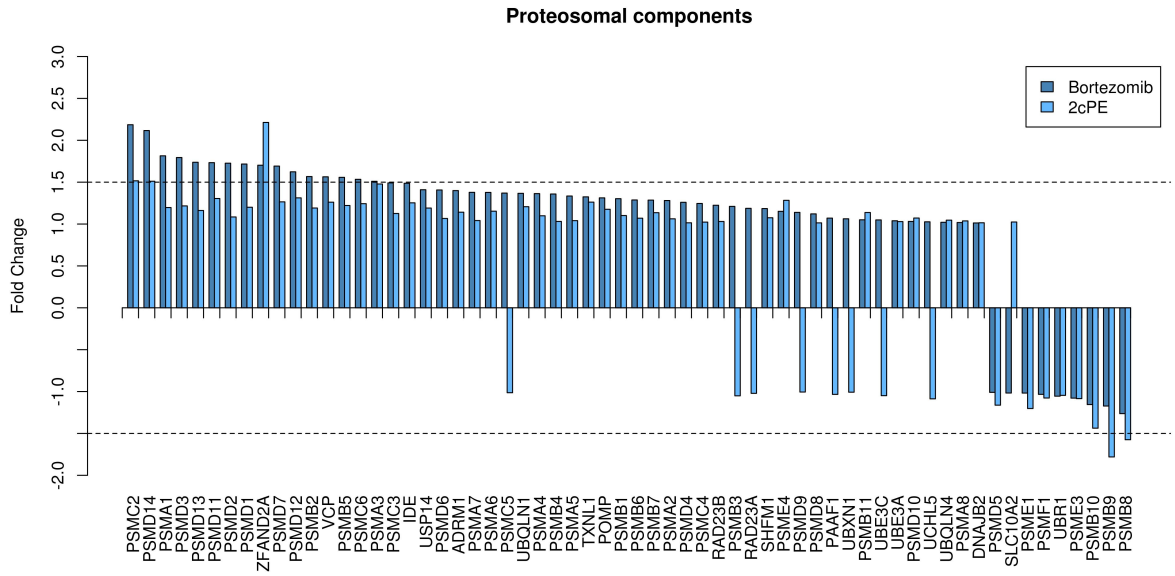


Figure 6. Fold change of the proteosomal components after 2cPE and Bortezomib treatments.

Clustering of no-treated samples

The analysis of genes differentially expressed following 2cPE treatment evidenced (with the exclusion of patient LLC37) a strong correlation between the magnitude of the response and the apoptotic susceptibility. Hence to identify putative predictors of the drug responsiveness we focused the analysis on the gene expression profiles characterizing CLL from different patients before drug-treatmet.

We used the *agnes* function in the bioconductor *cluster* package to clusterize samples using the average linkage method.

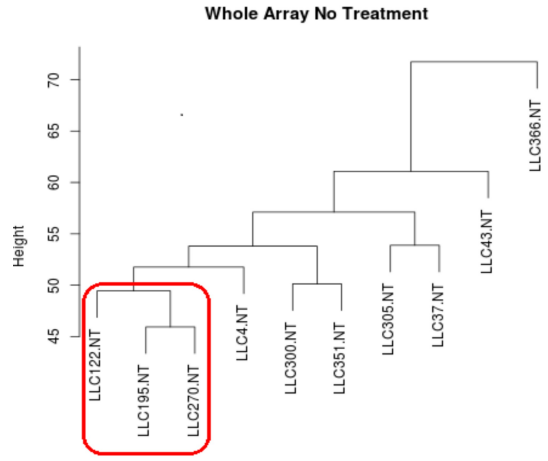


Figure 7. Hierarchical clustering of whole gene expression in CLL cells from different patients before the 2cPE treatment.

Importantly, three patients that showed low sensitivity to 2cPE treatment cluster together (Figure 7). In order to have information about which genes are much strongly responsible for this clustering separation, a standard *t-test* was performed. A the top of the rank the ATM gene was found (pvalue 2.9306E-005 FC -1.87).

This result is of particular interest, because all three patients under consideration are characterized by the 11q deletion, which affects the ATM locus (Stankovic, 2014).

Other two genes, NPAT and CUL3 lie in the same chromosome region and they score a significant reduction in terms of expression in the non-responsive patients (pvalue 0.0035 FC -1.69 and pvalue 0.0020 FC -1.64, respectively).

A subset of genes was extracted using pvalue and fold change (mean difference of expression between the two clusters) as selection criteria. This gene list is now under evaluation.

ATM is a protein kinase involved in DNA damage and oxidative responses. Several studies suggest that loss of ATM correlates with CLL resistance to chemotherapy and poor patient survival (Ripollés, 2006; Austen, 2007). The combined status of ATM and p53 is predictive of the response to chemotherapeutic treatments (Jiang, 2009). Preliminary studies have confirmed the activation of ATM following 2cPE treatment in the CLL cell line MEC-1 and further studies have been planned to evaluate the contribution of this serin-threonin kinase to the apoptotic response engaged by the isopeptidase inhibitor.

5. ADDITIONAL WORK

5.1 Next-Generation Sequencing Analysis of miRNAs Expression in Control and FSHD Myogenesis

Previous studies have demonstrated the involvement of miRNAs in muscle development, through the control of myoblast proliferation and differentiation. Raffaella Meneveri and coworkers used next-generation sequencing (NGS) to discover which miRNAs are specifically expressed during normal myogenesis, in comparison with a pathological condition. Specifically, they compared results obtained from normal tissues with those deriving from patients affected by facioscapulohumeral muscular dystrophy (FSHD), the third most common myopathy, with an incidence of 1 in 14.000 in the general population.

They obtained two list of miRNAs modulated exclusively in the normal and in the FSHD muscle. Control myogenesis showed the modulation of 38 miRNAs (4 down and 34 up-regulated) whereas FSDH myogenesis dysregulated miRNAs were only 15 (4 down and 11 up-regulated). During my PhD thesis I analyzed two different datasets, which have in common the same differentiation protocol and the platform to extrapolate a list of differentially expressed genes. TargetScan, a tool for the prediction of putative target genes, was used in order to define among all the differential regulated genes a set of potentially validated targets.

5.2 Synthesis, Characterization, and Optimization for in Vivo Delivery of a Nonselective Isopeptidase Inhibitor as New Antineoplastic Agent

The non-selective isopeptidase inhibitor G5 has a cytotoxic activity against different cancer cells. To improve its activity a small chemical library of G5-derivatives was generated. Among all the G5-like compounds, we selected 2c, which even though less active respect to G5, it was selected because capable of inducing apoptosis over necrosis, and for the presence of a OH group. This OH group allowed the addiction of PEG (Polyethyleneglycole) to increase 2c solubility in water. Conjugation of PEG to a low molecular weight drugs is a common strategy to increase the efficacy in vivo.

2c conjugated to PEG was named 2cPE and in this manner it behaves as a pro-drug. In fact 2cPE has to be converted into an active species through the action of secreted esterases.

It was observed that not all the cancer cell lines were capable of processing the prodrug. Furthermore, the conditioned medium of 2cPE responsive cell lines was able to transform unresponsive cells into responsive ones. Hence during this thesis, I applied genes expression profile studies to discover the secreted esterase involved in 2cPE maturation. A list of 173 putative esterases codified by the human genome was extracted from Gene Ontology and integrated through the use of HomoloGene. The enzyme capable to release the active form of 2c must have the expression pattern reflecting the difference in term of responsiveness of the cell lines used to assess the efficacy of the drug: higher level in the cell line with the highest responsiveness and lower in the case of the cell line less responsive.

Only seven esterases satisfied the expression profile. Among these enzymes, PLA2G7 was taken in consideration because of the structure of the interface binding area that appeared able to bind molecules like 2cPE.

6. CONCLUSIONS

The main purposes of my PhD were to get practice and skills in the analysis of data generated by the microarray technology and to develop a computational pipeline for this analysis. After a short period spent in the evaluation of freely available tools, we decided to use R to have a stronger control over all steps of the analysis. The applied pipeline has been adapted to different special cases of microarray data, including in house generated data and data recovered from public available databanks. When the experimental environment required it, the pipeline was expanded in order to extract different information from the data under consideration.

Several steps have been performed from the command line in order to speed up the results attainment and to increase their reliability and the correctness. A simple but more efficient script has been written to carry out many operations in sequence and AWK has been used to manage big files in a table format with substantial time saving. The computational tools have been applied to investigate different biological questions (the role of caspase-2; the anti-neoplastic response elicited by UPS inhibitors and the basis of genetic diseases).

Thanks to the analysis of different gene expression profiles I was able with this thesis:

I) to unveil correlations between cholesterol genes, regulative caspases and brain aging.

II) to propose the enzyme involved in the maturation of a pro-drug

III) to define survival pathways and resistance mechanisms engaged by a new UPS inhibitor.

The main goal of this thesis has been the creation of a computational service, available to all members of the Department, for the analysis of the microarray data, which is now fully operative.

7. Appendix

7.1 Scripts – part 1

```
R
library(affy)
library(maanova)
data<-ReadAffy()
eset <- rma(data)
design.table=data.frame(Array=row.names(pData(data)),Strain=c(
   "siC2mut","siC2mut","siC2wt","siC2wt","siPGAM5","siPGAM5","
   siUSP34","siUSP34"),Sample=c(1,1,2,2,3,3,4,4),Dye=c(1,1,1,1
   ,1,1,1,1))
write.table(design.table,"design.txt",sep="\t", row=F,quote=F)
datafile<-exprs(eset)
data<-data.frame(probeid=row.names(datafile),datafile)
write.table(data, "datix.txt",sep="\t",row=F,quote=F)
pdata <- read.madata(datafile, designfile="design.txt")
fit.full.mix <-fitmaanova(pdata,formula =
   ~Strain+Sample,random = ~Sample)
C = matrix(c(1,-1,0,0,1,0,-1,0,1,0,0,-1), ncol=4, byrow=T)
ftest.pair = matest(pdata, fit.full.mix, Contrast =
   C,term="Strain", n.perm=100)
summarytable(ftest.pair,outfile='shortsummarytablepair.csv')
annot = read.delim("annot.csv")
out= read.delim("shortsummarytablepair.csv")
out.annot = merge(annot,out,by.x="PROBE_ID", by.y="PROBE_ID",
   all.x=T, all.y=F, sort=F)
write.table(out.annot,"AnovaResults.txt",sep="\t",
   row=F,quote=F)
png("BoxplotNomi.png")
boxplot(data,title="Summary of data distribution",
   col=c("2","3","4","5","6","7","8","12","20"),lwd=0.9,adj=0.
   21,names=c("1","2","3","4","5","6","7","8"),cex.axis=0.8,la
   s=3,cex=1)
dev.off()
```

```
A<-read.table("datixCASP2.csv",header = TRUE, row.names = 1)
y1 <- (A[,1] + A[,2])/2
y2 <- (A[,3] + A[,4])/2
s2 <- ((A[,3] - A[,4])/2)^2
s1 <- ((A[,1] - A[,2])/2)^2
s <- sqrt((s1 +s2)/2)
s[s<0.08478]<-0.0847800
summary(s)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.0002991 0.0476900 0.0847800 0.1091000 0.1425000 1.2220000
d<-abs(y1-y2)
t<-d/s
pv<-1-pt(t,2)
A[,5]<-pv
colnames(A)[5]<- c("pvalue")
write.table(A,"tmp.txt",row.name=TRUE,col.name=TRUE,sep="\t")

library(affy)
data<-ReadAffy(cdfname ="HG-U133PLUS2_HS_ENTREZG")
eset <- mas5(data)
datafile<-exprs(eset)
datax<-data.frame(probeid=row.names(datafile),datafile)
write.table(datax, "datiBA.txt",sep="\t",row=F,quote=F)

library(devEMF)
A<-read.table("Brain_log.csv",header=TRUE,row.names = 1)
A<-t(A)
B<-read.table("Gliobl_log.csv",header=TRUE,row.names = 1)
B<-t(B)
emf("Overlay.emf",height=7, width=16)
boxplot(B,las = 2,outpch = NA,col="grey",at=c(2, 6, 10, 14,
   18, 22, 26, 30, 34, 38), xaxt = "n")
ll<-data.frame(B)
stripchart(ll,method="jitter",pch =19,cex=0.2,add =
   TRUE,vertical = TRUE,at=c(2, 6, 10, 14, 18, 22, 26, 30, 34,
   38))
boxplot(A,las = 2,outpch = NA,col="light blue", add=TRUE,
   at=c(1, 5, 9, 13, 17, 21, 25, 29, 33, 37))
```

```
l <-data.frame(A)
stripchart(l,method="jitter",pch =19,cex=0.2,add =
   TRUE,vertical = TRUE,at=c(1, 5, 9, 13, 17, 21, 25, 29, 33,
   37))
legend(2,2,c("Brain", "Glioblastoma"),pch = 21, pt.bg =
   c("light blue","grey"), pt.cex = 1.5)
axis(1,at=c(2, 6, 10, 14, 18, 22, 26, 30, 34,
   38),labels=c("CASP1","CASP2","CASP3","CASP4","CASP5","CASP6
   ","CASP7","CASP8","CASP9","CASP10"), tick=FALSE)
title(main="Caspases Expression in Brain and Glioblastoma",pch
   = 30)
dev.off()

A<-read.table("Brain_ages",row.names = 1)
plot(A[,1],A[,2])
fit<- glm(A[,1]~A[,2])
co <- coef(fit)
abline(fit)



head -1 datiBAGlioblplus2.txt >Tab_Glioblplus2.txt
for sonda in `cat probe_geni.txt`; do
echo $sonda
awk -v probe=$sonda '{if ($1 == probe){print $0;}}'
   datiBAGlioblplus2.txt >>Tab_Glioblplus2.txt
done

echo 'nome directory output?'
read outdir
elenco_categ_dir=categorie

mkdir $outdir
categorie=`ls ../$elenco_categ_dir/`
for categoria in $categorie; do
nome_categoria=`echo $categoria | sed s/.csv//`
echo $categoria
for gene in `cat ../$elenco_categ_dir/$categoria`; do
echo $gene
```

```
awk -v probe=$gene '{if ($1 == probe){print $0}}'
   corr_test.csv >> output_categorie/175_"$nome_categoria".csv
done
done
```

7.2 Scripts – part 2

```
R
library(affy)
data<-ReadAffy(cdfname ="HG-U219_HS_ENTREZG")
eset <- rma(data)
datafile<-exprs(eset)
datax<-data.frame(probeid=row.names(datafile),datafile)
write.table(datax, "datix.txt",sep="\t",row=F,quote=F)
#quality control
#differentially expressed genes determination
even_indexes<-seq(2,length(names_CEL),2)
odd_indexes<-seq(1,length(names_CEL),2)
ct=A[even_indexes]
tr=A[odd_indexes]
B<-cbind(tr,ct)
pvalue<-apply(B, 1,function(x)
   { t.test(x[1:length(even_indexes)], x[(length(odd_indexes)
   +1):(length(even_indexes)*2)],paired=TRUE)$p.value } )
ddd<-p.adjust(bbb, method ="fdr”)
write.table(ddd,"fdr.csv",sep="\t")


A<-read.table("SH3BP5_rips.csv",header=TRUE,row.names = 1)
A<-t(A)
fit<- glm(A[,2]~A[,1])
co <- coef(fit)
png("SH3BP5-survival.png")
plot(A[,1],A[,2],pch=20,main="correlation treatment response -
   single Fold Change",xlab="% of cell survival",ylab="SH3BP5
   FC")
text(80,0.7,"spearman correlation -0.9879", col="blue")
```

```
text(80,0.65,"pvalue_adj 0.0017", col="blue")
abline(fit, col="blue",lwd=2)
dev.off()

library(cluster)
A<-read.table("datiBA_NT_only.csv",header=TRUE,row.names = 1)
A<-t(A)
agn<-agnes(A,method="average",metric = "euclidean",stand =
   FALSE)
plot(agn)


#!/bin/csh -f
set outdir=OUT
mkdir $outdir
set i=1
set in=`(ls Bortezomib*_uniq.dat)`
while ($i <= $#in)
set nome_paziente1=`echo $in[$i] | sed s/_uniq.dat// | sed
   s/Bortezomib_//`
set ii=1
set n=`(ls 2FEP*_uniq.dat)`
while ($ii <= $#n)
set nome_paziente2=`echo $n[$ii] | sed s/_uniq.dat// | sed
   s/2FEP_//`
if ($nome_paziente1 == $nome_paziente2) then
echo $i 'contro' $ii
comm $in[$i] $n[$ii]> $outdir/"$nome_paziente1".dat
endif
@ ii = $ii + 1
end
@ i = $i + 1
end


#!/bin/csh
set i=1
set in=`(ls *.csv)`
```

```
while ($i <= $#in)
set nome_paziente=`echo $in[$i] | sed s/.csv//`
cat $in[$i] | awk '{print $2}' | grep -v NA | grep -v Symbol |
   sort | uniq > "$nome_paziente"_uniq.dat
@ i = $i + 1
end


#!/bin/csh -f
set outdir=OUT
mkdir $outdir
set i=1
set in=`(ls *_uniq.dat)`
while ($i <= $#in)
set ii=1
set n=`(ls *_uniq.dat)`
while ($ii <= $#n)
set nome_paziente=`echo $in[$i] | sed s/_uniq.dat//`
set paziente=`echo $n[$ii] | sed s/Bortezomib_// | sed
   s/_uniq.dat//`
echo $i 'contro' $ii
cat $in[$i] | grep -w -o -f  $in[$i] $n[$ii]>
   $outdir/"$nome_paziente"_"$paziente".dat
@ ii = $ii + 1
end
@ i = $i + 1
end


awk '{a[$0]++}END{for(i in a){if(a[i])print i,a[i]}}'
   cat.csv>count.csv
```

8. Bibliography

Abdi H, Williams LJ. **Partial least squares methods: partial least squares correlation and partial least square regression.** Methods Mol Biol. 2013;930:549-79.

Adams J. **The proteasome: a suitable antineoplastic target.** Nat Rev Cancer. 2004 May;4(5):349-60.

Alnemri ES, Livingston DJ, Nicholson DW, Salvesen G, Thornberry NA, Wong WW, Yuan J. **Human ICE/CED-3 protease nomenclature.** Cell. 1996 Oct 18;87(2):171.

Aleo E, Henderson CJ, Fontanini A, Solazzo B, Brancolini C. **Identification of new compounds that trigger apoptosome-independent caspase activation and apoptosis.** Cancer Res. 2006 Sep 15;66(18):9235-44.

Alter O, Brown PO, Botstein D. **Singular value decomposition for genome-wide expression data processing and modeling.** Proc Natl Acad Sci U S A. 2000 Aug 29;97(18):10101-6.

Andersen JL, Thompson JW, Lindblom KR, Johnson ES, Yang CS, Lilley LR, Freel CD, Moseley MA, Kornbluth S. **A biotin switch-based proteomics approach identifies 14-3-3ζ as a target of Sirt1 in the metabolic regulation of caspase-2.** Mol Cell. 2011 Sep 2;43(5):834-42.

Ando K, Kernan JL, Liu PH, Sanda T, Logette E, Tschopp J, Look AT, Wang J, Bouchier-Hayes L, Sidi S. **PIDD death-domain phosphorylation by ATM controls prodeath versus prosurvival PIDDosome signaling.** Mol Cell. 2012 Sep 14;47(5):681-93.

Austen B, Powell JE, Alvi A, Edwards I, Hooper L, Starczynski J, Taylor AM, Fegan C, Moss P, Stankovic T. **Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL.** Blood. 2005 Nov 1;106(9):3175-82.

Austen B, Skowronska A, Baker C, Powell JE, Gardiner A, Oscier D, Majid A, Dyer M, Siebert R, Taylor AM, Moss PA, Stankovic T. **Mutation status of the residual ATM allele is an important determinant of the cellular response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion.** J Clin Oncol. 2007 Dec 1;25(34):5448-57.

Baliga BC, Colussi PA, Read SH, Dias MM, Jans DA, Kumar S. **Role of prodomain in importin-mediated nuclear localization and activation of caspase-2.** J Biol Chem. 2003 Feb 14;278(7):4899-905.

Baliga BC, Read SH, Kumar S. **The biochemical mechanism of caspase-2 activation.** Cell Death Differ. 2004 Nov;11(11):1234-41.

Baptiste-Okoh N, Barsotti AM, Prives C. **A role for caspase 2 and PIDD in the process of**

p53-mediated apoptosis. Proc Natl Acad Sci U S A. 2008 Feb 12;105(6):1937-42.

Barnhart BC, Lee JC, Alappat EC, Peter ME. **The death effector domain protein family.** Oncogene. 2003 Nov 24;22(53):8634-44.

Benchoua A, Guégan C, Couriaud C, Hosseini H, Sampaïo N, Morin D, Onténiente B. **Specific caspase pathways are activated in the two stages of cerebral infarction.** J Neurosci. 2001 Sep 15;21(18):7127-34.

Bergeron L, Perez GI, Macdonald G, Shi L, Sun Y, Jurisicova A, Varmuza S, Latham KE, Flaws JA, Salter JC, Hara H, Moskowitz MA, Li E, Greenberg A, Tilly JL, Yuan J. **Defects in regulation of apoptosis in caspase-2-deficient mice.** Genes Dev. 1998 May 1;12(9):1304-14.

Bertilaccio MT, Scielzo C, Muzio M, Caligaris-Cappio F. **An overview of chronic lymphocytic leukaemia biology.** Best Pract Res Clin Haematol. 2010 Mar;23(1):21-32.

Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, Vaugier G, Potron G, Colona P, Oberling F, Thomas M, Tchernia G, Boivin P, Lesty C, Duault M, Monconduit M, Belabbes S, Gremy F. **A new prognostic classification of chronic lymphocytic leukaemia derived from a multivariate survival analysis.** Cancer. 1981; 48:198–206.

Bolstad B. **Probe-level model based test statistics for detecting differential expression.** PhD thesis. University of California, Berkeley; 2004.

Bonzon C, Bouchier-Hayes L, Pagliari LJ, Green DR, Newmeyer DD. **Caspase-2-induced apoptosis requires bid cleavage: a physiological role for bid in heat shock-induced death.** Mol Biol Cell. 2006 May;17(5):2150-7.

Boulesteix AL, Strimmer K. **Partial least squares: a versatile tool for the analysis of high-dimensional genomic data**. Brief Bioinform. 2007 Jan;8(1):32-44.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.** Nat Genet. 2001 Dec;29(4):365-71.

Burger JA, Chiorazzi N. **B cell receptor signaling in chronic lymphocytic leukemia.** Trends Immunol. 2013 Dec;34(12):592-601.

Cersosimo U, Sgorbissa A, Foti C, Drioli S, Angelica R, Tomasella A, Picco R, Semrau MS, Storici P, Benedetti F, Berti F, Brancolini C. **Synthesis, characterization, and optimization for in vivo delivery of a nonselective isopeptidase inhibitor as new antineoplastic agent.** J Med Chem. 2015 Feb 26;58(4):1691-704.

Chipman H, Hastie TJ, Tibshirani R. **Clustering microarray data** in Statistical Analysis of Gene Expression Microarray Data. Terry Speed, editor, Chapman & Hall/CRC, Boca Raton, Florida, 2003

Cleveland WS. **Robust locally weighted regression and smoothing scatterplots.** J. Am. Stat. Assoc. 1979; 74:829–836

Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F. **Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data**. Nucleic Acids Res. 2005 Nov 10;33(20):e175.

Dameshek W. **Chronic lymphocytic leukemia--an accumulative disease of immunolgically incompetent lymphocytes.** Blood. 1967 Apr;29(4):Suppl:566-84.

Denecker G, Ovaere P, Vandenabeele P, Declercq W. **Caspase-14 reveals its secrets.** J Cell Biol. 2008 Feb 11;180(3):451-8.

Deshmukh M, Johnson EM Jr. **Programmed cell death in neurons: focus on the pathway of nerve growth factor deprivation-induced death of sympathetic neurons.** Mol Pharmacol. 1997 Jun;51(6):897-906.

Dicker F, Herholz H, Schnittger S, Nakao A, Patten N, Wu L, Kern W, Haferlach T, Haferlach C. **The detection of TP53 mutations in chronic lymphocytic leukemia independently predicts rapid disease progression and is highly correlated with a complex aberrant karyotype.** Leukemia. 2009 Jan;23(1):117-24.

Do JH, Choi DK. **Clustering approaches to identifying gene expression patterns from DNA microarray data.** Mol Cells. 2008 Apr 30;25(2):279-88.

Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L, Döhner K, Bentz M, Lichter P. **Genomic aberrations and survival in chronic lymphocytic leukemia.** N Engl J Med. 2000, 343(26):1910-1916

Dopazo J, Zanders E, Dragoni I, Amphlett G, Falciani F. **Methods and approaches in the analysis of gene expression data.** J Immunol Methods. 2001 Apr;250(1-2):93-112.

Dorstyn L, Puccini J, Nikolic A, Shalini S, Wilson CH, Norris MD, Haber M, Kumar S. **An unexpected role for caspase-2 in neuroblastoma.** Cell Death Dis. 2014 Aug 21;5:e1383.

Dreger P, Corradini P, Kimby E, Michallet M, Milligan D, Schetelig J, Wiktor-Jedrzejczak W, Niederwieser D, Hallek M, Montserrat E; Chronic Leukemia Working Party of the EBMT. **Indications for allogeneic stem cell transplantation in chronic lymphocytic leukemia: the EBMT transplant consensus.** Leukemia. 2007 Jan;21(1):12-7.

Droin N, Rébé C, Bichat F, Hammann A, Bertrand R, Solary E. **Modulation of apoptosis by procaspase-2 short isoform: selective inhibition of chromatin condensation, apoptotic body formation and phosphatidylserine externalization.** Oncogene. 2001 Jan 11;20(2):260-9.

Eisen MB, Spellman PT, Brown PO, Botstein D. **Cluster analysis and display of genome-wide expression patterns.** Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):14863-8.

Fava LL, Bock FJ, Geley S, Villunger A. **Caspase-2 at a glance.** J Cell Sci. 2012 Dec

15;125(Pt 24):5911-5.

Ferrari F, Bortoluzzi S, Coppe A, Sirota A, Safran M, Shmoish M, Ferrari S, Lancet D, Danieli GA, Bicciato S. **Novel definition files for human GeneChips based on GeneAnnot.** BMC Bioinformatics. 2007 Nov 15;8:446.

Fontanini A, Foti C, Potu H, Crivellato E, Maestro R, Bernardi P, Demarchi F, Brancolini C. **The Isopeptidase Inhibitor G5 Triggers a Caspase-independent Necrotic Death in Cells Resistant to Apoptosis:** A COMPARATIVE STUDY WITH THE PROTEASOME INHIBITOR BORTEZOMIB. J Biol Chem. 2009 Mar 27;284(13):8369-81.

Foti C, Florean C, Pezzutto A, Roncaglia P, Tomasella A, Gustincich S, Brancolini C. **Characterization of caspase-dependent and caspase-independent deaths in glioblastoma cells treated with inhibitors of the ubiquitin-proteasome system.** Mol Cancer Ther. 2009 Nov;8(11):3140-50.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. **Bioconductor: open software development for computational biology and bioinformatics.** Genome Biol. 2004;5(10):R80.

Gollub J, Sherlock G. **Clustering microarray data.** Methods Enzymol. 2006;411:194-213.

Gribben JG. **Stem cell transplantation in chronic lymphocytic leukemia.** Biol Blood Marrow Transplant. 2009 Jan;15(1 Suppl):53-8.

Guo Y, Srinivasula SM, Druilhe A, Fernandes-Alnemri T, Alnemri ES. **Caspase-2 induces apoptosis by releasing proapoptotic proteins from mitochondria.** J Biol Chem. 2002 Apr 19;277(16):13430-7.

Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, Hillmen P, Keating MJ, Montserrat E, Rai KR, Kipps TJ; International Workshop on Chronic Lymphocytic Leukemia. **Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines.** Blood. 2008 Jun 15;111(12):5446-56.

Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. **Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia.** Blood. 1999 Sep 15;94(6):1848-54.

Hamblin T. **Chronic lymphocytic leukaemia: one disease or two?** Ann Hematol. 2002 Jun;81(6):299-303.

Haviv R, Lindenboim L, Li H, Yuan J, Stein R. **Need for caspases in apoptosis of trophic factor-deprived PC12 cells.** J Neurosci Res. 1997 Oct 1;50(1):69-80.

He Q, Huang Y, Sheikh MS. **Bax deficiency affects caspase-2 activation during ultraviolet radiation-induced apoptosis.** Oncogene. 2004 Feb 12;23(6):1321-5.

Henderson CJ, Aleo E, Fontanini A, Maestro R, Paroni G, Brancolini C. **Caspase activation and apoptosis in response to proteasome inhibitors.** Cell Death Differ. 2005 Sep;12(9):1240-54.

Ho LH, Taylor R, Dorstyn L, Cakouros D, Bouillet P, Kumar S. **A tumor suppressor function for caspase-2.** Proc Natl Acad Sci U S A. 2009 Mar 31;106(13):5336-41.

Hoheisel JD. **Microarray technology: beyond transcript profiling and genotype analysis.** Nat Rev Genet. 2006 Mar;7(3):200-10.

Huang L, Chen CH. **Proteasome regulators: activators and inhibitors.** Curr Med Chem. 2009;16(8):931-9.

Hyman BT, Yuan J. **Apoptotic and non-apoptotic roles of caspases in neuronal physiology and pathophysiology.** Nat Rev Neurosci. 2012 May 18;13(6):395-406.

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. **Summaries of Affymetrix GeneChip probe level data**. Nucleic Acids Res. 2003 Feb 15;31(4):e15.

Itoh K, Wakabayashi N, Katoh Y, Ishii T, Igarashi K, Engel JD, Yamamoto M. **Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain.** Genes Dev. 1999 Jan 1;13(1):76-86.

Janssens S, Tinel A, Lippens S, Tschopp J. **PIDD mediates NF-kappaB activation in response to DNA damage.** Cell. 2005 Dec 16;123(6):1079-92.

Jiang H, Reinhardt HC, Bartkova J, Tommiska J, Blomqvist C, Nevanlinna H, Bartek J, Yaffe MB, Hemann MT. **The combined status of ATM and p53 link tumor development with therapeutic response.** Genes Dev. 2009 Aug 15;23(16):1895-909.

Jin Z, El-Deiry WS. **Overview of cell death signaling pathways.** Cancer Biol Ther. 2005 Feb;4(2):139-63.

Kauffmann A, Gentleman R, Huber W. **arrayQualityMetrics--a bioconductor package for quality assessment of microarray data.** Bioinformatics. 2009 Feb 1;25(3):415-6.

Klein U, Dalla-Favera R. **New insights into the pathogenesis of chronic lymphocytic leukemia.** Semin Cancer Biol. 2010 Dec;20(6):377-83.

Kröber A, Seiler T, Benner A, Bullinger L, Brückle E, Lichter P, Döhner H, Stilgenbauer S. **V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia.** Blood. 2002 Aug 15;100(4):1410-6.

Krumschnabel G, Sohm B, Bock F, Manzl C, Villunger A. **The enigma of caspase-2: the laymen's view.** Cell Death Differ. 2009 Feb;16(2):195-207.

Kumar S, Tomooka Y, Noda M. **Identification of a set of genes with developmentally down-regulated expression in the mouse brain.** Biochem Biophys Res Commun. 1992 Jun 30;185(3):1155-61.

Lamkanfi M, Festjens N, Declercq W, Vanden Berghe T, Vandenabeele P. **Caspases in cell survival, proliferation and differentiation.** Cell Death Differ. 2007 Jan;14(1):44-55.

Launay S, Hermine O, Fontenay M, Kroemer G, Solary E, Garrido C. **Vital functions for lethal caspases.** Oncogene. 2005 Aug 4;24(33):5137-48.

Li J, Yuan J. **Caspases in apoptosis and beyond.** Oncogene. 2008 Oct 20;27(48):6194-206.

Li L, Prevette D, Oppenheim RW, Milligan CE. **Involvement of specific caspases in motoneuron cell death in vivo and in vitro following trophic factor deprivation.** Mol Cell Neurosci. 1998 Oct;12(3):157-67.

Li Z, Jo J, Jia JM, Lo SC, Whitcomb DJ, Jiao S, Cho K, Sheng M. **Caspase-3 activation via mitochondria is required for long-term depression and AMPA receptor internalization.** Cell. 2010 May 28;141(5):859-71.

Li Z, Sheng M. **Caspases in synaptic plasticity.** Mol Brain. 2012 May 14;5:15.

Lin Y, Ma W, Benchimol S. **Pidd, a new death-domain-containing protein, is induced by p53 and promotes apoptosis.** Nat Genet. 2000 Sep;26(1):122-7.

Logette E, Le Jossic-Corcos C, Masson D, Solier S, Sequeira-Legrand A, Dugail I, Lemaire-Ewing S, Desoche L, Solary E, Corcos L. **Caspase-2, a novel lipid sensor under the control of sterol regulatory element binding protein 2.** Mol Cell Biol. 2005 Nov;25(21):9621-31.

Lüthi AU, Martin SJ. **The CASBAH: a searchable database of caspase substrates.** Cell Death Differ. 2007 Apr;14(4):641-50.

Manzl C, Krumschnabel G, Bock F, Sohm B, Labi V, Baumgartner F, Logette E, Tschopp J, Villunger A. **Caspase-2 activation in the absence of PIDDosome formation.** J Cell Biol. 2009 Apr 20;185(2):291-303.

Manzl C, Peintner L, Krumschnabel G, Bock F, Labi V, Drach M, Newbold A, Johnstone R, Villunger A. **PIDDosome-independent tumor suppression by Caspase-2.** Cell Death Differ. 2012 Oct;19(10):1722-32.

Mehmood T, Liland KH, Snipen L, and Sæbø S **A review of variable selection methods in partial least squares regression.** Chemometrics and Intelligent Laboratory Systems. 2012;118:62-69

Mitsiades N, Mitsiades CS, Poulaki V, Chauhan D, Fanourakis G, Gu X, Bailey C, Joseph M, Libermann TA, Treon SP, Munshi NC, Richardson PG, Hideshima T, Anderson KC. **Molecular sequelae of proteasome inhibition in human multiple myeloma cells.** Proc Natl Acad Sci U S A. 2002 Oct 29;99(22):14374-9.

Mohamed AJ, Yu L, Bäckesjö CM, Vargas L, Faryal R, Aints A, Christensson B, Berglöf A, Vihinen M, Nore BF, Smith CI. **Bruton's tyrosine kinase (Btk): function, regulation, and transformation with special emphasis on the PH domain.** Immunol Rev. 2009 Mar;228(1):58-73.

Montillo M, Hamblin T, Hallek M, Montserrat E, Morra E. **Chronic lymphocytic leukemia: novel prognostic factors and their relevance for risk-adapted therapeutic strategies.** Haematologica. 2005 Mar;90(3):391-9.

Nutt LK, Margolis SS, Jensen M, Herman CE, Dunphy WG, Rathmell JC, Kornbluth S. **Metabolic regulation of oocyte cell death through the CaMKII-mediated phosphorylation of caspase-2.** Cell. 2005 Oct 7;123(1):89-103.

Nutt LK, Buchakjian MR, Gan E, Darbandi R, Yoon SY, Wu JQ, Miyamoto YJ, Gibbons JA, Andersen JL, Freel CD, Tang W, He C, Kurokawa M, Wang Y, Margolis SS, Fissore RA, Kornbluth S. **Metabolic control of oocyte apoptosis mediated by 14-3-3zeta-regulated dephosphorylation of caspase-2.** Dev Cell. 2009 Jun;16(6):856-66.

O'Reilly LA, Ekert P, Harvey N, Marsden V, Cullen L, Vaux DL, Hacker G, Magnusson C, Pakusch M, Cecconi F, Kuida K, Strasser A, Huang DC, Kumar S. **Caspase-2 is not required for thymocyte or neuronal apoptosis even though cleavage of caspase-2 is dependent on both Apaf-1 and caspase-9.** Cell Death Differ. 2002 Aug;9(8):832-41.

Olsson M, Vakifahmetoglu H, Abruzzo PM, Högstrand K, Grandien A, Zhivotovsky B. **DISC-mediated activation of caspase-2 in DNA damage-induced apoptosis.** Oncogene. 2009 May 7;28(18):1949-59.

Olsson M, Forsberg J, Zhivotovsky B. **Caspase-2: the reinvented enzyme.** Oncogene. 2014 Jun 2.

Park HH, Logette E, Raunser S, Cuenin S, Walz T, Tschopp J, Wu H. **Death domain assembly mechanism revealed by crystal structure of the oligomeric PIDDosome core complex.** Cell. 2007 Feb 9;128(3):533-46.

Paroni G, Henderson C, Schneider C, Brancolini C. **Caspase-2-induced apoptosis is dependent on caspase-9, but its processing during UV- or tumor necrosis factor-dependent cell death requires caspase-3.** J Biol Chem. 2001 Jun 15;276(24):21907-15.

Paroni G, Henderson C, Schneider C, Brancolini C. **Caspase-2 can trigger cytochrome C release and apoptosis from the nucleus.** J Biol Chem. 2002 Apr 26;277(17):15147-61.

Parsons MJ, McCormick L, Janke L, Howard A, Bouchier-Hayes L, Green DR. **Genetic deletion of caspase-2 accelerates MMTV/c-neu-driven mammary carcinogenesis in mice.** Cell Death Differ. 2013 Sep;20(9):1174-82.

Pavlidis P. **Using ANOVA for gene selection from microarray studies of the nervous system.** Methods. 2003 Dec;31(4):282-9.

Pérez-Galán P, Roué G, Villamor N, Montserrat E, Campo E, Colomer D. **The proteasome inhibitor bortezomib induces apoptosis in mantle-cell lymphoma through generation of ROS and Noxa activation independent of p53 status.** Blood. 2006 Jan 1;107(1):257-64.

Porpaczy E, Bilban M, Heinze G, Gruber M, Vanura K, Schwarzinger I, Stilgenbauer S, Streubel B, Fonatsch C, Jaeger U. **Gene expression signature of chronic lymphocytic**

leukaemia with Trisomy 12. Eur J Clin Invest. 2009 Jul;39(7):568-75.

Puccini J, Dorstyn L, Kumar S. **Caspase-2 as a tumour suppressor.** Cell Death Differ. 2013 Sep;20(9):1133-9.

Puccini J, Shalini S, Voss AK, Gatei M, Wilson CH, Hiwase DK, Lavin MF, Dorstyn L, Kumar S. **Loss of caspase-2 augments lymphomagenesis and enhances genomic instability in Atm-deficient mice.** Proc Natl Acad Sci U S A. 2013 Dec 3;110(49):19920-5.

Rai KI, Sawitsky A, Cronkite EP, Chanana A, Levy R, Pasternak B. **Clinical staging of chronic lymphocytic leukaemia**. Blood. 1975; 46:219–234.

Raychaudhuri S, Stuart JM, Altman RB. **Principal components analysis to summarize microarray experiments: application to sporulation time series.** Pac Symp Biocomput. 2000:455-66.

Ribe EM, Jean YY, Goldstein RL, Manzl C, Stefanis L, Villunger A, Troy CM. **Neuronal caspase 2 activity and function requires RAIDD, but not PIDD.** Biochem J. 2012 Jun 15;444(3):591-9.

Ripollés L, Ortega M, Ortuño F, González A, Losada J, Ojanguren J, Soler JA, Bergua J, Coll MD, Caballín MR. **Genetic abnormalities and clinical outcome in chronic lymphocytic leukemia.** Cancer Genet Cytogenet. 2006 Nov;171(1):57-64.

Rodríguez-Vicente AE, Díaz MG, Hernández-Rivas JM. **Chronic lymphocytic leukemia: a clinical and molecular heterogenous disease.** Cancer Genet. 2013 Mar;206(3):49-62.

Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, Yu X, Yang L, Pickeral OK, Rassenti LZ, Powell J, Botstein D, Byrd JC, Grever MR, Cheson BD, Chiorazzi N, Wilson WH, Kipps TJ, Brown PO, Staudt LM. **Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia.** J Exp Med. 2001 Dec 3;194(11):1639-47.

Rossi D, Cerri M, Deambrogi C, Sozzi E, Cresta S, Rasi S, De Paoli L, Spina V, Gattei V, Capello D, Forconi F, Lauria F, Gaidano G. **The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: implications for overall survival and chemorefractoriness.** Clin Cancer Res. 2009 Feb 1;15(3):995-1004.

Samraj AK, Sohn D, Schulze-Osthoff K, Schmitz I. **Loss of caspase-9 reveals its essential role for caspase-2 activation and mitochondrial membrane depolarization.** Mol Biol Cell. 2007 Jan;18(1):84-93.

Schena M, Shalon D, Davis RW, Brown PO. **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** Science. 1995 Oct 20;270(5235):467-70.

Segatto M, Trapani L, Lecis C, Pallottini V. **Regulation of cholesterol biosynthetic pathway in different regions of the rat central nervous system**. Acta Physiol (Oxf). 2012 Sep;206(1):62-71.

Seligmann M, Preud'Homme JL, Brouet JC. **B and T cell markers in human proliferative blood diseases and primary immunodeficiencies, with special reference to membrane bound immunoglobulins.** Transplant Rev. 1973;16:85-113.

Shalini S, Dorstyn L, Wilson C, Puccini J, Ho L, Kumar S. **Impaired antioxidant defence and accumulation of oxidative stress in caspase-2-deficient mice.** Cell Death Differ. 2012 Aug;19(8):1370-80.

Shanafelt TD, Witzig TE, Fink SR, Jenkins RB, Paternoster SF, Smoley SA, Stockero KJ, Nast DM, Flynn HC, Tschumper RC, Geyer S, Zent CS, Call TG, Jelinek DF, Kay NE, Dewald GW. **Prospective evaluation of clonal evolution during long-term follow-up of patients with untreated early-stage chronic lymphocytic leukemia.** J Clin Oncol. 2006 Oct 1;24(28):4634-41.

Sherman BT, Huang da W, Tan Q, Guo Y, Bour S, Liu D, Stephens R, Baseler MW, Lane HC, Lempicki RA. **DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis.** BMC Bioinformatics. 2007 Nov 2;8:426.

Shi M, Vivian CJ, Lee KJ, Ge C, Morotomi-Yano K, Manzl C, Bock F, Sato S, Tomomori-Sato C, Zhu R, Haug JS, Swanson SK, Washburn MP, Chen DJ, Chen BP, Villunger A, Florens L, Du C. **DNA-PKcs-PIDDosome: a nuclear caspase-2-activating complex with role in G2/M checkpoint maintenance.** Cell. 2009 Feb 6;136(3):508-20.

Shirakura H, Hayashi N, Ogino S, Tsuruma K, Uehara T, Nomura Y. **Caspase recruitment domain of procaspase-2 could be a target for SUMO-1 modification through Ubc9.** Biochem Biophys Res Commun. 2005 Jun 17;331(4):1007-15.

Slee EA, Harte MT, Kluck RM, Wolf BB, Casiano CA, Newmeyer DD, Wang HG, Reed JC, Nicholson DW, Alnemri ES, Green DR, Martin SJ. **Ordering the cytochrome c-initiated caspase cascade: hierarchical activation of caspases-2, -3, -6, -7, -8, and -10 in a caspase-9-dependent manner.** J Cell Biol. 1999 Jan 25;144(2):281-92.

Stankovic T, Skowronska A. **The role of ATM mutations and 11q deletions in disease progression in chronic lymphocytic leukemia.** Leuk Lymphoma. 2014 Jun;55(6):1227-39.

Stilgenbauer S, Lichter P, Döhner H. **Genetic features of B-cell chronic lymphocytic leukemia.** Rev Clin Exp Hematol. 2000 Mar;4(1):48-72.

Stilgenbauer S, Bullinger L, Lichter P, Döhner H; German CLL Study Group (GCLLSG). Chronic lymphocytic leukemia. **Genetics of chronic lymphocytic leukemia: genomic aberrations and V(H) gene mutation status in pathogenesis and clinical course.** Leukemia. 2002 Jun;16(6):993-1007.

