Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Metabolic Profiling of Cancer Cells and Correlations between Metabolism, Gene Expression and Drug Sensitivity

Diplom- Mariela Huichalaf Carbonell

Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Diplom- Mariela Huichalaf Carbonell

born in Santiago, Chile

Oral-examination 17 of October 2017

Metabolic Profiling of Cancer Cells and Correlations between Metabolism, Gene Expression and Drug Sensitivity

Referees: Prof. Dr. Stefan Wölfl
Prof. Dr. Tobias Werner

Table of Contents

List of Tables

List of Figures

Abbreviations

- ATP: Adenosine Triphosphate
- COX5B: Cytochrome C Oxidase Subunit 5b
- DHE: Dihydroethidium
- DMEM: Dulbecco's Modified Eagle Medium
- DMSO: Dimethyl sulfoxide
- DNA: Deoxyribonucleic Acid
- DPBS: Dulbecco’s Phosphate Buffered Saline
- F: Farads
- GCN1L1: EIF-2-Alpha Kinase Activator GCN1
- GDSC: Genomics of Drug Sensitivity in Cancer Project
- GEO: Gene Expression Omnibus
- GOT1: Glutamic-Oxaloacetic Transaminase 1
- HK1: Hexokinase 1
- HK2: Hexokinase 2
- HOXA7: Homeobox 7
- IDES sensor: Interdigitated Electrodes
- IFI16: Gamma-Interferon-Inducible Protein 16
- INPP5B: Inositol Polyphosphate-5-Phosphatase B
- ISFET sensors: Ion-Sensitive Field-Effect Transistors
- LDHA: Lactate Dehydrogenase A
- LGALS8: Galectin 8
- miRNA: microRNA
- MTT: 3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide
- NADH: Nicotinamide Adenine Dinucleotide
- NCBI: National Center for Biotechnology Information
- OXPHOS: Oxidative Phosphorylation
- PANP: Presence-Absence calls with Negative Probe sets
- PFKM: Phosphofructokinase-Muscle
- PGLS: 6-phosphogluconolactonase
- PPP: Pentose Phosphate Pathway
- qPCR: Real-Time Polymerase Chain Reaction
- r: Correlation Coefficient

- RHBDL2: Rhomboid-Related Protein 2
- ROS: Reactive Oxygen Species
- SLC2A1: Solute Carrier Family 2 Member 1
- SRB: Sulforhodamine B
- TCA: Tricarboxylic Acid

Summary

Detailed metabolic characterization of cancer cells provides an important cellular footprint in addition to well-studied genomic features and could play an important role for understanding sensitivity of cancer cells to drug treatment. Thus, detailed understanding of metabolism and gene expression profiles of cancer cells could be an important guideline supporting the steps of drug selection and a determining factor for designing treatment strategies with available anticancer drugs.

The aim of the presented study was to analyze the metabolic activities and cell-cell interaction and cell-matrix adhesion properties of a large number of cancer cell lines and correlate these cellular characteristics with drug sensitivity and compare this relation to established knowledge using gene expression and drug sensitivity prediction. For this, correlations among these different set of analytical information were analyzed to obtain cancer cell profiles linking metabolism, gene expression and drug IC_{50} values, by calculating Pearson correlation coefficients.

The results obtained clearly corroborate well-known relations but also provide important novel findings underlying the importance of cancer cell metabolism in drug sensitivity. As expected, the proliferation rate correlates well with the glycolytic activity of cancer cells ($r = 0.71^{**}$) as well as with HOXA7 expression ($r = 0.95^{***}$). The metabolism-related candidate genes are PGLS, COX5B, RHBDL2, IFI16, GCN1L1, INPP5B, and LGALS8. Furthermore, the calculated Pearson product-moment correlation coefficients show that the sensitivity of seven known drugs correlate with some of the analyzed cellular features of cancer cells. These compounds are elesclomol, Nutlin 3a, PF 4708671, EHT 1864, IPA3, RDEA119, and methotrexate.

The detailed metabolic and electrochemical profiling of cancer cells carried out for this study reveals new relationships between genes expression, drug sensitivity/tolerance, and basic metabolic features of cancer cells. New insights are provided into cancer metabolism, genomic regulation, and a potential application to designing new chemotherapeutic strategies.

Zusammenfassung

Die detaillierte metabolische Charakterisierung von Krebszellen liefert eine Reihe wichtiger Parameter zum zellulären Wachstumsverhalten, die bisher noch nicht genauer erfasst wurden. Diese metabolischen Informationen liefern eine wichtige Ergänzung zu bereits sehr gut untersuchten genomischen Veränderungen, die in Tumorzellen auftreten. Die metabolischen Informationen könnten daher eine neue ergänzende, wichtige Rolle für unser Verständnis zum Ansprechen der Krebszellen auf anti-tumor Wirkstoffe bieten. Langfristig könnte die Kombination aus genetischen Veränderungen, Genaktivität und Metabolismus ein wichtiger Leitfaden für die Auswahl geeigneter Wirkstoffen und ein bestimmender Faktor für die Planung und Festlegung der Therapiestrategie sein.