Stilgenbauer S, Sander S, Bullinger L, Benner A, Leupolt E, Winkler D, Kröber A, Kienle D, Lichter P, Döhner H. **Clonal evolution in chronic lymphocytic leukemia: acquisition of high-risk genomic aberrations associated with unmutated VH, resistance to therapy, and short survival.** Haematologica. 2007 Sep;92(9):1242-5.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

Thornberry NA, Lazebnik Y. **Caspases: enemies within.** Science. 1998 Aug 28;281(5381):1312-6.

Tinel A, Tschopp J. **The PIDDosome, a protein complex implicated in activation of caspase-2 in response to genotoxic stress.** Science. 2004 May 7;304(5672):843-6.

Tinel A, Janssens S, Lippens S, Cuenin S, Logette E, Jaccard B, Quadroni M, Tschopp J. **Autoproteolysis of PIDD marks the bifurcation between pro-death caspase-2 and pro-survival NF-kappaB pathway.** EMBO J. 2007 Jan 10;26(1):197-208.

Trevino V, Falciani F, Barrera-Saldaña HA. **DNA microarrays: a powerful genomic tool for biomedical and clinical research.** Mol Med. 2007 Sep-Oct;13(9-10):527-41.

Troy CM, Rabacchi SA, Friedman WJ, Frappier TF, Brown K, Shelanski ML. **Caspase-2 mediates neuronal cell death induced by beta-amyloid.** J Neurosci. 2000 Feb 15;20(4):1386-92.

Twomey JJ. **Infections complicating multiple myeloma and chronic lymphocytic leukemia.** Arch Intern Med. 1973 Oct;132(4):562-5.

Upton JP, Wang L, Han D, Wang ES, Huskey NE, Lim L, Truitt M, McManus MT, Ruggero D, Goga A, Papa FR, Oakes SA. **IRE1α cleaves select microRNAs during ER stress to derepress translation of proapoptotic Caspase-2.** Science. 2012 Nov 9;338(6108):818-22.

Vakifahmetoglu H, Olsson M, Orrenius S, Zhivotovsky B. **Functional connection between p53 and caspase-2 is essential for apoptosis induced by DNA damage.** Oncogene. 2006 Sep 14;25(41):5683-92.

Vakifahmetoglu H, Olsson M, Tamm C, Heidari N, Orrenius S, Zhivotovsky B. **DNA damage induces two distinct modes of cell death in ovarian carcinomas.** Cell Death Differ. 2008 Mar;15(3):555-66.

Vakifahmetoglu-Norberg H, Zhivotovsky B. **The unpredictable caspase-2: what can it do?** Trends Cell Biol. 2010 Mar;20(3):150-9.

Yamadori T, Baba Y, Matsushita M, Hashimoto S, Kurosaki M, Kurosaki T, Kishimoto T, Tsukada S. **Bruton's tyrosine kinase activity is negatively regulated by Sab, the Btk-SH3 domain-binding protein.** Proc Natl Acad Sci U S A. 1999 May 25;96(11):6341-6.

Wall ME, Dyck PA, Brettin TS. **SVDMAN--singular value decomposition analysis of microarray data.** Bioinformatics. 2001 Jun;17(6):566-8.

Wang L, Miura M, Bergeron L, Zhu H, Yuan J. **Ich-1, an Ice/ced-3-related gene, encodes both positive and negative regulators of programmed cell death.** Cell. 1994 Sep

9;78(5):739-50.

Williams DW, Kondo S, Krzyzanowska A, Hiromi Y, Truman JW. **Local caspase activity directs engulfment of dendrites during pruning.** Nat Neurosci. 2006 Oct;9(10):1234-6.

Winkler D, Schneider C, Kröber A, Pasqualucci L, Lichter P, Döhner H, Stilgenbauer S. **Protein expression analysis of chromosome 12 candidate genes in chronic lymphocytic leukemia (CLL).** Leukemia. 2005 Jul;19(7):1211-5.

Woyach JA, Johnson AJ, Byrd JC. **The B-cell receptor signaling pathway as a therapeutic target in CLL.** Blood. 2012 Aug 9;120(6):1175-84.

Zenz T, Kröber A, Scherer K, Häbe S, Bühler A, Benner A, Denzel T, Winkler D, Edelmann J, Schwänen C, Döhner H, Stilgenbauer S. **Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up.** Blood. 2008 Oct 15;112(8):3322-9.

Zenz T, Fröhling S, Mertens D, Döhner H, Stilgenbauer S. **Moving from prognostic to predictive factors in chronic lymphocytic leukaemia (CLL).** Best Pract Res Clin Haematol. 2010 Mar;23(1):71-84.

Zenz T, Mertens D, Döhner H, Stilgenbauer S. **Importance of genetics in chronic lymphocytic leukemia.** Blood Rev. 2011 May;25(3):131-7.

Zhivotovsky B, Samali A, Gahm A, Orrenius S. **Caspases: their intracellular localization and translocation during apoptosis.** Cell Death Differ. 1999 Jul;6(7):644-51.

Zhivotovsky B, Orrenius S. **Caspase-2 function in response to DNA damage.** Biochem Biophys Res Commun. 2005 Jun 10;331(3):859-67.

OPEN ACCESS Freely available online

PLOS ONE

Transcriptomic Analysis Unveils Correlations between Regulative Apoptotic Caspases and Genes of Cholesterol Homeostasis in Human Brain

Raffaella Picco, Andrea Tomasella, Federico Fogolari, Claudio Brancolini*

Department of Medical and Biological Sciences, Università degli Studi di Udine, Udine, Italy

Abstract

Regulative circuits controlling expression of genes involved in the same biological processes are frequently interconnected. These circuits operate to coordinate the expression of multiple genes and also to compensate dysfunctions in specific elements of the network. Caspases are cysteine-proteases with key roles in the execution phase of apoptosis. Silencing of caspase-2 expression in cultured glioblastoma cells allows the up-regulation of a limited number of genes, among which some are related to cholesterol homeostasis. Lysosomal Acid Lipase A (LIPA) was up-regulated in two different cell lines in response to caspase-2 down-regulation and cells silenced for caspase-2 exhibit reduced cholesterol staining in the lipid droplets. We expanded this observation by large-scale analysis of mRNA expression. All caspases were analyzed in terms of co-expression in comparison with 166 genes involved in cholesterol homeostasis. In the brain, hierarchical clustering has revealed that the expression of regulative apoptotic caspases (CASP2, CASP8 CASP9, CASP10) and of the inflammatory CASP1 is linked to several genes involved in cholesterol homeostasis. These correlations resulted in altered GBM (Glioblastoma Multiforme), in particular for CASP1. We have also demonstrated that these correlations are tissue specific being reduced (CASP9 and CASP10) or different (CASP2) in the liver. For some caspases (CASP1, CASP6 and CASP7) these correlations could be related to brain aging.

Citation: Picco R, Tomasella A, Fogolari F, Brancolini C (2014) Transcriptomic Analysis Unveils Correlations between Regulative Apoptotic Caspases and Genes of Cholesterol Homeostasis in Human Brain. PLoS ONE 9(10): e110610. doi:10.1371/journal.pone.0110610

Editor: Boris Zhivotovsky, Karolinska Institutet, Sweden

Received May 21, 2014; **Accepted** September 23, 2014; **Published** October 16, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. Microarray data have been deposited to the NCBI Gene Expression Omnibus (GEO) and the GEO accession number is GSE61388.

Funding: This work was supported by Associazione Italiana Ricerca sul Cancro (AIRC) (IG-10437) and Programma di ricerca scientifica di rilevante interesse nazionale (PRIN) (Progetto 2010W4J4RM_002) to CB. AT received a fellowship from AIRC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: claudio.brancolini@uniud.it

Introduction

Caspases were initially discovered as critical enzymes in the control of apoptosis. Quite soon it was evident that, they can supervise additional biological processes, such as inflammation and differentiation [1,2]. This discovery has granted the dichotomy between apoptotic and non-apoptotic caspases. More recently, it has been observed that caspases controlling apoptosis can also play specific roles unrelated to cell death [3–5].

Caspases can be divided into initiator and effector caspases depending on the presence of a long prodomain at their amino-terminal region. Initiator caspases act at the apex of a proteolytic cascade, whereas effector caspases act downstream and are involved in the cleavage of specific cellular proteins [1]. Caspase-2, caspase-8, caspase-9, and caspase-10 are the long prodomain caspases involved in the apoptotic process. Caspase-8 and caspase-10 have well-established roles in the engagement of the extrinsic pathway, whereas caspase-9 is the critical enzyme for the intrinsic/mitochondrial pathway. Caspase-8 can also play roles unrelated to apoptosis, such as in NF-kB activation or in limiting necroptosis and caspase-10 has been recently shown to control autophagy [6].

Caspase-2 is still a mysterious caspase. It shows a peculiar nuclear localization that is regulated by two different NLSs [7]. A CARD domain, at the amino-terminal region is responsible for the homotypic interaction with adaptor molecules. Although a molecular platform controlling its activation has been described [8,9], the contribution of caspase-2 to apoptosis is still debated. Different, sometimes controversial results have been published and multiple functions have been attributed to caspase-2.

Mice deficient for caspase-2 and for its adaptor protein RAIDD have proved absent or only very limited defects in apoptosis [8,9]. Further studies with cells from caspase-2 $-/-$ mice have indicated that caspase-2 could be considered a tumor suppressor, since its absence can favor oncogene-mediated transformation [10,11].

Difficulties in defining a gene function in specific biological context could arise from the presence of regulative circuits that compensate the experimental alteration. There are several examples of genes down-regulated by siRNA approaches, or through homologous recombination in mice, which have generated only minimal phenotypes [12–14].

The still obscure impact of caspase-2 on cell functions could be masked by compensatory mechanisms engaged following its inactivation. In this manuscript we have investigated the gene

expression profile of human cells silenced for caspase-2 expression. Our goal was to unveil whether perturbation of caspase-2 levels could influence the expression of genes involved in specific cellular functions, either as part of common regulative circuits or of compensatory mechanisms.

Materials and Methods

Cell culture, siRNA, reagents and antibodies

U87MG and IMR90-E1A cells were grown in DMEM supplemented with 10% FBS, penicillin (100 U/mL), glutamine (2 mmol/L), and streptomycin (100 μg/mL) at 37°C in 5% CO_2 atmosphere. RNA oligos for interference (RNAi) were purchased from Dharmacon: CASP2 RNAi1, AAACAGCUGUUGUUGAGCGAA; Control1 (CASP2 mutated) RNAi, AAACAAUUGUUGUUGAGCGAA; or Qiagen: CASP2 RNAi2 CAUCUUCUGGAGAAGGACATT and a non-targeting siRNA UUCUCCGAACGUGUCACGU, Control2. Cells were transfected 24 hours after plating by adding the Opti-MEM medium containing Lipofectamine 2000 (Invitrogen) plus RNAi oligos. BSA and Filipin (Sigma), anti-LBPA [15], anti-transferrin receptor (Tnf-R) (OKT9), anti-GM-130 (BD biosciences), anti-caspase-2 [16]. Secondary anti-mouse and anti-rabbit antibodies were Alexa Fluor 488 and Alexa Fluor 546 conjugated (Invitrogen).

Western blotting

Proteins obtained after an SDS denaturating lysis and sonication were transferred to a 0.2-μm-pore-sized nitrocellulose membrane and incubated with the specific primary antibodies. After several washes, blots were incubated with peroxidase-conjugated goat anti-rabbit or (Euroclone Milano I) for 1 h at room temperature. Finally, blots were developed with Super Signal West Dura, as recommended by the vendor (Pierce).

RNA expression array and data analysis

Total RNA was isolated using RNeasy Mini kit (Qiagen). RNA sample was labeled according to the standard one cycle amplification and labeling protocol (Affymetrix). Labeled cRNA was hybridized on Affymetrix GeneChip Human Gene 1.0 ST Array. Robust Multi-Array Average (RMA) normalization was applied [17]. Data analysis was performed using the t-test as implemented in the R statistical package. A minimum standard deviation was assumed corresponding to the median percentile of all standard deviations in order to avoid fortuitously large t statistics. Differentially expressed genes were selected based on$>$ 1.5 fold change and $P<0.05$. The analysis of Gene Ontology terms was performed using the DAVID server [18,19]. Microarray data have been deposited in NCBI Gene Expression Omnibus (GEO) and the GEO accession number is GSE61388.

RNA extraction and quantitative qRT-PCR (quantitative reverse transcription polymerase chain reaction)

Cells were lysed using Tri-Reagent (Molecular Research Center). A total of 1μg of total RNA was retrotranscribed by using 100 U of Moloney murine leukemia virus reverse transcriptase (Invitrogen). Quantitative reverse transcription-PCR (qRT-PCR) analyses were performed using Bio-Rad CFX96 and SYBR Green technology. The data were analyzed by use of a comparative threshold cycle using HPRT (hypoxanthine phosphoribosyltransferase) and β-actin as normalizer genes. All reactions were done in triplicate.

Immunofluorescence microscopy

Cells were fixed with 3% paraformaldehyde and permeabilized with 0.5% Triton X-100 and blocked in PBS 3% BSA for 1 h RT. After washes coverslips were incubated with Filipin (100μg/mL) and relative primary antibodies for 2 hrs. After several washes coverslips were incubated with secondary antibodies. Cells were imaged with a Leica confocal scanner SP equipped with a 488 λ Ar laser and a 543 to 633 λ HeNe laser. Cell images for deconvolution were taken using the Leica AF6000 LX microscope.

Data preparation and analysis

29 human microarray datasets were included in this study, totaling 726 arrays. Brain (11 datasets 293 microarrays) GBM (13 datasets 327 microarray) Liver (5 datasets 106 microarrays) were used and in all cases. All datasets were downloaded manually from GEO [20] and ArrayExpress databases [21]. We analyzed only expression data obtained using the most comprehensive human expression platform HG U133 Plus 2.0. For GBM (GSE11100; GSE13041; GSE15824; GSE19728; GSE23806; GSE23935; GSE29796; GSE30563; GSE32374; GSE4290; GSE7696; GSE9171) for normal human brain (GSE5281; GSE7307; GSE17612; GSE21935; GSE15824; E-MEXP-2351; GSE21354; E-MEXP-2280; GSE15209; GSE7692; GSE4290) for human liver (E-GEOD-40873; E-MTAB-950; E-GEOD-23343; E-AFMX-11; E-MEXP-2128). We processed all the CEL files together by using standard tools available within the affy package in R [22].

We use a UniGene ID centered CDF (Chip Description file) in order to have only one intensity value per gene. CDFs were downloaded from the Molecular and Behavioral Neuroscience Institute Microarray Lab (URL:http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/genomic_curated_CDF.asp) [23]. All annotation information were downloaded from the same website. The normalization step was done with the standard MAS5.0 algorithm, described in the Statistical Algorithms Description Document available from Affymetrix (URL: http://www.affymetrix.com/support/technical/whitepapers/sadd_whitepaper.pdf).

We converted all microarray data to log values. We extracted the data regarding the genes of interest in an automatic way. Correlations among gene expression levels were calculated using the library psych in R choosing the Pearson correlation method. p-values were adjusted for multiple testing using Benjamini and Hochberg's method [24]. Genes were clustered using hierarchical clustering using the complete linkage method according to similarity in correlation patterns, as measured by euclidean distance [25]. Heat maps were generated with R with positive correlation scores (values) colored by blue while negative ones colored by dark green. GO (Gene Ontology) annotations and knowledge from the literature was used to create a list of genes involved with the cholesterol metabolism. Gene expression levels were correlated with age using the Spearman's correlation making minimal assumptions about the relationship between the two diverse variables.

Results and Discussion

Gene expression profile studies in cells with down-regulated caspase-2 expression

To identify genes and pathways under the influence of caspase-2 we silenced its expression in the glioblastoma cell line U87MG. We selected glioblastoma cells since important apoptotic functions have been attributed to caspase-2 in the brain and because CASP2 deficits elicit compensatory mechanisms in this tissue [26,27].

Caspase-2 deficient cells did not display overt alterations in terms of cell proliferation, cell cycle and apoptosis (data not

A Genes up-regulated >1.5 fold in U87MG cells silenced for caspase-2

Gene symbol	Fold difference	Gene name	Function	Localization	ProbeSet ID	P value
CDK11B	3.6720	Cyclin-dependent kinase 11B	cell proliferation apoptosis	Nucleus cytoplasm	8180291	0.0288
GTF2IRD2B	2.2199	GTF2I repeat domain containing 2B	Transcription	Nucleus	8133549	0.0300
PCDHA5	2.1444	Protocadherin alpha 5	Cell adhesion	Plasma membrane	8180279	0.0249
MOG	2.0705	Myelin oligodendrocyte glycoprotein	Cell adhesion	Plasma membrane	8177709	0.0355
TRAPPC6B	1.9015	Trafficking protein particle complex 6B	Vesicle-mediated transport	ER/Golgi	7978739	0.0189
IGHG1	1.7586	Immunoglobulin heavy constant gamma 1	Complement activation	Extracellular	7981720	0.0071
CYP51A1	1.7426	Cytochrome P450	Cholesterol biosynthesis	ER	8140864	0.0370
EFCAB4B	1.6646	EF-hand calcium binding domain 4B	Ca++ signalling	Pancellular	7960365	0.0309
PIK3R3	1.6365	Phosphoinositide-3-kinase, regulatory subunit 3 (gamma)	Signalling	Plasma membrane	7915787	0.0256
LIPA	1.6097	Lipase A	Cholesterol release	Lysosome	7934920	0.0304
KCTD16	1.5742	Potassium channel tetramerisation domain containing 16	Signalling	Plasma membrane	8108905	0.0280
C10orf10	1.5707	chromosome 10 open reading frame 10	Unknown	Membrane	7933204	0.0149
XCL1	1.5613	chemokine (C motif) ligand 1	Immune response	Extracellular	7907156	0.0263
OR1F2P	1.5450	olfactory receptor, family 1, subfamily F, member 2	Signalling	Plasma membrane	7992865	0.0439
STK19	1.5399	serine/threonine kinase 19	Unknown	Nucleus	8178164	0.0416
CYP1B1	1.5108	Cytochrome P450	xenobiotic, steroid hormon metbolism	ER	8051583	0.0365
LCN1	1.5036	Lipocalin 1	Lipids binding	Extracellular	8158995	0.0348
*CASP2	-4.1734	Caspase-2	apoptosis	Nucleus	8136869	0.0095

* Downregulation of caspase-2 as resulted from the microarray experiment

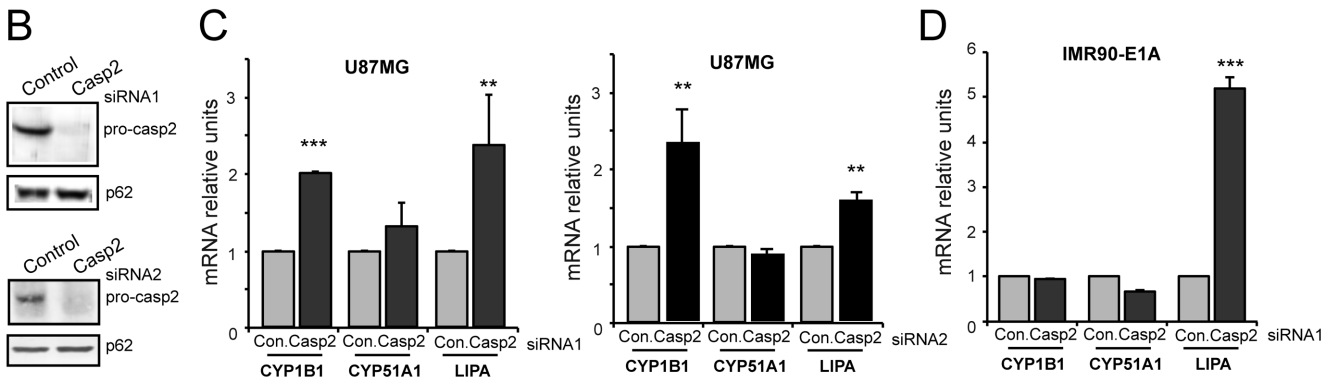


Figure 1. Transcriptomic variations in cells silenced for caspase-2. A. List of the top up-regulated genes (cut-off 1.5) in U87MG cells silenced for caspase-2 expression. B. Cellular lysates from U87MG cells transfected with the indicated siRNAs against caspase-2 or the relative control siRNAs were generated and after immunoblot were probed with an anti-caspase-2 antibody. P62, nucleoporin was used as loading control. C. mRNA expression levels of CYP1B1, CYP51A1 and LIPA were measured using qRT-PCR in U87MG cells transfected with the indicated siRNAs against caspase-2 or the relative control siRNAs. Data are presented as mean ± SD; $n = 3$. D. mRNA expression levels of CYP1B1, CYP51A1 and LIPA were measured using qRT-PCR in IMR90-E1A cells transfected with siRNA1 against caspase-2 or the control siRNA. Data are presented as mean ± SD; $n = 3$.
doi:10.1371/journal.pone.0110610.g001

shown). Next transcriptional expression profiles of cells transfected with caspase-2 siRNA and control siRNA were compared. We selected a 1.5 fold cut-off and globally 24 genes were significantly down-regulated (Table S1), whereas 17 genes were significantly up-regulated in caspase-2 silenced cells (Fig. 1A). This number is particularly small, since in parallel experiments after silencing of other genes, such as USP34 or PGAM5 fluctuations in the expression of more than 200 and 800 genes, respectively were observed (data not shown). Microarray and immunoblot analysis proved the effective down-regulation of CASP2 mRNA (Fig. 1A) and protein (Fig. 1B).

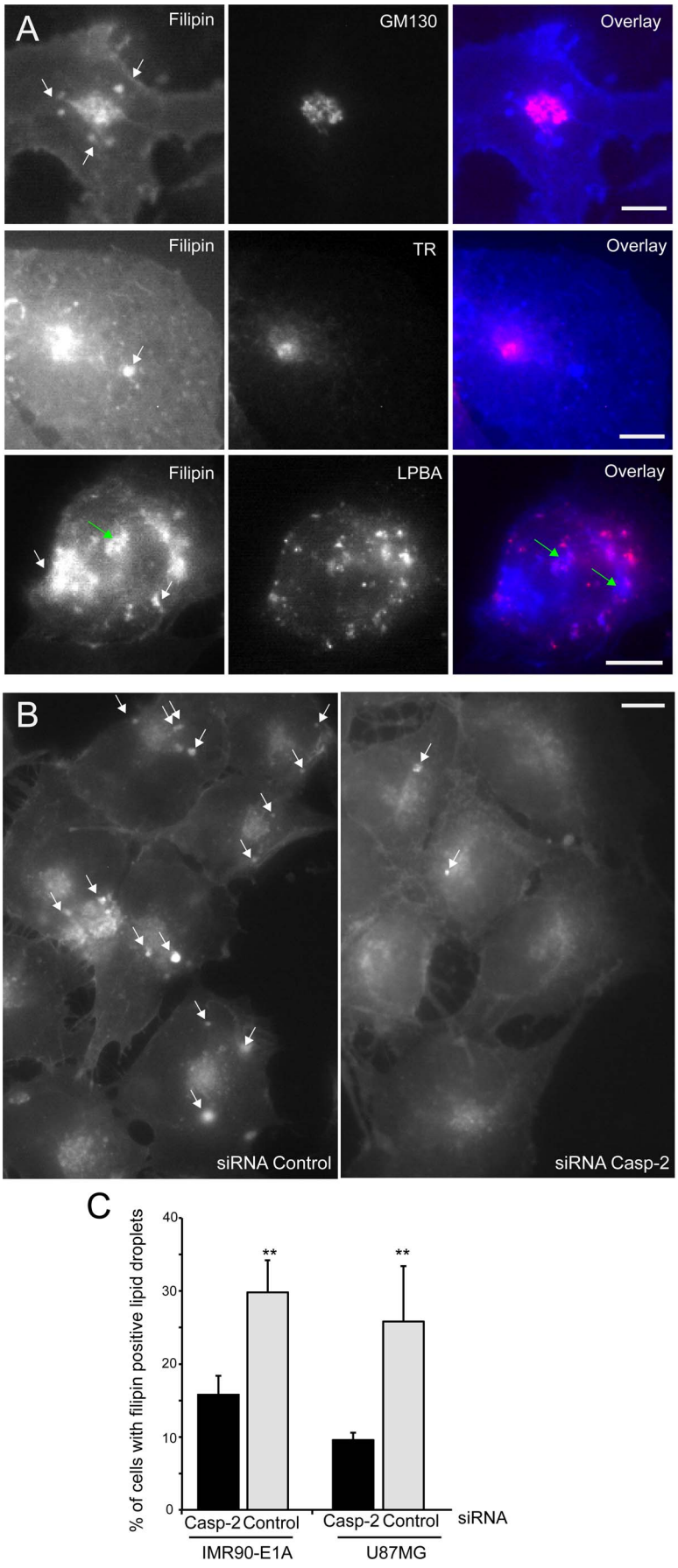

A
Filipin
GM130
Overlay
Filipin
TR
Overlay
Filipin
LPBA
Overlay
B
siRNA Control
siRNA Casp-2
C
% of cells with filipin positive lipid droplets
40
30
20
10
0
**
**
Casp-2
Control
Casp-2
Control
siRNA
IMR90-E1A
U87MG

Figure 2. Cholesterol distribution in cells silenced for caspase-2 expression. A. U87MG cells were fixed and processed for immunofluorescence. Epifluorescence microscopy followed by deconvolution analysis was used to visualize different subcellular compartments and cholesterol distribution. With arrows indicate LDs and green arrows LDs encircled by lysosomes. Scale bar 40 μM. B. U87MG cells were transfected with siRNA against caspase-2 or control siRNA as indicated. 48 h later cells fixed and stained with filipin to visualize the intracellular distribution of cholesterol. Arrows point to LDs. Scale bar 30 μM. C. Quantitative analysis of LDs presence in U87MG and IMR90-E1A cells transfected with siRNA against caspase-2 or control siRNA. Data are presented as mean ± SD; $n = 3$.
doi:10.1371/journal.pone.0110610.g002

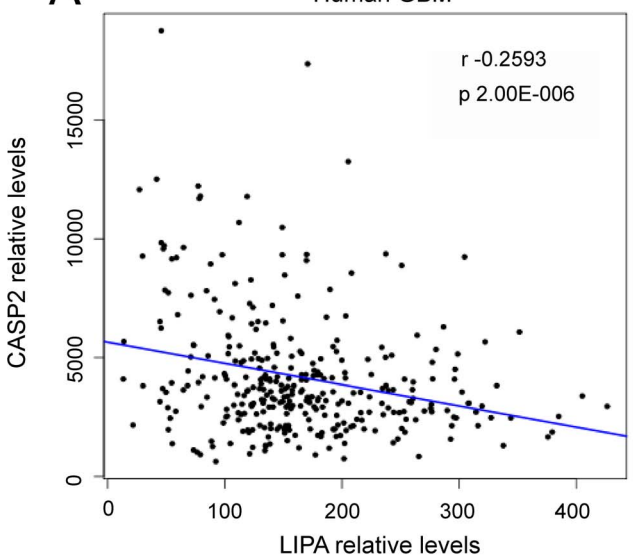
Next we focused our attention on the 17 genes up-regulated after caspase-2 down-regulation, which could be entangled in compensatory responses. Analysis of associated GO terms indicated that these genes are involved in different biological functions, including cell cycle control, inflammation and membrane trafficking. Particularly, 3 of them: CYP51A1, CYP1B1 and LIPA are linked to cholesterol metabolism. Since caspase-2 expression can be influenced by SREBPs and a previous study proposed a role of caspase-2 in the control of cholesterol and triacylgycerol levels [28,29], we investigated in more detail the relationships between caspase-2 and cholesterol genes.

CYP51A1 and CYP1B1 are cytochrome P450 family members involved respectively in cholesterol/sterol biosynthetic processes and in the metabolism of a wide range of structurally diverse substrates, including cholesterol [30,31]. LIPA encodes for the key enzyme responsible for acidic hydrolysis of cholesteryl esters and triglycerides delivered from lipoproteins to lysosomes [32].

Furthermore also PIK3R3 (phosphatidylinositol-3 kinase regulatory subunit p55γ), LCN1 (tears lipocalin), other two caspase-2 influenced genes are in some relations with cholesterol and lipid metabolism. The first is a target of SREBPs [33] and the second can bind an assortment of lipids including cholesterol [34]. For all these reasons we decided to study the relationships between caspase-2 and CYP51A1, CYP1B1 and LIPA.

qRT-PCR analysis was performed to validate the microarray experiments. The expression of CYP1B1 and LIPA was augmented in U87MG cells silenced for caspase-2, whereas for CYP51A1 the increase was minimal and not statistically significant (Fig. 1C). To confirm these results we used a second siRNA against caspase-2 (siRNA2) and a second control oligos, from a different provider (Fig. 1B). The results were similar. CYP1B1 and LIPA expression was augment in U87MG cells silenced for caspase-2. We also investigated whether this up-regulation could be observed in other cell lines. Figure 1D shows that only LIPA induction can be observed in human fibroblasts expressing the E1A oncogene after caspase-2 silencing. In this cell line CYP51A1 and CYP1B1 mRNA levels resulted unchanged. However LIPA up-regulation was much more pronounced compared to U87MG cells. These findings indicate that correlations between caspase-2 and cholesterol genes expression could vary in different cell types.

A



B

Correlations of LIPA mRNA levels in GBM

Gene symbol	*correlation r*	*P*
CASP2	-0.259346792	2.00E-006
CASP1	0.313388802	2.01E-008
CASP3	0.096062625	0.056364589
CASP4	0.263308072	0.000396117
CASP5	0.026602733	0.114237324
CASP6	-0.031962855	0.984267902
CASP7	0.078981007	0.463581545
CASP8	-0.00682104	0.827404576
CASP9	-0.132493862	0.022734265
CASP10	0.044043382	0.427324628
CRADD	-0.067381793	0.232317656
PIDD	-0.349707067	3.38E-014

C

CASP2 - LIPA mRNA levels correlations

Tissue	*correlation r*	*P*
Cerebellum	-0.5669	0.00025384
Hippocampus	0.4556	0.03790958
Brain	0.29696	1.263E-006

PIDD - LIPA mRNA levels correlations

Tissue	*correlation r*	*P*
Cerebellum	-0.5602	0.0019
Hippocampus	0.06936	0.76514
Brain	0.11113	0.051748

Figure 3. Analysis of LIPA and caspases expression in GBM and different areas of the CNS. A. Plot of CASP2 versus LIPA expression levels in GBM. Linear regression is reported. B. Correlations in expression levels between LIPA and the indicated genes in GBM. C. Correlations in expression levels between LIPA and the indicated genes in different CNS areas.
doi:10.1371/journal.pone.0110610.g003

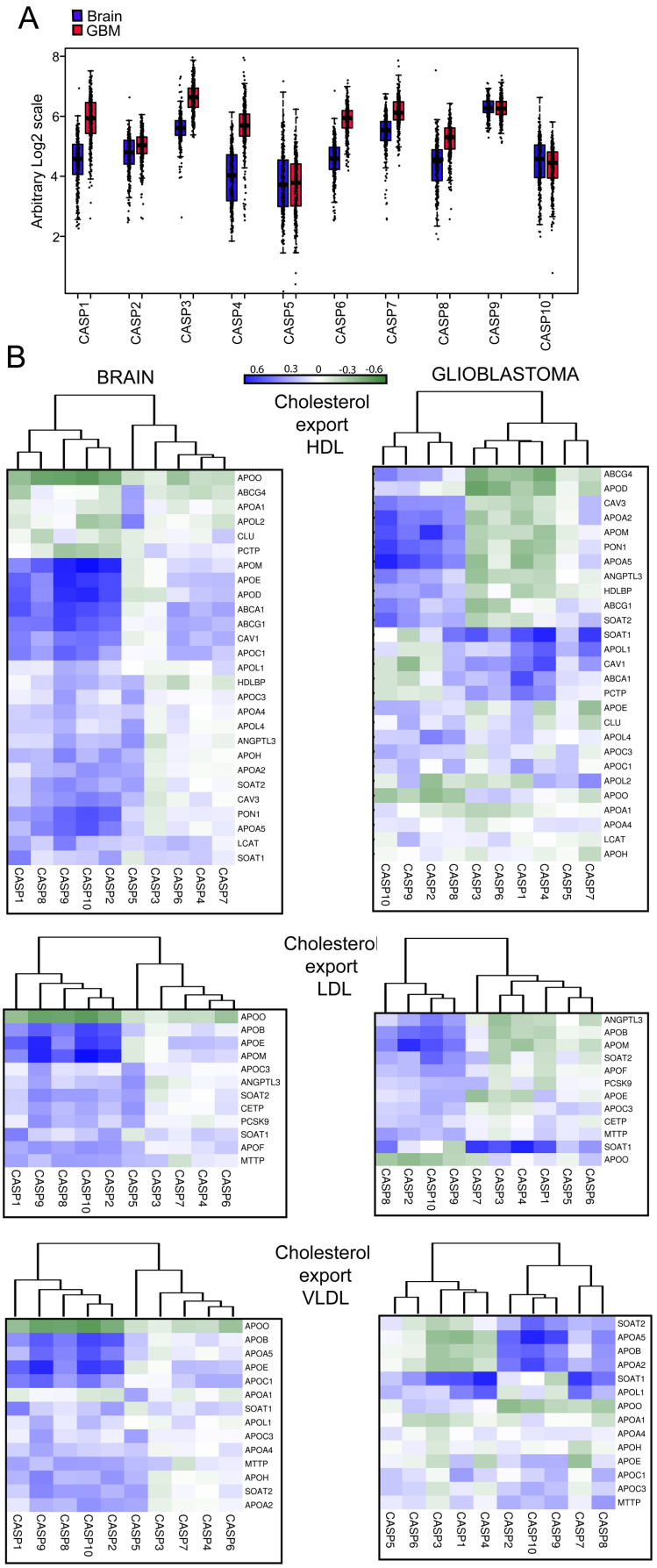

A
Brain
GBM
Arbitrary Log2 scale
8
6
4
2
CASP1
CASP2
CASP3
CASP4
CASP5
CASP6
CASP7
CASP8
CASP9
CASP10
B
BRAIN
0.6
0.3
0
-0.3
-0.6
GLIOBLASTOMA
Cholesterol export HDL
APOO
ABCG4
APOA1
APOL2
CLU
PCTP
APOM
APOE
APOD
ABCA1
ABCG1
CAV1
APOC1
APOL1
HDLBP
APOC3
APOA4
APOL4
ANGPTL3
APOH
APOA2
SOAT2
CAV3
PON1
APOA5
LCAT
SOAT1
CASP1 CASP8 CASP9 CASP10 CASP2 CASP5 CASP3 CASP6 CASP4 CASP7
ABCG4
APOD
CAV3
APOA2
APOM
PON1
APOA5
ANGPTL3
HDLBP
ABCG1
SOAT2
SOAT1
APOL1
CAV1
ABCA1
PCTP
APOE
CLU
APOL4
APOC3
APOC1
APOL2
APOO
APOA1
APOA4
LCAT
APOH
CASP10 CASP9 CASP2 CASP8 CASP3 CASP6 CASP1 CASP4 CASP5 CASP7
Cholesterol export LDL
APOO
APOB
APOE
APOM
APOC3
ANGPTL3
SOAT2
CETP
PCSK9
SOAT1
APOF
MTTP
CASP1 CASP9 CASP8 CASP10 CASP2 CASP5 CASP3 CASP7 CASP4 CASP6
ANGPTL3
APOB
APOM
SOAT2
APOF
PCSK9
APOE
APOC3
CETP
MTTP
SOAT1
APOO
CASP8 CASP2 CASP10 CASP9 CASP7 CASP3 CASP4 CASP1 CASP5 CASP6
Cholesterol export VLDL
APOO
APOB
APOA5
APOE
APOC1
APOA1
SOAT1
APOL1
APOC3
APOA4
MTTP
APOH
SOAT2
APOA2
CASP1 CASP9 CASP8 CASP10 CASP2 CASP5 CASP3 CASP7 CASP4 CASP6
SOAT2
APOA5
APOB
APOA2
SOAT1
APOL1
APOO
APOA1
APOA4
APOH
APOE
APOC1
APOC3
MTTP
CASP5 CASP6 CASP3 CASP1 CASP4 CASP2 CASP10 CASP9 CASP7 CASP8

Figure 4. Co-expression analysis of caspases and cholesterol genes in human brain and GBM. A. Expression levels of the different caspases in GBM and in the normal brains. Box plots depicted in red mark tumors whereas blue was used for normal brain. mRNA levels of CASP1, CASP3, CASP4, CASP6, CASP7 and CASP8 were significantly augmented in tumor samples. p-value $<2.2^{e-16}$. B. Correlations of expression levels between caspases and cholesterol genes in different brain and GBM samples in *Cholesterol export, lipoproteins* category. Data obtained were used to calculate the correlation values with the Pearson method. In the heat map positive values are displayed in blue and negative in dark green. The dendrograms displayed on the top are based on hierarchical clustering using the complete linkage method.
doi:10.1371/journal.pone.0110610.g004

LIPA generates unesterified cholesterol, which can be used as substrate for steroidogenesis or be re-esterified for storage in lipid droplets by acyl-CoA:cholesterol acyl transferase [32]. The increase of LIPA levels in cells with down-regulated caspase-2 might compensate a deficit in Srebp2-driven lipid synthesis/accumulation in human cells, as previously suggested [29].

To understand whether mRNA levels of LIPA, CYP51A1 and CYP1B1 are influenced by other caspases, we silenced caspase-8 expression in U87MG cells. We selected caspase-8 because, like caspase-2, it is a regulative apoptotic caspase, and because it is expressed in U87MG cells (Figure S1A). The designed siRNA silenced caspase-8 expression (Figure S1B). When Caspase-8 expression was down-regulated only CYP1B1 mRNA levels were clearly augmented. LIPA and CYP51A1 expression was not significantly changed.

Analysis of intracellular cholesterol distribution in caspase-2 silenced cells

Having evidences of the existence of a connection linking caspase-2 to some genes of the cholesterol pathway, we compared the subcellular distribution of cholesterol after labeling of IMR90-E1A and U87MG cells with filipin. Fluorescence staining was detectable in the plasma membrane (PM) and in intracellular membrane structures (Fig 2A). As expected co-localization studies using GM130, transferrin receptor and LBPA, as markers respectively of, Golgi apparatus, late endosomal compartment and endosomal/recycling compartment evidenced that cholesterol can be detected in all these organelles. In addition, intense filipin fluorescence staining was present in regular spherical structures that do not co-localize with the used markers and that can be identified as LDs (lipid droplets).

Next we compared filipin staining between caspase-2 silenced and control cells. Remarkably, the number of LDs detectable in a cell, as well as the percentage of cells holding them was reduced after caspase-2 silencing (Fig. 2B). Quantitative analysis proved that in both cell lines, the percentage of cells presenting LDs was reduced after caspase-2 silencing (Fig. 2C). This observation further indicates that down-regulation of caspase-2 could influence in cholesterol homeostasis.

Correlation studies of LIPA and caspase-2 expression in glioblastoma

Taking into account the complex regulative networks influencing cholesterol homeostasis in vivo, the use of the cell culture models to investigate correlations between caspase-2 and cholesterol genes expression is limiting. Hence, to further expand our study we decided to interrogate public available gene expression datasets of glioblastomas. In principle, if caspase-2 controls a circuit that influence LIPA expression in cultured glioblastoma cells the expression of these two genes should be inversely correlated in tumors. Gene expression profiles from 13 datasets including 327 microarrays of GBM were interrogated. Figure 3A illustrates that in glioblastoma, mRNA levels for CASP2 and LIPA evidence a weak, but significant inverse Pearson correlation (r -0.2593; p-value 2.00E-00.6).

Since also caspase-8 down-regulation can influence the expression of certain cholesterol genes (Figure S1), we extended this study to all caspases. as well as, to elements of the molecular platform involved in caspase-2 activation (CRADD/RAIDD and PIDD) [8,9]. With the exclusion of CASP1 mRNA, which level weakly but positively correlates with LIPA, all the other caspases do not evidence significant correlations with LIPA expression (Fig. 3B). Concerning PIDD, its expression negatively correlates with LIPA mRNA levels (r -0.3497; p-value 3.38E-014) in GBM.

Expression correlations between CASP2/LIPA and PIDD/LIPA were also evaluated in normal tissues. A good and significant inverse correlation was scored for both genes in cerebellum. On the contrary, in the hippocampus and with lowest score in the brain, a significant positive correlation only for CASP2/LIPA was observed (Fig. 3C). Interestingly, differences in terms of expression of cholesterol genes between different brain areas are known [35].

Expression correlations among caspases and cholesterol genes in normal brain and in glioblastoma

Although in cultured cell lines we noted an inverse correlation in terms of expression between CASP2 and LIPA, in vivo the situation is heterogeneous, as suggested by the analysis of different CNS (Central Nervous System) areas. Moreover, since silencing of caspase-2 affects also other cholesterol genes, it is evident that these correlations cannot be simply analyzed without taking into account other genes involved in cholesterol homeostasis. For example down-regulation of LIPA provoked the up-regulation of several genes implicated in cholesterol biosynthesis, as compensatory mechanism [36]. Therefore, we decided to investigate with a more comprehensive approach the correlations between expression of caspase-2 and the expression of genes involved in cholesterol homeostasis. We extended this study to all caspases and we also included the normal tissues. Gene expression profile data from 11 datasets including 293 microarrays of normal brain were recovered from public available databases.

We first evaluated changes in the mRNA levels for the different caspases between normal brain and GBM. Surprisingly the mRNA levels of several caspases, both inflammatory (CASP1 and CASP4) and apoptotic (CASP3, CASP6, CASP7 and CASP8) were significantly augmented in tumor samples (Fig. 4A).

Next, genes involved in cholesterol homeostasis, including steroidogenesis were extracted from Gene Ontology (GO) and integrated from literature data. In total we selected 256 genes, of which 166 were grouped into 5 categories: *biosynthesis*, *adsorption/import*, *export*, *steroid and bile acid synthesis* and *transcriptional regulators*. The category *export* was subdivided into HDL (High Density Lipoproteins), LDL (Low Density Lipoproteins) and VLDL (Very Low Density Lipoproteins). At this point we exploited microarray datasets to evaluate correlations in the expression levels between caspases and cholesterol genes. Pearson correlation between the expression level of each caspase and each cholesterol gene was computed for human brain samples. Based on the similarity in correlation patterns caspases and cholesterol genes were hierarchically clustered using the complete linkage method.

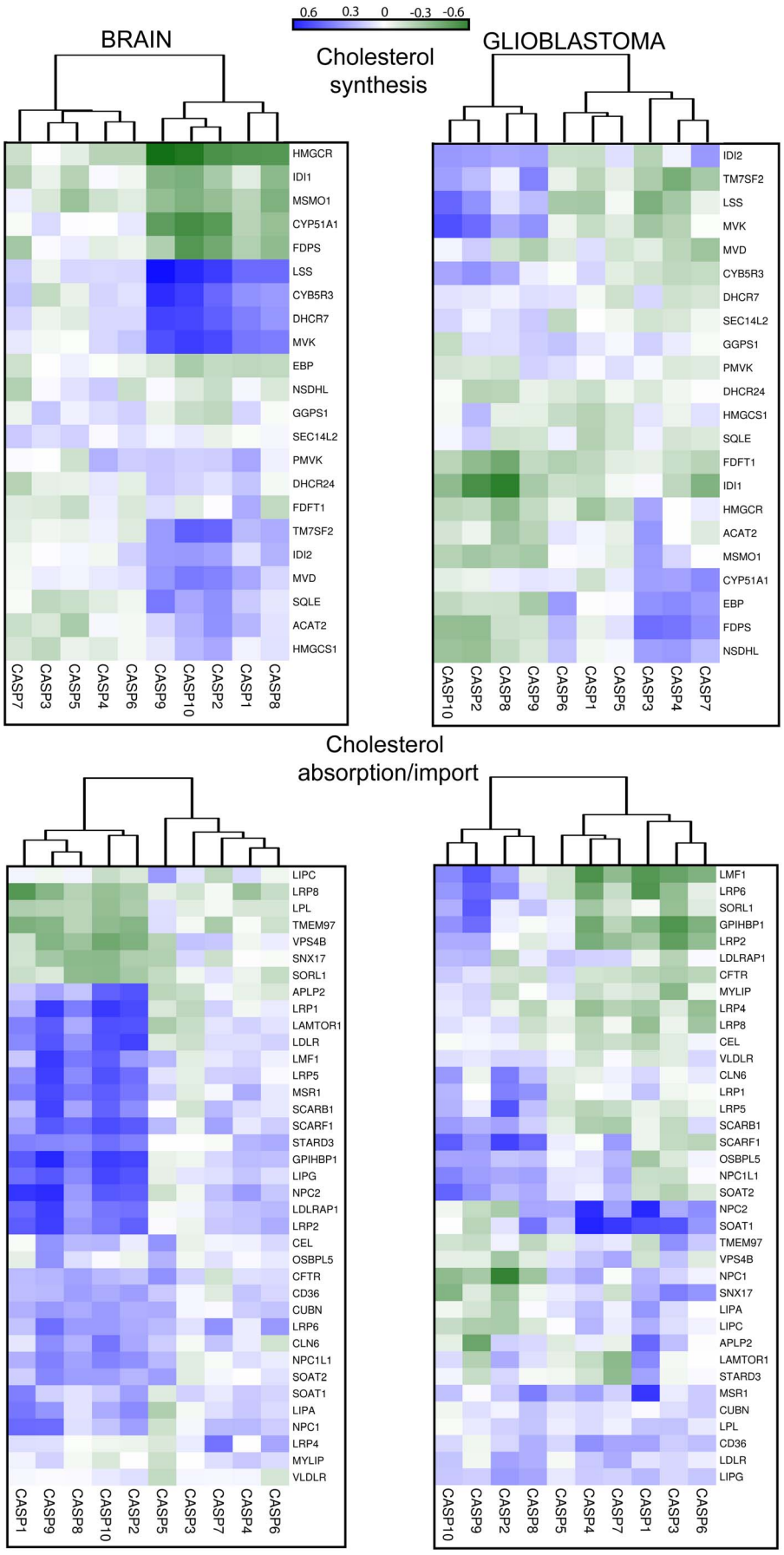

0.6 0.3 0 -0.3 -0.6
BRAIN
Cholesterol synthesis
GLIOBLASTOMA
HMGCR
IDI1
MSMO1
CYP51A1
FDPS
LSS
CYB5R3
DHCR7
MVK
EBP
NSDHL
GGPS1
SEC14L2
PMVK
DHCR24
FDFT1
TM7SF2
IDI2
MVD
SQLE
ACAT2
HMGCS1
CASP7 CASP3 CASP5 CASP4 CASP6 CASP9 CASP10 CASP2 CASP1 CASP8
IDI2
TM7SF2
LSS
MVK
MVD
CYB5R3
DHCR7
SEC14L2
GGPS1
PMVK
DHCR24
HMGCS1
SQLE
FDFT1
IDI1
HMGCR
ACAT2
MSMO1
CYP51A1
EBP
FDPS
NSDHL
CASP10 CASP2 CASP8 CASP9 CASP6 CASP1 CASP5 CASP3 CASP4 CASP7
Cholesterol absorption/import
LIPC
LRP8
LPL
TMEM97
VPS4B
SNX17
SORL1
APLP2
LRP1
LAMTOR1
LDLR
LMF1
LRP5
MSR1
SCARB1
SCARF1
STARD3
GPIHBP1
LIPG
NPC2
LDLRAP1
LRP2
CEL
OSBPL5
CFTR
CD36
CUBN
LRP6
CLN6
NPC1L1
SOAT2
SOAT1
LIPA
NPC1
LRP4
MYLIP
VLDLR
CASP1 CASP9 CASP8 CASP10 CASP2 CASP5 CASP3 CASP7 CASP4 CASP6
LMF1
LRP6
SORL1
GPIHBP1
LRP2
LDLRAP1
CFTR
MYLIP
LRP4
LRP8
CEL
VLDLR
CLN6
LRP1
LRP5
SCARB1
SCARF1
OSBPL5
NPC1L1
SOAT2
NPC2
SOAT1
TMEM97
VPS4B
NPC1
SNX17
LIPA
LIPC
APLP2
LAMTOR1
STARD3
MSR1
CUBN
LPL
CD36
LDLR
LIPG
CASP10 CASP9 CASP2 CASP8 CASP5 CASP4 CASP7 CASP1 CASP3 CASP6

Figure 5. Co-expression analysis of caspases and cholesterol genes in human brain and GBM. Correlations in terms of expression levels between caspases and cholesterol genes in different brain and GBM samples, for the indicated categories. Data obtained were used to calculate the correlation values with the Pearson method. In the heat map positive values are displayed in blue and negative in dark green. The dendrograms displayed on the top are based on hierarchical clustering using the complete linkage method.
doi:10.1371/journal.pone.0110610.g005

Cholesterol export, lipoproteins. The CNS produces its own lipoprotein transport system that is distinct from the plasma [37]. CASP1, 2, 8, 9, and 10 share correlations in terms of changes in the mRNA levels with several genes involved in HDL biogenesis. In particular expression of CASP2, 9 and 10 show the highest correlations with APOD, APOM, APOE, APOA5, PON1, CAV1, CAV3, ABCA1 and ABCG1 (Fig. 4B and Table S2).

APOD and APOM are apolipoprotein belonging to the lipocalin protein superfamily [38]. APOM expression is under the influence of hormones and cytokines [39]. In *apoM*-deficient mice plasma HDL are reduced by approximately 17–21% [40]. Moreover, APOM is an important carrier of shingosine-1-phosphate (S1P), a signaling molecule controlling several processes including inflammation [41]. APOD not only contributes to HDL formation but can also act as antioxidant [42]. This function could explain the up-regulation of APOD expression with aging in human prefrontal cortex, as part of a protective circuit [38,43].

APOE is the major apolipoprotein for CNS HDL and is mainly synthetized by astrocytes [37,44]. It is possible that ApoE-containing lipoproteins are involved in delivering cholesterol to neurons for growth, repair and synaptogenesis [37,45]. Fluctuations in APOE mRNA could be related to repair. In fact, in astrocytes ApoE synthesis increases dramatically after nerve injury [46,47]. APOE and in particular the ε4 allele is the major known genetic risk factor for late-onset Alzheimer's disease [48].

PON1 belongs to paraoxonase genes family and is secreted into the extracellular environment where it binds HDL. This antioxidant enzyme confers to HDL some of the anti-atherogenic properties, such as HDL-mediated cholesterol efflux from macrophages, and the inhibition of LDL oxidation [49].

Although in the brain their activities are less characterized, ATP-binding cassette (ABC) transporters ABCA1 and ABCG1, but also CAV1 and CAV3 mediate cholesterol efflux and play important roles in the transfer of phospholipids and cholesterol to apolipoproteins such as ApoE and ApoM [37,50].

In GBM expression correlations among constituents of HDL and pro-apoptotic regulative CASP2, 8, 9, and 10 are still present but reduced. In particular CASP2, 9 and 10 do not show significant correlations with APOD, ABAC1 and CAV1 and also in the case of ABCG1 the correlation is diminished. Instead, another transporter, ABCG4 exhibits correlations.

In GBM, CASP1 reveals profound different correlations with HDL genes and clusters together with CASP4, another inflammatory caspase, and the effector caspases. HDL genes showing strongest correlations with CASP1 are APOL1, SOAT1, CAV1, ABCA1 and PCTP. Interestingly, cholesterol efflux associates more strongly with the expression of ABCG1 than of ABCA1 [51].

Cholesterol biosynthesis. Almost all the cholesterol found in the CNS is produced from local biosynthesis [52]. Also in the case of the category "cholesterol biosynthesis", correlation analysis revealed that the regulative apoptotic caspases, with the addition of CASP1 cluster together. In particular strong positive correlations (Fig. 5 and Table S3) with mRNA levels of a group of genes involved in cholesterol biosynthesis (LSS, CYB5R3, DHCR7, MVK) emerged. By contrast, HMGCR the gene encoding for the enzyme converting the 3-hydroxyl-3-methyl-glutarylCoA (HMG-CoA) into mevalonate scores a good but negative correlation. Since it represents the limiting step in cholesterol biosynthesis, the implications of the correlations between certain enzymes of cholesterol biosynthesis and regulative apoptotic caspases are unclear at the moment. Significant correlations were not observed for all the other caspases. Expression of CYP51A1 in the brain achieves a strong inverse correlation compared to CASP2, CASP9 and CASP10. CYP51A1 encodes for lanosterol 14α-demethylase, which in addition to being a key enzyme of the cholesterol biosynthetic pathway [31] is also involved in the steroidogenesis [53].

In GBM correlations among regulative apoptotic caspases and genes involved in cholesterol biosynthesis are less evident. Again CASP1 clustered with effector caspases in the absence of significant correlation scores. Interestingly, CASP3 shows good correlations with some genes involved in cholesterol biosynthesis including HMGCR.

Cholesterol absorption/import. Lipoproteins produced by astrocytes can be internalized, after binding to LDL receptor superfamily by neurons and glia cells [37]. Also in the category of absorption/import, analysis of the correlations at mRNA levels resulted in two clusters where the regulative apoptotic caspases plus CASP1 segregate from all the others. CASP2 CASP9 and CASP10 show the strongest correlations with APLP2, LRP1, LDLR, LRP2, LRP5, MSR1, LIPG, SCARF1, GPIHBP1, NPC2, LAMTOR1 and LDLRAP1 (Fig. 5 and Table S4).

LDLR, LRP1, LRP2, LRP5 are all members of the low-density lipoprotein receptors family [54]. These receptors can bind a large number of extracellular ligands but a common ligand for all is the ApoE protein, which mediates internalization and catabolism of lipoprotein particles [55]. The LDLR is highest expressed in glial cells than in neurons, on the opposite LRP1 is highest expressed in neurons than in glia [56,57]. These observations suggest that correlations with regulative apoptotic caspases are not limited to a specific cell type. APLP2 an APP homologous can influence LRP1 expression and the toxicity mediated by beta-amyloid oligomers [58]. LDLRAP1 encodes for an adaptor cytosolic protein that interacts with and is involved in the endocytosis of LDLR. MSR1, SCARF1 and SCARB1 are involved in the uptake of lipoproteins. Endothelial lipase (LIPG) regulates the circulating level of HDL [59]. It is expressed in different areas of the CNS including CA3 pyramidal cells of the hippocampus, ependymal cells in the ventral part of the third ventricle [60]. LIPG is also expressed in brain capillary endothelial cells, major constituents of the blood brain barrier [61]. GPIHBP1, (glycosylphosphatidylinositol-anchored high-density lipoprotein binding) can be considered a platform for lipolysis [62] and it is also involved in the transport of lipoprotein lipase (LPL) [63]. LAMTOR1 and NPC2 work downstream to the internalization of lipoproteins. The first as key regulator of endosome dynamics and lysosome biogenesis, the second by removing unesterified cholesterol from late endosomes/lysosomes [64,65].

It is evident that several genes involved in cholesterol and lipids up-take show a coordinate expression with regulative apoptotic caspases. High positive scores in terms of correlation among pro-apoptotic regulative caspases and genes involved in cholesterol internalization are in agreement with the negative score observed in the case HMGCR [66]. In GBM, as above observed for other categories, these correlations are less robust and CASP1 distinctly

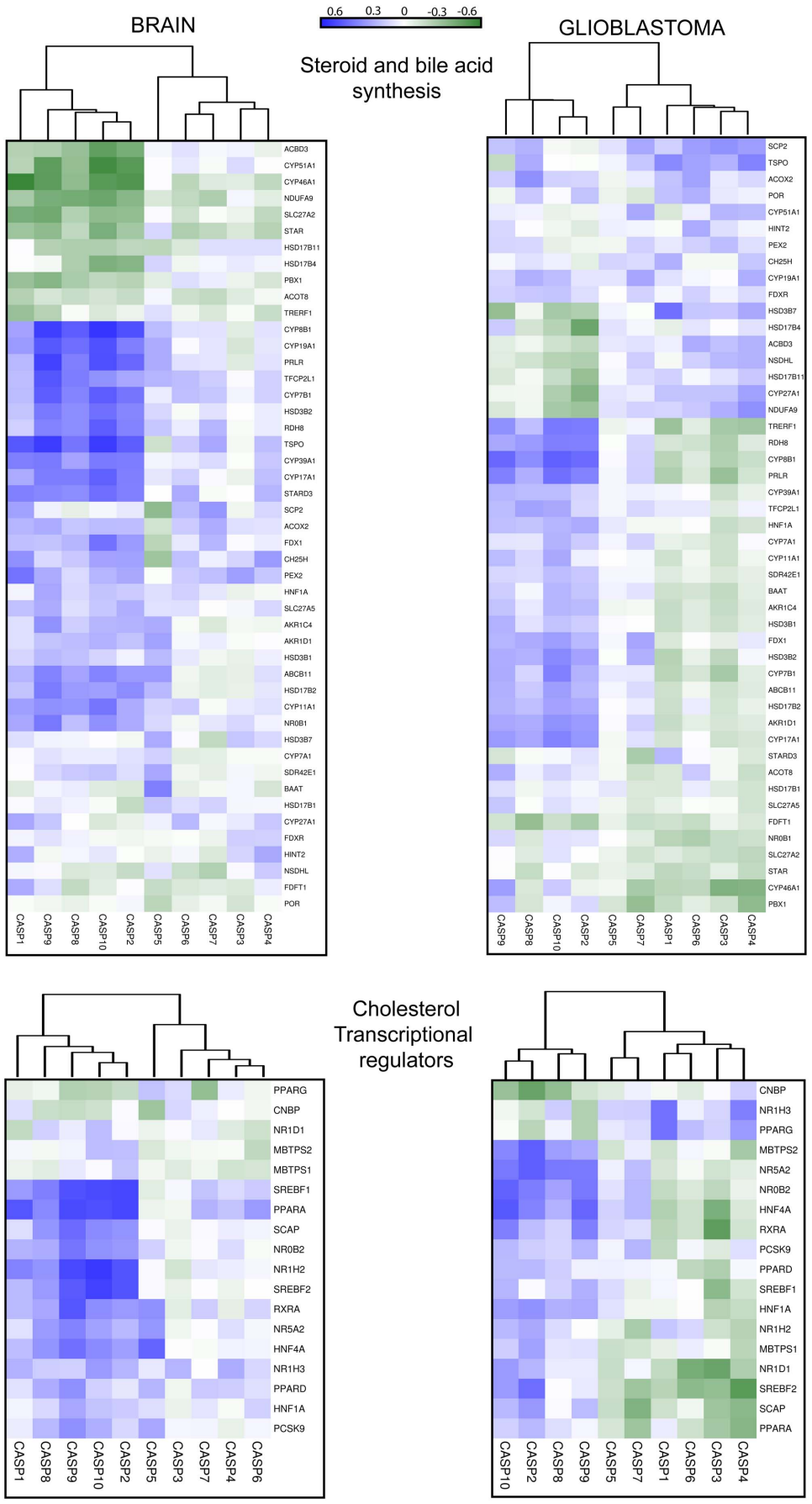

BRAIN
0.6 0.3 0 -0.3 -0.6
Steroid and bile acid synthesis
GLIOBLASTOMA
ACBD3
CYP51A1
CYP46A1
NDUFA9
SLC27A2
STAR
HSD17B11
HSD17B4
PBX1
ACOT8
TRERF1
CYP8B1
CYP19A1
PRLR
TFCP2L1
CYP7B1
HSD3B2
RDH8
TSPO
CYP39A1
CYP17A1
STARD3
SCP2
ACOX2
FDX1
CH25H
PEX2
HNF1A
SLC27A5
AKR1C4
AKR1D1
HSD3B1
ABCB11
HSD17B2
CYP11A1
NR0B1
HSD3B7
CYP7A1
SDR42E1
BAAT
HSD17B1
CYP27A1
FDXR
HINT2
NSDHL
FDFT1
POR
CASP1 CASP9 CASP8 CASP10 CASP2 CASP5 CASP6 CASP7 CASP3 CASP4
SCP2
TSPO
ACOX2
POR
CYP51A1
HINT2
PEX2
CH25H
CYP19A1
FDXR
HSD3B7
HSD17B4
ACBD3
NSDHL
HSD17B11
CYP27A1
NDUFA9
TRERF1
RDH8
CYP8B1
PRLR
CYP39A1
TFCP2L1
HNF1A
CYP7A1
CYP11A1
SDR42E1
BAAT
AKR1C4
HSD3B1
FDX1
HSD3B2
CYP7B1
ABCB11
HSD17B2
AKR1D1
CYP17A1
STARD3
ACOT8
HSD17B1
SLC27A5
FDFT1
NR0B1
SLC27A2
STAR
CYP46A1
PBX1
CASP9 CASP8 CASP10 CASP2 CASP5 CASP7 CASP1 CASP6 CASP3 CASP4
Cholesterol Transcriptional regulators
PPARG
CNBP
NR1D1
MBTPS2
MBTPS1
SREBF1
PPARA
SCAP
NR0B2
NR1H2
SREBF2
RXRA
NR5A2
HNF4A
NR1H3
PPARD
HNF1A
PCSK9
CASP1 CASP8 CASP9 CASP10 CASP2 CASP5 CASP3 CASP7 CASP4 CASP6
CNBP
NR1H3
PPARG
MBTPS2
NR5A2
NR0B2
HNF4A
RXRA
PCSK9
PPARD
SREBF1
HNF1A
NR1H2
MBTPS1
NR1D1
SREBF2
SCAP
PPARA
CASP10 CASP2 CASP8 CASP9 CASP5 CASP7 CASP1 CASP6 CASP3 CASP4

Figure 6. Co-expression analysis of caspases and cholesterol genes in human brain and GBM. Correlations of expression levels between caspases and cholesterol genes in different brain and GBM samples, for the indicated categories. Data obtained were used to calculate the correlation values with the Pearson method. In the heat map positive values are displayed in blue and negative in dark green. The dendrograms displayed on the top are based on hierarchical clustering using the complete linkage method.
doi:10.1371/journal.pone.0110610.g006

changes relationships with cholesterol genes, thus clustering with the effector caspases.

Steroid and bile acid synthesis. This category collects heterogeneous gene families. Nevertheless, the apoptotic regulative caspases cluster again together, whereas CASP1 is here found together with other inflammatory caspases and with the effector caspases. TSPO, CYP7B1, CYP8B1 and CYP19A1 show the highest positive correlation scores (Fig. 6 and Table S5).

Steroid biosynthesis begins with the transfer of free cholesterol from intracellular stores into mitochondria [67]. TSPO encodes for a translocator located on the outer mitochondrial membrane, which binds cholesterol with high affinity and transfers it in the inner mitochondrial membrane. TSPO activity is the rate-limiting step in the synthesis of all steroids [68]. Interestingly, alterations in TSPO expression has been found in various psychiatric disorders, including social phobia, post-traumatic stress disorder, adult separation anxiety and schizophrenia [69].

CYP7B1 catalyzes 7-hydroxylation of C_{19} and C_{21} steroids and in the brain it is involved in the metabolism of neurosteroids and oxysterols [70]. Defects of CYP7B1 in humans have been linked to spastic paraplegia [71,72].

CYP8B1 is a sterol 12α–hydroxylase. In mice expression of CYP7A1 and of CYP8B1 is integrated [73]. In the liver, circadian signals can influence CYP7A, CYP8B, and CYP51A1 expression [74].

CYP19A1 encodes for the cytochrome P450 aromatase an enzyme responsible for the synthesis of all oestrogens from androgen precursors [75]. Oestradiol synthesis in the brain regulates several functions of the adult CNS, from neural plasticity to injury responses [76]. Increased expression of CYP19A1 during neurodegeneration could interfere with apoptotic pathways and to decrease the extent of brain damage [77].

In glioblastoma the correlations affect other genes and are in general attenuated. TSPO expression is not longer linked to regulative caspases, whereas it shows positive correlations with effectors and inflammatory caspases.

Transcriptional regulators. Finally, we investigated the correlations among caspases and the expression of TFs (Transcription Factors) involved in the control of cholesterol genes expression. Figure 6 documents the results, which photocopy those previously obtained with the other categories. Apoptotic regulative caspases and CASP1 share similar relationships, in terms of expression correlations, with transcriptional regulators of cholesterol genes. SREBF1, SREBF2, PPARA and NR1H2 display the highest score with CASP2, CASP9 and CASP10 (Figure 6 and Table S6). In the case of SREBF2, Pearson correlation coefficients with caspases were similar to those described for cholesterol genes [78].

SREBF1 and SREBF2 (sterol regulatory element binding proteins) are TFs that control cellular lipid homeostasis. SREBF1 encodes for two proteins Srebp1a and Srebp1c produced via alternative transcription start sites. Srebp1c preferentially influences expression of fatty acid biosynthesis genes, whereas Srebp2 is devoted to transcribe genes involved in cholesterol homeostasis, lipoproteins import and lipids trafficking. Serbp1a can support transcription of both SREBF1 and SREBF2 genes [79]. In this respect it is interesting to note that only SREBF2 accomplishes a good correlation with CASP2 in GBM.

PPARA belongs to the family of peroxisome proliferator-activated receptors, which includes (PPARα, PPARβ/δ and PPARγ). These TFs function as obligate heterodimers with retinoid-X receptors (RXRs). PPARα supervises energy homeostasis by stimulating fatty acids and cholesterol breakdown and gluconeogenesis [80–82]. PPARβ/δ is mainly engaged in fatty acid oxidation. PPARγ principal activity is to drive storage of lipids, in particular by controlling adipocyte differentiation [83]. In the brain PPARD is the most abundant and quite ubiquitously expressed member, whereas PPARA and PPARG are expressed in more restricted areas and cell types [84,85]. In addition to their metabolic role, PPARs in the CNS have been implicated in the control of neuronal differentiation, death, inflammation and neurodegeneration [83]. In GBM expression correlation between PPARA and regulative caspases are abrogated. On the other side a good correlation between PPARG and CASP1 appears (Fig. 6).

NR1H2 (liver X receptor-beta) is another master TF orchestrating the expression of genes of the cholesterol homeostasis [86]. Similarly to SREBF1, SREBF2 and PPARA correlations between NR1H2 and CASP2, CASP9 and CASP10 are abrogated in GBM. Interestingly, in GBM new correlations emerged between the second nuclear hormone receptors liver X receptor alpha (LXRalpha/NR1H3) and inflammatory caspases CASP1 and CASP4.

In summary, the correlation among regulative apoptotic caspases and certain cholesterol genes observed in this study could be orchestrated by selected TFs, well-known master regulators of cholesterol metabolism such as SREBF1, SREBF2, PPARA and NR1H2.

Expression correlations among caspases and cholesterol genes in the liver

To understand whether our discoveries are limited to CNS or can be observed also in other tissues, we decided to compare variations in the expression levels of caspases and of cholesterol genes in human liver, an essential organ for cholesterol homeostasis. Gene expression profiles from 5 datasets including 106 microarrays of normal human liver were interrogated. Figure 7 shows that correlations among cholesterol genes and expression of regulative apoptotic caspases are less pronounced and only in some categories: *transcriptional regulators*, *cholesterol biosynthesis*, *steroid and bile acid synthesis*, these genes cluster together. In the liver CASP7 reaches the highest scores both for positive and negative correlations. For example CASP7 expression is strongly inversely correlated with those of several apolipoprotein genes, but it positively correlates with NPC1, VLDLR, SOAT1, SNX17 and VPS4B (Table S7); which are genes involved in cholesterol up-take and storage. It is important to note that similarly to caspase-2, caspase-7 expression is under the influence of Srebp1/2 and of statins [87].

CASP9 and CASP10 in terms of correlations with cholesterol genes maintain in the liver a pattern resembling that remarked in the brain and they cluster in almost all categories. By contrast CASP2 exhibits a correlation pattern rather different from the brain and it frequently clusters with CASP5. This observation suggests that the correlations between caspase-2 and cholesterol genes are related to a specific brain endeavor [88,89]. Remark-

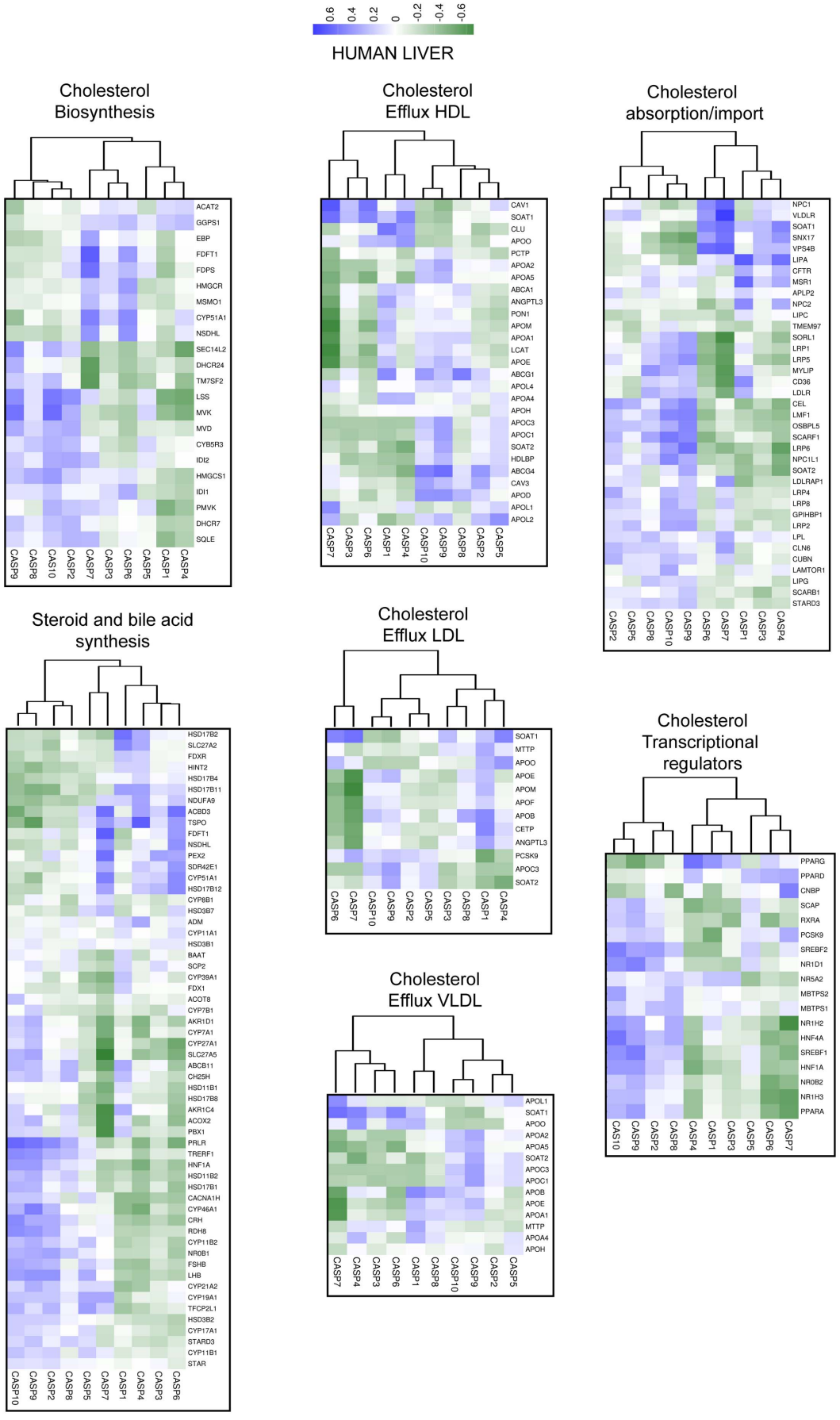

0.6
0.4
0.2
0
-0.2
-0.4
-0.6
HUMAN LIVER
Cholesterol Biosynthesis
ACAT2
GGPS1
EBP
FDFT1
FDPS
HMGCR
MSMO1
CYP51A1
NSDHL
SEC14L2
DHCR24
TM7SF2
LSS
MVK
MVD
CYB5R3
IDI2
HMGCS1
IDI1
PMVK
DHCR7
SQLE
CASP9 CASP8 CAS10 CASP2 CASP7 CASP3 CASP6 CASP5 CASP1 CASP4
Cholesterol Efflux HDL
CAV1
SOAT1
CLU
APOO
PCTP
APOA2
APOA5
ABCA1
ANGPTL3
PON1
APOM
APOA1
LCAT
APOE
ABCG1
APOL4
APOA4
APOH
APOC3
APOC1
SOAT2
HDLBP
ABCG4
CAV3
APOD
APOL1
APOL2
CASP7 CASP3 CASP6 CASP1 CASP4 CASP10 CASP9 CASP8 CASP2 CASP5
Cholesterol absorption/import
NPC1
VLDLR
SOAT1
SNX17
VPS4B
LIPA
CFTR
MSR1
APLP2
NPC2
LIPC
TMEM97
SORL1
LRP1
LRP5
MYLIP
CD36
LDLR
CEL
LMF1
OSBPL5
SCARF1
LRP6
NPC1L1
SOAT2
LDLRAP1
LRP4
LRP8
GPIHBP1
LRP2
LPL
CLN6
CUBN
LAMTOR1
LIPG
SCARB1
STARD3
CASP2 CASP5 CASP8 CASP10 CASP9 CASP6 CASP7 CASP1 CASP3 CASP4
Steroid and bile acid synthesis
HSD17B2
SLC27A2
FDXR
HINT2
HSD17B4
HSD17B11
NDUFA9
ACBD3
TSPO
FDFT1
NSDHL
PEX2
SDR42E1
CYP51A1
HSD17B12
CYP8B1
HSD3B7
ADM
CYP11A1
HSD3B1
BAAT
SCP2
CYP39A1
FDX1
ACOT8
CYP7B1
AKR1D1
CYP7A1
CYP27A1
SLC27A5
ABCB11
CH25H
HSD11B1
HSD17B8
AKR1C4
ACOX2
PBX1
PRLR
TRERF1
HNF1A
HSD11B2
HSD17B1
CACNA1H
CYP46A1
CRH
RDH8
CYP11B2
NR0B1
FSHB
LHB
CYP21A2
CYP19A1
TFCP2L1
HSD3B2
CYP17A1
STARD3
CYP11B1
STAR
CASP10 CASP9 CASP2 CASP8 CASP5 CASP7 CASP1 CASP4 CASP3 CASP6
Cholesterol Efflux LDL
SOAT1
MTTP
APOO
APOE
APOM
APOF
APOB
CETP
ANGPTL3
PCSK9
APOC3
SOAT2
CASP6 CASP7 CASP10 CASP9 CASP2 CASP5 CASP3 CASP8 CASP1 CASP4
Cholesterol Efflux VLDL
APOL1
SOAT1
APOO
APOA2
APOA5
SOAT2
APOC3
APOC1
APOB
APOE
APOA1
MTTP
APOA4
APOH
CASP7 CASP4 CASP3 CASP6 CASP1 CASP8 CASP10 CASP9 CASP2 CASP5
Cholesterol Transcriptional regulators
PPARG
PPARD
CNBP
SCAP
RXRA
PCSK9
SREBF2
NR1D1
NR5A2
MBTPS2
MBTPS1
NR1H2
HNF4A
SREBF1
HNF1A
NR0B2
NR1H3
PPARA
CAS10 CASP9 CASP2 CASP8 CASP4 CASP1 CASP3 CASP5 CASP6 CASP7

Figure 7. Co-expression analysis of caspases and cholesterol genes in human liver. Correlations of expression levels between caspases and cholesterol genes in different liver samples, for the indicated categories. Data obtained were used to calculate the correlation values with the Pearson method. In the heat map positive values are displayed in blue and negative in dark green. The dendrograms displayed on the top are based on hierarchical clustering using the complete linkage method.
doi:10.1371/journal.pone.0110610.g007

ably, several reports have proposed specific biological activities for caspase-2 in the CNS [26,89–91].

Caspases, cholesterol genes and aging

Changes in the expression of cholesterol genes in the brain are linked to several physiological/pathological conditions including, regeneration, plasticity, circadian rhythms, diet, degeneration, social behavior and also the presence of several different cell linages [37,38,43,58,69,74,77,83,92]. To understand the significance of the described correlations, we decided to interrogate more precisely the public available datasets used in our studies. We restricted our inquiries to one dataset (GSE 17612) in which, information about the specific brain area (the anterior prefrontal cortex/Brodmann area 10) and subjects age were available. 23 microarray of normal brain met these requirements. Age distribution is shown in figure 8A. Cholesterol and caspase genes were ranked accordingly to the Spearman method, respect to the age of the subjects. Figure 8B illustrates that the expression of a set of cholesterol genes shows a good inverse correlation respect to age (green dots) and on the opposite, the expression of a different set of genes increases with aging (blue dots).

Among genes, which mRNAs levels are down-regulated with aging, we have found CYP46A1 and HMGCR (Fig. 8B and 8C). CYP46A1 encodes for the hydroxylase, which converts cholesterol to 24S-hydroxycholesterol and provides the major route in cholesterol excretion from the brain [93]. In the case of HMGCR our observations are in agreement with previous results pointing to a decline in brain cholesterol synthesis with age [43,94,95]. On the opposite previous study reported an up-regulation of cholesterol-24-hydroxylase mRNA in mouse brain with aging [96]. Indeed in their studies the authors analyzed the hippocampus from day 10 up to 21 months. After an up-regulation at 3 months, a trend to decrease in both mRNA and protein levels of CYP46A1 can be appreciated. In humans 24S-hydroxycholesterol levels are highest in the first decade of life and then decline with age [97]. Furthermore, also in human brain, analysis of cholesterol 24-hydroxylase protein evidenced some reduction during aging [98]. Undoubtedly, further studies are necessary to clarify this point. Taking into account that mice lacking 24-hydroxylase exhibit severe learning and memory defects [99], clarification of CYP46A1 expression during aging is of particular interest.

We have included in this analysis also the different caspases. Figure 8C reveals that expression of inflammatory CASP1 and also of the effector caspases CASP6 and CASP7 positively correlates with aging.

Several studies have indicated a contribution of caspase-6 to neurodegeneration including Alzheimer and Huntington diseases [100–104]. A more recent report has proposed that caspase-6 activity can predict lower episodic memory ability in aged individuals [105]. CASP1 is a key regulator of inflammation, via the generation of IL-1β and involved in the regulation of age-related cognitive-dysfunctions. CASP1 genetic variations have been associated with cognitive function [106] and several data have linked caspase-1 to brain aging [107,108]. Deficiency in Nlrp3 inflammasome-mediated caspase-1 activity improved cognitive function and motor performance in aged mice [109]. Furthermore, caspase-1 also influences in a still undetermined manner lipid metabolism [110].

Conclusions

Our initial aim, of discovering compensatory pathways to solve the caspase-2 mystery has led us to unveil large and shared correlations among different caspases and cholesterol genes. Expression correlations studies have attracted attention to uncover new biological circuits [111,112] and dedicated methodological tools have been developed [113,114]. We have applied hierarchical clustering of gene expression correlations to hypothesize new functions for caspases and caspase-2 in particular. An unexpected finding was that the regulative apoptotic caspases (CASP9 and CASP10 in particular) share a correlation pattern with cholesterol genes, similarly to CASP2. The reported strong correlations among these caspases and certain cholesterol genes in a heterogeneous tissue, such as the human brain and in a heterogeneous population, suggest that expression of these genes is influenced by common signaling networks linked to specific biological processes. Hence, expression of certain cholesterol genes and of regulative apoptotic caspases in the brain should be under the control of the same regulative circuits. Since in GBM these correlations are in general less prominent, it is possible that genetic lesions, altering proliferation also impact on cholesterol homeostasis. Although the contribution of cholesterol genes to GBM development is largely unexplored, recently a survival pathway engaged by the LDL receptor, through the EGFRvIII/PI3K/SREBP-1 axis has been discovered [115]. This observation suggests that oncogenic driven changes in gene expression could revise the harmonic co-regulation of caspases and cholesterol genes.

This scenario is well exemplified by CASP1, which evidences the most overt changes in terms of correlations with cholesterol genes between brain and GBM. Taking into account that its expression is augmented in GBM, these changes could reflect the establishment of an inflammatory microenvironment.

When the same analysis were performed using microarray data obtained from cultured glioblastoma cells, correlations were in general weaker and apoptotic regulative caspases clustered separately (data not shown). This observation indicates that complex networks of environmental signals control the expression of these genes. Networks that cannot be easily replaced under *in vitro* culture conditions. A consideration that could explain the low number of genes, which expression is influenced by caspase-2 siRNA in cultured glioblastoma cells.

It is evident that, correlations in terms of expression could have different biological origins/implications: from the presence of different cell lineages in the sample, up to different neuronal activities, different inflammatory/degenerative states, different cognitive functions and different hormonal signaling. We are consciousness that this complexity deserves further experimental work. Here we have discovered that for some caspases, in a specific brain's area, the correlations with certain cholesterol genes could be related to aging. In this case the correlation could be linked to the induction of degenerative processes. In fact, caspase signaling engaged by the ordered activation of caspase-8 and caspase-3/7 controls microglia activation and neurotoxicity with implications in Parkinson's and Alzheimer's diseases [116].

For other caspases, including caspase-2 the reason for this correlation is unclear at the moment. Taking into account that

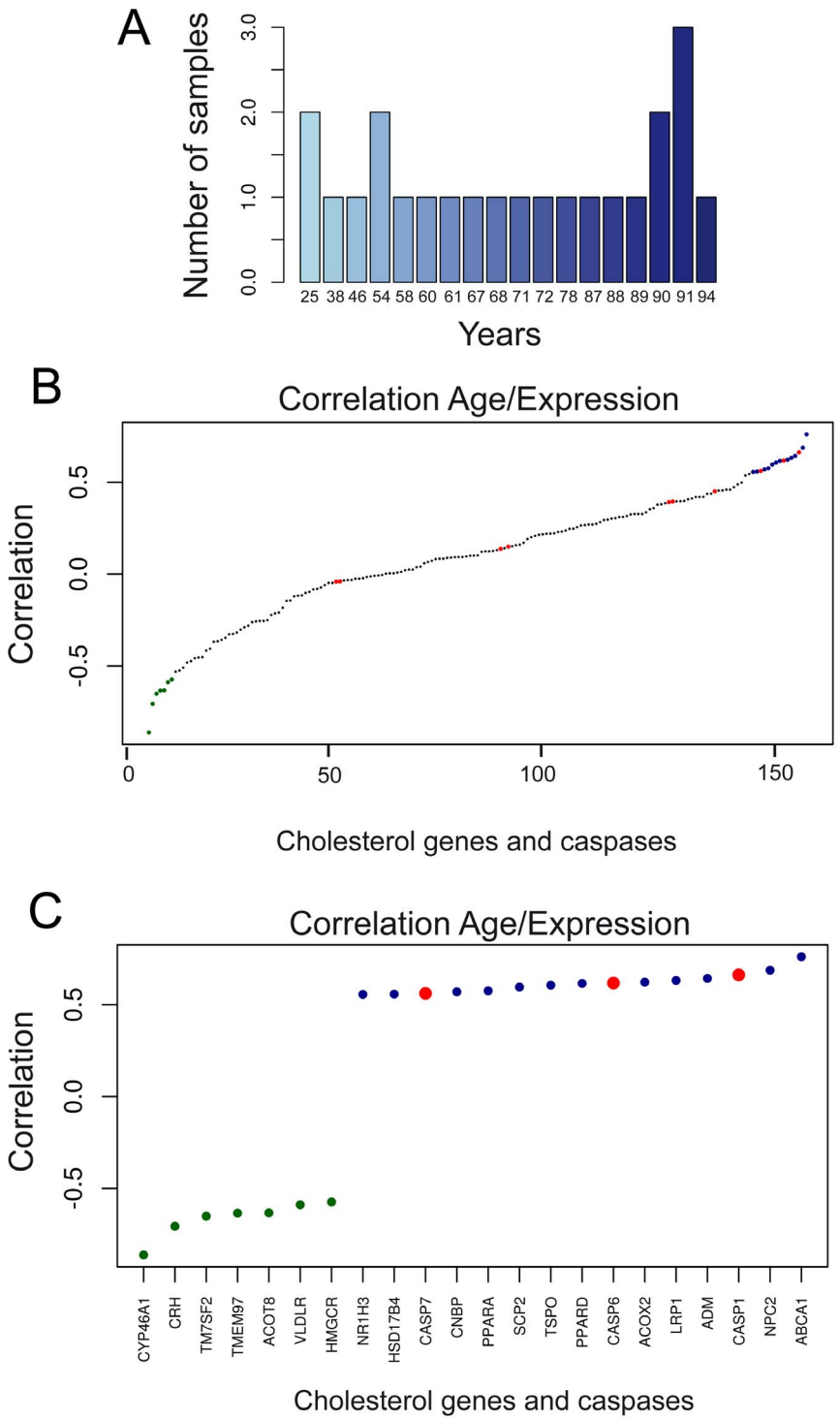


Figure 8. Analysis of cholesterol and caspase genes expression with aging in anterior prefrontal cortex. A. Bar plot displaying the age frequencies of the samples used in the analysis. B. Dot plot of ranked correlations of the 129 genes involved in cholesterol homeostasis/steroidogenesis and caspases respect to age. C. Dot plot of ranked correlations of genes involved in cholesterol homeostasis/steroidogenesis and caspases that scored significant expression correlation with respect to age. We applied the Spearman method (Benjamini and Hockberg correction). $P<0.05$.

doi:10.1371/journal.pone.0110610.g008

evidences sustaining non-apoptotic roles of caspases in the CNS are accumulating [117] several hypothesis concerning caspase-2 could be formulated. Previous studies [28], here confirmed, linking SREBFs levels to CASP2 expression suggest a direct involvement of this enzyme in cholesterol homeostasis. Although the mechanisms need to be elucidated, our data further encourage investigating towards this direction.

Supporting Information

Figure S1 A. Expression profiles in U87MG cells of the different caspases. Microarray expression data were obtained from GSE14889 [118]. B. mRNA expression levels of CASP8 were measured using qRT-PCR in U87MG cells transfected with siRNA against Caspase-8 or a control siRNA. Data are presented as mean ± SD; $n = 3$. C. mRNA expression levels of CYP1B1, CYP51A1 and LIPA were measured using qRT-PCR in U87MG cells transfected with siRNA against Caspase-8 or a control siRNA. Data are presented as mean ± SD; $n = 3$.
(JPG)

Table S1 Genes down-regulated>1.5 fold in U87MG cells silenced for Caspase-2.
(XLS)

Table S2 Top score cholesterol genes of the subclass *cholesterol export* in terms of correlations of expression levels with regulative apoptotic caspases in human brain. Values are shown for all caspases.
(XLS)

Table S3 Top score cholesterol genes of the subclass *cholesterol biosynthesis* in terms of correlations of expression levels with regulative apoptotic caspases in human brain. Values are shown for all caspases.
(XLS)

Table S4 Top score cholesterol genes of the subclass *cholesterol adsorption/import* in terms of correlations of expression levels with regulative apoptotic caspases in human brain. Values are shown for all caspases.
(XLS)

Table S5 Top score cholesterol genes of the subclass *steroid and bile acid synthesis* in terms of correlations of expression levels with regulative apoptotic caspases in human brain. Values are shown for all caspases.
(XLS)

Table S6 Top score cholesterol genes of the subclass *transcriptional regulators* in terms of correlations of expression levels with regulative apoptotic caspases in human brain. Values are shown for all caspases.
(XLS)

Table S7 Top score cholesterol genes in terms of correlations of expression levels with caspase-7 in human liver. Values are shown for all caspases.
(XLS)

Acknowledgments

We thank Ivana Manini for helping in some experiments. AT received a fellowship from AIRC.

Author Contributions

Conceived and designed the experiments: CB. Performed the experiments: RP AT. Analyzed the data: RP AT FF CB. Contributed reagents/materials/analysis tools: RP AT FF CB. Wrote the paper: RP AT FF CB.

References

1. Thornberry NA, Lazebnik Y (1998) Caspases: enemies within. Science 1998, 281:1312–1316.
2. Martinon F, Burns K, Tschopp J (2002) The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. Mol Cell 10:417–426.
3. Kanuka H, Kuranaga E, Takemoto K, Hiratou T, Okano H, et al. (2005) Drosophila caspase transduces Shaggy/GSK-3beta kinase activity in neural precursor development. EMBO J 24:3793–3806.
4. Huesmann GR, Clayton DF (2006) Dynamic role of postsynaptic caspase-3 and BIRC4 in zebra finch song-response habituation. Neuron 52:1061–1072.
5. Wang JY, Chen F, Fu XQ, Ding CS, Zhou L, et al. (2014) Caspase-3 cleavage of dishevelled induces elimination of postsynaptic structures. Dev Cell 28:670–684.
6. Lamy L, Ngo VN, Emre NC, Shaffer AL 3rd, Yang Y, et al. (2013) Control of autophagic cell death by caspase-10 in multiple myeloma. Cancer Cell 23:435–449.
7. Paroni G, Henderson C, Schneider C, Brancolini C (2002) Caspase-2 can trigger cytochrome C release and apoptosis from the nucleus. J Biol Chem 277:15147–15161.
8. Vakifahmetoglu-Norberg H, Zhivotovsky B (2010) The unpredictable caspase-2: what can it do? Trends Cell Biol 20:150–159
9. Fava LL, Bock FJ, Geley S, Villunger A (2012) Caspase-2 at a glance. J Cell Sci 125:5911–5915.
10. Puccini J, Dorstyn L, Kumar S (2013) Caspase-2 as a tumour suppressor. Cell Death Differ 20:1133–1139.
11. Puccini J, Shalini S, Voss AK, Gatei M, Wilson CH, et al. (2013) Loss of caspase-2 augments lymphomagenesis and enhances genomic instability in Atm-deficient mice. Proc Natl Acad Sci U S A 110:19920–19925.
12. Senapedis WT, Kennedy CJ, Boyle PM, Silver PA (2011) Whole genome siRNA cell-based screen links mitochondria to Akt signaling network through uncoupling of electron transport chain. Mol Biol Cell 22:1791–1805.
13. Scarpulla RC (2011) Metabolic control of mitochondrial biogenesis through the PGC-1 family regulatory network. Biochim Biophys Acta 1813:1269–1278.
14. Baggs JE, Price TS, DiTacchio L, Panda S, Fitzgerald GA, et al. (2009) Network features of the mammalian circadian clock. PLoS Biol 7:e52.
15. Chies R, Nobbio L, Edomi P, Schenone A, Schneider C (2003) Alterations in the Arf6-regulated plasma membrane endosomal recycling pathway in cells overexpressing the tetraspan protein Gas3/PMP22. J Cell Sci 116:987–999.
16. Paroni G, Henderson C, Schneider C, Brancolini C (2001) Caspase-2-induced apoptosis is dependent on caspase-9, but its processing during UV- or tumor necrosis factor-dependent cell death requires caspase-3. J Biol Chem 276:21907–21915.
17. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Speed TP. Biostatistics 4:249–64.
18. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57.
19. Huang da W, Sherman BT, Zheng X, Yang J, Imamichi T, et al. (2009) Extracting biological meaning from large gene lists with DAVID. Curr Protoc Bioinformatics Chapter 13:Unit 13 11.
20. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, et al. (2011) NCBI GEO: archive for functional genomics data sets—10 years on. Nucleic Acids Res 39:D1005–1010.
21. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, et al. (2007) ArrayExpress a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 35:D747–750.
22. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004): affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20:307–315.
23. Dai M, Wang P, Boyd AD, Kostov G, Athey B, et al. (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33:e175.
24. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met 57:289–300.
25. Ihmels J, Bergmann S, Berman J, Barkai N (2005) Comparative gene expression analysis by differential clustering approach: application to the Candida albicans transcription program. PLoS Genet 1:e39.
26. Troy CM, Rabacchi SA, Friedman WJ, Frappier TF, Brown K, et al. (2000) Caspase-2 mediates neuronal cell death induced by beta-amyloid. 20:1386–1392.
27. Tiwari M, Lopez-Cruzan M, Morgan WW, Herman B (2011) Loss of caspase-2-dependent apoptosis induces autophagy after mitochondrial oxidative stress in primary cultures of young adult cortical neurons. J Biol Chem. 286:8493–506.
28. Logette E, Le Jossic-Corcos C, Masson D, Solier S, Sequeira-Legrand A, et al. (2005) Caspase-2, a novel lipid sensor under the control of sterol regulatory element binding protein 2. Mol Cell Biol 25:9621–9631.
29. Logette E, Solary E, Corcos L (2005) Identification of a functional DNA binding site for the SREBP-1c transcription factor in the first intron of the human caspase-2 gene. Biochim Biophys Acta 1738:1–5.
30. Nebert DW, Dalton TP (2006) The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. Nat Rev Cancer 6:947–960.
31. Lepesheva GI, Waterman MR (2004) CYP51 the omnipotent P450. Mol Cell Endocrinol 215:165–170.

32. Fouchier SW, Defesche JC (2013) Lysosomal acid lipase A and the hypercholesterolaemic phenotype. Curr Opin Lipidol 24:332–338.
33. Kallin A, Johannessen LE, Cani PD, Marbehant CY, Essaghir A, et al. (2007) SREBP-1 regulates the expression of heme oxygenase 1 and the phosphatidylinositol-3 kinase regulatory subunit p55 gamma. J Lipid Res 48:1628–1636.
34. Glasgow BJ, Abduragimov AR, Farahbakhsh ZT, Faull KF, Hubbell WL (1995) Tear lipocalins bind a broad array of lipid ligands. Curr Eye Res 14:363–372.
35. Segatto M, Trapani L, Lecis C, Pallottini V (2012) Regulation of cholesterol biosynthetic pathway in different regions of the rat central nervous system. Acta Physiol (Oxf) 206:62–71.
36. Yan C, Ding X, Dasgupta N, Wu L, Du H (2012) Gene profile of myeloid-derived suppressive cells from the bone marrow of lysosomal acid lipase knock-out mice. PLoS One. 7:e30701.
37. Vance JE, Hayashi H (2010) Formation and function of apolipoprotein E-containing lipoproteins in the nervous system. Biochim Biophys Acta 1801:806–818.
38. Dassati S, Waldner A, Schweigreiter R (2014) Apolipoprotein D takes center stage in the stress response of the aging and degenerative brain. Neurobiol Aging 35:1632–1642.
39. Hu YW, Zheng L, Wang Q (2010) Characteristics of apolipoprotein M and its relation to atherosclerosis and diabetes. Biochim Biophys Acta 1801:100–105.
40. Christoffersen C, Jauhiainen M, Moser M, Porse B, Ehnholm C, et al. (2008) Effect of apolipoprotein M on high density lipoprotein metabolism and atherosclerosis in low density lipoprotein receptor knock-out mice. J Biol Chem 283:1839–1847.
41. Christoffersen C, Nielsen LB (2013) Apolipoprotein M: bridging HDL and endothelial function. Curr Opin Lipidol 24:295–300.
42. Tsukamoto K, Mani DR, Shi J, Zhang S, Haagensen DE, et al. (2013) Identification of apolipoprotein D as a cardioprotective gene using a mouse model of lethal atherosclerotic coronary artery disease. Proc Natl Acad Sci U S A 110:17023–17028.
43. Kim WS, Wong J, Weickert CS, Webster MJ, Bahn S, et al. (2009) Apolipoprotein-D expression is increased during development and maturation of the human prefrontal cortex. J Neurochem 109:1053–1066.
44. DeMattos RB, Brendza RP, Heuser JE, Kierson M, Cirrito JR, et al. (2001) Purification and characterization of astrocyte-secreted apolipoprotein E and J-containing lipoproteins from wild-type and human apoE transgenic mice. Neurochem Int 39:415–425.
45. Hayashi H, Campenot RB, Vance DE, Vance JE (2004) Glial lipoproteins stimulate axon growth of central nervous system neurons in compartmented cultures. J Biol Chem 279:14009–14015.
46. Boyles JK, Zoellner CD, Anderson LJ, Kosik LM, Pitas RE, et al. (1989) A role for apolipoprotein E, apolipoprotein A-I, and low density lipoprotein receptors in cholesterol transport during regeneration and remyelination of the rat sciatic nerve. J Clin Invest 83:1015–1031.
47. Ignatius MJ, Gebicke-Harter PJ, Skene JH, Schilling JW, Weisgraber KH, et al. (1986) Expression of apolipoprotein E during nerve degeneration and regeneration. Proc Natl Acad Sci USA 83:1125–1129.
48. Wolf AB, Caselli RJ, Reiman EM, Valla J (2013) APOE and neuroenergetics: an emerging paradigm in Alzheimer's disease. Neurobiol Aging 34:1007–1017.
49. Aviram M, Vaya J (2013) Paraoxonase 1 activities, regulation, and interactions with atherosclerotic lesion. Curr Opin Lipidol 24:339–344.
50. Arakawa R, Abe-Dohmae S, Asai M, Ito JI, Yokoyama S (2000) Involvement of caveolin-1 in cholesterol enrichment of high density lipoprotein during its assembly by apolipoprotein and THP-1 cells. J Lipid Res 41:1952–1962.
51. Karten B, Campenot RB, Vance DE, Vance JE (2006) Expression of ABCG1, but not ABCA1, correlates with cholesterol release by cerebellar astroglia. J Biol Chem 281:4049–4057.
52. Morell P, Jurevics H (1996) Origin of cholesterol in myelin. Neurochem Res 21:463–470.
53. Byskov AG, Andersen CY, Nordholm L, Thogersen H, Xia G, et al. (1995) Chemical structure of sterols that activate oocyte meiosis. Nature 374:559–562.
54. Wagner T, Pietrzik CU (2012) The role of lipoprotein receptors on the physiological function of APP. Exp Brain Res 217:377–387.
55. Koo C, Wernette-Hammond ME, Garcia Z, Malloy MJ, Uauy R, et al. (1988) Uptake of cholesterol-rich remnant lipoproteins by human monocyte-derived macrophages is mediated by low density lipoprotein receptors. J Clin Invest 81:1332–1340.
56. Rebeck GW, Reiter JS, Strickland DK, Hyman BT (1993) Apolipoprotein E in sporadic Alzheimer's disease: allelic variation and receptor interactions. Neuron 1:575–580.
57. Rapp A, Gmeiner B, Huttinger M (2006): Implication of apoE isoforms in cholesterol metabolism by primary rat hippocampal neurons and astrocytes. Biochimie 88:473–483.
58. Rushworth JV, Griffiths HH, Watt NT, Hooper NM (2013) Prion protein-mediated toxicity of amyloid-beta oligomers requires lipid rafts and the transmembrane LRP1. J Biol Chem 288:8935–8951.
59. Brown RJ, Lagor WR, Sankaranarayanan S, Yasuda T, Quertermous T, et al. (2010) Impact of combined deficiency of hepatic lipase and endothelial lipase on the metabolism of both high-density lipoproteins and apolipoprotein B-containing lipoproteins. Circ Res 107:357–364.
60. Paradis E, Clavel S, Julien P, Murthy MR, de Bilbao F, et al. (2004) Lipoprotein lipase and endothelial lipase expression in mouse brain: regional distribution and selective induction following kainic acid-induced lesion and focal cerebral ischemia. Neurobiol Dis 15:312–325.
61. Sovic A, Panzenboeck U, Wintersperger A, Kratzer I, Hammer A, et al. (2005) Regulated expression of endothelial lipase by porcine brain capillary endothelial cells constituting the blood-brain barrier. J Neurochem 94:109–119.
62. Young SG, Zechner R (2013) Biochemistry and pathophysiology of intravascular and intracellular lipolysis. Genes Dev 27:459–484.
63. Davies BS, Goulbourne CN, Barnes RH 2nd, Turlo KA, Gin P, et al. (2011) Assessing mechanisms of GPIHBP1 and lipoprotein lipase movement across endothelial cells. J Lipid Res 53:2690–2697.
64. Vance JE, Peake KB (2011) Function of the Niemann-Pick type C proteins and their bypass by cyclodextrin. Curr Opin Lipidol 22:204–209.
65. Takahashi Y, Nada S, Mori S, Soma-Nagae T, Oneyama C, et al. (2012) The late endosome/lysosome-anchored p18-mTORC1 pathway controls terminal maturation of lysosomes. Biochem Biophys Res Commun 417:1151–1157.
66. Brown MS, Goldstein JL (1979) Receptor-mediated endocytosis: insights from the lipoprotein receptor system. Proc Natl Acad Sci USA 76:3330–3337.
67. Miller RH, Harper AE (1988) Regulation of valine and alpha-ketoisocaproate metabolism in rat kidney mitochondria. Am J Physiol 255:E475–481.
68. Hauet T, Yao ZX, Bose HS, Wall CT, Han Z, et al. (2005) Peripheral-type benzodiazepine receptor-mediated action of steroidogenic acute regulatory protein on cholesterol entry into leydig cell mitochondria. Mol Endocrinol 19:540–554.
69. Da Pozzo E, Costa B, Martini C (2012) Translocator protein (TSPO) and neurosteroids: implications in psychiatric disorders. Curr Mol Med 12:426–442.
70. Martin C, Bean R, Rose K, Habib F, Seckl J (2001) cyp7b1 catalyses the 7alpha-hydroxylation of dehydroepiandrosterone and 25-hydroxycholesterol in rat prostate. Biochem J 355:509–515.
71. Salinas S, Proukakis C, Crosby A, Warner TT (2008) Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms. Lancet Neurol 7:1127–1138.
72. Lorbek G, Lewinska M, Rozman D (2012) Cytochrome P450s in the synthesis of cholesterol and bile acids—from mouse models to human diseases. FEBS J 279:1516–1533.
73. Murphy C, Parini P, Wang J, Bjorkhem I, Eggertsen G, et al. (2005) Cholic acid as key regulator of cholesterol synthesis, intestinal absorption and hepatic storage in mice. Biochim Biophys Acta 1735:167–175.
74. Noshiro M, Kawamoto T, Furukawa M, Fujimoto K, Yoshida Y, et al. (2004) Rhythmic expression of DEC1 and DEC2 in peripheral tissues: DEC2 is a potent suppressor for hepatic cytochrome P450s opposing DBP. Genes Cells 9:317–329.
75. Wu YG, Bennett J, Talla D, Stocco C (2011) Testosterone, not 5alpha-dihydrotestosterone, stimulates LRH-1 leading to FSH-independent expression of Cyp19 and P450scc in granulosa cells. Mol Endocrinol 25:656–668.
76. Cornil CA, Charlier TD (2010) Rapid behavioural effects of oestrogens and fast regulation of their local synthesis by brain aromatase. J Neuroendocrinol 22:664–673.
77. Saldanha CJ, Duncan KA, Walters BJ (2009) Neuroprotective actions of brain aromatase. Front Neuroendocrinol 30:106–118.
78. Brignull LM, Czimmerer Z, Saidi H, Daniel B, Villela I (2013) Reprogramming of lysosomal gene expression by interleukin-4 and Stat6. BMC Genomics 14:853.
79. Horton JD, Goldstein JL, Brown MS (2002) SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. J Clin Invest 109:1125–1131.
80. Kliewer SA, Umesono K, Noonan DJ, Heyman RA, Evans RM (1992) Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathways through heterodimer formation of their receptors. Nature 358:771–774.
81. Desvergne B, Wahli W (1999) Peroxisome proliferator-activated receptors: nuclear control of metabolism. Endocr Rev 20:649–688.
82. Varga T, Czimmerer Z, Nagy L (2011) PPARs are a unique set of fatty acid regulated transcription factors controlling both lipid metabolism and inflammation. Biochim Biophys Acta 1812:1007–1022.
83. Heneka MT, Landreth GE (2007) PPARs in the brain. Biochim Biophys Acta 1771:1031–1045.
84. Basu-Modak S, Braissant O, Escher P, Desvergne B, Honegger P, et al. (1999) Peroxisome proliferator-activated receptor beta regulates acyl-CoA synthetase 2 in reaggregated rat brain cell cultures. J Biol Chem 274:35881–35888.
85. Moreno S, Farioli-Vecchioli S, Ceru MP (2004) Immunolocalization of peroxisome proliferator-activated receptors and retinoid X receptors in the adult rat CNS. Neuroscience 123:131–145.
86. Zhao C, Dahlman-Wright K (2010) Liver X receptor in cholesterol metabolism. J Endocrinol 204:233–240.
87. Gibot L, Follet J, Metges JP, Auvray P, Simon B, et al. (2009) Human caspase 7 is positively controlled by SREBP-1 and SREBP-2. Biochem J 420:473–483.
88. Bhattacharyya R, Kovacs DM (2010) ACAT inhibition and amyloid beta reduction. Biochim Biophys Acta 1801:960–965.
89. Chae SS, Yoo CB, Jo C, Yun SM, Jo SA, et al. (2010) Caspases-2 and -8 are involved in the presenilin1/gamma-secretase-dependent cleavage of amyloid precursor protein after the induction of apoptosis. J Neurosci Res 88:1926–1933.

90. Carlsson Y, Schwendimann L, Vontell R, Rousset CI, Wang X, et al. (2011) Genetic inhibition of caspase-2 reduces hypoxic-ischemic and excitotoxic neonatal brain injury. Ann Neurol 70:781–789.
91. Pozueta J, Lefort R, Ribe EM, Troy CM, Arancio O, et al. (2013) Caspase-2 is required for dendritic spine and behavioural alterations in J20 APP transgenic mice. Nat Commun 4:1939.
92. Talwar P, Silla Y, Grover S, Gupta M, Agarwal R, et al. (2014). Genomic convergence and network analysis approach to identify candidate genes in Alzheimer's disease. BMC Genomics 15:199.
93. Russell DW, Halford RW, Ramirez DM, Shah R, Kotti T (2009) Cholesterol 24-hydroxylase: an enzyme of cholesterol turnover in the brain. Annu Rev Biochem 78:1017–1040.
94. Svennerholm L, Bostrom K, Jungbjer B, Olsson L (1994) Membrane lipids of adult human brain: lipid composition of frontal and temporal lobe in subjects of age 20 to 100 years. J Neurochem 63:1802–1811.
95. Martin M, Dotti CG, Ledesma MD (2010) Brain cholesterol in normal and pathological aging. Biochim Biophys Acta 1801:934–944.
96. Martin MG, Perga S, Trovo L, Rasola A, Holm P, et al. (2008) Cholesterol loss enhances TrkB signaling in hippocampal neurons aging in vitro. Mol Biol Cell 19:2101–2112.
97. Lutjohann D, Breuer O, Ahlborg G, Nennesmo I, Siden A, et al. (1996) Cholesterol homeostasis in human brain: evidence for an age-dependent flux of 24S-hydroxycholesterol from the brain into the circulation. Proc Natl Acad Sci USA 93:9799–9804.
98. Lund EG, Guileyardo JM, Russell DW (1999) cDNA cloning of cholesterol 24-hydroxylase, a mediator of cholesterol homeostasis in the brain. Proc Natl Acad Sci USA 96:7238–7243.
99. Kotti TJ, Ramirez DM, Pfeiffer BE, Huber KM, Russell DW (2006) Brain cholesterol turnover required for geranylgeraniol production and learning in mice. Proc Natl Acad Sci USA 103:3869–3874.
100. Albrecht S, Bourdeau M, Bennett D, Mufson EJ, Bhattacharjee M, et al. (2007) Activation of caspase-6 in aging and mild cognitive impairment. Am J Pathol 170:1200–1209.
101. Guo H, Albrecht S, Bourdeau M, Petzke T, Bergeron C, et al. (2004) Active caspase-6 and caspase-6-cleaved tau in neuropil threads, neuritic plaques, and neurofibrillary tangles of Alzheimer's disease. Am J Pathol 165:523–531.
102. Graham RK, Deng Y, Carroll J, Vaid K, Cowan C, et al. (2010) Cleavage at the 586 amino acid caspase-6 site in mutant huntingtin influences caspase-6 activation in vivo. J Neurosci 30:15019–15029.
103. Uribe V, Wong BK, Graham RK, Cusack CL, Skotte NH, et al. (2012) Rescue from excitotoxicity and axonal degeneration accompanied by age-dependent behavioral and neuroanatomical alterations in caspase-6-deficient mice. Hum Mol Genet 21:1954–1967.
104. Graham RK, Ehrnhoefer DE, Hayden MR (2011) Caspase-6 and neurodegeneration. Trends Neurosci 34:646–656.
105. Ramcharitar J, Afonso VM, Albrecht S, Bennett DA, LeBlanc AC (2013) Caspase-6 activity predicts lower episodic memory ability in aged individuals. Neurobiol Aging 34:1815–1824.
106. Trompet S, de Craen AJ, Slagboom P, Shepherd J, Blauw GJ, et al. (2008) Genetic variation in the interleukin-1 beta-converting enzyme associates with cognitive function. The PROSPER study. Brain 131:1069–1077.
107. Gemma C, Bachstetter AD, Cole MJ, Fister M, Hudson C, et al. (2007) Blockade of caspase-1 increases neurogenesis in the aged hippocampus. Eur J Neurosci 26:2795–2803.
108. Ojala J, Alafuzoff I, Herukka SK, van Groen T, Tanila H, et al. (2009) Expression of interleukin-18 is increased in the brains of Alzheimer's disease patients. Neurobiol Aging 30:198–209.
109. Youm YH, Grant RW, McCabe LR, Albarado DC, Nguyen KY, et al. (2013) Canonical Nlrp3 inflammasome links systemic low-grade inflammation to functional decline in aging. Cell Metab 18:519–532.
110. Kotas ME, Jurczak MJ, Annicelli C, Gillum MP, Cline GW, et al. (2013) Role of caspase-1 in regulation of triglyceride metabolism. Proc Natl Acad Sci USA 2013, 110:4810–4815.
111. Freeman TC, Goldovsky L, Brosch M, van Dongen S, Maziere P (2007) Construction, visualisation, and clustering of transcription networks from microarray expression data. PLoS Comput Biol 3:2032–2042.
112. Tseng GC, Ghosh D, Feingold E (2012) Comprehensive literature review and statistical considerations for microarray meta-analysis. Nucleic Acids Res 40:3785–3799.
113. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P (2004) Coexpression analysis of human genes across many microarray data sets. Genome Res 14:1085–1094.
114. Watson M (2006) CoXpress: differential co-expression in gene expression data. BMC Bioinformatics 7:509.
115. Guo D, Reinitz F, Youssef M, Hong C, Nathanson D, et al. (2011) An LXR agonist promotes glioblastoma cell death through inhibition of an EGFR/AKT/SREBP-1/LDLR-dependent pathway. Cancer Discov 1:442–56.
116. Burguillos MA, Deierborg T, Kavanagh E, Persson A, Hajji N, et al. (2011) Caspase signalling controls microglia activation and neurotoxicity. Nature 2011 472:319–324.
117. Hyman BT, Yuan J (2012) Apoptotic and non-apoptotic roles of caspases in neuronal physiology and pathophysiology. Nat Rev Neurosci 13:395–406.
118. Foti C, Florean C, Pezzutto A, Roncaglia P, Tomasella A et al. (2009) Characterization of caspase-dependent and caspase-independent deaths in glioblastoma cells treated with inhibitors of the ubiquitin-proteasome system. Mol Cancer Ther 8:3140–3150.

OPEN ACCESS Freely available online

PLOS ONE

Next-Generation Sequencing Analysis of MiRNA Expression in Control and FSHD Myogenesis

Veronica Colangelo[1☯], Stéphanie François[1☯], Giulia Soldà[2], Raffaella Picco[3], Francesca Roma[2], Enrico Ginelli[2], Raffaella Meneveri[1]*

1 Department of Health Sciences, University of Milano-Bicocca, Monza, Italy, **2** Department of Medical Biotechnology and Translational Medicine, University of Milan, Milan, Italy, **3** Department of Medical and Biological Sciences, University of Udine, Udine, Italy

Abstract

Emerging evidence has demonstrated that miRNA sequences can regulate skeletal myogenesis by controlling the process of myoblast proliferation and differentiation. However, at present a deep analysis of miRNA expression in control and FSHD myoblasts during differentiation has not yet been derived. To close this gap, we used a next-generation sequencing (NGS) approach applied to *in vitro* myogenesis. Furthermore, to minimize sample genetic heterogeneity and muscle-type specific patterns of gene expression, miRNA profiling from NGS data was filtered with FC≥4 (log_2FC≥2) and p-value<0.05, and its validation was derived by qRT-PCR on myoblasts from seven muscle districts. In particular, control myogenesis showed the modulation of 38 miRNAs, the majority of which (34 out 38) were up-regulated, including myomiRs (miR-1, -133a, -133b and -206). Approximately one third of the modulated miRNAs were not previously reported to be involved in muscle differentiation, and interestingly some of these (i.e. miR-874, -1290, -95 and -146a) were previously shown to regulate cell proliferation and differentiation. FSHD myogenesis evidenced a reduced number of modulated miRNAs than healthy muscle cells. The two processes shared nine miRNAs, including myomiRs, although with FC values lower in FSHD than in control cells. In addition, FSHD cells showed the modulation of six miRNAs (miR-1268, -1268b, -1908, 4258, -4508- and -4516) not evidenced in control cells and that therefore could be considered FSHD-specific, likewise three novel miRNAs that seem to be specifically expressed in FSHD myotubes. These data further clarify the impact of miRNA regulation during control myogenesis and strongly suggest that a complex dysregulation of miRNA expression characterizes FSHD, impairing two important features of myogenesis: cell cycle and muscle development. The derived miRNA profiling could represent a novel molecular signature for FSHD that includes diagnostic biomarkers and possibly therapeutic targets.

Citation: Colangelo V, François S, Soldà G, Picco R, Roma F, et al. (2014) Next-Generation Sequencing Analysis of MiRNA Expression in Control and FSHD Myogenesis. PLoS ONE 9(10): e108411. doi:10.1371/journal.pone.0108411

Editor: Atsushi Asakura, University of Minnesota Medical School, United States of America

Received March 18, 2014; **Accepted** August 28, 2014; **Published** October 6, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. Sequencing raw data are available from the SRA database (accession number SRP034654).

Funding: This work was supported by grants from the AFMTéléthon (reference Number 16547), http://www.afm-telethon.com, and the Italian Ministry of University and Research (PRIN 2008BEYKL8_002), www.prin.miur.it. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: raffaella.meneveri@unimib.it

☯ These authors contributed equally to this work.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is the third most common myopathy, with an incidence of 1 in 14.000 in the general population. Signs of FSHD become visible in an individual's 20's (men) or 30's (women) and include loss of muscle strength in the face, shoulders, and upper arms before eventually attaining the abdomen, legs and feet. FSHD is transmitted as an autosomal dominant trait and it is thought to be mainly associated to an epigenetic alteration leading to transcriptional imbalance of the responsible genes [1,2]. Almost all FSHD patients carry rearrangements reducing the copy number of a 3.3 kb tandemly repeated sequence (D4Z4) located at 4q35, and containing a conserved open reading frame for a double homeobox gene (DUX4). D4Z4 copy number is highly polymorphic in healthy individuals ranging between 11 and >100copies while FSHD patients carry fewer than 11 repeats [3]. Notably, although the number of D4Z4 repeats seems to be a critical determinant of the age of onset and clinical severity of FSHD, patients without D4Z4 contraction (phenotypic FSHD or FSHD2) as well as healthy individuals with D4Z4 contraction (carrier) have been also identified [4,5]. All these observations strongly suggests that FSHD derives from the interplay of more complex genetic and epigenetic events than those already described; these additional events might take place at either 4q35 or elsewhere in the human genome.

Recently a unifying genetic model [6] that provides the expression of D4Z4 as a major cause of FSHD has been proposed. Another recent paper [7] defining the epigenetic regulation of 4q35 gene expression, demonstrated that D4Z4 deletion is associated to reduced epigenetic repression by Polycomb silencing in FSHD patients. Furthermore, *DBE-T*, a chromatin associated non-coding RNA is produced selectively in FSHD patients and it coordinates the de-repression of 4q35 genes. However, another study evaluating a large-scale population analysis of healthy and

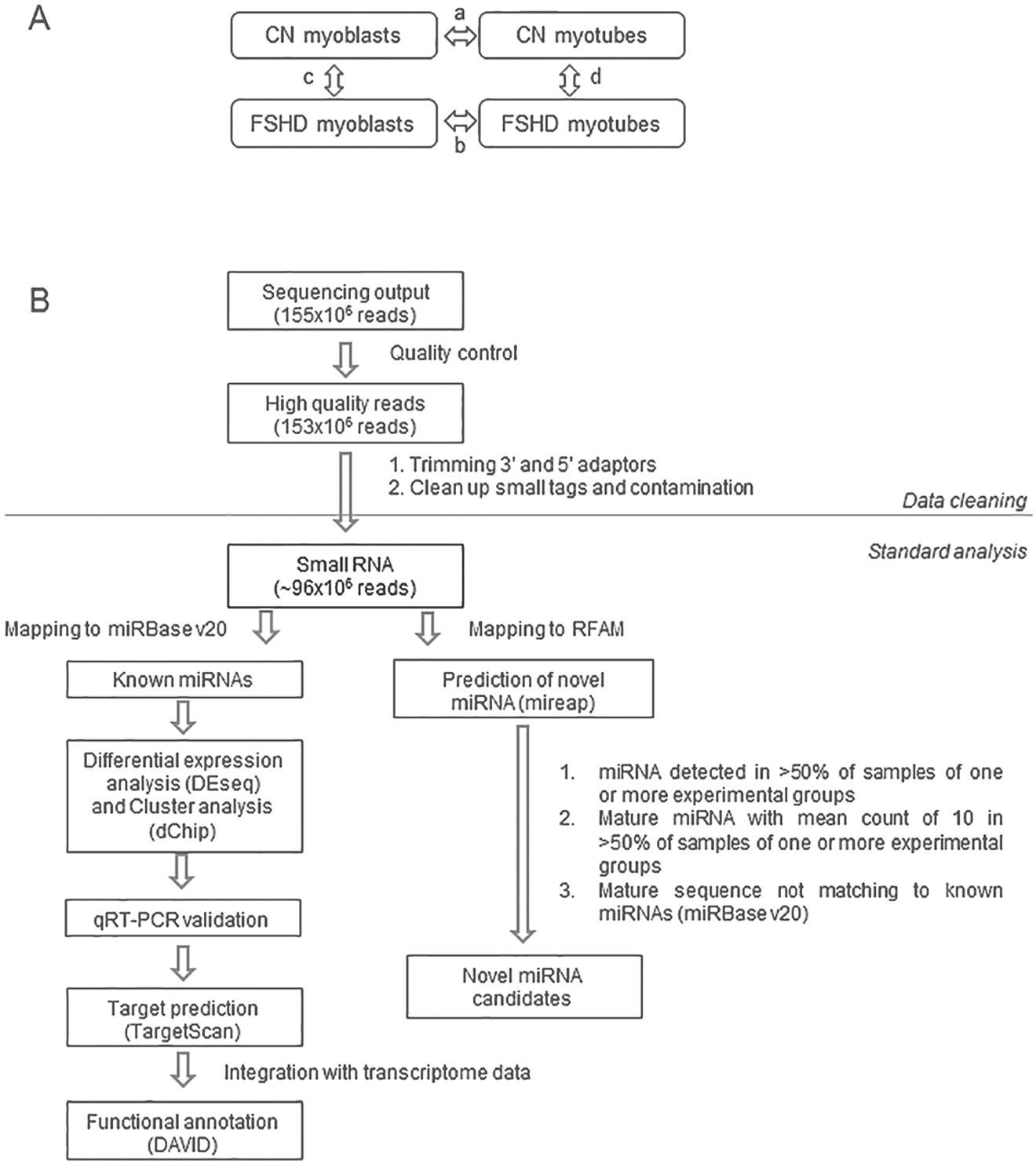


Figure 1. Study design and data analysis. A) Study design: Next-generation Sequencing (NGS) on three control and three FSHD myoblast cell lines before and after *in vitro* myogenic differentiation was used in order to derive miRNA modulation in: a) control myogenesis (CN myotubes vs CN myoblasts; arrow a); b) FSHD myogenesis (FSHD myotubes vs FSHD myoblasts; arrow b); c) FSHD myoblasts versus control myoblast (arrow c), and d) FSHD myotubes vs control myotubes (arrow d). B) Flow chart of filtering and analysis of NGS data. NGS generated a total of 153×10^6 high quality reads, that were filtered for rRNA, tRNA, snRNA, snoRNA, repeat associated RNAs and intron/exon. The filtered reads (approx. 99×10^6 reads, an average of 8×10^6/sample) were analyzed to derive known miRNAs (R/Bioconductor) and novel miRNAs (mireap). Differentially expressed miRNAs between samples were derived by $\log_2FC \geq 2$ and p-value<0.05 parameters. The homogeneity of miRNA modulation among samples was evaluated by cluster analysis (dChip). miRNAs were then validated by qRT-PCR. Finally, target genes were predicted for modulated miRNAs and functionally annotated by DAVID.
doi:10.1371/journal.pone.0108411.g001

unrelated FSHD patients reports that the genetic criteria in order to manifest FSHD (D4Z4 contraction associated with a specific chromosomal background 4A-161-p(A)- pathogenic haplotype) occur in 63.7% of the analyzed FSHD patients and in 1.3% of healthy subjects [8]. Although these data certainly represent a major advance toward the definition of the molecular basis of FSHD, many questions on the disease etiology remain unexplained. Also the reported high degree of variability of the disease, in term of onset, progression and severity strongly suggests that other mechanism(s) linked to the 4q subtelomere and/or to other regions of the human genome may play a role in the disease pathogenesis.

Various recent studies have demonstrated that both FSHD myoblasts and myotubes are characterized by an extensive gene expression dysregulation mainly affecting the myogenesis and including genes linked to cell cycle control, particularly G1/S and G2/M transitions, muscle structure, mitochondrial function, oxidative stress response, and cholesterol biosynthesis [9,10,11].

The deciphering of the molecular basis of FSHD has been further complicated by the finding that microRNAs (miRNAs) are involved in both control and pathological myogenesis [12,13,14]. MiRNAs are evolutionarily conserved short non-coding RNAs (~22 nts) that regulate the stability and/or the translational efficiency of target mRNAs. They have a very pervasive role since it is estimated that a single miRNA has the potential to regulate hundreds of target genes, and therefore, >90% of all human genes could be under miRNAs regulation [15]. MiRNAs are essential for normal mammalian development and are involved in fine-tuning of many biological processes, such as differentiation, proliferation and apoptosis [16,17]. Emerging evidence has demonstrated that miRNA sequences can regulate skeletal myogenesis by controlling the process of myoblast proliferation and differentiation, in particular, microRNA-1, -206 and -133a/b were defined as myomiRNAs to emphasize their crucial role in myogenesis [18,19]. More recently, a simultaneous microRNA/mRNA expression profiling of healthy myogenic cells during differentiation allowed to identify the involvement of miRNAs in the regulation of various biological processes such as cell cycle, transcription, transport, apoptosis and DNA damage [20]. Given these assumptions it was not surprising that miRNAs dysregulation was found to be involved in muscle dysfunctions [9,12,21].

To date, miRNA studies reported for FSHD were essentially based on the analysis of a restricted number of known miRNA sequences, thus not allowing the derivation of the full miRNA-based dysregulation network. To close this gap, here we report miRNAs expression analysis, derived by next-generation sequencing (NGS), in primary muscle cells from healthy and FSHD subjects during differentiation.

Results

Study design and NGS general results

In order to determine the entire small non coding RNAs (< 35 nts) transcriptome in control (CN) and FSHD primary myoblast cell lines, before and after *in vitro* myogenic differentiation, we used next-generation sequencing (NGS). Study design was organized to allow the comparison of small non-coding RNA expression profiles between FSHD and CN myoblasts and myotubes (Fig. 1A, arrows c and d respectively) and of the two differentiation processes (Fig. 1A, arrows a and b, respectively). In order to derive biological markers (i.e. miRNA dysregulation) commonly manifested by different affected muscle districts, we used two FSHD primary myoblasts cell lines deriving from rhomboid and one from ilio-psoas muscles, and three control myoblasts from tensor fascia lata, quadriceps and vastus intermedius (Table S1).

As shown in the flow chart reported in Fig. 1B, small RNA sequencing generated a total of 153×10^6 high quality reads. Mature miRNAs make up the majority of sequences in the 18 to 25 nts size range (65% average), with a clear peak at 22 nts in all samples. The average of known miRNAs per sample was of 556, whereas un-annotated small RNAs (new miRNA candidates) per sample were 28.

The differential expression of known miRNAs was analyzed in the different stages of muscle differentiation by DEseq analysis. Furthermore, in order to assess the robustness of our approach, some of the miRNAs identified as differentially expressed were validated by qRT-PCR using specific TaqMan miRNA assays in primary FSHD and healthy myoblasts. For these experiments we employed the same cell lines used for NGS and additional ones from different muscles, including biceps and deltoid (Table S1). As reported in Materials and Methods, the nine control and the seven FSHD cell lines showed a highly comparable extent of Desmin-positive cells and of myogenic markers modulation upon differentiation (Fig. S1). Gene targets of differentially expressed miRNAs were predicted in both control and FSHD cellular systems by using the TargetScan algorithm. Derived gene targets were filtered on two independent transcriptome profiling experiments carried out on control and FSHD myogenesis [9,10], and shared targets were then functionally annotated by DAVID. Novel miRNAs were predicted by mireap and considered as novel candidates only if detected with a mean reads of ten in at least two out of three samples of one or more experimental groups (CN and FSHD myoblasts; CN and FSHD myotubes).

Modulation of miRNA expression during physiological and FSHD myoblast differentiation

We first analyzed the data regarding physiological myogenesis (control myotubes vs control myoblasts; Fig. 1A, arrow a). Filtered miRNA reads (mapping to miRBase v20) from the three control myoblasts samples and the corresponding myotubes were analyzed for differential expression by DEseq analysis, setting the $\log_2$ Fold Change ($\log_2$FC) at ≥ 2 and p-value<0.05. From this analysis we evidenced that during the control myogenesis 38 miRNAs showed a modulation in their expression, and that the great majority of them (34 out of 38) were up-regulated (Fig. 2A and B).

The hierarchical clustering analysis clearly separated proliferating from differentiated cells independently of the muscle district used (tensor fascia lata, quadriceps and vastus intermedius). As expected, the muscle specific miRNAs (myomiRs) hsa-miR-1, -133a, -133b and -206, were among the most up-regulated (Fig. 2B and Table S3). Twenty-six miRNAs were already reported to be involved in muscle differentiation either in human or in mouse cells, whereas 12 miRNAs, ten up-regulated (hsa-miR-95, -146a, -874, -1246, -1290, -3164, -4488, -208a, -944 and -3144) and two down-regulated (hsa-miR-3934 and -3165), were not previously known to be involved in muscle differentiation. The full list of the miRNAs modulated during control myoblasts differentiation with corresponding FC and p-value is reported in Table S3.

The same analysis was carried out on FSHD myogenesis (Fig. 1A, arrow b). As shown in Fig. 3A, the DEseq analysis evidenced the modulation of only 15 miRNAs during pathological muscle differentiation. Even in this case the hierarchical clustering analysis clearly separated proliferating from differentiated cells, independently of the muscle district (Fig. 3B). The majority of miRNAs was up-regulated (11 out of 15), including myomiR-1 and -206, although with a FC lower than that showed in control

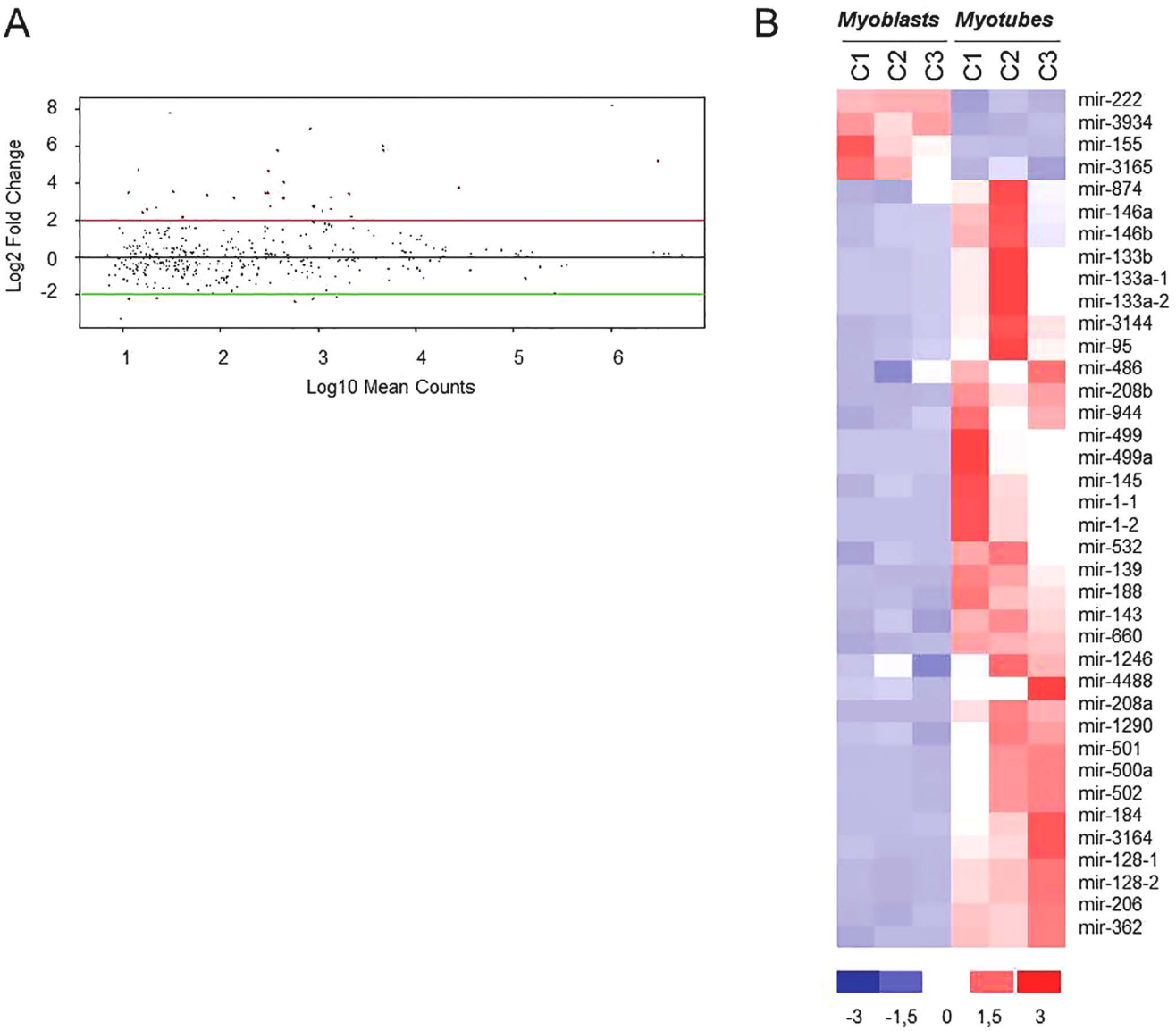


Figure 2. MiRNA modulation in control myogenesis. A) DEseq analysis of miRNAs differentially expressed in control myotubes vs control myoblasts (control differentiation). MiRNAs showing a modulation with $log_2FC \geq 2$ and a p-value<0.05 are shown as red dots. B) Hierarchical clustering of the 38 modulated miRNAs (34 up-regulated and 4 down-regulated) in regard to the analyzed samples. C1:MX01010MBS; C2: MX03609MBS; C3: MX01110MBS, Control cell lines (see Table S1).
doi:10.1371/journal.pone.0108411.g002

myogenesis (Table S4). MyomiR-133a and -133b showed up-regulation trend ($log_2FC>5$) without reaching significance (p-value = 0.33). The full list of the miRNAs modulated in FSHD myogenesis, with corresponding FC and p-value, is reported in Table S4. Scatter plots of the reads of modulated miRNAs (for each control and FSHD proliferating and differentiated cell line) are reported in Fig. S2. To further support the results obtained by the sequencing approach, the same control and FSHD myoblast and myotube RNAs were used to analyze the expression of myomiRs (miR-1, miR-133a and miR-206) by qRT-PCR (Fig. S3). In both control and FSHD myotubes, we confirmed the general trend of myomiRs up-regulation derived by sequencing, with the pathological samples showing a lower extent of up-regulation than the normal ones.

Dysregulation of miRNA expression in FSHD myoblasts and myotubes

We next performed DEseq analysis of miRNAs differentially expressed in FSHD myoblasts and myotubes vs controls (Fig. 1, arrows c and d). No miRNAs were found significantly dysregulated ($log_2FC \geq 2$ and p-value<0.05) in FSHD versus control myoblasts (Fig. 1, arrows c); this result was probably due to the high variability of miRNA expression observed in myoblasts. Conversely, 21 miRNAs were found dysregulated in FSHD myotubes (Table S5 and Fig. 4A), among these 12 miRNAs were up-regulated. The hierarchical clustering analysis clearly separated the pathological samples from the control ones and the three analyzed samples of each group resulted homogeneous in miRNAs dysregulation (Fig. 4B).

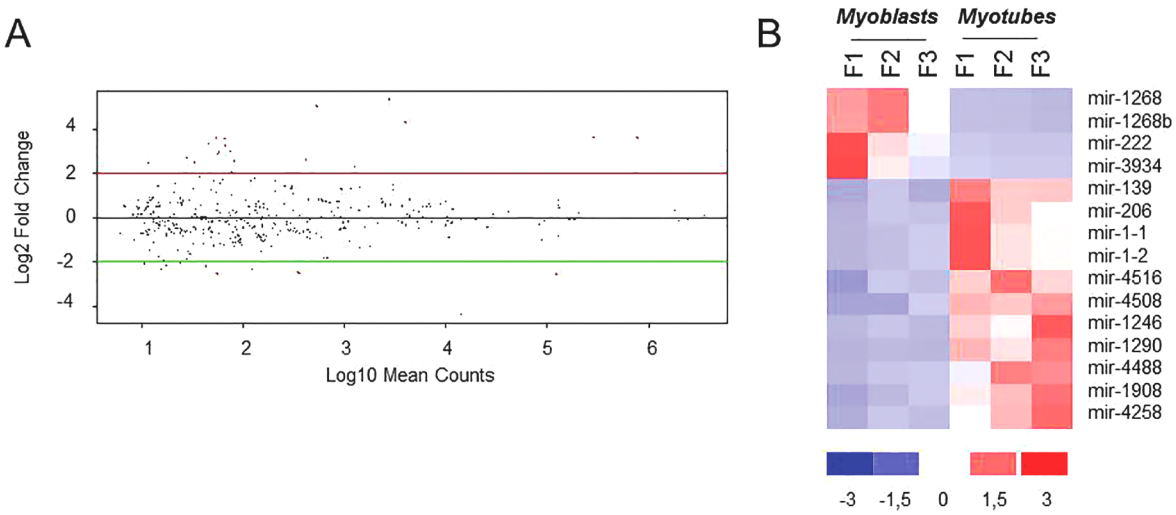


Figure 3. MiRNA modulation in FSHD myogenesis. A) DEseq analysis of miRNAs differentially expressed in FSHD myotubes vs FSHD myoblasts (FSHD differentiation). MiRNAs showing a modulation with $\log_2 FC \geq 2$ and a p-value<0.05 are shown as red dots. B) Hierarchical clustering of the 15 modulated miRNAs (11 up-regulated and 4 down-regulated) in regard to the analyzed samples. F1:MX00409MBS; F2: MX03010MBS; F3:MX04309MBS, FSHD cell lines (see Table S1).
doi:10.1371/journal.pone.0108411.g003

qRT-PCR Validation

The effective validation of deep sequencing results was performed by the TaqMan miRNA assay on all the cell lines listed in Table S1, including those already used for the NGS experiment. Particularly, for myomiR-1, -133a and -206 the assay was carried out at different time points during myogenic differentiation (0, 3 and 7 days of differentiation) (Fig. 5A). In control cells, the myomiRs progressively increased their expression with the proceeding of time of differentiation, reaching the maximum of expression at seven days, with FC values ranging from approximately 350 folds (miR-1) to 28 folds (miR-206). In FSHD cells myomiRs showed an up-regulation significantly lower than that observed in controls, reaching at day seven an expression value similar to or slightly lower than that showed by control cells at day three. Comparable fusion indexes and expression values of myogenic markers in healthy and FSHD myoblasts and myotubes (see Fig. S1) support that the obtained results are not related to a

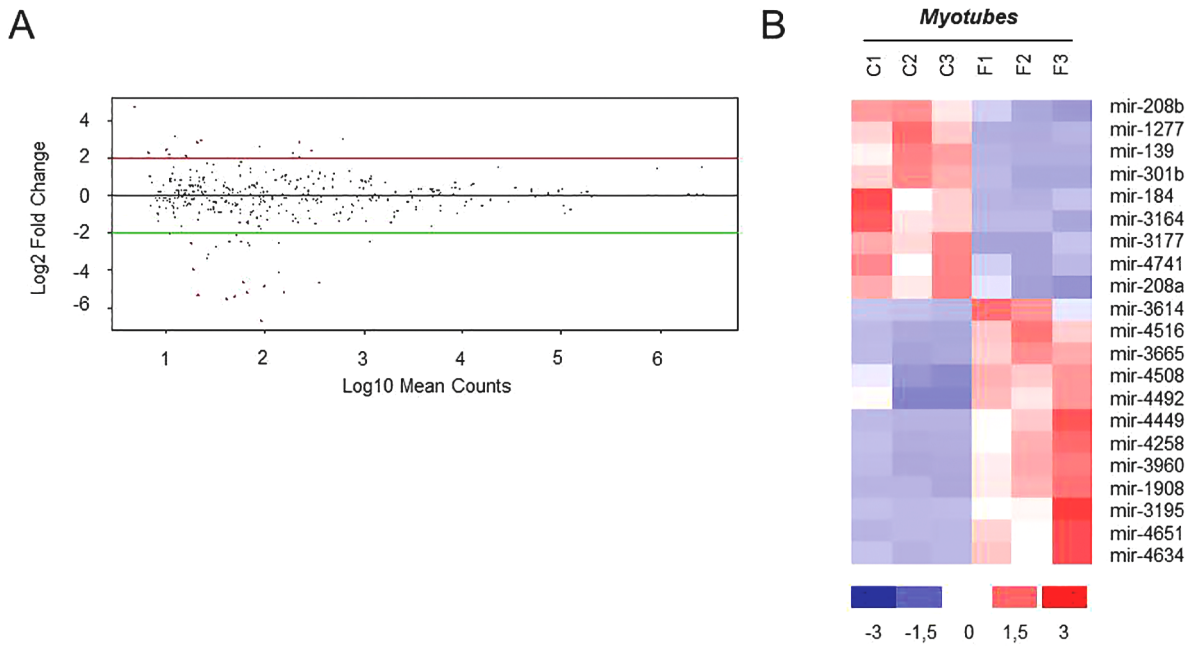


Figure 4. MiRNA dysregulation in FSHD myotubes. A) DEseq analysis of miRNAs differentially expressed in FSHD myotubes vs control myotubes. MiRNAs showing a differential expression of $\log_2 FC \geq 2$ and a p-value<0.05 are shown as red dots. B) Hierarchical clustering of the 21 modulated miRNAs (12 up-regulated and 9 down-regulated) in regard to the analyzed samples. C1:MX01010MBS; C2: MX03609MBS; C3: MX01110MBS, Control cell lines; F1:MX00409MBS; F2: MX03010MBS; F3:MX04309MBS, FSHD cell lines (see Table S1).
doi:10.1371/journal.pone.0108411.g004

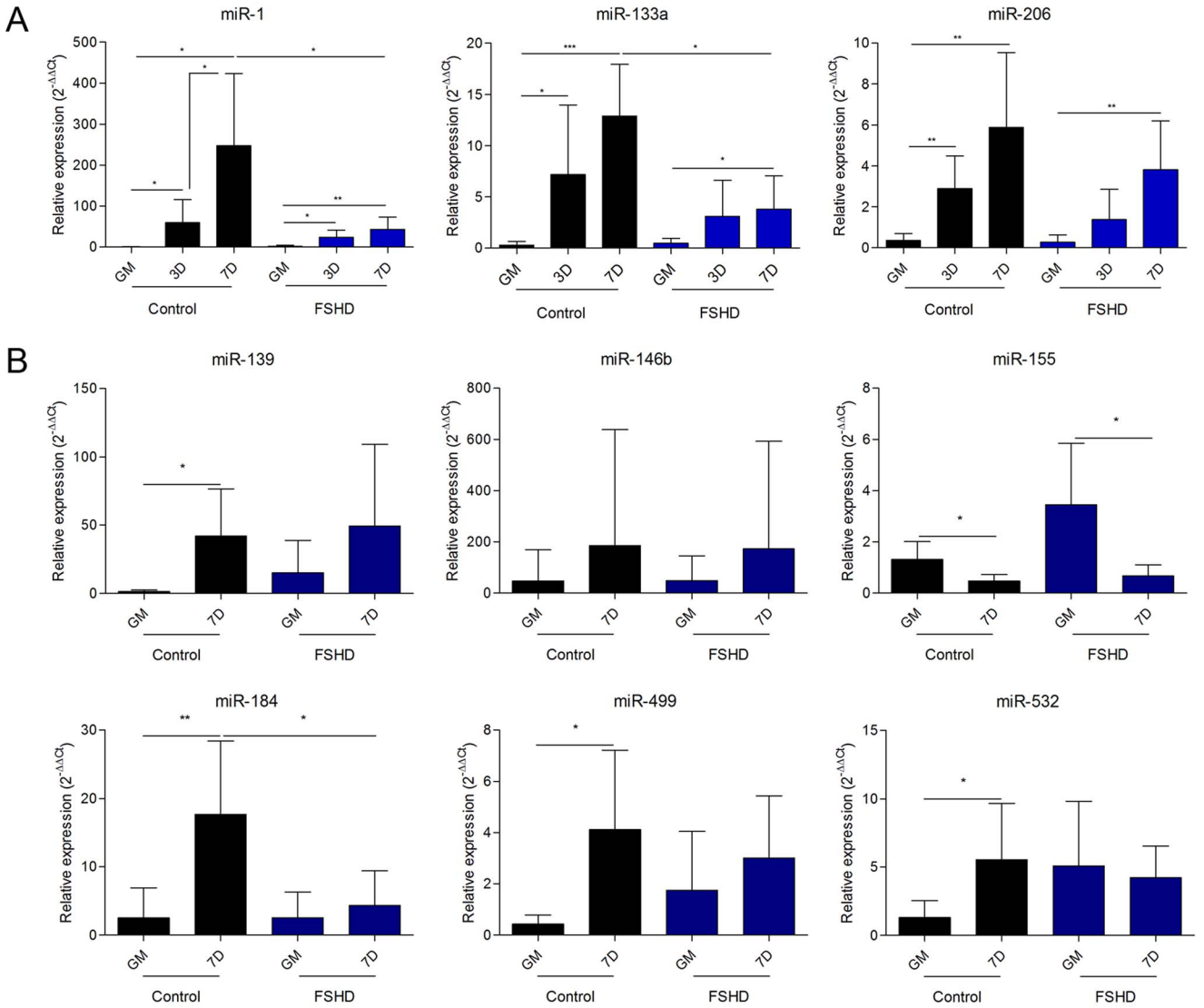


Figure 5. Validation of NGS data. A) qRT-PCR analysis of myomiRs (miR-1, miR-133a and miR-206) during control and FSHD myogenesis at 0, 3 and 7 days of differentiation. B) qRT-PCR analysis of six microRNAs modulated in control and/or FSHD myogenesis. GM: growth medium; 3D: 3 days of differentiation; 7D: 7 days of differentiation. * p-value$<$0.05; ** p-value$<$0.01; *** p-value$<$0.001.
doi:10.1371/journal.pone.0108411.g005

different extent of differentiation between control and pathological samples.

Six additional miRNAs were evaluated for their expression by qRT-PCR (Fig. 5B). As shown in Fig. 5 and summarized in Table 1, the qRT-PCR assays validated about the 70% of the analyzed NGS data. Particularly, the up-regulation of hsa-miR-139 and hsa-miR-146b during, respectively, FSHD and control myogenesis, and the down-regulation of hsa-miR-206 in FSHD vs CN myotubes did not reach the statistical significance showed by NGS results, while maintaining the same trend. On the contrary, the up-regulation of miR-133a in FSHD myogenesis, the down-regulation of hsa-miR-1 and hsa-miR-133a in FSHD vs CN myotubes, and the down-regulation of hsa-miR-155 in FSHD myogenesis already observed in the NGS analysis became significant in the qRT-PCR analysis.

Comparison of FSHD and control myogenesis

The comparison of miRNA modulation between control and FSHD differentiation processes is reported in Fig. 6A, where black and striped bars identify the Fold Change of miRNAs up- and down-regulated, respectively, in control and FSHD myogenesis. From this comparison it was possible to derive that FSHD differentiation lacks the modulation of 29 miRNAs, the majority of which (27/29) was up-regulated in control differentiation (black bars in Fig. 6A, and Fig. 6B); while six miRNAs (4 up- and 2 down-regulated) were modulated only during the FSHD differentiation process (striped bars in Fig. 6A and Fig. 6B). Nine miRNAs showed the same trend in both processes (Fig. 6A and B), but with differences in Fold Change values. Among these, miRNAs pivotal for the myogenic process, such as hsa-miR-1, -206 and -222, were included. Thus, FSHD myogenesis differs from control myogenesis for the complete (35) or partial (9) dysregulation of a total of 44 miRNAs.

Table 1. qRT-PCR validation of NGS data.

miRNA	Control myogenesis				FSHD myogenesis				FSHD vs control myotubes			
	Deep seq		qRT-PCR		Deep seq.		qRT-PCR		Deep seq		qRT-PCR	
	FC	p-value	FC	p-value	FC	p-value	FC	p-value	FC	p-value	FC	p-value
miR-1	293.3	2E-05	352.2	0.007	12.4	3E-06	25.8	0.007	−2.8	0.510*	−5.7	0.033
miR-133a	64.5	2E-05	44.8	1E-04	40.7	0.335	7.9	0.03	−1.2	0.978*	−3.4	0.007
miR-139	24.7	4E-07	28.9	0.018	5.7	0.021	3.3	0.189*	−7.3	0.002	1.2	0.905*
miR-146b	5.5	0.016	3.9	0.419*	2.2	0.858	3.6	0.460	4	0.510	−1.1	0.955
miR-155	−5.6	0.003	−2.8	0.031	−1.4	0.869*	−5.2	0.019	2.7	0.263	1.4	0.413
miR-184	53.9	9E-11	7.1	0.003	7.4	0.145	1.7	0.493	−5.3	0.002	−4.1	0.035
miR-206	36.1	4E-17	28.7	0.009	12.03	1E-06	11.8	0.008	−2.9	0.002	−3.1	0.07*
miR-499a	122.1	0.005	9.7	0.031	5.9	0.143	1.7	0.359	−8.3	0.338	−1.4	0.525
miR-532	9.3	3E-06	4.3	0.037	3.9	0.131	1.2	0.699	−1.99	0.456	−1.3	0.511

Fold Change and p-value of nine miRNAs derived by deep sequencing and subsequently analyzed by qRT-PCR. Asterisked values refer to miRNAs that did not reach significance, although showing the same trend of variation in both analyses.

doi:10.1371/journal.pone.0108411.t001

Prediction of miRNA target genes

To understand the functional impact of miRNA dysregulation during FSHD myogenesis we used TargetScan prediction software to derive potentially affected targets. In order to improve target prediction accuracy, a common approach is to combine the output of two or more prediction algorithms, however this strategy has been proved inefficient [20]. Therefore, we have used a single algorithm, TargetScan, which uses many parameters to predict target scoring without omitting miRNAs with multiple target sites [22]. Since the binding of a miRNA to the 3′ UTR of its mRNA target predominantly act to decrease target mRNA levels [23] we decide to essentially focalize our attention on mRNA targets showing an opposite expression value compared to the analyzed miRNA. Normally, this approach has been carried out on mRNA expression profile derived by using the same cells from which the miRNA expression profile has been derived [11,20,21]. However, the comparison of mRNA expression profiles derived by myoblast cell lines or biopsies from different FSHD patients and controls clearly evidenced a certain variability in the obtained results [5,9,10,11,24,25]. In addition, mRNA expression differences were also found by analyzing different muscles, such as biceps and deltoids [11]. To reduce sample variability, we filtered the predicted mRNA targets on two chip expression data (GSE26061 [9]; GSE26145 [10]), sharing *in vitro* myogenic differentiation protocol and platform although using primary FSHD and control cell lines different from those analyzed in this work. Functional classes corresponding to the filtered mRNAs were assigned by DAVID Gene Ontology Database (Table 2). As shown in Fig. 6, control myogenesis showed the modulation of 38 miRNAs (4 down- and 34 up-regulated), whereas FSHD myogenesis was characterized by 15 dysregulated miRNAs (4 down- and 11 up-regulated) and the lack of modulation of 29 miRNAs. Applying the rationale described above, we derived a total of 139 and 78 down- and up-modulated mRNAs in control myogenesis (potentially "validated" target, Table S6), and a total of 37 down- and 18 up-regulated transcripts in FSHD myogenesis (potentially "validated" target, Table S7). In control myogenic differentiation, the majority of down-regulated genes belonged to cell cycle (27 entries), DNA metabolic process (17 entries), cytoskeleton organization (11 entries), angiogenesis (8 entries) and signal transduction (19 entries); genes involved in cell adhesion (9 entries), regulation of cell migration (5 entries), muscle development (7 entries), lipid biosynthetic process (6 entries) and response to insulin (4 entries) were found up-regulated (Table 2). Conversely, in FSHD myogenesis genes belonging to muscle development (3 entries) and cell adhesion (5 entries) were down-regulated, whereas those involved in regulation of signal transduction (3 entries) were up-regulated. All the identified biological processes, except the down-regulation of cell adhesion in FSHD samples, showed a significant p-value ranging from 3.4E-10 to 3.2E-02 (see Table 2). It is noteworthy that target genes involved in two important biological processes of myogenesis (i.e. cell cycle and striated muscle development) subjected to miRNA control were, as expected, down- and up-regulated, respectively, in control cells. In FSHD myogenesis, on the contrary, the cell cycle was not down-regulated, and control of striated muscle development was down-regulated. It is important to notice that this analysis did not take into account the different FC showed by the nine miRNAs shared by control and FSHD myogenesis.

Identification of novel miRNAs

To identify novel potential miRNAs involved in human muscle system, the unclassified tags were further processed by mireap (http://sourceforge.net/projects/mireap). We considered only

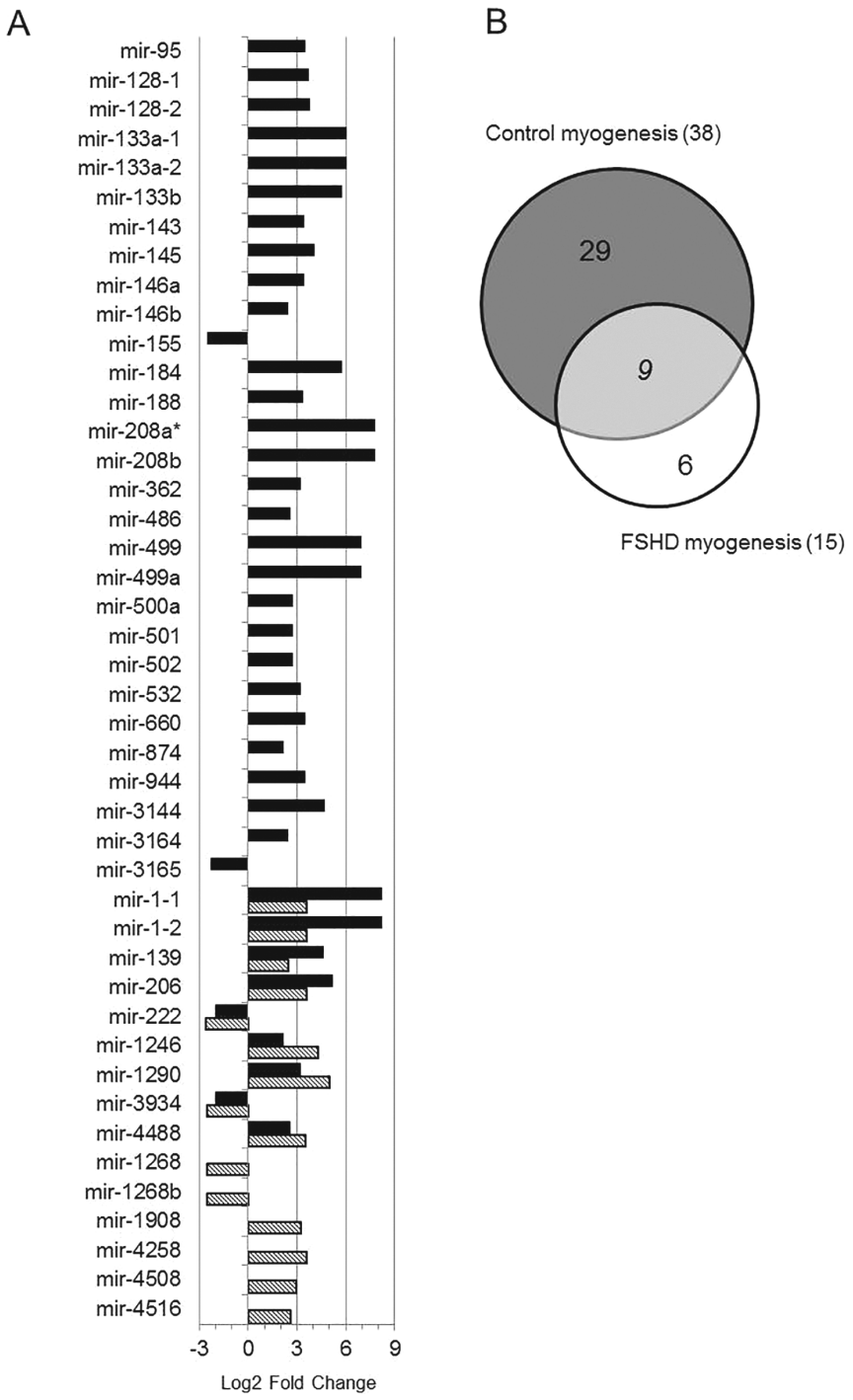


Figure 6. Comparison of miRNA modulation in control and FSHD myogenesis. A) Black and striped bars identify the Fold Change of miRNAs modulated respectively, in control and FSHD myogenesis. Bars on the left and on the right represent, respectively, down- and up-regulated miRNAs. *hsa-mir-208a showed infinite FC value (see Table S3). B) Venn diagram showing the number of miRNAs unique to FSHD (white) or control (grey), and shared (light grey) by FSHD and control differentiation processes.
doi:10.1371/journal.pone.0108411.g006

tags meeting the default parameters, expressed in all experimental groups or preferentially expressed in one or more sample groups (i.e. proliferating vs differentiated cells, or FSHD vs control cells) and with mean read counts per group greater than ten. By using these criteria we identified a total of 13 novel candidate miRNA genes. In Table S8 are reported the main features of these novel miRNA genes, including chromosome location and genomic organization, mfe (minimum free energy), sequence and structure of hairpin precursor, and sequence of 5p or 3p. A summary of these data is reported in Table 3: six miRNAs showed a preferential expression in myoblasts (both in FSHD and control) and four miRNAs seemed to be specific for myotubes. The remaining three miRNAs characterized all the considered groups (both control and FSHD myoblasts and myotubes). Among the 13

Table 2. Functional classification of predicted target genes in control and FSHD myogenesis.

Biological processes	CONTROL MYOGENESIS		FSHD MYOGENESIS	
	Down (p-value)	*Up (p-value)*	*Down (p-value)*	*Up (p-value)*
Cell cycle	+(3.4E-10)			
DNA metabolic process	+(2.8E-06)			
Cytoskeleton organization	+(2.7E-03)			
Angiogenesis	+(1.2E-03)			
Signal transduction	+(8.2E-03)			+(3.2E-02)
Cell migration		+(6.0E-03)		
Cell adhesion		+(1.2E-02)	+(7.4E-02)	
Striated muscle development		+(1.6E-02)	+(5.8E-03)	
Sterol biosynthetic process		+(1.0E-02)		+(5.0E-04)
Response to insulin		+(3.4E-03)		

Functional classification of predicted target genes of modulated miRNAs in control and FSHD myogenesis, filtered on GSE26061 [9] and GSE26145 [10]. For full lists of considered miRNAs and predicted target genes refer to Tables S3, S4 and S6, S7, respectively.
doi:10.1371/journal.pone.0108411.t002

novel miRNAs, two miRNAs (namely hsa-miR-m1-3p and hsa-miR-m13-5p) had already been detected by analyzing prostate and breast tumor cells [26,27] and the mature hsa-mir-m9-3p showed 100% sequence similarity with hsa-mir-574 whose gene however differs in genomic location [28].

Furthermore, it is interesting to note that no sample showed reads generated from the D4Z4 region. This observation, derived either by the analysis of the filtered out repeats or by the re-mapping NGS raw data to specific D4Z4-bearing chromosome regions such as 4q and 10q, suggests that short transcribed sequences from D4Z4 array may have a length greater than 35 nts, the threshold used to build our libraries.

Discussion

The paper reports the first complete analysis of miRNA modulation during *in vitro* differentiation in both control and FSHD-derived myogenic cells. Myogenesis is a complex process that includes proliferation, differentiation, and formation of myotubes and myofibers. These molecular events are regulated by myogenic factors and miRNAs. MiRNAs specifically expressed in skeletal and cardiac muscles are called myomiRs, to imply their important roles in the regulation of muscle development and differentiation [13,19,29]. Recently miRNA dysregulation has been reported in FSHD [9,12,21]. However, due to the approaches used, these studies were limited for the number and type of miRNAs that could be simultaneously investigated; in addition they would not detect miRNAs expressed at low level and excluded discovery of novel miRNAs. Thus, to get the whole pattern of miRNA dysregulation in FSHD we used a next-generation sequencing (NGS) approach. Previous work aimed at identifying biomarkers in FSHD by the transcriptional profiling found muscle-type specific patterns of gene expression [11]. Similarly, DUX4-fl expression was found to vary between myotubes derived from different muscle groups [30]. Therefore, we tailored the experimental protocol to derive FSHD and control miRNA profiles common to different muscles. To this aim, due to inter-individual genetic heterogeneity, from deep sequencing data we considered only miRNA modulation with $FC \geq 4$ ($\log_2 FC \geq 2$) and p-value<0.05. Then the derived miRNA expression in both FSHD and control myogenesis was validated by qRT-PCR in all the available FSHD and control cell lines.

Control myogenesis showed the modulation of 38 miRNAs, the majority of which (34 out 38) were up-regulated. The up-regulated miRNAs included those previously identified as key regulators of both proliferation and differentiation of myogenic cells and for this reason called myomiRs: hsa-miR-1, -133a, -133b and -206 [19,31,32]. The obtained results are in agreement but also expand what is known about miRNA modulation during *in vitro* human myogenic differentiation. Among the modulated miRNAs, 27 were in fact already reported to be involved in muscle differentiation either in human or in mouse cells [20,33]. Conversely, 12 miRNAs, ten up-regulated (hsa-miR-95, -146a, -874, -1246, -1290, -3164, -4488, -208a, -944 and -3144) and two down-regulated (hsa-miR-3934 and -3165), were not previously detected to be differentially expressed during control myogenesis. In comparison with a previous work [20], the reduced number of modulated miRNAs during control myogenesis that we derived is probably due to the choice of higher FC value ($FC \geq 4$). Furthermore, the observed up-regulation of myomiRs strongly supports the validity of used cell lines and differentiation protocol. Interestingly, some up-regulated miRNAs not previously reported to be involved in muscle differentiation, were previously shown to affect cell proliferation by targeting HDAC1 (hsa-miR-874), impairing cytokinesis (hsa-miR-1290), inhibiting cell growth (hsa-miR-95) and regulating differentiation of smooth muscle cells (hsa-miR-146a) [34,35,36,37].

Control myogenesis also showed the possible involvement of some of the novel miRNAs we derived by NGS. In this regard, six out of the 13 identified novel miRNAs (see Table 3) seem to characterize the proliferating status of muscle cells (myoblasts, miR-m2-3p, -m3-3p, -m4-5p, -m7-5p, -m12-3p, and –m13-5p) and one the differentiated status (myotubes, miR-m6-3p). Two, hsa-miR-m1-3p, and hsa-miR-m13-5p, have been previously identified by the NGS approach and validated in breast and prostate cancer cells (identified respectively as hsa-miR-B19 and hsa-novel-miR-08) [26,27]. Further experiments are thus necessary to validate and determine the possible involvement in muscle cells differentiation of these novel miRNAs.

The comparison of control and FSHD myogenesis clearly evidenced a reduced number of modulated miRNAs in FSHD than in control muscle cells, thus suggesting that a complex dysregulation of miRNA expression characterizes the dystrophy. In total, nine miRNAs were shared between the two processes and

Table 3. Novel miRNAs predicted by mireap.

Name	Chromosome location	Mature miRNA sequence	Length	Genomic context	Expression n.samples	Other evidence
hsa-miR-m1-3p	chr11:122022800–122022877	AAAAGGGGGCTGAGGTGGAGG	21	intronic	12/12 (higher expression in myoblasts)	PMID:21346806
hsa-miR-m2-3p	chr11:125757935–125758025	AGGGGCGCGGCCCAGGAGCTCAGA	24	intronic	5/6 myoblasts	no
hsa-miR-m3-3p	chr13:111102986–111103008	AGCTGGGGATGGAAGCTGAAGCC	23	intronic	4/6 myoblasts	no
hsa-miR-m4-5p	chr14:74998697–74998789	CTGCTCTGATGTCTGGCTGAGC	22	intronic	5/6 myoblasts	No
hsa-miR-m5-5p	chr15:41592311–41592403	ATCATTTGGCAGGGGGTAGAGTA	23	intergenic	3/3 FSHD myotubes	No
hsa-miR-m6-3p	chr15:45493361–45493452	TTGTGGAAACAATGGTACGGCA	22	overlaps repeat/tRNA	4/6 myotubes	No
hsa-miR-m7-5p	chr17:8042708–8042779	GAGTTAGCGGGGAGTGATATATT	23	overlaps repeat/tRNA	4/6 myoblasts	No
hsa-miR-m8-3p	chr:6:28918819–28918903	TCGGGCGGGAGTGGTGGCTTTT	22	overlaps repeat/tRNA	12/12	No
hsa-miR-m9-3p	chr8:79679467–79679541	TGAGTGTGTGTGTGTGAGTGTGA	23	intronic	9/12 (all groups)	mature miRNA identical to hsa-mir-574, different genomic location PMID:17604727
hsa-miR-m10-5p	chrX:18651329–18651427	AACTTTGGAATGTGGTAGGGTA	22	intronic	3/3 FSHD myotubes	No
hsa-miR-m11-5p	chrX:40478974–40479066	ATCATTTGGCAGGGGGTAGAGTA	23	intergenic	3/3 FSHD myotubes	No
hsa-miR-m12-3p	chr13:111102941–111103018	AGCTGGGGATGGAAGCTGAAGCC	23	intronic	4/6 myoblasts	No
hsa-miR-m13-5p	chr20:3194751–3194835	CAAAATGATGAGGTACCTGATA	22	Intronic	6/6 myoblasts	PMID:21152091

doi:10.1371/journal.pone.0108411.t003

these included myomiR-1 and -206, with FC values of up-regulation during differentiation lower than those derived for control cells. Moreover, qRT-PCR analysis proved that in control cells the up-regulation of myomiRs is higher than in FSHD ones by a FC ranging from 2.4–5.5× for hsa-miR-206 and 133a to 13× for hsa-miR-1. Furthermore, the kinetic of myomiRs up-regulation during FSHD myogenesis strongly suggests a defect in late stages of the differentiation process. Other differences between control and FSHD differentiation are represented by six miRNAs (i.e. hsa-miR-1268, -1268b, -1908, 4258, -4508- and -4516) not modulated in control cells and that therefore could be considered FSHD-specific, likewise three novel miRNAs (hsa-miR-m5-5p, hsa-miR-m10-5p and hsa-miR-m11-5p) that seem to be specifically expressed in FSHD myotubes (see Table 3). Of interest, hsa-miR-1268 exhibited a significant differential expression during the differentiation of pluripotent human embryonic stem cells into embryoid bodies [38]. In summary, FSHD myogenesis differed from control myogenesis by the loss of modulation of 29 miRNAs (black bars in Fig. 6A, and Fig. 6B) and the acquisition of modulation of six miRNAs, two down-regulated and four up-regulated (striped bars in Fig. 6A, and Fig. 6B). Among the nine miRNAs shared by the two differentiation processes (black and striped double bars in Fig. 6A), the myomiRs showed a significant deficit of expression in late phases of FSHD differentiation. Moreover, the comparison of miRNA expression between control and FSHD myoblasts or myotubes detected 21 dysregulated miRNAs only in myotubes (12 up-regulated and 9 down-regulated). The lack of differentially expressed miRNAs in FSHD myoblasts may be explained both by a high variance of miRNA expression showed by myoblasts and by the high FC used.

Some discrepancies between the data we derived and those recently reported in a similar cellular system [21] require several considerations. First, the methodological approach (NGS against transcriptome profiling), and consequently the cut-off used make the results obtained not comparable; second, both healthy and FSHD myoblast cell lines characterized by a high percentage of DES+ cells were induced to differentiate for three days [20] and for seven days (herein). Lastly, our study design was set up in order to derive a FSHD miRNA profiling possibly shared by different muscle districts and including all the microRNAs present in miRBase (release 20), as well as novel miRNAs. In this regard, it is noteworthy that if we had used the microRNA panel version 1.0 (a TaqMan low density array containing 365 miRNAs) instead of the NGS approach, we would have only detected the modulation of five miRNAs during differentiation of FSHD myoblasts (namely hsa-miR-1-1, 1-2, -206, -222 and -139), instead of the fifteen effectively found (see Table S4). Thus, as previously shown in other cellular systems [26,27,39,40] the deep sequencing approach allowed us to derive a more complete view of miRNA dysregulation in FSHD.

Our data strongly suggest that, in addition to the recently reported up-regulation in proliferating FSHD vs control cells, which however did not result in a complete down-regulation of the corresponding target genes [21], a defect of myomiRs expression also characterize late stages of FSHD differentiation. In fact, the extent of myomiRs expression in FSHD cells after seven days of differentiation was similar to or lower than that found at three days in control cells. Thus, besides the reported up-regulation of myomiRs in FSHD myoblasts due to the early euchromatization of their promoters [21] other defects could be responsible of their down-regulation during late stages of differentiation. In this regard it is possible to hypothesize a defect in FSHD myotubes at the myomiRs transcriptional or post-transcriptional levels, such as a decrease of myogenic differentiation factors controlling their transcription (i.e. *MEF2*) [41], or of factors controlling their processing. The latter hypothesis agrees with previous results showing that FSHD myotubes are characterized by the down-

regulation of a gene (*Dicer1*) controlling the cytoplasmic maturation of pre-miRNAs [10].

Our data allowed us to confirm a few miRNAs previously found dysregulated by independent analysis of ten major skeletal muscle disorders, including FSHD [12,42]. Among the miRNAs we derived to be deregulated during FSHD muscle differentiation, four miRNAs (miR-146a, -146b, -155, -222) were consistently found up-regulated in six or more muscular disorders, including FSHD, whereas miR-501 was found dysregulated in five muscle diseases, but not in FSHD. Furthermore miRNA-486, a muscle enriched miRNA, previously found significantly reduced in patients with DMD [12], was found up-regulated in the present study. Interestingly overexpression of this miRNA in mouse primary myoblasts resulted in increased proliferation and thus in altered cell-cycle kinetics [43].

In order to understand the functional outcome of miRNA dysregulation in FSHD, the derived up- and down-regulated target genes were functionally clustered into biological processes. This approach when applied to healthy muscle differentiation evidenced two important features of myogenesis: cell cycle and muscle development. Effectively, as muscle differentiation proceeds, sustained by the up-regulation of myogenic markers (due to the down-regulation of the corresponding miRNA regulators), the cell proliferation program must slow down due to the up-regulation of miRNA controlling genes involved in this process. An opposite trend of the two biological processes was found to characterize FSHD myogenesis. In fact, down-regulated genes were essentially involved in the regulation of striated muscle tissue development, and no regulation of cell cycle was observed. Thus in FSHD cells miRNA dysregulation affects two important aspects of differentiation leading to a defect in myogenesis. These data are in agreement with previously reported studies [9,10,20].

By the NGS approach we derived that FSHD myogenesis is characterized by a profound dysregulation of miRNA expression showing the involvement of at least 38 known miRNAs, including the myomiRs and possibly three novel miRNAs, but excluding small RNAs previously reported to derive from the D4Z4 array [14]. This and previous works have clearly demonstrated that FSHD cells are characterized by a global dysregulation of mRNA, miRNA and protein expression essentially affecting the myogenic process [9,10,11,21,24,44].

The up-regulation of the last *DUX4* gene in individual showing a reduced numbers ($\leq$8) of D4Z4 repeats at 4q35 combined with a specific molecular signature (4A(159/161/168) DUX4 polyadenylation signal (PA) haplotype) is supposed to underlie FSHD pathophysiology [6]. However, it has been recently reported that 1.3% of healthy individuals carry the same molecular signature and 19% of subjects affected by FSHD do not carry alleles with eight or fewer D4Z4 repeats [8]. Furthermore, a dysregulation of genes involved in myogenesis has been recently observed in FSHD fetuses; importantly, the DUX4-fl pathogenic transcript was detected in both FSHD and control samples [45], as well as in unaffected individuals, but not in all FSHD cases [8]. These data suggest that the molecular basis of FSHD might not be simply based on the overexpression of the single *DUX4* gene, but rather from a cascade of dysregulation mediated by the D4Z4 array contraction. This structural alteration, as previously shown, might induce conformational changes in the 4q35 region itself, and perhaps elsewhere in the human genome [46,47]. Furthermore, in the dysregulation cascade could also play a role lncRNAs, such as DBE-T [7].

Conclusions

By using the NGS approach, we derived the complete pattern of miRNAs regulating *in vitro* control and FSHD myogenesis. In addition to confirming previously reported FSHD-related miRNAs, we identified additional known and novel miRNAs that are differentially expressed between FSHD and control myogenesis and thus potentially contributing to the FSHD pathogenic mechanism. In general, the comparison of control and FSHD myogenesis reveals that the dystrophy is characterized by a complex alteration of miRNA expression, which also includes the significant down-regulation of myomiRs at late stages of differentiation, thus essentially affecting muscle differentiation and development.

Thus, the full range of molecular alteration(s) at the basis of FSHD is not yet fully deciphered and the miRNA profiling we derive could represent a novel molecular signature for FSHD that includes diagnostic biomarkers and possibly therapeutic targets.

Materials and Methods

Cell lines

Primary FSHD and control cell lines were obtained from Myobank-AFM (Institut de Myologie-Groupe Hospitalier Pitié-Salpetrière, Paris) and Boston Biomedical Research Institute (BBRI, Senator Paul D. Wellstone Muscular Dystrophy Cooperative, Research Center for FSHD). Six cell lines derived from biopsies of different healthy and FSHD muscles including vastus, tensor fascia lata, quadriceps femoris (controls) and ilio-psoas and rhomboid (FSHD) (Table S1) were used for deep small RNA sequencing. In addition, to these cell lines, five control and four FSHD cell lines from deltoid and biceps [48] (Table S1) were used to validate deep sequencing data by qRT-PCR. FSHD primary cell lines were derived from biopsies of mild or not affected muscles and showed a D4Z4 array contraction ranging from 5.9 to 28 kb as determined by Southern Blot after EcoRI/BnlI digestion. The results reported below were derived by the analysis of all the cell lines listed in Table S1, comprising nine controls and seven FSHD and thus including also the cells used for NGS. Control and FSHD myoblasts were at low population doubling (from 2 to 7) and highly comparable for the expression of the muscular marker Desmin (96–97%) and the proliferation marker Ki67 (62–65%), as determined by immunofluorescence (Fig. S1). Furthermore, control and FSHD cell lines showed a comparable extent of differentiation as demonstrated by the down-regulation of the proliferation marker Ki67 (by immunofluorescence) and of MYF5 (by qRT-PCR), and by the up-regulation of MYOG (by qRT-PCR), MYOD (by Western blot) and MHC (by qRT-PCR and Western blot), as well as a comparable extent of fusion index (40–45%) (Fig. S1). In addition, FSHD and control myoblasts and myotubes appeared similar when analyzed by immunofluorescence. The cell lines used for NGS originated results in the average comparable to those shown in Fig. S1. Cells were cultured as described in guidelines of BBRI and Cheli et al [9].

Immunofluorescence, image acquisition and analysis

Cell immunofluorescence was performed as described [49], with antibodies specific for Desmin (rAb, Sigma Aldrich), ki67 (rAb, Vector) and sarcomeric myosin MHC (MF20, from Developmental Studies Hybridoma Bank). Appropriate secondary antibodies conjugated with Alexa 488 (green, Cell Signalling) or Alexa 568 (red; Cell Signalling) were used for fluorescence detection, Nuclei were stained with Hoechst Stain Solution (H6024, SIGMA).

Fluorescent images were taken on confocal laser scanning microscope (Zeiss Lsm 01, Biorad mrc 600, Biorad 1024) using 12× magnification. Images showing double or triple fluorescence were separately acquired using appropriate filters, and the different layers were merged with ImageJ software.

For all control and FSHD cell lines used in this study, the quantification of Desmin and ki67 positive cells has been performed on myoblasts and myotubes. Furthermore, for all control and FSHD cell lines, the absolute fusion index has been calculated as the percentage of MHC-positive nuclei over total number of nuclei after 7 days in differentiation medium.

An average value was determined by counting cells (200–300 cells/field) in at least 5 microscopic fields per sample at 12× magnification.

RNA isolation and deep sequencing

Total RNA was isolated with the mirVana miRNA isolation kit (cat.# AM1560, Life Technologies) from myoblast cell lines derived from 3 FSHD patients and 3 control subjects, before and after *in vitro* differentiation. RNA was quantified by Nanodrop spectrophotometer (Thermo Scientific) and its integrity was evaluated on an Experion automated electrophoresis system (Bio-Rad); all samples had a RNA Quality Indicator (RQI) value ≥9.

20 micrograms of total RNA were used for PAGE purification of small RNA molecules shorter than 35 nucleotides, adaptor ligation, and small RNA library preparation. The obtained libraries were sequenced on a HiSeq 2000 platform (Illumina) at BGI, Hong Kong, giving approximately 12 million high quality reads per sample (submitted to SRA database under acc. number SRP034654).

Sequencing data analysis

MicroRNA differential expression analysis was performed using R/Bioconductor, by following the workflow implemented in the oneChannelGUI interface [50,51]. Briefly, adaptor sequences were trimmed from fastq files using a specific perl script, and then sequences were aligned to the reference human miRBase v.20 precursor dataset (www.mirbase.org) using bowtie 1.0.0. Data were filtered for count threshold (>8 reads in 50% of samples analyzed) and pairwise comparisons of differential miRNA expression were performed using DEseq ($\log_2 FC \geq 2$; p-value< 0.05). Hierarchical clustering of differentially expressed miRNA was performed with dChip (version 2010.01; https://sites.google.com/site/dchipsoft/).

Identification of novel miRNAs

After excluding all reads that matched known small RNA classes annotated in miRBase v.20 (known miRNAs) and Rfam (e.g. tRNA, snRNA, snoRNA), putative novel miRNAs were predicted using mireap (http://sourceforge.net/projects/mireap/). The program predicts novel miRNAs from deep sequenced small RNA libraries by taking into consideration miRNA biogenesis, sequencing depth, and structural features (hairpin structure and stability) to improve the sensitivity and specificity of miRNA identification. Among predicted novel miRNAs, we considered as plausible candidates those matching the following criteria: 1) the detection in several samples (at least 2 out of 3 samples of one or more experimental groups); 2) the mature miRNA had sufficient sequence support (at least a mean of 10 reads for each experimental group); 3) the sequence did not match to known miRNAs in miRBase v.20.

Quantitative Real-time PCR

Quantitative RT-PCR (qRT-PCR) analysis was performed on 7900 HT Fast Real-Time PCR System (Applied Biosystems) by TaqMan small RNA Assays to validate the miRNA sequencing data. The miRNA specific probes were from Applied Biosystems. 150 ng RNA was reverse transcribed by TaqMan MicroRNA Reverse Transcription Kit (cat.# 4366596; Applied Biosystems) at 16°C for 30 min, 42°C for 30 min and 85°C for 5 min. Each amplicon was analyzed in duplicate in 96-well plates. TaqMan small RNA Assays reactions were performed following manufacturer's protocol (cat.# 4440048; Applied Biosystems). RNU48 was used for normalization. Thermal cycling conditions for real time PCR were 2 min at 95°C, followed by 40 cycles at 95°C for 10 s and 60°C for 30 s. Results were analyzed using the comparative $2^{-\Delta\Delta Ct}$ method. qRT-PCR experiments for MYF5, MYOG, MHC and GAPDH gene expression analysis were performed as described [9]. The statistical analysis was performed using a two-tail unpaired t-test and the error bars on the graphs are referred to standard deviation. qRT-PCR probes and primers are listed in Table S2.

Derivation of target genes

The putative miRNAs target genes were predicted by TargetScan Human (http://www.targetscan.org/) [52]. The prediction tool is based on different parameters such as complementarity to the seed region, 3′ complementarity, local AU content, position contribution and conservation in different species [22]. Predicted target genes were then filtered on the basis of their inverse correlation with the expression of mRNAs of two different chip analysis on Affymetrix human exon 1.0 ST array [9,10], using a FC≥1.5 and a p-value<0.05.

Pathway and functional annotation analysis

The derived predicted target genes, inversely correlated to the miRNAs expression, were subjected to the analysis of Gene Ontology terms (biological processes) by DAVID (Database for Annotation, Visualization and Integrated Discovery, v6.7) [53,54]. The target genes were mapped to the GO annotation dataset, and the enriched biological processes were extracted using the EASE score, a modified Fisher exact p-value.

Protein extracts and Immunoblot analysis

Cells were collected in RIPA Buffer (50 mM TrisHCl pH = 7,4, 150 mM NaCl, 0,1% SDS, 0,5% Deoxycholate Sodium, 1% NP-40 and protease inhibitor cocktail 1X-cat.# P2714-1BTL, Sigma MO, USA), and centrifuged 15 minutes at 13000 rpm at 4°C to discard cellular debris. Sample preparation and Western blot analyses were performed as described in Pisconti et al [55]. After electrophoresis, polypeptides were electrophoretically transferred to nitrocellulose filters (Thermo Scientific) and antigens revealed by the respective primary Abs and the appropriate secondary Abs, through autoradiography using enhanced chemiluminescence (LiteAblot Plus, cat.# EMP011005, Euroclone). In Western blot analyses, primary antibodies against MHC (MF20, from Developmental Studies Hybridoma Bank), MYOD (cat.# sc-31942, Santa Cruz) and housekeeping gene GAPDH (cat.# G8795; Sigma) were used.

Supporting Information

Figure S1 Characterization of control and FSHD myoblasts cell lines. A) Example of immunostaining experiment on proliferating and differentiated primary myoblasts (control: MX01010MBS; FSHD: MX04309MBS). Images have been taken at confocal laser

scanning microscope at 12× magnification. Nuclei were stained with Hoescht (blue). Panels I–IV show localization of Desmin and Ki67 in proliferating myoblasts; panels I–II show immunostaining experiment using the polyclonal anti-Desmin (red); panels II–IV show immunostaining experiment using the polyclonal anti-Ki67 (red). Panels V–VIII show co-localization of Desmin or Ki67 and MHC on differentiated primary myoblasts: panels V–VI show immunostaining with polyclonal anti-Desmin and monoclonal anti-MHC (Ab-Desmin-red and Ab-MHC-green); panels VII–VIII show immunostaining with polyclonal anti-Ki67 and monoclonal anti-MHC (Ab-Ki67-red and Ab-MHC-green). Scale bar = 100 μm. B) Percentage of Desmin and Ki67 positive cells in myoblasts and myotubes after 7 days of differentiation derived from immunostaining with appropriate antibodies (Ab-Desmin and Ab-Ki67). Results are expressed as mean±SD of independent experiments performed on all cell lines described in Table S1. C) Absolute fusion index was determined at day 7 of differentiation (D7), counting the percentage of MHC- positive nuclei over the total number of nuclei. An average value was determined by counting cells in at least 5 microscopic fields (200–300 cells/field). Results are expressed as mean±SD of independent experiments performed on all cell lines (see Table S1). *$p<0.05$. D) Myogenic differentiation was evaluated by qRT-PCR analysis for MYF5, MYOG, MHC expression. All data points were calculated in triplicate as gene expression relative to endogenous GAPDH expression. Data are represented as the mean±SD of independent experiments performed on all cell lines described in Table S1. GM: growth medium; 7D: seven days of differentiation. *$p<0.05$, **$p<0.01$. E) Example of Western blot analysis with specific antibodies against MYOD and MHC in control and FSHD myoblasts at different time points during myogenic differentiation (GM: growth medium; 3D: three days of differentiation; 7D: seven days of differentiation). GAPDH protein level was used as an internal loading control. Graphs show mean values ±SD obtained from the ratio of densitometric values of protein/GAPDH bands. Data are representative of independent experiments performed on all cell lines described in Table S1. The Western blot in E shows a representative experiment (control: MX01010MBS; FSHD: MX04309MBS). *$p<0.05$, **$p<0.01$.
(TIF)

Figure S2 Scatter plots of the reads of miRNAs modulated in control and FSHD myogenesis. C1: MX01010MBS; C2: MX03609MBS; C3: MX01110MBS, Control cell lines; F1:MX00409MBS; F2: MX03010MBS; F3:MX04309MBS, FSHD cell lines (see Table S1).
(PDF)

Figure S3 Authentication of NGS data by qRT-PCR. qRT-PCR analysis of myomiRs (miR-1, miR-133a and miR-206) during control and FSHD myogenesis at 0 and 7 days of differentiation on the three control and three FSHD cell lines used in the NGS experiment (MX01010MBS; MX03609MBS; MX01110MBS, MX00409MBS; MX03010MBS; MX04309MBS). GM: growth medium; 7D: seven days of differentiation. *$p<0.05$; **$p<0.01$.
(TIF)

Table S1 Primary myoblasts cell lines used in this study. Cell lines have been obtained from Myobank-AFM Istitut de Myologie (Paris)*and Boston Biomedical Research Institute (BBRI, Boston).
(XLSX)

Table S2 Taqman probes and primers used in qRT-PCR experiments.
(XLSX)

Table S3 List of microRNAs modulated in control myogenesis resulting by DEseq analysis.
(XLSX)

Table S4 List of microRNAs modulated in FSHD myogenesis resulting by DEseq analysis.
(XLSX)

Table S5 List of microRNAs modulated in FSHD vs control myotubes resulting by DEseq analysis.
(XLSX)

Table S6 Potentially "validated" targets. List of predicted target genes of miRNAs modulated in control myogenesis, filtered on GSE26061 [9] and GSE26145 [10].
(XLSX)

Table S7 Potentially "validated" targets. List of predicted target genes of miRNAs modulated in FSHD myogenesis, filtered on GSE26061 [9] and GSE26145 [10].
(XLSX)

Table S8 Novel miRNAs predicted by mireap.
(XLSX)

Acknowledgments

We would like to thank Dr. J. Chen for the cell lines used in the study. We are grateful to Cristina D'Orlando for her technical support and to Donatella Barisani for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: RM EG VC SF. Performed the experiments: VC SF GMS RP FR. Analyzed the data: VC SF GMS RP. Contributed reagents/materials/analysis tools: FR. Wrote the paper: RM EG.

References

1. Padberg GW, Lunt PW, Koch M, Fardeau M (1992) Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscul Disord 1:231–4
2. Tupler R, Gabellini D (2004) Molecular basis of facioscapulohumeral muscular dystrophy. Cell Mol Life Sci 61:557–66
3. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, et al. (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nature Genetics 2:26–30
4. Yamanaka G, Goto K, Ishihara T, Oya Y, Miyajima T, et al. (2004) FSHD-like patients without 4q35 deletion. J Neurol Sci 219:89–93
5. Arashiro P, Eisenberg I, Kho AT, Cerqueira AM, Canovas M, et al. (2009) Transcriptional regulation differs in affected facioscapulohumeral muscular dystrophy patients compared to asymptomatic related carriers. Proc Natl Acad Sci USA 106:6220–5
6. Lemmers RJ, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, et al. (2010) A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329:1650–3
7. Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, et al. (2012) A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell 149:819–31
8. Scionti I, Greco F, Ricci G, Govi M, Arashiro P, et al. (2012) Large-scale population analysis challenges the current criteria for the molecular diagnosis of facioscapulohumeral muscular dystrophy. Am J Hum Genet 90:628:35
9. Cheli S, François S, Bodega B, Ferrari F, Tenedini E, et al. (2011) Expression profiling of FSHD-1 and FSHD-2 cells during myogenic differentiation evidences common and distinctive gene dysregulation patterns. PLoS One 6(6)
10. Tsumagari K, Chang SC, Lacey M, Baribault C, Chittur SV, et al. (2011) Gene expression during normal and FSHD myogenesis. BMC Med Genomics 4:67
11. Rahimov F, King OD, Leung DG, Bibat GM, Emerson CP Jr, et al. (2012) Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. Proc Natl Acad Sci USA 109:16234–9

12. Eisenberg I, Eran A, Nishino I, Moggio M, Lamperti C, et al. (2007) Distinctive patterns of microRNA expression in primary muscular disorders. Proc Natl Acad Sci USA 104:17016–21
13. Ge Y, Chen J (2011) MicroRNAs in skeletal myogenesis. Cell Cycle 10:441–8
14. Snider L, Asawachaicharn A, Tyler AE, Geng LN, Petek LM, et al. (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. Hum Mol Genet 18:2414–30
15. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215–33
16. Kloosterman WP, Plasterk RH (2006) The diverse functions of microRNAs in animal development and disease. Dev Cell 11:441–50
17. Mendell JT (2005) MicroRNAs: critical regulators of develompment, cellular physiology and malignancy. Cell Cycle 4:1179–84
18. Zhao Y, Samal E, Srivastava D (2005) Serum response factor regulates a muscle-specif microRNA that targets Hand2 during cardiogenesis. Nature 436:214–20
19. Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, et al. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 38:228–33
20. Dmitriev P, Barat A, Polesskaya A, O'Connell MJ, Robert T, et al. (2013) Simultaneous miRNA and mRNA transcriptome profiling of human myoblasts reveals a novel set of myogenic differentiation-associated miRNAs and their target genes. BMC Genomics 14:265
21. Dmitriev P, Stankevicins L, Ansseau E, Petrov A, Barat A, et al. (2013) Defective regulation of microRNA target genes in myoblasts from facioscapulohumeral dystrophy patients. J Biol Chem 288:34989–5002
22. Witkos TM, Koscianska E, Krzyzosiak WJ (2011) Practical Aspects of microRNA Target Prediction. Current Molecular Medicine 11:93–109
23. Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466:835–40
24. Winokur ST, Chen YW, Masny PS, Martin JH, Ehmsen JT, et al. (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. Hum Mol Genet 12:2895–907
25. Osborne RJ, Welle S, Venance SL, Thornton CA, Tawil R (2007) Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. Neurology 68:569–77
26. Ryu S, Joshi N, McDonnell K, Woo J, Choi H, et al. (2011) Discovery of novel human breast cancer microRNAs from deep sequencing data by analysis of pri-microRNA secondary structures. PLoS One 6:e16403
27. Xu G, Wu J, Zhou L, Chen B, Sun Z, et al. (2010) Characterization of the small RNA transcriptomes of androgen dependent and independent prostate cancer cell line by deep sequencing. PLoS One 5:e15519
28. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129:1401–14
29. Luo W, Nie Q, Zhang X (2013) MicroRNAs involved in skeletal muscle differentiation. J Genet Genomics 40:107–16
30. Ferreboeuf M, Mariot V, Bessières B, Vasiljevic A, Attié-Bitach T, et al. (2014) DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. Hum Mol Genet 23:171–81
31. McCarthy JJ (2008) MicroRNA-206: the skeletal muscle-specific myomiR. Biochim Biophys Acta 1779:682–91
32. Townley-Tilson WH, Callis TE, Wang D (2010) MicroRNAs 1, 133, and 206: critical factors of skeletal and cardiac muscle development, function, and disease. Int J Biochem Cell Biol 42:1252–5
33. Callis TE, Deng Z, Chen JF, Wang DZ (2008) Muscling through the microRNA world. Experimental biology and medicine 233:131–138
34. Wu J, Ji X, Zhu L, Jiang Q, Wen Z, et al. (2013) Up-regulation of microRNA-1290 impairs cytokinesis and affects the reprogramming of colon cancer cells. Cancer Lett 329:155–63
35. Nohata N, Hanazawa T, Kinoshita T, Inamine A, Kikkawa N, et al. (2013) Tumor-suppressive microRNA-874 contributes to cell proliferation through targeting of histone deacetylase 1 in head and neck squamous cell carcinoma. Br J Cancer 108:1648–58
36. Cheng AM, Byrom MW, Shelton J, Ford LP (2005) Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. Nucleic Acids Res 33:1290–7
37. Dong S, Xiong W, Yuan J, Li J, Liu J, et al. (2013) MiRNA-146a regulates the maturation and differentiation of vascular smooth muscle cells by targeting NF-kB expression. Mol Med Rep 8:407–12
38. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, et al. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18:610:21
39. Yang Q, Hua J, Wang L, Xu B, Zhang H, et al. (2013) MicroRNA and piRNA profiles in normal human testis detected by next generation sequencing. PLoS One 8:e66809
40. Schee K, Lorenz S, Worren MM, Günther CC, Holden M, et al. (2013) Deep sequencing the microRNA transcriptome in colorectal cancer. PLoS One 8:e66165
41. Rao PK, Kumar RM, Farkhondeh M, Baskerville S, Lodish HF (2006) Myogenic factors that regulate expression of muscle-specific microRNAs. Proc Natl Acad Sci USA 103:8721–6
42. Goljanek-Whysall K, Sweetman D, Münsterberg AE (2012) microRNAs in skeletal muscle differentiation and disease. Clin Sci (Lond) 123:611–25
43. Alexander MS, Casar JC, Motohashi N, Myers JA, Eisenberg I, et al. (2011) Regulation of DMD pathology by an ankyrin-encoded miRNA. Skelet Muscle 1:27.
44. Celegato B, Capitanio D, Pescatori M, Romualdi C, Pacchioni B, et al. (2006) Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. Proteomics 6:5303–21
45. Broucqsault N, Morere J, Gaillard MC, Dumonceaux J, Torrents J, et al. (2013) Dysregulation of 4q35- and muscle-specific genes in fetuses with a short D4Z4 array linked to facio-scapulo-humeral dystrophy. Hum Mol Genet 22:4206–14
46. Ottaviani A, Schluth-Bolard C, Rival-Gervier S, Boussouar A, Rondier D, et al. (2009) Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. EMBO J 28:2428–36
47. Bodega B, Ramirez GD, Grasser F, Cheli S, Brunelli S, et al. (2009) Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. BMC Biol 7:41
48. Homma S, Chen JC, Rahimov F, Beermann ML, Hanger K, et al. (2012) A unique library of myogenic cells from facioscapulohumeral muscular dystrophy subjects and unaffected relatives: family disease and cell function. Eur J Hum Genet 20:404–10
49. Brunelli S, Tagliafico E, De Angelis FG, Tonlorenzi R, Baesso S, et al. (2004) Msx2 and necdin combined activities are required for smooth muscle differentiation in mesoangioblast stem cells. Circ Res 94: 1571–8
50. Sanges R, Cordero F, Calogero RA (2007) oneChannelGUI: a graphical interface to Bioconductor tools, designed for life scientists who are not familiar with R language. Bioinformatics 23:3406–8
51. Cordero F, Beccuti M, Arigoni M, Donatelli S, Calogero RA (2012) Optimizing a massive parallel sequencing workflow for quantitative miRNA expression analysis. PLoS One 7:e31630
52. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115:787–798
53. da Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57
54. da Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13
55. Pisconti A, Brunelli S, Di Padova M, De Palma C, Deponti D, et al. (2006) Follistatin induction by nitric oxide through cyclic GMP: a tightly regulated signaling pathway that controls myoblast fusion. J Cell Biol 172:233–44

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

Synthesis, Characterization, and Optimization for in Vivo Delivery of a Nonselective Isopeptidase Inhibitor as New Antineoplastic Agent

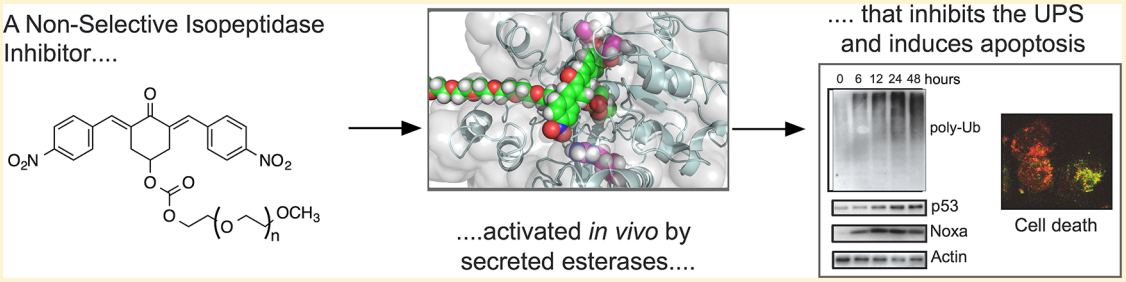
Ulma Cersosimo,[†,∥] Andrea Sgorbissa,[‡,∥] Carmen Foti,[‡] Sara Drioli,[†] Rosario Angelica,[‡] Andrea Tomasella,[‡] Raffaella Picco,[‡] Marta Stefania Semrau,[§] Paola Storici,[§] Fabio Benedetti,[†] Federico Berti,*[,†] and Claudio Brancolini*[,‡]

[†]Dipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Trieste, Via Giorgieri 1, 34127 Trieste, Italy

[‡]Dipartimento di Scienze Mediche e Biologiche, Università degli Studi di Udine, P.le Kolbe 4, 33100 Udine, Italy

[§]Structural Biology Laboratory, Elettra-Sincrotrone Trieste S.C.p.A., Area Science Park, 34149 Basovizza, Trieste, Italy

Supporting Information



ABSTRACT: Bis-arylidenecycloalkanones structurally related to the nonselective isopeptidase inhibitor **G5** were synthesized and tested for cytotoxic activity against glioblastoma cells. Cytotoxicities correlate well with Hammett σ constants for substituted arylidene groups, confirming the proposed inhibition mechanism. A new inhibitor (**2c**) based on the 4-hydroxycyclohexanone scaffold, which favors apoptosis over necrosis, was selected for further development. **2c** inhibited representative deubiquitinases with micromolar IC_{50}, and its proapoptotic activity was studied on several cancer cell lines. Inhibitor **2c** was conjugated to PEG via dicarbamate and diester linkers. While the dicarbamate was inactive, the diester (2cPE) behaves like a prodrug and is converted into the active species **2c** by secreted esterase activities. Finally, 2cPE was also tested in vivo on A549 lung carcinoma xenografts generated in mice. Intravenous treatment with 2cPE led to a significant reduction in primary tumor growth, without appreciable toxicity to mice.

INTRODUCTION

Protein modification by the addition of the 8 kDa ubiquitin or Ub-like (Ubl) proteins is a well-known and widespread post-translation modification not limited to influencing protein destruction but also their subcellular localization or the assembling into multiprotein complexes.[1−3] Ub and Ubl peptides are generally ligated to proteins by the sequential action of three enzymes: a ubiquitin-activating enzyme (E1), a Ub-carrier enzyme (E2), and a Ub-protein ligase (E3). E3 enzymes show specificity in substrate choice and represent critical players in several signaling pathways.[3,4]

Ub and Ubl linkages are reversible, and their cleavage entails the involvement of a large family of enzymes, known as isopeptidases. Although isopeptidases can be viewed as E3-ligase antagonists, additional functions, such as maturation of Ub or Ubl peptides, are under their supervision. The isopeptidase family includes deubiquitinating enzymes (DUBs), which in principle should be specifically devoted to the rupture of Ub linkages, and other proteolytic enzymes, which target additional Ubl proteins.[5,6]

The discovery of bortezomib[7] and its approval for the treatment of relapsed multiple myeloma and mantle cell lymphoma has opened the field to new inhibitors targeting, more specifically, critical enzymes of the ubiquitin−proteasome system (UPS) and showing less adverse side effects.[8,9] In this scenario E3 ligases and DUBs/isopeptidases have gained increasing attention as targets for drug development. These enzymes represent good candidates to influence the function of proteins controlling the transformed phenotype or to induce cellular stresses, which can kill cancer cells.[9−12]

In the past decade several reports have described the discovery, synthesis, and characterization of isopeptidase inhibitors.[13−23] On the basis of the specific target selectivity, they can be divided into two classes: selective inhibitors, acting on a specific enzyme or on a limited number of enzymes, and nonselective isopeptidase inhibitors (N-SIIs), which in principle can affect the activity of several isopeptidases. The two classes offer different and complementary advantages for the develop-

Received: September 1, 2014

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

ment of new anticancer treatments. The first class guarantees advantages in terms of selectivity when growth/survival of a specific tumor depends on a specific isopeptidase. On the other hand, the second class, affecting more enzymes and multiple pathways, may offer advantages in terms of effectiveness on different tumors. Hence, further studies to improve the antineoplastic activities of inhibitors of both classes are of primary importance.

A subclass of N-SIIs includes molecules characterized by the presence of a cross-conjugated α,β-unsaturated dienone with two sterically accessible electrophilic β-carbons[13] that can act as Michael acceptors to target nucleophiles, like the catalytic cysteine of several isopeptidases.[13−17] Recently, a molecule of this family, RA190, has been engineered to exert specificity against proteasome ubiquitin receptor RPN13 (Figure 1).[23]

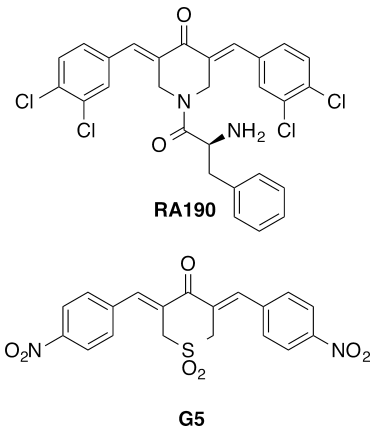


Figure 1. Cross-conjugated dienone N-SIIs RA190 and **G5**.

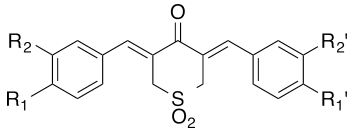
Optimization of these inhibitors in terms of in vivo efficacy is a fundamental step toward their possible use in the clinic. In this work, starting from **G5**, a N-SII previously identified by us (Figure 1),[14] we performed structure−activity studies that lend support to the proposed mechanism of inhibition and allowed us to develop a **G5** derivative optimized for in vivo antineoplastic activity.

RESULTS AND DISCUSSION

Structure−Activity Studies. The effect on tumor cell survival of two series of dienones was investigated in glioblastoma U87MG cells, which in response to **G5** enter both apoptosis and necrosis.[24] The first series (Table 1) is based on the bis(arylidene)tetrahydrothiapyran-4-one 1,1-dioxide scaffold of **G5** with variations in the substituents on the aromatic rings. In the second series (Table 2), the sulfone group of the **G5** scaffold is replaced by other groups, while the 4-nitro substituents on the aromatic rings are fixed.

With the exception of the fluoro derivative **1f**, symmetrical compounds **1a−i** and **G5** were obtained by the acid catalyzed Knoevenagel condensation of tetrahydrothiapyran-4-one 1,1-dioxide (Scheme 1, X = SO_2) with aromatic aldehydes.[25,26] An alternative approach was adopted for the synthesis of **1f** (Scheme 1), consisting of the condensation between the corresponding aldehyde and tetrahydrothiapyran-4-one followed by mCPBA oxidation of the resulting sulfide **3**. The oxidation of the *p*-fluoro derivative **3** to the desired sulfone **1f** was accompanied by the formation of the Baeyer−Villiger

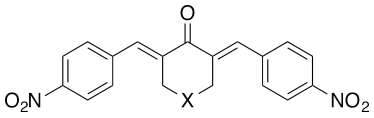
Table 1. Cytotoxicity of Bis(arylidene)tetrahydrothiapyran-4-one 1,1-Dioxides against U87MG Glioblastoma Cells



compd	R_1	R_1'	R_2	R_2'	IC_{50} [μM][a,b]	sd[b]
1a	H	H	H	H	5.11	1.67
1b	CH_3	CH_3	H	H	4.74	0.21
1c	OH	OH	H	H	6.73	0.15
1d	OCH_3	OCH_3	H	H	3.57	0.72
1e	OPh	OPh	H	H	2.94	0.57
1f	F	F	H	H	2.21	0.23
1g	CN	CN	H	H	1.35	0.15
1h	OCH_3	OCH_3	NO_2	NO_2	1.14	0.18
1i	H	H	NO_2	NO_2	0.76	0.07
1j	NO_2	OH	H	H	2.17	0.27
1k	NO_2	OCH_3	H	H	2.09	0.42
1l	NO_2	CH_3	H	H	1.39	0.34
G5	NO_2	NO_2	H	H	0.77	0.21

[a]From cell viabilities measured with a resazurin assay, 48 h after treatments. [b]Experiments were carried out in triplicate and are presented as mean values and standard deviations (sd).

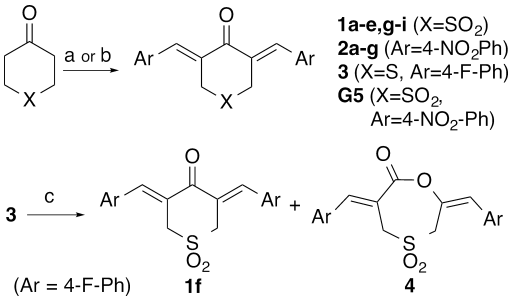
Table 2. Cytotoxicity of Nitrobenzylidene Dienones against U87MG Glioblastoma Cells



compd	X	IC_{50} [μM][a,b]	sd[b]
2a	CH_2	27.5	2.0
2b	$CH(OCH_2CH_2O)$	15.33	0.71
2c	CHOH	4.62	1.96
2d	$CHCOOC_2H_5$	2.47	0.25
2e	NH	7.15	2.90
2f	O	50.1	14.3
2g	S	11.6	2.01
2h	SO	1.79	0.81
G5	SO_2	0.77	0.21

[a]From cell viabilities measured with a resazurin assay, 48 h after treatments. [b]Experiments were carried out in triplicate and are presented as mean values and standard deviations (sd).

Scheme 1. Synthesis of Symmetrical Derivatives of Series 1 and 2[a]



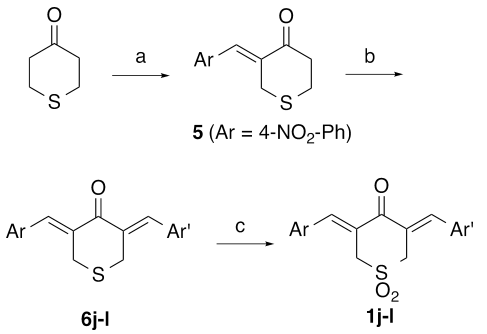
[a]Reagents and conditions: (a) ArCHO, EtOH, aq HCl, 15−78%; (b) ArCHO, EtOH, $Ba(OH)_2$, 52% (for **2b**); (c) mCPBA, CH_2Cl_2, 41−63%.

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

byproduct **4** (Scheme 1). The two products, in ratio 3:1, were easily separated by crystallization from acetone/hexane.

Unsymmetrical compounds **1j–l** were obtained as shown in Scheme 2. Tetrahydrothiapyran-4-one was initially converted

Scheme 2. Synthesis of Unsymmetrical Derivatives of Series 1[a]



[a]Reagents and conditions: (a) (i) ArCHO, NaOH, $MgSO_4$, H_2O, (ii) HCl, EtOH 56%; (b) ArCHO, EtOH, aq HCl, 44–95%; (c) mCPBA, CH_2Cl_2, 21–90%.

into the monoalkylidene derivative **5** by an aldol reaction with 4-nitrobenzaldehyde followed by dehydration of the aldol. Acid-catalyzed Knoevenagel reaction of **5** with the appropriate aromatic aldehyde was then followed by mCPBA oxidation of the sulfides **6j–l**.

All compounds of series **2** (Table 2) are symmetrical and, with the exception of **2h**, were obtained by the Knoevenagel condensation of the appropriate ketone and 4-nitrobenzaldehyde (Scheme 1). For compound **2b**, containing an acid-labile acetal group, the condensation was carried out in basic conditions.[27] Sulfoxide **2h** (Table 2) was obtained by mCPBA oxidation of the corresponding sulfide **2g**. Even under controlled conditions (1 equiv of mCPBA, 0 °C) the reaction gave a 1:1 mixture of sulfoxide **2h** and sulfone **G5**, but the desired product **2h** was readily purified by flash chromatography on silica gel.

^{1}H and ^{13}C NMR spectroscopy showed that all compounds of series **1** and **2** were single stereoisomers with the aryl groups trans with respect to the carbonyl.

Data in Table 1 indicate that all the sulfones of series **1** are active against U87MG glioblastoma cells with IC_{50} values in the range 0.8–7 μM, the dinitroderivatives **G5** and **1i** being the most active compounds of the series. Variations of hydrophobicity (log *P* values) and solvent accessible area are small in this series, reflecting the limited structural changes, and no general correlation is observed between these parameters and citotoxicity. However, a clear correlation can be observed between cytotoxicity and the electronic effect of the substituents on the aromatic ring. A plot of the logarithm of the observed IC_{50} against the sum of the Hammett σ constants for the substituents on the aromatic rings[28] (Figure 2) reveals that cytotoxicity depends on the electron-withdrawing ability of the substituents, suggesting that the cytotoxic activity is directly related to the electrophilicity of the enone β-carbon. This is consistent with the observation that alcohol **7**, obtained by $NaBH_4$ reduction of **G5**, is devoid of cytotoxic activity and strongly suggests that cytotoxicity may indeed result from the alkylation of cysteine residues present in the catalytic site of isopeptidases by the dienone moiety.[13,29] Similar correlations have been observed also for cytotoxic *N*-acyl-bis(arylidene)-4-

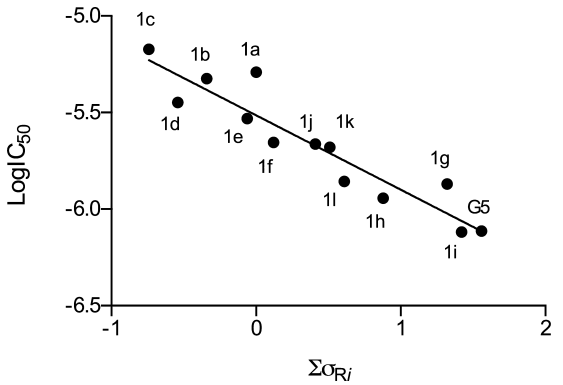


Figure 2. Linear correlation between U87MG glioblastoma cells cytotoxicity (log IC_{50}) and Hammett σ constants for compounds of series **1** (slope of -0.39 ± 0.04; $r^2 = 0.8808$). $\sum\sigma_{Ri}$ is the sum of the σ_p and/or σ_m for all the substituents on both rings.

piperidones,[30] and arylidenecyclohexanones,[31,32] which have been postulated to selectively react as Michael acceptors with cellular thiols.[29]

Finally, the contribution of the X group was investigated in a series of dinitro derivatives **2** (Table 2). While all compounds in this series are cytotoxic, the presence of a polar, electron withdrawing group in the six-membered ring appears in general to be beneficial to achieve strong activity.

Assessment of the Dienone-Induced Apoptotic and Necrotic Responses. In cells resistant to apoptosis, such as mouse fibroblasts defective for Bax and Bak (Bax/Bak DKO) or glioblastoma U87MG cells, **G5** can also activate a caspase-independent death.[24,33] Although caspase-independent death can engage different mechanisms, previous studies have excluded the involvement of autophagy and have demonstrated that **G5** triggers a necrotic cell death.[33,34] Hence, we decided to study, on a small set of dienones and using **G5** as a reference, whether variations in the structure affect the pronecrotic and proapoptotic activities. To this purpose, we selected derivatives **1b**, **2c**, and **2h** and also used T98G glioblastoma cells, which preferentially die by apoptosis in response to **G5** (Figure 3).[24,34]

As illustrated in Figure 3A,B, the sulfoxide **2h** behaves similarly to **G5** while compounds **1b** and **2c** show a higher propensity to trigger apoptosis compared to **G5** and **2h**. In fact, unlike the latter compounds, **1b** and **2c** are more effective in killing T98G than U87MG cells, as proved by trypan blue staining and caspase activities. Cell death in T98G cells, induced by **1b** and **2c**, is characterized by a robust caspase engagement. On the other side in U87MG cells, during cell death induced by **G5** and **2h**, caspases are much less involved, thus testifying to the existence of two different types of cell death. The differential induction of necrosis and apoptosis in the two cell lines was confirmed, for **2c**, by using the caspase inhibitor zVAD-fmk (Figure 3C), which was effective in counteracting cell death only in T98G cells; similar results were obtained with the other compounds (Supporting Information Figure S1).

In summary, the simultaneous presence of the sulfone/sulfoxide and nitro groups in **2h** and **G5** appears to promote a necrotic response; replacement of either group, as in **2c** and **1b**, lowers the necrotic activity of the N-SII and favors the induction of apoptosis.

The ability of the compound **2c** to kill preferentially through apoptosis suggests that it could represent a better choice for in

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX–XXX

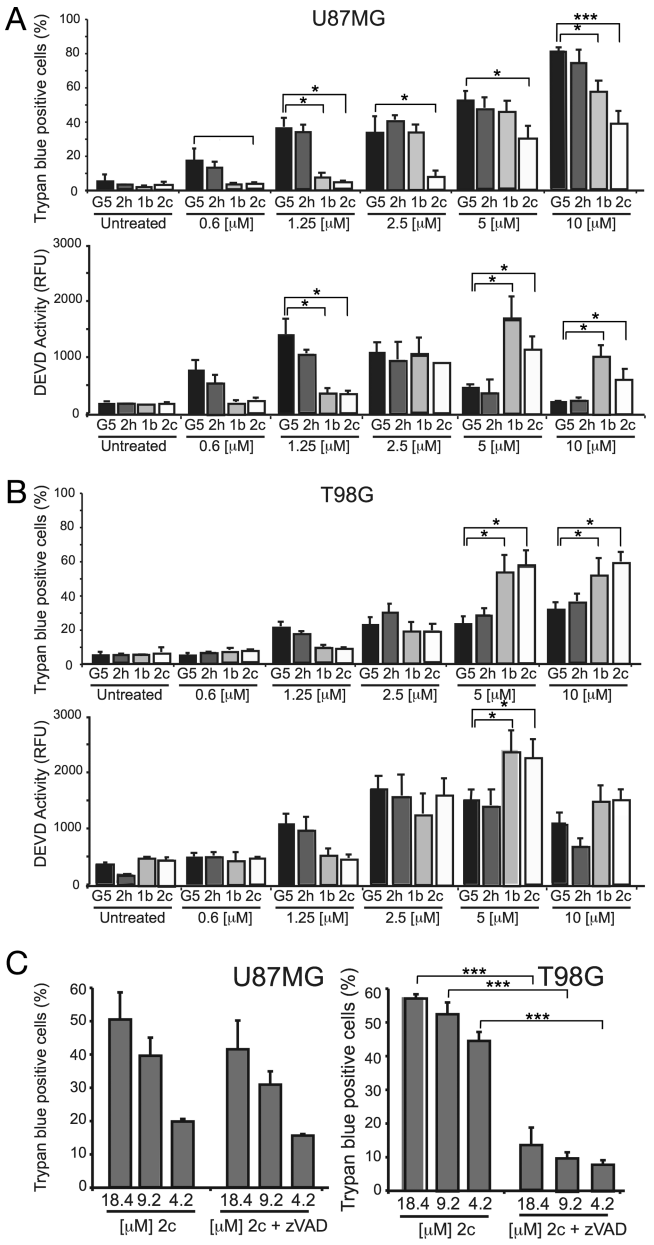


Figure 3. Apoptotic and necrotic responses induced by different dienones in U87MG (A) and T98G cells (B). Cells were treated with the indicated concentrations of the different compounds for 24 h. Upper: cell death as measured by percent of trypan-blue-positive cells. Lower: caspase activity measured with rhodamine 110 bis-(*N*-*Z*-L-aspartyl-L-glutamyl-L-valyl-aspartic acid amide) fluorogenic substrate (relative fluorescence units) (black, **G5**; gray, **2h**; light gray, **1b**; white, **2c**). (C) Effect of caspase inhibitor zVAD-fmk on cell death induced by **2c** (percent of trypan-blue-positive cells).

vivo applications, with respect to **G5**. Moreover, compound **2c**, possessing a reactive OH group, appears to be a good compromise between cytotoxicity and the possibility to introduce modifications on the core to improve the inhibitor's druglike properties. For these reasons **2c** was selected for further development.

Deubiquitinase Inhibition by G5 and 2c. We evaluated the inhibitory profile of the two compounds against purified isopeptidases using ubiquitin−AMC as a substrate. Previous studies demonstrated that dienones are broad inhibitors, capable of inhibiting deubiquitinases but also deSUMOylase activities.[35] Thus, we investigated **G5** and **2c** activities against ubiquitin carboxy-terminal hydrolases UCHL1, UCHL5 and the ubiquitin-specific protease USP2.

UCHL5/UCH37 is a proteasome-associated UCH, which, in conjunction with POH1/PSMD14 and USP14, provides the proteasome the deubiquitinase activity.[3] Unlike UCHL1 and UCHL3, UCHL5 can also process poly-ubiquitin chains, and previous studies identified UCHL5 as a target of dienone-based N-SIIs.[16]

Both **G5** and **2c** weakly inhibit UCHL1 (Figure 4A) but are much less effective than other small-molecule inhibitors.[36] Conversely, both compounds exhibit a more pronounced inhibitory activity against UCHL5. **G5** is considerably more potent than **2c**, as shown by the IC_{50} values of 13.58 μM and 42.99 μM, respectively. Thus, UCHL5 inhibition may play an important role in eliciting the accumulation of poly-ubiquitinated proteins in cells treated with these compounds.

When USP2 activity was analyzed, again **G5** showed some inhibitory activity (IC_{50} = 32.56 μM) whereas **2c** was almost inactive. This IC_{50} value is in agreement with previous studies where a compound related to **G5** was used.[14,35]

Finally, we also tested **G5** and **2c** inhibitory potency against the isopeptidase USP18, which processes the Ub-like ISG15 protein and is a key regulator of the interferon response.[37] Both compounds inhibit USP18 activity, and **2c** appears to be slightly more active than **G5**, with IC_{50} of 21.20 and 33,23 μM, respectively. The ability of **2c** to inhibit USP18 was also verified in cells by coexpressing USP18 and a fusion between ISG15 and GFP.[38] When ISG15-GFP was coexpressed with USP18, accumulation of free GFP is predominant (Figure 4B). By contrast, in the presence of increasing concentrations of **G5** or **2c**, accumulation of the uncleaved ISG15-GFP chimera can be appreciated. In the same assay, bortezomib failed in influencing USP18-dependent cleavage of ISG15-GFP.

Analysis of 2c Proapoptotic Activity. Next, to evaluate whether **2c** shows a strong and broad spectrum of antiproliferative activities, we measured its IC_{50} against several other cancer cell lines. Data in Table 3 indicate that **2c** antiproliferative activity is indeed wide and mainly in the low micromolar range, as already observed for U87MG and T98G cell lines.

We then evaluated in more detail the effect of **2c** on the lung cancer cell line A549 that was later chosen for in vivo experimentation (vide infra). Initially, we compared the apoptotic response elicited by **2c**, **G5**, and bortezomib to identify the concentration ranges of the three inhibitors that induce comparable extents of cell death (Figure 5A). We then used these concentrations to compare the effects of the three inhibitors on the activity of the UPS and on the accumulation of the BH3-only protein Noxa, a sensor of endoplasmic reticulum stress and key element of the death pathway elicited by UPS inhibitors.[14] Immunoblot analysis (Figure 5B) demonstrated that at low concentrations bortezomib is more effective in eliciting accumulation of poly-ubiquitinated proteins and in stabilizing p53 with respect to **2c** or **G5**. Interestingly, when cell death is induced with similar intensity, accumulation of Noxa is comparable in bortezomib and **2c** treatments, thus suggesting that ER stress is similarly evoked by the two compounds. By contrast, high doses of **G5** less efficiently promote Noxa up-regulation, consistent with the possible engagement of a necrotic response (Figure 5B).

Synthesis and Activity of a 2c Derivative Optimized for in Vivo Studies. **2c**, like many anticancer compounds, is

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

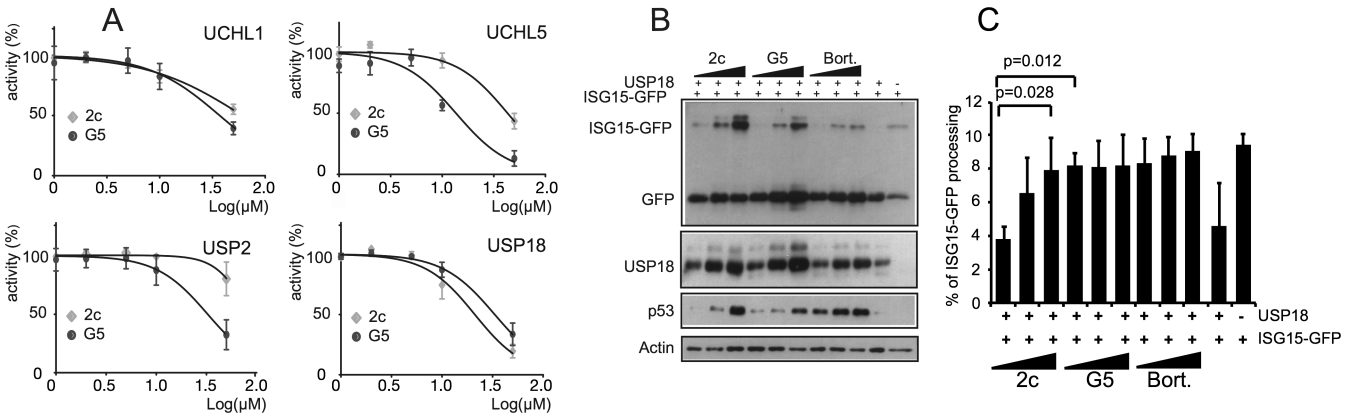


Figure 4. (A) Inhibition of ubiquitin-specific hydrolases UCHL1, UCHL5, USP2, and USP18 by **G5** and **2c**. (B) Immunoblot analysis showing inhibition of USP18 in A549 cells coexpressing USP18 and the ISG15-GFP chimera. p53 levels were used to monitor the inhibition of the UPS. Inhibitor concentrations were the following: **2c** (2.5, 5, 10 μM); **G5** (1, 2.5, 5 μM); bortezomib (25, 50, 100 nM). Cellular lysates were generated after 24 h of treatment. Under these concentrations comparable % of death in A549 lung cancer cells can be observed 48 h later. (C) Quantitative densitometric analysis of immunoblots. Data are presented as the mean of two experiments.

Table 3. Antiproliferative Activity of 2c and PEGylated Derivative 2cPC (11) in Different Cell Lines

		IC_{50} (μM)[a]	
cell line		**2c**	**11**
U87MG	glioblastoma	4.62	6.9
A375	melanoma	6.1	5.7
A549	lung cancer	16	22
HT29	colorectal adenocarcinoma	7.6	>100
Hep3B	hepatocellular carcinoma	8.5	
Mia PaCa-2	pancreatic cancer	35.2	

[a]From cell viabilities measured with a resazurin assay, 48 h after treatments.

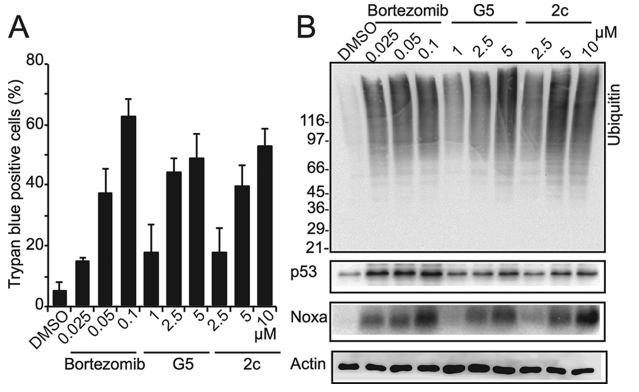


Figure 5. (A) Inhibitor concentrations eliciting comparable death percentage in A549 lung cancer cells after 48 h of treatment. (B) Immunoblot analysis of extracts from A549 cells treated as in (A) and monitored for UPS inhibition, (poly-ubiquitin and p53 accumulation) and induction of ER stress (Noxa induction) after 24 h of treatment.

poorly soluble in aqueous solutions. A common strategy for improving drug solubility and bioavailability consists in the conjugation of the compound with polyethylene glycol (PEG). Despite the intrinsic limitation of a low drug/carrier mass ratio when used with low molecular weight drugs, polyethylene glycol offers some advantages as a water-soluble carrier for anticancer compounds.[39−41] We thus synthesized two soluble derivatives by conjugating **2c** to monomethoxy PEG (5000 Da) via the OH group and suitable linkers (Scheme 3).

For the synthesis of the first conjugate (**9**), with a dicarbamate linker (2cPC, 2c-PEG-carbamate), **2c** was converted into the mixed carbonate **8** and then coupled with mono-PEGylated 1,3-diaminopropane (41% overall yield). When tested on the panel of cancer cell lines, however, this conjugate was inactive, incapable of triggering cell death and reducing cell proliferation (data not shown). We thus synthesized a second derivative (**11**), in 57% overall yield, by conjugating **2c** with PEG through a succinate linker (Scheme 3). We reasoned that this conjugate (2cPE, 2c-PEG-ester), containing a more reactive diester linker, could act as a prodrug, releasing the active species **2c** upon the action of cellular hydrolases,[42] and indeed, we found that **11** was active on three of the four cell lines tested (Table 3).

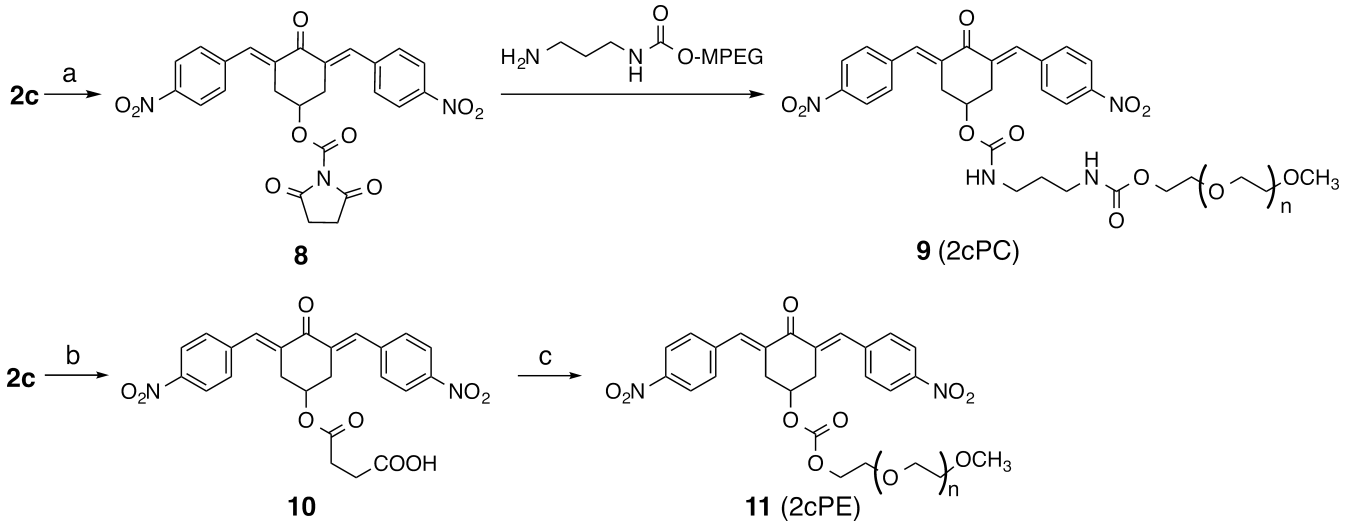
In melanoma A375 and glioblastoma U87MG cells the antiproliferative activity of 2cPE (**11**) is identical to that of the parent molecule **2c** (Table 3). By contrast, in A549 and, even more so, in HT29, respectively lung and colon cancer derived cell lines, the PEGylated molecule is less effective in suppressing proliferation. To verify this observation, we compared the response of A549 cells to 2cPE (**11**) and **2c** by scoring cell death with trypan-blue assay. Dose dependent studies (Figure 6A) and time course analysis (Figure 6C) confirm that 2cPE is less effective than **2c** in triggering cell death in A549 cells, an observation further corroborated by the analysis of poly-ubiquitin accumulation, p53 stabilization, and Noxa induction (Supporting Information Figure S2). Similar results were obtained in HT29 cells (Supporting Information Figure S3).

On the opposite, in A375 cells, which show similar sensitivity to **2c** and 2cPE in resazurin assay (Table 3), cell death (Figure 6B), accumulation of poly-ubiquitinated proteins, p53 stabilization, and Noxa up-regulation (Supporting Information Figure S2) confirm that 2cPE behaves like the active compound **2c**. In summary, in A375 cells **2c** and the PEGylated derivative 2cPE are undistinguishable as UPS inhibitors and inducers of cell death, while A549 and HT29 cells show some resistance to the PEGylated derivative 2cPE (**11**).

Activation of the 2cPE Prodrug by Secreted Esterases. A different pattern of expression of enzymes capable of hydrolytically releasing the active species **2c** from the PEG conjugate might explain the differential responsiveness of the

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

Scheme 3. PEGylation of 2c[a]



[a]Reagents and conditions: (a) *N,N′*-disuccinimidyl carbonate (75%); (b) succinic anhydride, DMAP (82%); (c) MPEG-OH, EDC, HOBT, TEA (70%).

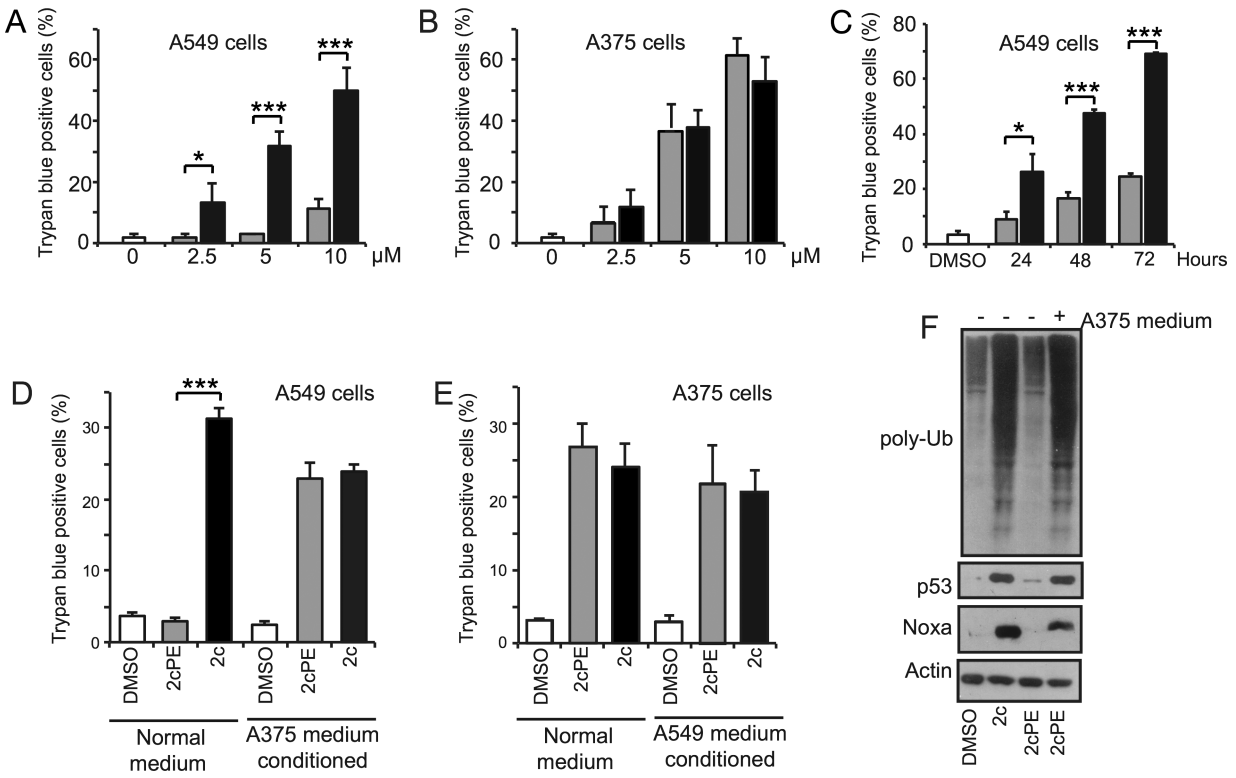


Figure 6. Cell death, after 48 h, as a function of concentration in A549 (A) and A375 (B) cells treated with 2.5−10 μM **2c** (black bars) and 2cPE (**11**) (gray bars). (C) Cell death as a function of time in A549 cells treated with 10 μM **2c** (black bars) and **11** (gray bars) for 24−72 h. (D) Response to 10 μM **2c** and 2cPE (**11**) of A549 cells grown in normal and A375 conditioned medium. (E) Response to 10 μM **2c** and 2cPE (**11**) of A375 cells grown in normal and A549 conditioned medium. (F) Immunoblot analysis showing poly-ubiquitin accumulation, p53 stabilization, and Noxa induction in A549 cells treated with 10 μM **2c** (second lane) and 2cPE (**11**) in normal and A375 conditioned medium (third and fourth lane, respectively). DMSO is the negative control throughout.

four cell lines to 2cPE (Table 3). This hypothesis is supported by the observation that unresponsive A549 cells become fully responsive to 2cPE when they are grown in the presence of A375 conditioned medium (Figure 6D), while A375 cells retain their responsiveness to 2cPE also in the presence of A549 conditioned medium (Figure 6E). Similar results were obtained with HT29 cells (Supporting Information Figure S3). Accumulation of poly-ubiquitin chains, stabilization of p53, and Noxa induction (Figure 6F) confirm that in the presence of conditioned medium from A375 cells, the activities of **2c** and 2cPE are indistinguishable, thus indicating that the prodrug is readily converted into the active molecule by this medium.

These results indicate that the response to 2cPE (Table 3) of different cell lines depends on their ability to secrete an esterase

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

activity capable of releasing **2c** from the prodrug. To gain insight into the enzymes that might be responsible for such processing, we applied a bioinformatics analysis.

Initially, the different esterases encoded by the human genome were extracted using GEO categories and HomoloGene.[43] In this manner we generated a list of 173 putative esterases. Next, to understand which secreted esterase exhibited an expression profile compatible with the responsiveness to 2cPE, we interrogated gene expression profiles available for A375, A549, and HT29 cells[44] with our list. The expression of the candidate esterase is expected to be high in A375, reduced in A549 cells, and even more reduced in HT29 cells. Seven esterases satisfied this expression profile, as illustrated in Figure 7A. Among these potential candidates, we have focused our attention on the phospholipase A2 group 7 (encoded by the PLA2G7 gene). This enzyme, also known as PAF-AH (platelet activating factor acetylhydrolase), has been extensively characterized,[45] and the crystal structures of both the ligand-free and of the paraoxon covalently inhibited enzyme are reported.[46] PLA2G7, which hydrolyzes the ester bond at the sn-2 position of phospholipid substrates shows a canonical σ/β hydrolase fold, and it has an interface binding area that makes it capable of binding to the hydrophobic portions of lipoproteins or to the internal hydrophobic region of cell membranes. The interface is located at the end of a long hydrophobic channel that leads to the catalytic serine and to the oxyanion hole. Polyethylene glycol is known to activate the phospholipase activity of the enzyme, probably by stabilizing the interface complex between the lipase and the hydrophobic inner regions of cell membranes.[47] To verify that PLA2G7 can accept 2cPE as a substrate, we have built a model of the tetrahedral covalent intermediate for the hydrolysis of **2c** conjugated to an ethylene glycol dodecamer through the same succinate linker present in 2cPE. The crystal structure of the paraoxon covalently inhibited enzyme was used as a starting point and stepwise mutated and optimized to build the model (see Supporting Information). The unique conformation corresponding to a stable and productive intermediate according to the stereoelectronic theory of Deslongchamps[48,49] is reported in Figure 7B. In this conformation, the first five units of the PEG chain lie inside the hydrophobic channel of the enzyme, and **2c** is partially buried inside the catalytic site, which appears to be broad and flexible enough to host this rather large structure. The two nitro groups have a key role in stabilizing this intermediate: one of them is found at hydrogen bond distance with the hydroxy groups of serines 319 and 349, while the aromatic ring establishes π-stacking interactions with phenylalanine 322 and further hydrophobic contacts with histidine 351 (omitted for clarity in the figure). The other nitro group is in contact with the side chain of arginine 218. Interactions of arginine and serine with the nitro groups of a ligand, such as those found in this model, are well-known in antibodies and receptor complexes.[50]

The model confirms that 2cPE could be a substrate for phospholipase A2 secreted by A375 cells and thus supports the hypothesis that lack of this enzyme, or of other esterases that can activate the prodrug, might explain the different low response of A549 and HT29 cell lines (Table 3). To prove the involvement of PLA2G7 in prodrug maturation, we generated by retroviral infection A549 cells expressing PLA2G7 isolated from A375 cells. As control we used A549 cells expressing only the resistance gene PURO. In the presence of PLA2G7, A549 cells dramatically increase responsiveness to the prodrug in terms of cell death (Figure 7C). Further evidence was obtained by incubating 2cPE in the presence of a commercially available preparation of PLA2G7; the enzyme was able to fully hydrolyze 2cPE in phosphate buffer at room temperature, yielding **2c** and PEG5000, as indicated by mass spectrometry (see Supporting Information p S31). Conversely, no hydrolysis was observed in the absence of the enzyme upon the same incubation times.

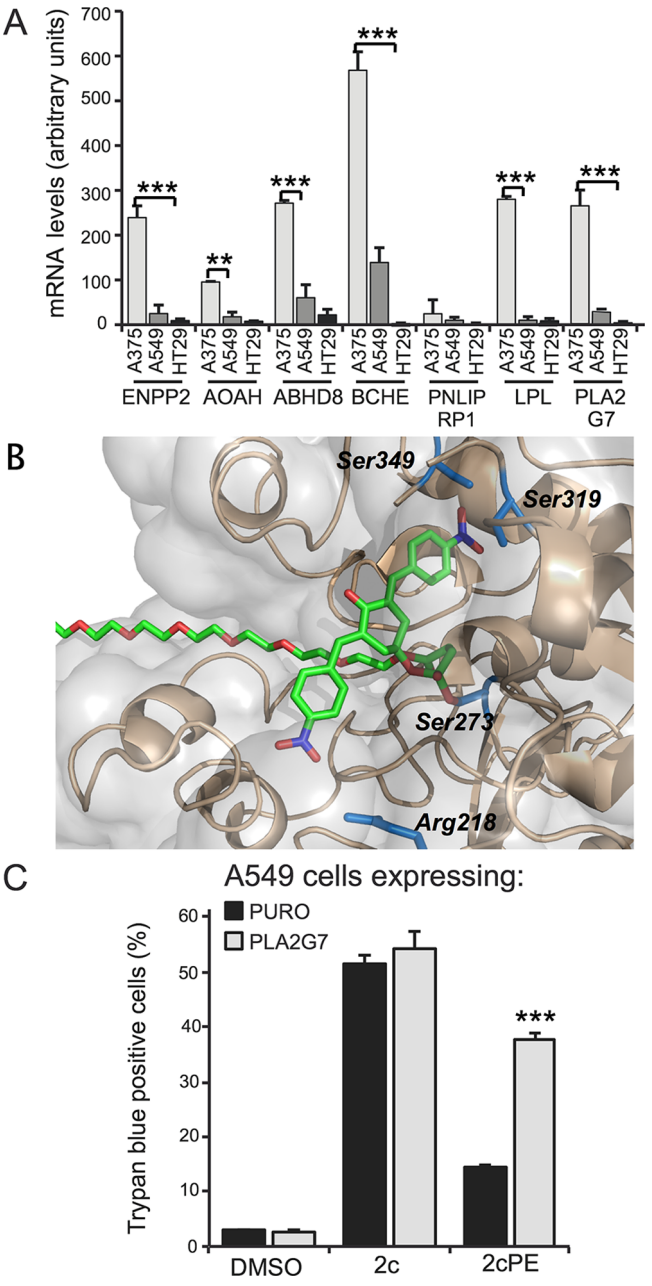


Figure 7. (A) Expression profiles in A375, A549, and HT29 cells of seven candidate esterases for the hydrolysis of 2cPE. (B) Optimized model of the rate determining tetrahedral intermediate for hydrolysis of 2cPE catalyzed by phospholipase A (PLA2G7). The model was built from the crystal structure of PLA2G7 covalently inhibited by paraoxon (PDB code 3D5E). (C) Response to 10 μM **2c** and **2cPE** of A549 cells engineered to express PLA2G7 or the control gene PURO. Trypan blue analysis was performed 48 h after treatment.

2cPE Inhibits Tumor Growth in Mice. With convincing evidence in hand that the prodrug 2cPE (**11**) is converted into the active form **2c** by secreted hydrolases and PLA2G7 in

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

particular, we set to analyzing whether the PEG conjugate displays antitumor activity in vivo. To this end we selected A549 lung cancer cells, even if these cells showed a low response to the prodrug in culture. However, we expected that, in animals, the prodrug should be efficiently processed by mouse secreted 2cPE activator PLA2G7 and, possibly, other secreted esterases.

Initial toxicity tests indicated that 2cPE is well tolerated with transient minimal effect on the animal's body weight up to 800 mg/kg. All treated groups showed slight (2%) body weight loss over the course of the first week, independent from the doses. Subsequently, the body weight of all animals steadily increased until the end of the study with final 6−8% gain in body weight (Supporting Information Figure S4). No gross toxicity was observed throughout the treatment, and 2cPE did not induce any behavioral change or grossly visible pathological changes.

A549 lung carcinoma xenografts were generated in immune-compromised mice, and when the tumors reached the size of 0.1 cm^3, 2cPE (170 mg/kg) was administered three times with four days intervals. After 1 week from the last injection (day 15) the percentage of tumor volume inhibition (TGI) was 72% and 78%, respectively, for subcutaneous and intravenous treatments (Figure 8A). At the end of the experiment, the reduction of primary tumor growth was particularly marked (TGI = 79%) in the intravenously treated group. No significant adverse effects were observed correlated to drug administration.

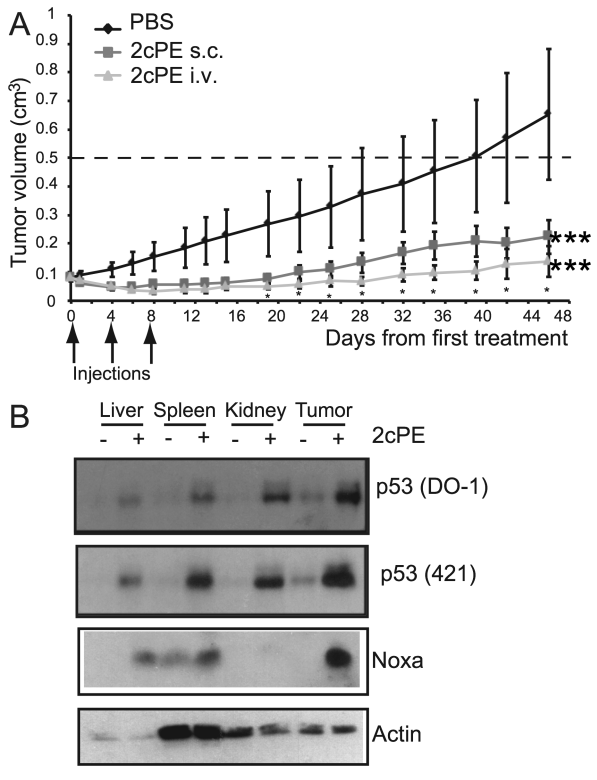


Figure 8. (A) Variations of A549 lung carcinoma xenografts volume in mice following subcutaneous (sc) and intravenous (iv) treatments with 2cPE (170 mg/kg). (B) Immunoblot analysis showing up-regulation of p53 and Noxa proteins in tissues of mice treated with 2cPE. Mice in which the tumor mass reached the size of at least 0.5 cm^3 were treated for the first time with 170 mg/kg 2cPE, and 48 h later they were sacrificed for the generation of tissue extracts. p53 was detected with DO-1 and 421 antibodies, showing no significant difference.

To further characterize the in vivo antitumor activity of 2cPE, we evaluated the cellular responses in normal mouse tissues (liver, kidney, and spleen) and in the tumor. Immunoblot analysis of tissue extracts also confirms in vivo the up-regulation of p53 and Noxa following 2cPE treatment (Figure 8B). This up-regulation was much more evident in the tumor compared to normal tissue. To exclude species-specific enrichment of the antibody against p53, we used, in addition to DO-1 antibody, the 421 antibody, which recognizes with similar affinity the human and murine proteins.

CONCLUSIONS

Several studies have established that compounds characterized by the presence of the 1,5-diaryl-3-oxo-1,4-pentadienyl pharmacophore exhibit cytotoxic activities in vitro against multiple cancer cell lines.[29] While different cellular thiols may in principle react with these Michael acceptors, isopeptidases, a heterogeneous family of cysteine proteases, are important targets for their antineoplastic activity. These inhibitors, also called N-SIIs, can trigger both necrosis and apoptosis, depending on the cellular context.[24,33,34] In this work we have shown that the cytotoxicity of a family of dienones structurally related to sulfone **G5** is directly correlated with the electrophilicity of the dienone system. This provides indirect evidence that the biological activity of these compounds is due to their ability to act as mechanism-based inhibitors of cysteine proteases. We have also demonstrated that by varying the substituents on the aromatic rings and the structure of the cyclic scaffold, it is possible to limit necrosis and to favor cell death by apoptosis. This has led to the identification of compound **2c**, based on a 4-hydroxycyclohexanone scaffold, as a good candidate for further development.

Analysis of the inhibitor activities has indicated that **G5** is in general a more potent inhibitor of DUBs such as UCHL1, UCHL5, and USP2 compared to **2c**. By contrast, **2c** functions, to some extent, as a better inhibitor for the deISGylase USP18. In this scenario the dichotomy apoptosis/necrosis might reflect different patterns of cysteine protease inhibition. Certainly, in view of the existence of additional cellular targets, alternative hypothesis are possible and further studies will be necessary to clarify this point.

It has been suggested that similar N-SIIs, although characterized by a broad spectrum of activity, can nevertheless exhibit different preferences for certain targets.[15−17,35] In this view, the lower IC_{50} for proteasome-associated deubiquitinase UCHL5 compared to other USPs, displayed by both **G5** and **2c**, can explain their action as UPS inhibitors,[13,14,16] even if additional mechanisms must also be taken into account.[23]

In this study we have also designed and synthesized the first N-SII prodrug (2cPE) by conjugating compound **2c** with PEG, through a cleavable diester linker. This strategy has overcome the solubility issue and has provided effective antitumoral activity in vivo, after administration of the prodrug by intravenous injections. From the diverse response in vitro of different cell lines to the prodrug and by applying bioinformatics analysis and molecular modeling, we have identified in PLA2G7/PAF-AH a key esterase involved in prodrug maturation. This esterase has been proposed as an activating factor for anticancer prodrugs and delivery systems due to its overexpression in breast, stomach, colorectal, pancreatic, prostate, and liver cancers. Anticancer lipids have been included in liposomes to be hydrolyzed by phospholipase A2,[51] and designed phospholipids containing PEG350 or

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

PEG2000 chains linked to phosphatidylethanolamine by a carbamate linkage have proven to be substrates for the enzyme.[52] PLA2G7 is primarily produced by macrophages and circulates in plasma in active form as a complex with LDL and HDL.[53] This pattern of expression explains why A549 cells, which are unresponsive to 2cPE in vitro, become fully responsive when evaluated in vivo.

The antineoplastic activity of 2cPE in vivo was accompanied by the up-regulation of p53 and of the proapoptotic protein Noxa. This up-regulation was more marked in tumor cells compared to liver, spleen, and kidney. Importantly, although proapoptotic genes were up-regulated also in normal tissues at some extent, the animals did not show any effect of toxicity. Both the reduced up-regulation of the apoptotic program, in normal cells and their higher intrinsic resistance to stress, can explain the absence of evident toxicity.

Isopeptidase inhibitors represent interesting tools in antineoplastic therapy. Understanding mechanisms of actions, improving the delivery, for instance by the generation of prodrugs through a cleavable PEGylation as described here for the first time, and further medicinal chemistry approaches to improve activity are efforts that need pursuing to provide additional anticancer therapeutics.

EXPERIMENTAL SECTION

All biologically evaluated compounds had a purity >95% as determined by HPLC analyses (see Supporting Information).

Chemistry: General Protocols. Melting points are uncorrected and are given in Celsius degrees. ^{1}H NMR and ^{13}C NMR spectra were recorded in $CDCl_3$, unless otherwise stated, on Jeol EX400 (400 MHz) and Varian X500 (500 MHz) spectrometers. Chemical shifts are given in ppm relative to tetramethylsilane. IR spectra were obtained as Nujol mulls with a Thermo-Nicolet AVATAR 320 FT instrument. Electron impact mass spectra were obtained on a Varian Saturn 2200 spectrometer equipped with a direct insertion probe. Electrospray mass spectra were obtained with a Bruker Daltonics Esquire 4000 spectrometer. Flash chromatography was performed on silica gel 60 (Merck, 230–400 mesh). 4*H*-Thiopyran-4-one 1,1-dioxide,[25] and compounds **1a**,[26] **2a**,[27] **2b**,[27] **2e**[30] were synthesized according to the literature. Polyethylene glycol methyl ether (Sigma-Aldrich) and all the PEGylated derivatives were coevaporated twice with dry dichloromethane and dried under vacuum immediately prior to use.

(3*Z*)-3-[(4-Nitrophenyl)methylene]-4*H*-tetrahydrothiapyran-4-one (5). A suspension of tetrahydrothiapyran-4-one (23.2 g, 0.2 mol) and 4-nitrobenzaldehyde (15.1 g, 0.1 mol) in 100 mL of water containing $MgSO_4 \cdot 7H_2O$ (14.8 g, 0.06 mol) and NaOH (5 g, 0.125 mol) was stirred at room temperature for 20 h. The solid was collected, dried, and washed repeatedly with diethyl ether to remove the excess tetrahydrothiapyran-4-one. The product, containing approximately equal amounts of the aldol diastereoisomers, was dehydrated for 1 h in refluxing ethanol (100 mL) containing conc HCl (10 mL). The mixture was cooled to 25 °C and filtered, giving the crude enone **5** (13.9 g, 56%); mp 105–106 °C (from ethanol). ^{1}H NMR: δ = 2.05 (t, 2H), 2.96 (t, 2H), 3.76 (s, 2H), 7.48 (s, 1H), 7.51 (d, 2H), 8.27 ppm (d, 2H). ^{13}C NMR: δ = 26.0, 28.5, 41.5, 123.8, 130.3, 132.4, 137.2, 141.3, 147.6, 198.5 ppm. IR: ν = 1672 (C═O), 1515 and 1344 cm^{-1} (NO_2). MS: *m*/*z* 249 (M^+ 45%), 232 (100%). Elemental analysis calculated (%) for $C_{12}H_{11}NO_3S$: C 57.8, H 4.45, N 5.62. Found: C 57.8, H 4.34, N 5.30.

General Procedure for the Synthesis of Dienones 1, 2, and 6 by the Acid-Catalyzed Knoevenagel Condensation. A solution, or mixture, of the appropriate arylaldehyde (0.1 mol) and cyclic ketone (0.05 mol) in 30 mL of ethanol containing 3 mL of 37% HCl was heated at gentle reflux for 1–2 h.[26] Alternatively, the reaction mixture, in a stoppered vial, was heated at 120 °C for 20 min in a microwave reactor. The solution was cooled in an ice bath, and the product was collected by filtration, washed with ethanol, and dried.

(3*Z*,5*Z*)-3,5-Bis[(4-nitrophenyl)methylene]tetrahydro-4*H*-thiopyran-4-one 1,1-Dioxide (G5). 45% from tetrahydro-4*H*-thiopyran-4-one 1,1-dioxide and 4-nitrobenzaldehyde; mp 226–230 °C (lit.[54] 233.5–234.5 °C). ^{1}H NMR (DMSO-d_6): δ = 4.70 (s, 4H, CH_2), 7.70 (d, 4H, ArH), 7.95 (s, 2H, ═CH) 8.3 ppm (d, 4H, ArH). ^{13}C NMR (DMSO-d_6): δ = 52.5, 123.8, 131.8, 130.7, 139.9, 140.8, 147.6, 184.4 ppm. IR: ν = 1677 (C═O), 1513 and 1350 (NO_2), 1320 and 1136 cm^{-1} (SO_2). MS: *m*/*z* 414 (M^+ 5%), 333 (100%).

(3*Z*,5*Z*)-3,5-Bis[(4-methyphenyl)methylene]tetrahydro-4*H*-thiopyran-4-one 1,1-Dioxide (1b). 69% from tetrahydro-4*H*-thiopyran-4-one 1,1-dioxide and *p*-tolualdehyde; mp 186–188 °C. ^{1}H NMR: δ = 2.40 (s, 6H), 4.46 (s, 4H), 7.25 (d, 4H), 7.31 (d, 4H), 7.98 ppm (s, 2H). ^{13}C NMR: δ = 21.6, 53.2, 129.8, 129.9, 125.9, 130.8, 140.7, 144.3, 186.1 ppm. IR: ν = 1664 (C═O), 1333 and 1127 cm^{-1} (SO_2). MS: *m*/*z* 352 (M^+ 20%), 273 (100%).

(3*Z*,5*Z*)-3,5-Bis[(4-hydroxyphenyl)methylene]tetrahydro-4*H*-thiopyran-4-one 1,1-Dioxide (1c). 23% from tetrahydro-4*H*-thiopyran-4-one 1,1-dioxide and 4-hydroxybenzaldehyde; mp 244–246 °C. ^{1}H NMR ($(CD_3)_2CO$): δ = 3.0 (s, 2H), 4.57 (s, 4H), 6.94 (d, 4H), 7.41 (d, 4H), 7.87 ppm (s, 2H). ^{13}C NMR ($(CD_3)_2CO$): δ = 52.8, 116.1, 132.6, 125.6, 125.8, 142.3, 159.5, 184.2 ppm. IR: ν = 3388 (OH), 1659 (C═O), 1304 and 1128 cm^{-1} (SO_2). MS: *m*/*z* 356 ([M-2H]$^+$ 58%).

(3*Z*,5*Z*)-3,5-Bis[(4-methoxyphenyl)methylene]tetrahydro-4*H*-thiopyran-4-one 1,1-Dioxide (1d). 50% from tetrahydro-4*H*-thiopyran-4-one 1,1-dioxide and *p*-anisaldehyde; mp 175–177 °C. ^{1}H NMR: δ = 3.83 (s, 6H), 4.46 (s, 4H), 6.96 (d, 4H), 7.40 (d, 4H), 7.97 ppm (s, 2H). ^{13}C NMR: δ = 53.3, 55.5, 114.6, 131.9, 124.6, 126.2, 143.9, 161.2, 185.8 ppm. IR: ν = 1659 (C═O), 1318 and 1125 cm^{-1} (SO_2). MS: *m*/*z* 384 (M^+ 75%), 88 (100%).

(3*Z*,5*Z*)-3,5-Bis[(4-phenoxyphenyl)methylene]tetrahydro-4*H*-thiopyran-4-one 1,1-Dioxide (1e). 50% from tetrahydro-4*H*-thiopyran-4-one 1,1-dioxide and 4-phenoxybenzaldehyde; mp 165–168 °C. ^{1}H NMR: δ = 4.40 (s, 4H), 7.00–7.08 (m, 8H), 7.17 (t, 2H), 7.35–7.41 (m, 8H), 7.96 ppm (s, 2H). ^{13}C NMR: δ = 53.2, 118.3, 120.0, 124.5, 125.5, 130.1, 131.8, 128.0, 143.6, 155.8, 159.5, 186.0 ppm. IR: ν = 1666 (C═O), 1308 and 1133 cm^{-1} (SO_2). MS: *m*/*z* 443 ([M − SO_2]$^+$ 100%).

(3*Z*,5*Z*)-3,5-Bis[(4-cyanophenyl)methylene]tetrahydro-4*H*-thiopyran-4-one 1,1-Dioxide (1g). 20% from tetrahydro-4*H*-thiopyran-4-one 1,1-dioxide and 4-cyanobenzaldehyde; mp 260–262 °C. ^{1}H NMR (DMSO-d_6): δ = 4.69 (s, 4H), 7.93 (s, 2H), 7.72 (d, 4H), 8.00 ppm (d, 4H). ^{13}C NMR (DMSO-d_6): δ = 52.1, 111.9, 118.6, 130.2, 140.0, 130.7, 132.6, 138.1, 184.6 ppm. IR: ν = 2232 (CN), 1692 (C═O), 1327 and 1123 cm^{-1} (SO_2). MS: *m*/*z* 374 (M^+ 30%), 309 (100%). Elemental analysis calculated (%) for $C_{21}H_{14}N_2O_3S$: C 67.4, H 3.77, N 7.48. Found: C 67.0, H 3.81, N, 7.37.

(3*Z*,5*Z*)-3,5-Bis[(3-nitro-4-methoxyphenyl)methylene]-tetrahydro-4*H*-thiopyran-4-one 1,1-Dioxide (1h). 54% from tetrahydro-4*H*-thiopyran-4-one 1,1-dioxide and 3-nitro-4-methoxybenzaldehyde; mp 259–260 °C. ^{1}H NMR (DMSO-d_6): δ = 4.00 (s, 6H), 4.74 (s, 4H), 7.45 (s, 2H), 7.85 (s, 2H), 7.88 (d, 2H), 8.09 ppm (d, 2H). ^{13}C NMR (DMSO-d_6): δ = 51.9, 57.0, 114.7, 126.6, 136.4, 125.9, 128.1, 139.5, 152.7, 184.2 ppm. IR: ν = 1696 (C═O), 1531 and 1340 (NO_2), 1311 and 1131 cm^{-1} (SO_2). MS: *m*/*z* 474 (M^+ 1%), 393 (100%). Elemental analysis calculated (%) for $C_{21}H_{18}N_2O_9S$: C 53.16, H 3.82, N 5.90. Found: C 52.77, H 3.80, N 5.54.

(3*Z*,5*Z*)-3,5-Bis[(3-nitrophenyl)methylene]tetrahydro-4*H*-thiopyran-4-one 1,1-Dioxide (1i). 21% from tetrahydro-4*H*-thiopyran-4-one 1,1-dioxide and 3-nitrobenzaldehyde; mp 230–231 °C. ^{1}H NMR (DMSO-d_6): δ = 4.78 (s, 4H), 7.78 (t, 2H), 7.97 (s, 2H), 7.99 (s, 2H), 8.29 (d, 2H), 8.34 ppm (d, 2H). ^{13}C NMR (DMSO-d_6): δ = 52.5, 124.7, 125.1, 130.6, 130.9, 135.6, 136.7, 140.3, 148.6, 185.5 ppm. IR: ν = 1690 (C═O), 1530 and 1350 (NO_2), 1320 and 1123 cm^{-1} (SO_2). MS: *m*/*z* 414 (M^+ 10%), 333 (100%). Elemental analysis calculated (%) for $C_{19}H_{14}N_2O_7S$: C 55.1, H 3.41, N 6.76, S 7.74. Found: C 54.7, H 3.38, N 6.44, S 7.55.

(2*E*,6*E*)-2,6-Bis[(4-nitrophenyl)methylene]-4-hydroxycyclohexanone (2c). 78% from 4-hydroxycyclohexanone and 4-nitrobenzaldehyde; mp 210–213 °C. ^{1}H NMR (DMSO-d_6): δ = 2.98 (dd, 2H), 3.09 (dd, 2H)), 4.07–4.13 (m, 1H), 5.04 (d, 1H), 7.7 (s, 2H),

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX–XXX

7.8 (d, 4H), 8.3 ppm (d, 4H). ^{13}C NMR (DMSO-d_6): δ = 36.0, 63.0, 123.8, 131.4, 134.0, 137.1, 142.2, 147.2, 188.9 ppm. IR: ν = 3539 (OH), 1668 (C═O), 1508 and 1342 cm^{-1} (NO_2). ESI-MS: m/z 403 (MNa^+), 381 (MH^+). Elemental analysis calculated (%) for $C_{20}H_{16}N_2O_6$: C 63.2, H 4.24, N 7.37. Found: C 62.9, H 4.19, N 7.32.

Ethyl (3*E*,5*E*)-3,5-Bis[(4-nitrophenyl)methylene]-4-oxocyclohexanecarboxylate (2d). 36% from ethyl 4-oxocyclohexanecarboxylate and 4-nitrobenzaldehyde; mp 172−174 °C. ^{1}H NMR: δ = 1.16 (t, 3H), 2.76−2.84 (m, 1H), 3.06−3.24 (m, 4H), 4.12 (q, 2H), 7.56 (d, 4H), 7.85 (s, 2H), 8.27 ppm (d, 4H). ^{13}C NMR: δ = 14.2, 30.5, 3961.4, 123.8, 130.9, 135.9, 136.4, 141.7, 147.5, 173.2, 187.7 ppm. IR: ν = 1723 (C═O), 1673 (C═O), 1512 and 1340 cm^{-1} (NO_2). MS: m/z 436 (M^+ 50%), 419 (100%).

(3*E*,5*E*)-Tetrahydro-3,5-bis[(4-nitrophenyl)methylene]-4*H*-pyran-4-one (2f). 24% from tetrahydro-4*H*-pyran-4-one and 4-nitrobenzaldehyde; mp 263−265 °C (lit.[54] 272−273 °C). ^{1}H NMR (DMSO-d_6): δ = 4.96 (s, 4H), 7.73 (d, 4H), 7.77 (s, 2H), 8.28 ppm (d, 4H). ^{13}C NMR (DMSO-d_6): δ = 68.2, 124.2, 132.1, 133.5, 136.8, 140.9, 148.0 ppm. IR: ν = 1682 (C═O), 1514 and 1347 cm^{-1} (NO_2). MS: m/z 366 (M^+ 10%), 349 (100%).

(3*Z*,5*Z*)-Tetrahydro-3,5-bis[(4-nitrophenyl)methylene]-4*H*-thiopyran-4-one (2g). 41% from tetrahydro-4*H*-thiopyran-4-one and 4-nitrobenzaldehyde; mp 220−223 °C (lit.[54] 224−225 °C). ^{1}H NMR: δ = 3.87 (s, 4H), 7.55 (d, 4H), 7.77 (s, 2H), 8.29 ppm (d, 4H). ^{13}C NMR: δ = 30.1, 124.0, 130.6, 134.6, 136.2, 141.3, 147.7, 188.0 ppm. IR: ν = 1667 (C═O), 1509 and 1342 cm^{-1} (NO_2). MS: m/z 382 (M^+, 20%).

(3*Z*,5*Z*)-Tetrahydro-3,5-bis[(4-fluorophenyl)methylene)]-4*H*-tetrahydrothiapyran-4-one (3). 54% from tetrahydro-4*H*-thiapyran-4-one and 4-fluorobenzaldehyde; mp 122−124 °C. ^{1}H NMR: δ = 3.88 (s, 4H), 7.11−7.38 (m, 4H), 7.73 ppm (s, 2H). ^{13}C NMR: δ = 30.0, 115.4 (d, J_{C-F} = 22 Hz), 130.8 (d, J_{C-F} = 3 Hz), 131.6 (d, J_{C-F} = 8 Hz), 133.2 (d, J_{C-F} = 1 Hz), 135.4, 162.5, (d, J_{C-F} = 251 Hz), 188.4 ppm. IR: ν = 1662 (C═O), 1510 and 1330 cm^{-1} (NO_2). MS: m/z 328 (M^+ 100%). Elemental analysis calculated (%) for $C_{19}H_{14}F_2OS$: C 69.5, H 4.30. Found: C 69.3, H 4.41%.

(3*Z*,5*Z*)-3-[(4-Hydroxyphenyl)methylene]-5-[(4-nitrophenyl)-methylene]-4*H*-tetrahydrothiapyran-4-one (6j). 44% from **5** and p-hydroxybenzaldehyde; mp 178−180 °C. ^{1}H NMR (DMSO-d_6): δ = 3.93 (s, 2H), 3.98 (s, 2H), 6.85 (d, 2H), 7.42 (d, 2H), 7.56 (s, 1H), 7.61 (s, 1H), 7.73 (d, 2H), 8.25 (d, 2H), 10.0 ppm (broad, 1H). ^{13}C NMR (DMSO-d_6): δ = 29.0, 29.6, 115.7, 123.6, 125.4, 130.8, 131.2, 131.8, 132.8, 136.9, 137.3, 141.6, 146.9, 158.9, 187.7 ppm. IR: ν = 3451 (OH), 1660 (C═O), 1519 and 1347 cm^{-1} (NO_2). MS: m/z 353 (M^+ 100%).

(3*Z*,5*Z*)-3-[(4-Methoxyphenyl)methylene]-5-[(4-nitrophenyl)-methylene]-4*H*-tetrahydrothiapyran-4-one (6k). 95% from **5** and *p*-methoxybenzaldehyde; mp 152−154 °C. ^{1}H NMR: δ = 3.74 (s, 2H), 3.84 (s, 3H), 3.96 (s, 2H), 6.96 (d, 2H), 7.39 (d, 2H), 7.51 (d, 2H), 7.73 (s, 1H), 7.77 (s, 1H), 8.27 ppm (d, 2H). ^{13}C NMR: δ = 29.9, 30.4, 55.5, 114.3, 123.9, 127.5, 130.6, 131.3, 132.4, 133.1, 136.9, 138.4,, 142.0, 147.5, 160.7, 188.4 ppm. IR: ν = 1653 (C═O), 1514 and 1344 (NO_2). MS: m/z 367 (M^+ 100%). Elemental analysis calculated (%) for $C_{20}H_{17}NO_4S$: C 65.38, H 4.66, N 3.81. Found: C 65.01, H 4.64, N 3.72.

(3*Z*,5*Z*)-3-[(4-Methylphenyl)methylene]-5-[(4-nitrophenyl)-methylene]-4*H*-tetrahydrothiapyran-4-one (6l). 73% from **5** and *p*-tolualdehyde; mp 139−142 °C. ^{1}H NMR (DMSO-d_6): δ = 2.32 (s, 3H), 3.95 (s, 2H), 3.98 (s, 2H), 7.26 (d, 2H), 7.42 (d, 2H), 7.58 (s, 1H) 7.61 (s, 1H), 7.75 (d, 2H), 8.25 ppm (d, 2H). ^{13}C NMR (DMSO-d_6): δ = 21.0, 29.2, 29.5, 123.7, 129.4, 130.5, 131.3, 131.7, 132.3, 133.2, 136.1, 137.2, 138.0, 141.4, 147.0, 188.0 ppm. IR: ν = 1662 (C═O), 1517 and 1344 cm^{-1} (NO_2). MS: m/z 351 (M^+ 30%), 336 (100%).

General Procedure for mCBPA Oxidations. A solution of mCPBA (1 mmol for oxidations to sulfoxides; 2.5 mmol for oxidations to sulfones) in CH_2Cl_2 (5 or 12.5 mL, respectively) was added dropwise, at 0 °C, to a suspension of the sulfide (1 mmol) in dichloromethane (5 mL) and stirred at 25 °C for 16 h. The mixture was diluted with 20 mL of dichloromethane and extracted with 10% aqueous $Na_2S_2O_5$, saturated $NaHCO_3$, and brine. The organic layer was dried over Na_2SO_4, and the solvent was removed in vacuo to give the crude product that was crystallized from ethanol.

(3*Z*,5*Z*)-Tetrahydro-3,5-bis[(4-fluorophenyl)methylene)-4*H*-tetrahydrothiapyran-4-one 1,1-Dioxide (1f). Oxidation of sulfide **3** with 2 equiv of mCPBA gave a mixture of sulfone **1f** and the Baeyer−Villiger product **4** which were separated by crystallization from acetone/hexane. **1f** (41%) had mp 179−180 °C. ^{1}H NMR: δ = 4.43 (s, 4H), 7.12−7.45 (m, 4H), 7.96 ppm (s, 2H). ^{13}C NMR: δ = 53.1, 116.6 (d, J_{C-F} = 22 Hz), 126.7 (d, J_{C-F} = 1 Hz), 129.8 (d, J_{C-F} = 3 Hz), 132.0 (d, J_{C-F} = 9 Hz), 143.4, 163.9 (d, J_{C-F} = 252 Hz), 186.4 ppm. IR: ν = 1664 (C═O), 1333 and 1128 cm^{-1} (SO_2). MS: m/z 360 (M^+ 20%), 294 (100%). Elemental analysis calculated (%) for $C_{19}H_{14}F_2O_3S$: C 63.3, H 3.92, S 8.90. Found: C 63.2, H 3.81, S 8.74. **4** (10%) had mp 176−178 °C. ^{1}H NMR: δ = 4.27 (s, 4H), 6.89 (s, 1H), 7.07−7.14 (m, 4H), 7.53−7.72 (m, 4H), 8.05 ppm (s, 1H). ^{13}C NMR: δ = 53.1, 55.7, 115.8, 116.13, 116.16, 116.5, 119.9, 128.0, 128.8, 137.4, 126.1, 130.5, 130.6, 131.8, 131.9, 149.7, 162.4, 165.5 ppm. IR: ν = 1721 (C═O), 1315 and 1120 cm^{-1} (SO_2). MS: m/z 376 (M^+ 43%), 312 (100%). Elemental analysis calculated (%) for $C_{19}H_{14}F_2O_4S$: C 60.6, H 3.72. Found: C 60.2, H 3.72.

(3*Z*,5*Z*)-3-[(4-Hydroxyphenyl)methylene]-5-[(4-nitrophenyl)-methylene]-4*H*-tetrahydrothiapyran-4-one 1,1-Dioxide (1j). 60% from sulfide **7j**; mp 221−224 °C. ^{1}H NMR (DMSO-d_6): δ = 4.68 (s, 4H), 6.87 (d, 2H), 7.46 (d, 2H), 7.70 (s, 1H), 7.77 (d, 2H), 7.84 (s, 1H), 8.29 (d, 2H), 10.22 ppm (broad, 1H). ^{13}C NMR (DMSO-d_6): δ = 51.9, 52.5, 116.0, 123.8, 131.0, 133.0, 138.7, 142.9, 123.7, 124.2, 124.4, 140.4, 147.4, 159.7, 184.0 ppm. IR: ν = 3450 (OH), 1511 and 1349 (NO_2), 1308 and 1121 cm^{-1} (SO_2). MS: m/z 385 (M^+ 5%), 320 (100%).

(3*Z*,5*Z*)-3-[(4-Methoxyphenyl)methylene]-5-[(4-nitrophenyl)-methylene]-4*H*-tetrahydrothiapyran-4-one 1,1-Dioxide (1k). 60% from sulfide **7k**; mp 158−160 °C (from acetone−hexane). ^{1}H NMR: δ = 3.88 (s, 3H), 4.35 (s, 2H), 4.51 (s, 2H), 6.99 (d, 2H), 7.42 (d, 2H), 7.58 (d, 2H), 7.98 (s, 1H), 8.02 (s, 1H) 8.32 ppm (d, 2H). ^{13}C NMR: δ = 52.7, 53.7, 55.5, 114.7, 124.2, 130.2, 132.2, 140.6, 145.2, 123.5, 126.0, 130.0, 140.0, 147.0, 161.6, 185.7 ppm. IR: ν = 1663 (C═O), 1510 and 1346 (NO_2), 1319 and 1126 cm^{-1} (SO_2). MS: m/z 399 (M^+ 10%), 304 (100%).

(3*Z*,5*Z*)-3-[(4-Methylphenyl)methylene]-5-[(4-nitrophenyl)-methylene]-4*H*-tetrahydrothiapyran-4-one 1,1-Dioxide (1l). 90% from sulfide **7l**; mp 155−158 °C. ^{1}H NMR: δ = 2.44 (s, 3H), 4.35 (s, 2H), 4.49 (s, 2H), 7.27 (d, 2H), 7.32 (d, 2H), 7.58 (d, 2H), 7.97 (s, 1H), 8.01 (s, 1H) 8.30 ppm (d, 2H). ^{13}C NMR: δ = 21.6, 52.8, 53.6, 124.2, 125.0, 129.9, 130.0, 130.2, 140.0, 140.8, 141.3, 145.4, 148.2, 185.8 ppm. IR: ν = 1663 (C═O), 1519 and 1346 (NO_2), 1326 and 1133 cm^{-1} (SO_2). MS: m/z 383 (M^+ 20%), 304 (100%).

(3*Z*,5*Z*)-Tetrahydro-3,5-bis[(4-nitrophenyl)methylene]-4*H*-thiopyran-4-one 1-Oxide (2h). Oxidation of the sulfide **2g** with 1 equiv of mCPBA gave a 1:1 mixture of sulfone 2G and sulfoxide **2h** that were separated by flash chromatography on silica gel with 8:2 dichloromethane/ethyl acetate as eluent, giving pure **2h** (34%); mp 185−190 °C. ^{1}H NMR: δ = 3.93 (d, 2H, J = 12 Hz), 4.21 (d, 2H, J = 12 Hz), 7.58 (d, 4H), 8.03 (s, 2H), 8.3 ppm (d, 4H). ^{13}C NMR (DMSO-d_6): δ = 48.7, 124.3, 131.0, 131.6, 140.8, 141.4, 147.9, 187.0 ppm. IR: ν = 1676 (C═O), 1506 and 1349 cm^{-1} (NO_2). MS: m/z 398 (M^+, 10%), 365 (100%). Elemental analysis calculated (%) for $C_{19}H_{14}N_2O_6S$: C 57.3, H 3.54, N 7.03, S 8.05. Found: C 57.2, H 3.75, N 7.11, S 7.92.

(3*Z*,5*Z*)-3,5-Bis[(4-nitrophenyl)methylene]-4-hydroxy-tetrahydro-4*H*-thiopyran 1,1-Dioxide (7). $NaBH_4$ (0.016 g, 0.44 mmol) was added in small portions at 0 °C with constant stirring to a solution of **G5** (0.15 g, 0.4 mmol) in 25 mL of a 10:1 (v/v) THF/MeOH mixture. After 20 min the reaction was quenched by the addition of ice-cold, saturated brine (25 mL) and the product was extracted with diethyl ether (4 × 25 mL). The combined organic layers were washed with brine, dried over $MgSO_4$ and the solvent was removed under reduced pressure, giving the alcohol **7** (90%): mp 216−220 °C. ^{1}H NMR ($(CD_3)_2CO$): δ = 2.84 (s, 1H), 4.24 (d, 2H), 4.26 (d, 2H), 5.22 (d, 1H), 7.21 (s, 1H), 7.69 (d, 2H), 8.28 ppm (d, 2H). ^{13}C NMR ($(CD_3)_2CO$): δ = 51.7, 76.7, 123.8, 130.1, 128.5, 134.5, 142.4, 147.3

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

ppm. IR: ν = 3399 (OH), 1517 and 1351 (NO_2), 1329 and 1130 cm^{-1} (SO_2). Elemental analysis calculated (%) for $C_{19}H_{16}N_2O_7S$: C 54.80, H 3.87, N 6.73. Found: C 55.02, H 3.99, N 6.67.

(3*E*,5*E*)-3,5-Bis[(4-nitrophenyl)methylene]-4-oxocyclohexyl 2,5-dioxo-1-pyrrolidinyl Carbonate (8). *N*,*N*′-Disuccinimidyl carbonate (1.0 g, 3.89 mmol) was added to a solution of **2c** (0.74 g, 1.95 mmol) in 20 mL of 1:1 dichloromethane/acetonitrile. Pyridine was added until the solution became clear (1 mL), and the mixture was stirred at 25 °C for 18 h. The solution was cooled in an ice bath, and the mixed carbonate **8** was precipitated with diethyl ether, filtered, and dried (75%). Mp: 172−175 °C. ^{1}H NMR (DMSO-d_6): δ = 2.71 (s, 4H), 3.30−3.40 (m, 4H), 5.31−5.38 (m, 1H), 7.79 (d, 4H), 7.85 (s, 2H), 8.27 ppm (d, 4H). ^{13}C NMR (DMSO-d_6): δ = 25.3, 32.5, 74.5, 123.6, 131.2, 133.9, 136.7, 141.3, 147.1, 151.0, 169.6, 187.1 ppm. IR: ν = 1810 (C═O), 1788 (C═O), 1740 (C═O), 1670 (C═O), 1518 and 1347 cm^{-1} (NO_2).

2cPC (9). *N*,*N*′-Disuccinimidyl carbonate (205 mg, 0.8 mmol) was added to mPEG5000-OH (1.0 g, 0.2 mmol) in a mixture of dry dichloromethane (3.5 mL), acetonitrile (1 mL), and pyridine (0.5 mL), and the solution was stirred for 18 h at 25 °C under an argon atmosphere, cooled in an ice bath, and the crude product (mPEG-OSu, 94%) was precipitated with diethyl ether, filtered, and recrystallized from ethanol. ^{1}H NMR (DMSO-d_6): δ = 2.80 (s, 4H), 3.30−3.55 (m, PEG), 4.25−4.55 (m, 2H, PEG-CH_2OSu). 1,3-Diaminopropane (48 μL, 0.57 mmol) and mPEG-OSu (1.0 g, 0.19 mmol), in a minimum amount of dry dichloromethane, were stirred for 18 h at 25 °C under an argon atmosphere. The solution was cooled in an ice-bath, and the crude mono-PEGylated 1,3-diaminopropane (90%) was precipitated with diethyl ether, filtered, and recrystallized from ethanol. ^{1}H NMR (DMSO-d_6): δ = 1.48−1.54 (m, 2H, CH_2), 2.94−2.99 (m, 2H), 3.30−3.55 (m, PEG + CH_2), 3.98−4.04 (m, 2H, PEG-C$\underline{H_2}$OCONH), 7.17 ppm (m, 1H, NH). Triethylamine was added to a solution of the activated carbonate **8** (0.52 g, 1.0 mmol) and PEGylated 1,3-diaminopropane (1.0 g, 0.2 mmol) in the minimum amount of dry pyridine, to pH 8. The mixture was stirred for 18 h at 25 °C under an argon atmosphere and then cooled in an ice bath, and the crude product **9** (67%) was precipitated with diethyl ether, filtered, and recrystallized from ethanol. ^{1}H NMR (DMSO-d_6): δ = 1.35−1.38 (m, 2H), 2.77−2.83 (m, 4H), 3.20−3.70 (m, PEG), 3.97−4.01 (m, 2H), 4.97−5.03 (m, 1H), 7.02−7.09 (m, 2H), 7.74 (d, 4H), 7.77 (s, 2H), 8.30 (d, 4H).

3-{(3*E*,5*E*)-3,5-Bis[(*p*-nitrophenyl)methylene]-4-oxocyclohexyloxycarbonyl}propionic Acid (10). Succinic anhydride (2.63 g, 26.3 mmol) and DMAP (0.32 g, 2.63 mmol) were added to a solution of alcohol **2c** (1.0 g, 2.63 mmol) and pyridine (1 mL) in 25 mL of 1:1 dichloromethane/acetonitrile, and the reaction mixture was stirred at room temperature for 16 h and then extracted with 0.1 M HCl. The solid emisuccinate (370 mg) was collected, while the organic layer was dried over anhydrous sodium sulfate and evaporated. The solid residue was washed with hot methanol, giving a second crop of product **10** (664 mg). The crude emisuccinate (82% overall) was used without further purification. Mp: 195−198 °C. ^{1}H NMR (DMSO-d_6): δ = 2.31−2.36 (m, 4H), 3.13 (dd, 2H), 3.20−3.25 (m, 2H), 5.14−5.18 (m, 1H), 7.76 (s, 2H), 7.79 (d, 2H), 8.27 (d, 4H), 12.1 ppm (s, 1H). ^{13}C NMR (DMSO-d_6): δ = 28.8, 32.3, 35.7, 67.0, 123.6, 131.2, 135.0, 135.9, 141.5, 147.0, 171.5, 173.3, 187.0 ppm. IR: ν = 3502 (COOH), 1734 (C═O), 1710 (C═O), 1668 (C═O), 1517 and 1347 cm^{-1} (NO_2). MS: m/z 480 (M^+ 8%), 345 (100%).

2cPE (11). Emisuccinate **10** (0.96 g, 2 mmol), EDC (0.38 g, 2 mmol), and HOBT (0.27 g, 2 mmol) were added, in the order, to mPEG5000-OH (1.0 g, 0.2 mmol) in a minimum amount of dry dichloromethane. Triethylamine was added to pH 8 and the mixture was stirred for 18 h, at 25 °C, under an argon atmosphere. The solution was cooled in an ice bath and the crude product was precipitated by adding diethyl ether, filtered, and recrystallized from ethanol, giving 0.76 g of the conjugate **11** (70%). ^{1}H NMR (DMSO-d_6): δ = 2.41 (2H), 2.50 (2H), 3.25−3.70 (m, PEG), 3.95−4.00 (m, 2H), 5.13−5.20 (m, 1H), 7.81 (d, 4H), 7.82 (s, 2H), 8.30 ppm (d, 4H).

Modeling. Calculations were carried out on a workstation equipped with a Dual Opteron ASUS KFSN4-DRE/IKVM/IST mother board and two AMD Opteron SixCore 2427 2.2 GHz &MB 75 W processors. The MMFF94 force field[55] [as implemented in the Spartan '14 parallel suite (Wavefunction Inc.)] was used in all energy minimizations. The crystallographic coordinates of the reference covalent complex of phospholipase A2 with paraoxon[46] were obtained from the Protein Data Bank, Brookhaven National Laboratory (PDB code 3D5E). Full details on the construction of the model for the PLA2G7 catalyzed hydrolysis of 2cPE (**11**) are in the Supporting Information.

Hydrolysis of 2cPE by PLA2G7. The enzyme (recombinant human PLA2G7 from R&D Systems, code 5106-PL-010) was added at a final 0.2 nM concentration to a 25 mM solution of 2cPE in phosphate buffer, pH 7. The reaction was followed by ESI-MS on a Bruker Esquire 400 instrument (see Supporting Information).

Biological Assays. Reagents and Antibodies. The following reagents were used: bortezomib (LC Laboratories), DMSO (Sigma-Aldrich), resazurin (Invitrogene), GST-UCHL5 (Ubiquigent), ISG15-AMC (BostonBiochem), ubiquitin-AMC (Ubiquigent), zVAD-fmk (Bachem). Primary antibodies were anti-actin (Sigma-Aldrich), anti-ubiqutin (Covance), anti-p53 DO-1 (Santa Cruz), anti-Noxa (Merck Millipore).

Cell Culture, Cell Death, and Caspase Activity. All cell lines were grown in DMEM supplemented with 10% FBS, penicillin (100 U/mL), glutamine (2 mmol/L), and streptomycin (100 μg/mL) at 37 °C in 5% CO_2 atmosphere. In all trypan blue exclusion assays, at least 400 cells from three independent experiments were counted. Data are presented as mean value ± SD. A549 cells engineered to express PLA2G7 were generated by retroviral infection using the pLPC vector and selected for puromycin resistance. The cDNA of PLA2G7 was obtained after retrotranscription of RNAs isolated form A375 cells and PCR amplification using oligonucleotides covering the ORF. The cDNA of PLA2G7 was cloned *blunt/SalI*. Sequencing of the PLA2G7 cDNA was performed to scrutinize for the presence of mutations. The RT-PCR analysis confirming the expression of recombinant PLA2G7 in A549 cells is reported in the Supporting Information.

The caspase activity was evaluated using the Apo-ONE caspase-3/7 homogeneous assay (Promega). The assay includes a profluorescent caspase-3/7 consensus substrate, rhodamine 110 bis-(*N*-*Z*-L-aspartyl-L-glutamyl-L-valylaspartic acid amide). Cells grown in 96-well plates were treated with the different compounds and tested for caspase activity as recommended by the producer. Data are presented as mean value ± sd, n = 3.

Western Blotting. Tissue and cellular extracts were generated in Laemmli sample buffer. Before lysis, frozen small tissue fragments were pulverized using a dedicated apparatus (TissueLyser Qiagen). Protein extracts, after SDS−PAGE were transferred to a 0.2 μm nitrocellulose membrane and incubated with the specific primary antibodies as previously described.[14] After washing, blots were incubated with secondary antibody−peroxidase conjugates (Sigma). For primary antibody stripping, blots were incubated for 30 min at 60 °C in stripping solution (62.5 mM Tris-HCl, pH 6.8, 2% SDS, 100 mM β-mercaptoethanol).

Isopeptidase Inhibition. In a 96-well-plate, 20 nM UCHL1, 20 nM UCHL5, fused to GST, and 40 nM His-tagged USP2 catalytic domain or 270 nM USP18 were preincubated with **G5**, **2c**, or control for 30 min before substrate Ub-AMC (200 nM) addition. ISG15-AMC (400 nM) was used as substrate for USP18. Reaction buffer was 50 mM Tris, pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 1 mM DTT. Enzymes relative activities were determined by measuring the RFU values as average of the maximum values obtained within the initial linear range. Enzymatic activities were determined every minute for 60 min, at 37 °C. Fluorescence values were read with an Enspire 2300 multilabel reader (excitation 380 nm, emission 460 nm). The RFU values were normalized to 100% of activity for the isopeptidases incubated with vehicle alone. IC_{50} values were calculated with GraphPad Prism software using a nonlinear fit. Recombinant full-length UCHL1 and USP2 catalytic domain were expressed in bacteria and purified by affinity chromatography using GST or histidine

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

binding resin, respectively. The human USP18 sequence encoding the region 16-372 corresponding to the translated full-length protein starting at the rare codon CUG was cloned using the Gateway system in a baculovirus expression vector based on pVL1393 (GE Healthcare). The recombinant protein was expressed in Sf9 insect cells. For protein purification, cells were homogenized (Emulsiflex, Avestin) in lysis buffer (500 mM NaCl, 10% glycerol, 0.01% Tween, 20 mM DTT, 20 μg/mL DNaseI, protease inhibitors in 50 mM Na phosphate, pH 8.0). Cleared lysates were affinity purified on gluthatione-Sepharose 4B (GE Healthcare) and eluted by cleavage with HRV C3 protease. The protein was then purified on Superdex 200 10/30 GL (GE Healthcare) equilibrated in buffer (500 mM NaCl, 10% glycerol, 5 mM DTT in 50 mM Na phosphate, pH 8.0). The eluted monomeric protein was concentrated to 3.5 mg/mL, and aliquots were flash-frozen in liquid nitrogen. The quality assessment of recombinant protein was performed by ESI LC/MS analysis using a single quadrupole instrument with an electrospray ion source (1100 HPLC G1946 MSD Agilent system).

Gene Expression Studies. Microarray data sets used in this study were downloaded manually from GEO.[43a] We analyzed expression data obtained from GSE48433 (platform HG-U133_Plus_2, GPL570 array). We processed all the CEL files together by using standard tools available within the Affy package.[56] We used a UniGene ID centered CDF (Chip Description file) in order to have only one intensity value per gene.[57] Details on microarray data processing and analysis are reported in the Supporting Information. The list of lipases was downloaded from Gene Ontology and integrated with HomoloGene; only genes with direct enzymatic activity were considered, filtering out lipases regulators and cofactors. In the end we obtained a list of 173 genes of which 159 were present in the array's platform.

Statistical Analysis. All assays and control experiments were carried out in triplicate. Results are expressed as the mean ± sd. Statistical analysis was performed by Student's t test and P values of ≤0.05 were considered statistically significant. Asterisks in the figures indicate the following: (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.005$.

A549 Human Lung Carcinoma Xenograft Tumors in Mice. Six-week-old female athymic nude-foxn1nu mice (Harlan Ud Italy) were utilized for in vivo xenograft experiments, whereas 9-week-old female Balb/C OlaHsd mice were used for toxicity studies (Harlan Ud Italy). Animal studies were carried out according to the guidelines enforced in Italy (DDL 116 of Feb 21, 1992 and subsequent addenda) and in compliance with the Guide for the Care and Use of Laboratory Animals, Department of Health and Human Services Publication No. 86-23 (National Institutes of Health, Bethesda, MD, 1985). In vivo xenograft tumor model was established from initial sc injection of in vitro expanded A549 tumor cells. After tumor mass establishment, solid mass was surgically removed, under sterile conditions, and cut into fragments (2−3 mm), in sterile PBS. Tumor fragments obtained from donor mice were then serially expanded or injected for experimental purpose. Fifteen mice were implanted sc with A549 tumor fragments, obtained from donors at passage p8.

2cPE Toxicity Analysis. Animal's body weights were recorded for all animals three times a week from the first day of treatment. All animals were checked daily, and eventual behavioral changes, ill health, or mortality was recorded for each animal. On day 0, each animal was weighed and the volume of the dose was adjusted to ensure proper dosage in mg/kg. At the end of the experiments mice were sacrificed and basic autopsy was performed on all mice: thoracic and abdominal cavity was open, and all major organs were macroscopically examined. Livers from all animals were collected and weighted.

2cPE Administration to Mice. Mice were treated iv or sc every 4 days, for three times with 170 mg/kg of 2cPE, starting from day 16 after tumor fragments injection (day 0), when tumor size was 0.1 cm^3. Each animal received one dose of 2cEP (170 mg/kg) on day 0, another dose on day 4, and another dose on day 8. For both iv and sc, 2cPE was dissolved in PBS. Injected volume of compound solution was 200 μL for each treatment. Control mice received PBS. Before treatment, mice were randomly assigned to experimental groups ($n = 3$ with 4 mice/group). Tumor volume and body weights were recorded for all animals beginning from the day of first treatment and then twice or three times per week, until study termination. The percentage of tumor growth inhibition (TGI) was calculated according to the following equation:

$$\%\mathrm{TGI} = 100 - \left(\frac{\%\mathrm{TG,\ treated\ group}}{\%\mathrm{TG,\ control\ group}} \times 100\right)$$

■ ASSOCIATED CONTENT

Supporting Information

Additional figures illustrating cytotoxic activity, immunoblot analysis of selected markers, and 2cPE toxicity; details of animal studies, microarray data processing and analysis, modeling; additional analytical and spectral data; HPLC analysis of tested inhibitors; csv file containing SMILES identifiers of compounds discussed in the manuscript. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*F.B.: phone, +390405583920; e-mail, fberti@units.it.

*C.B.: phone, +39 0432494482; e-mail, claudio.brancolini@uniud.it.

Author Contributions

[‖]U.C. and A.S. contributed equally.

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by AIRC (Grant IG-10437) and FIRB (Progetto RBAP11S8C3_002) to C.B., by the Cross-Border Cooperation Program Italy−Slovenia 2007−2013, by the European Regional Development Fund and national funds to C.B. and P.S., by PRIN (Progetto 20109Z2XRJ_011), and by the University of Trieste (FRA2013) to F.B. A.T. received a fellowship from AIRC. We thank Cristina Degrassi (MRIT Lab) for assistance with the in vivo experiments and Dr. Pierluigi Polese (University of Udine) for the elemental analyses.

■ ABBREVIATIONS USED

BH3, BCL-2 homology domain 3; mCPBA, meta-chloroperoxybenzoic acid; DMAP, dimethylaminopyridine; DUB, deubiquitinating enzyme; EDC, *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide; GEO, Gene Expression Omnibus; GFP, green fluorescent protein; HDL, high density lipoprotein; HOBT, 1-hydroxybenzotriazole (hydrate); ISG15, interferon-stimulated gene 15; LDL, low density lipoprotein; MPEG5000, polyethylene glycol methyl ether MW 5000; N-SII, non-selective isopeptidase inhibitor; PAF-AH, platelet-activating factor acetyl hydrolase; PLA2G7, phospholipase A_2 group 7; SUMO, small ubiquitin-like modifier; TEA, triethylamine; TP53, tumor protein 53; Ub, ubiquitin; Ubl, ubiquitin-like; UCH, ubiquitin carboxy-terminal hydrolase; UPS, ubiquitin−proteasome system; USP, ubiquitin-specific protease

■ REFERENCES

(1) Chen, Z. J.; Sun, L. J. Nonproteolytic functions of ubiquitin in cell signaling. *Mol. Cell* **2009**, *33*, 275−286.

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

(2) van der Veen, A. G.; Ploegh, H. L. Ubiquitin-like proteins. *Annu. Rev. Biochem.* **2012**, *81*, 323−357.
(3) Ciechanover, A.; Stanhill, A. The complexity of recognition of ubiquitinated substrates by the 26S proteasome. *Biochim. Biophys. Acta* **2014**, *1843*, 86−96.
(4) Satija, Y. K.; Bhardwaj, A.; Das, S. A portrayal of E3 ubiquitin ligases and deubiquitylases in cancer. *Int. J. Cancer* **2013**, *133*, 2759−2768.
(5) Komander, D.; Clague, M. J.; Urbé, S. Breaking the chains: structure and function of the deubiquitinases. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 550−563.
(6) Eletr, Z. M.; Wilkinson, K. D. Regulation of proteolysis by human deubiquitinating enzymes. *Biochim. Biophys. Acta* **2014**, *1843*, 114−128.
(7) Chen, D.; Frezza, M.; Schmitt, S.; Kanwar, J.; Dou, K. P. Bortezomib as the first proteasome inhibitor anticancer drug: current status and future perspectives. *Curr. Cancer Drug Targets* **2011**, *11*, 239−253.
(8) Skrott, Z.; Cvek, B. Linking the activity of bortezomib in multiple myeloma and autoimmune diseases. *Crit. Rev. Oncol. Hematol.* **2014**, *9*, 61−70.
(9) Weathington, N. M.; Mallampalli, R. K. Emerging therapies targeting the ubiquitin proteasome system in cancer. *J. Clin. Invest.* **2014**, *124*, 6−12.
(10) da Silva, S. R.; Paiva, S. L.; Lukkarila, J. L.; Gunning, P. T. Exploring a new frontier in cancer treatment: targeting the ubiquitin and ubiquitin-like activating enzymes. *J. Med. Chem.* **2013**, *56*, 2165−2177.
(11) Sgorbissa, A.; Potu, H.; Brancolini, C. Isopeptidases in anticancer therapy: looking for inhibitors. *Am. J. Transl. Res.* **2010**, *2*, 235−247.
(12) Sippl, W.; Collura, V.; Colland, F. Ubiquitin-specific proteases as cancer drug targets. *Future Oncol.* **2011**, *7*, 619−632.
(13) Verbitski, S. M.; Mullally, J. E.; Fitzpatrick, F. A.; Ireland, C. M. Punaglandins, chlorinated prostaglandins, function as potent Michael receptors to inhibit ubiquitin isopeptidase activity. *J. Med. Chem.* **2004**, *47*, 2062−2070.
(14) Aleo, E.; Henderson, C. J.; Fontanini, A.; Solazzo, B.; Brancolini, C. Identification of new compounds that trigger apoptosome-independent caspase activation and apoptosis. *Cancer Res.* **2006**, *66*, 9235−9244.
(15) Kapuria, V.; Peterson, L. F.; Fang, D.; Bornmann, W. G.; Talpaz, M.; Donato, N. J. Deubiquitinase inhibition by small-molecule WP1130 triggers aggresome formation and tumor cell apoptosis. *Cancer Res.* **2010**, *70*, 9265−9276.
(16) D'Arcy, P.; Brnjic, S.; Olofsson, M. H.; Fryknäs, M.; Lindsten, K.; De Cesare, M.; Perego, P.; Sadeghi, B.; Hassan, M.; Larsson, R.; Linder, S. Inhibition of proteasome deubiquitinating activity as a new cancer therapy. *Nat. Med.* **2011**, *17*, 1636−1640.
(17) Coughlin, K.; Anchoori, R.; Iizuka, Y.; Meints, J.; MacNeill, L.; Vogel, R. I.; Orlowski, R. Z.; Lee, M. K.; Roden, R. B.; Bazzaro, M. Small-molecule RA-9 inhibits proteasome-associated DUBs and ovarian cancer in vitro and in vivo via exacerbating unfolded protein responses. *Clin. Cancer Res.* **2014**, *20*, 3174−3186.
(18) Lee, B. H.; Lee, M. J.; Park, S.; Oh, D. C.; Elsasser, S.; Chen, P. C.; Gartner, C.; Dimova, N.; Hanna, J.; Gygi, S. P.; Wilson, S. M.; King, R. W.; Finley, D. Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* **2010**, *467*, 179−184.
(19) Altun, M.; Kramer, H. B.; Willems, L. I.; McDermott, J. L.; Leach, C. A.; Goldenberg, S. J.; Kumar, K. G.; Konietzny, R.; Fischer, R.; Kogan, E.; Mackeen, M. M.; McGouran, J.; Khoronenkova, S. V.; Parsons, J. L.; Dianov, G. L.; Nicholson, B.; Kessler, B. M. Activity-based chemical proteomics accelerates inhibitor development for deubiquitylating enzymes. *Chem. Biol.* **2011**, *18*, 1401−1412.
(20) Reverdy, C.; Conrath, S.; Lopez, R.; Planquette, C.; Atmanene, C.; Collura, V.; Harpon, J.; Battaglia, V.; Vivat, V.; Sippl, W.; Colland, F. Discovery of specific inhibitors of human USP7/HAUSP deubiquitinating enzyme. *Chem. Biol.* **2012**, *19*, 467−77.
(21) Okada, K.; Ye, Y. Q.; Taniguchi, K.; Yoshida, A.; Akiyama, T.; Yoshioka, Y.; Onose, J.; Koshino, H.; Takahashi, S.; Yajima, A.; Abe, N.; Yajima, S. Vialinin A is a ubiquitin-specific peptidase inhibitor. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 4328−4331.
(22) Liang, Q.; Dexheimer, T. S.; Zhang, P.; Rosenthal, A. S.; Villamil, M. A.; You, C.; Zhang, Q.; Chen, J.; Ott, C. A.; Sun, H.; Luci, D. K.; Yuan, B.; Simeonov, A.; Jadhav, A.; Xiao, H.; Wang, Y.; Maloney, D. J.; Zhuang, Z. A selective USP1-UAF1 inhibitor links deubiquitination to DNA damage responses. *Nat. Chem. Biol.* **2014**, *10*, 298−304.
(23) Anchoori, R. K.; Karanam, B.; Peng, S.; Wang, J. W.; Jiang, R.; Tanno, T.; Orlowski, R. Z.; Matsui, W.; Zhao, M.; Rudek, M. A.; Hung, C.; Chen, X.; Walters, K. J.; Roden, R. B. S. A bis-benzylidine piperidone targeting proteasome ubiquitin receptor RPN13/ADRM1 as a therapy for cancer. *Cancer Cell* **2013**, *24*, 791−805.
(24) Foti, C.; Florean, C.; Pezzutto, A.; Roncaglia, P.; Tomasella, A.; Gustincich, S.; Brancolini, C. Characterization of caspase-dependent and caspase-independent deaths in glioblastoma cells treated with inhibitors of the ubiquitin-proteasome system. *Mol. Cancer Ther.* **2009**, *8*, 3140−3150.
(25) Fehenel, E. A.; Carmack, M. Studies in the thiapyran series. The preparation, properties and reactions of 1,4-thiapyrone-1-dioxide. *J. Am. Chem. Soc.* **1948**, *70*, 1813−1817.
(26) Puar, M. S.; Rovnyak, G. C.; Cohen, A. I.; Toeplitz, B.; Gougoutas, J. Z. Orientation of the sulfoxide bond as a stereochemical probe. Thiopyrano[4,3-*c*]pyrazoles. *J. Org. Chem.* **1979**, *44*, 2513−2518.
(27) Dimmock, J. R.; Padmanilayam, M. P.; Zello, G. a; Nienaber, K. H.; Allen, T. M.; Santos, C. L.; De Clercq, E.; Balzarini, J.; Manavathu, E. K.; Stables, J. P. Cytotoxic analogues of 2,6-bis(arylidene)-cyclohexanones. *Eur. J. Med. Chem.* **2003**, *38*, 169−177.
(28) Hansch, C.; Leo, A.; Taft, R. W. A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991**, *91*, 165−195.
(29) Das, U.; Sharma, R. K.; Dimmock, J. R. 1,5-Diaryl-3-oxo-1,4-pentadienes: a case for antineoplastics with multiple targets. *Curr. Med. Chem.* **2009**, *16*, 2001−2020.
(30) Dimmock, J. R.; Padmanilayam, M. P.; Puthucode, R. N.; Nazarali, A. J.; Motaganahalli, N. L.; Zello, G. A.; Quail, J. W.; Oloo, E. O.; Kraatz, H. B.; Prisciak, J. S.; Allen, T. M.; Santos, C. L.; Balzarini, J.; De Clercq, E.; Manavathu, E. K. A conformational and structure−activity relationship study of cytotoxic 3,5-bis(arylidene)-4-piperidones and related *N*-acryloyl analogues. *J. Med. Chem.* **2001**, *44*, 586−593.
(31) Dimmock, J. R.; Kandepu, N. M.; Nazarali, A. J.; Kowalchuk, T. P.; Motaganahalli, N.; Quail, J. W.; Mykytiuk, P. A.; Audette, G. F.; Prasad, L.; Perjési, P.; Allen, T. M.; Santos, C. L.; Szydlowski, J.; De Clercq, E.; Balzarini, J. Conformational and quantitative structure−activity relationship study of cytotoxic 2-arylidenebenzocycloalkanones. *J. Med. Chem.* **1999**, *42*, 1358−1366.
(32) Dimmock, J. R.; Sidhu, K. K.; Chen, M.; Li, J.; Quail, J. W.; Allen, T. M.; Kao, G. Y. Synthesis and cytotoxic evaluation of some cyclic arylidene ketones and related oximes, oxime esters, and analogs. *J. Pharm. Sci.* **1994**, *83*, 852−858.
(33) Fontanini, A.; Foti, C.; Potu, H.; Crivellato, E.; Maestro, R.; Bernardi, P.; Demarchi, F.; Brancolini, C. The isopeptidase inhibitor G5 triggers a caspase-independent necrotic death in cells resistant to apoptosis: a comparative study with the proteasome inhibitor bortezomib. *J. Biol. Chem.* **2009**, *284*, 8369−8381.
(34) Tomasella, A.; Blangy, A.; Brancolini, C. A receptor-interacting protein 1 (RIP1)-independent necrotic death under the control of protein phosphatase PP2A that involves the reorganization of actin cytoskeleton and the action of cofilin-1. *J. Biol. Chem.* **2014**, *289*, 25699−25710.
(35) Nicholson, B.; Leach, C. A.; Goldenberg, S. J.; Francis, D. M.; Kodrasov, M. P.; Tian, X.; Shanks, J.; Sterner, D. E.; Bernal, A.; Mattern, M. R.; Wilkinson, K. D.; Butt, T. R. Characterization of ubiquitin and ubiquitin-like-protein isopeptidase activities. *Protein Sci.* **2008**, *17*, 1035−1043.

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX

(36) Liu, Y.; Lashuel, H. A.; Choi, S.; Xing, X.; Case, A.; Ni, J.; Yeh, L. A.; Cuny, G. D.; Stein, R. L.; Lansbury, P. T., Jr. Discovery of inhibitors that elucidate the role of UCH-L1 activity in the H1299 lung cancer cell line. *Chem. Biol.* **2003**, *10*, 837−846.
(37) Sgorbissa, A.; Brancolini, C. IFNs, ISGylation and cancer: cui prodest? *Cytokine Growth Factor Rev.* **2012**, *23*, 307−314.
(38) Potu, H.; Sgorbissa, A.; Brancolini, C. Identification of USP18 as an important regulator of the susceptibility to IFN-alpha and drug-induced apoptosis. *Cancer Res.* **2010**, *70*, 655−665.
(39) Pasut, G.; Veronese, F. M. State of the art in PEGylation: the great versatility achieved after forty years of research. *J. Controlled Release* **2012**, *161*, 461−472.
(40) Pasut, G.; Veronese, F. M. PEG conjugates in clinical development or use as anticancer agents: an overview. *Adv. Drug Delivery Rev.* **2009**, *61*, 1177−1188.
(41) Safavy, A.; Raisch, K. P.; Mantena, S.; Sanford, L. L.; Sham, S. W.; Krishna, N. R.; Bonner, J. A. Design and development of water-soluble curcumin conjugates as potential anticancer agents. *J. Med. Chem.* **2007**, *50*, 6284−6288.
(42) Kolate, A.; Baradia, D.; Sushilkumar, P.; Vhora, I.; Kore, G.; Misra, A. PEG—a versatile ligand for drugs and drug delivery systems. *J. Controlled Release* **2014**, *192*, 67−81.
(43) (a) GEO (Gene Expression Omnibus). http://www.ncbi.nlm.nih.gov/geo/. (b) HomoloGene. http://www.ncbi.nlm.nih.gov/homologene.
(44) Hollingshead, M. G.; Stockwin, L. H.; Alcoser, S. Y.; Newton, D. L.; Orsburn, B. C.; Bonomi, C. A.; Borgel, S. D.; Divelbiss, R.; Dougherty, K. M.; Hager, E. J.; Holbeck, S. L.; Kaur, G.; Kimmel, D. J.; Kunkel, M. W.; Millione, A.; Mullendore, M. E.; Stotler, H.; Collins, J. Gene expression profiling of 49 human tumor xenografts from in vitro culture through multiple in vivo passages—strategies for data mining in support of therapeutic studies. *BMC Genomics* **2014**, *15*, 393.
(45) Stafforini, D. M.; McIntyre, T. M.; Zimmerman, G. A.; Prescott, S. M. Platelet-activating factor acetylhydrolases. *J. Biol. Chem.* **1997**, *272*, 17895−17898.
(46) Samanta, U.; Nahnson, B. J. Crystal structure of human plasma platelet-activating factor acetylhydrolase—structural implication to lipoprotein binding and catalysis. *J. Biol. Chem.* **2008**, *283*, 31617−31624.
(47) Vijayalakshmi, A.; KrishnaKumari, V.; Madhusudhana, R. N. Probing polyethylene glycol−phospholipid membrane interactions using enzymes. *J. Colloid Interface Sci.* **1999**, *219*, 190−194.
(48) Deslongchamps, P. *Stereoelectronic Effects in Organic Chemistry*; Pergamon Press: Oxford, U.K., 1984; pp 54−100.
(49) Berti, F.; Forzato, C.; Nitti, P.; Pitacco, G.; Valentin, E. A study of the enantiopreference of lipase PS (*Pseudomonas cepacia*) towards diastereomeric dihydro-5-alkyl-4-hydroxymethyl-2(3*H*)-furanones. *Tetrahedron: Asymmetry* **2005**, *16*, 1091−1102.
(50) See, for example, the following: (a) Yuhasz, S. C.; Parry, C.; Strand, M.; Amzel, L. M. Structural analysis of affinity maturation: the three-dimensional structures of complexes of an anti-nitrophenol antibody. *Mol. Immunol.* **1995**, *32*, 1143−1155. (b) Kleinstein, S. H.; Singh, J. P. Why are there so few key mutant clones? The influence of stochastic selection and blocking on affinity maturation in the germinal center. *Int. Immunol.* **2003**, *15*, 871−884. (c) Yang, P. L.; Schultz, P. G. Mutational analysis of the affinity maturation of antibody 48G7. *J. Mol. Biol.* **1999**, *294*, 1191−1201.
(51) Andresen, T. L.; Jensen, S. S.; Madsen, R.; Jørgensen, K. Synthesis and biological activity of anticancer ether lipids that are specifically released by phospholipase A2 in tumor tissue. *J. Med. Chem.* **2005**, *48*, 7305−7314.
(52) Andresen, T. L.; Davidsen, J.; Begtrup, M.; Mouritsen, O. G.; Jørgensen, K. Enzymatic release of antitumor ether lipids by specific phospholipase A2 activation of liposome-forming prodrugs. *J. Med. Chem.* **2004**, *47*, 1694−1703.
(53) Rosenson, R. S.; Stafforini, D. M. Modulation of oxidative stress, inflammation, and atherosclerosis by lipoprotein-associated phospholipase A2. *J. Lipid Res.* **2012**, *53*, 1767−1782.
(54) Leonard, N. J.; Choudhury, D. γ-Pyrones by isomerization. Substituted 3,5-dibenzyl-4*H*-pyran-4-ones. *J. Am. Chem. Soc.* **1957**, *79*, 156−160.
(55) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(56) Gautier, L.; Cope, L.; Bolstad, B. M.; Irizarry, R. A. Affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **2004**, *20*, 307−315.
(57) Dai, M.; Wang, P.; Boyd, A. D.; Kostov, G.; Athey, B.; Jones, E. G.; Bunney, W. E.; Myers, R. M.; Speed, T. P.; Akil, H.; Watson, S. J.; Meng, F. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **2005**, *33*, e175.

DOI: 10.1021/jm501336h
J. Med. Chem. XXXX, XXX, XXX−XXX