Das Ziel der vorliegenden Arbeit war es die metabolische Aktivität, Zell-Zell-Interkation und Zell-Matrix-Adhäsion-Eigenschaften einer großen Zahl von Tumorzelllinien zu bestimmen und diese mit anderen zellulären Eigenschaften wie Empfindlichkeit auf Wirkstoffe zu verknüpfen und diesen Zusammenhang mit bekannten Informationen zur Wirkung von Arzneistoffen in Abhängigkeit Zell-spezifischer Genexpressionsprofile zu vergleichen. Dazu wurden die Korrelationen zwischen diesen verschiedenen analytischen Datensätzen ermittelt und Tumorzell-spezifische Profile ermittelt, die Metabolismus, Genexpression und IC_{50} Werte der Wirkstoffe durch Berechnung von Pearson Korrelationskoeffizienten verknüpfen.

Die erhaltenen Ergebnisse unterstützen sehr gut bereits bekannte Beziehungen und liefern zusätzlich eine Reihe neuer Ergebnisse, die die Bedeutung des Zellmetabolismus für das Ansprechen von Tumorzellen auf Wirkstoffe bestätigen. Wie erwartet korreliert die Proliferationsrate sehr gut mit der glykolytischen Aktivität der Krebszellen (r = 0,71**) sowie mit der Expression von HOXA7 (r =0,95***). Neue hier gefundene Metabolismus-bezogene Markergene sind PGLS, COX5B, RHBDL2, IFI16, GCN1L1, INPP5B und LGALS8. Darüber hinaus zeigen die errechnete Pearson Produkt-Moment Korrelationskoeffizienten, dass das Ansprechen auf sieben bekannte Wirkstoffe sehr gut mit den zellulären Eigenschaften der Krebszellen übereinstimmt. Diese Wirkstoffe sind Elesclomol, Nutlin 3a, PF 4708671, EHT 1864, IPA3, RDE119 und Methotrexat.

Die hier vorgelegten detaillierten metabolischen und elektrochemischen Untersuchungen von Krebszellen zeigen bisher nicht bekannte Verknüpfungen zwischen Genexpression, Empfindlichkeit auf Wirkstoffe, und der grundlegenden metabolischen Aktivität von Krebszellen. Die Ergebnisse lassen auch erwarten, dass die neuen Einsichten in den Tumormetabolismus und dessen Verknüpfung mit genomischen Veränderungen eine vielversprechende Option für die Entwicklung neuer chemotherapeutischer Strategien liefern können.

1. Introduction

1.1 Cellular Features

1.1.1 Glycolysis and Gluconeogenesis

Glycolysis is an ancient pathway preserved in evolution and present in almost all living organisms (Berg et al. 2002, Zheng 2012). There are variations of glycolysis like the Entner-Doudoroff pathway that occurs in some bacteria and which produces just one molecule of ATP. Other organisms like phototrophs use an alternative glycolysis pathway where no ATP is produced at all and that simply serves to form carbon skeletons (Bar-Even et al. 2012). On the other hand, some microorganisms show variants of glycolysis with a high yield of ATP, producing three or four ATP molecules at the end of the pathway (Bar-Even et al. 2012). Glycolysis and variations are found in eukaryotes, bacteria, and archaeans (Romano & Conway 1996).

The term glycolysis comes from the Greek "to split sugar" in where the monosaccharide glucose is converted into two three-carbon molecules (pyruvate), two ATP, and two NADH molecules. Glycolysis takes place in the cytosol. The synthesis of glucose from pyruvate, lactic acid, or glycerol is called gluconeogenesis. It is used in animals and humans to balance the sugar level in the blood. Gluconeogenesis shares most of the enzymes with glycolysis but is not exactly the reverse process (Berg et al. 2002) and different enzymes in gluconeogenesis perform the irreversible steps in glycolysis. For example, the conversion of fructose 6-phosphate to fructose 1,6-phosphate is an irreversible step catalyzed by phosphofructokinase in glycolysis and the reversion is done by fructose biphosphatase in gluconeogenesis.

Elucidating the entire glycolysis pathway took more than 80 years and began with Louis Pasteur in 1860 who discovered that fermentation always involves a life process, either by yeast or some other microorganism. The Buchner brothers accidentally discover that fermentation can also occur by yeast extracts, meaning that the process can also take place outside a living cell. Gustav Embden eventually compiled the contributions of generations of researchers and elucidated the complete pathway in 1940.

The aim of glycolysis is to provide energy to the cells and although it is not an efficient way to produce ATP, it is the only way for cells to do not have mitochondria, like erythrocytes, and the most used pathway for this purpose for some other cells like

those derived from brain, skin, and cancer cells. In glycolysis, just two ATP molecules are synthesized in comparison with the 36 ATPs produced by oxidative phosphorylation (OXPHOS). OXPHOS is the process by which the cells produce ATP from ADP plus inorganic phosphate in mitochondria. Cancer cells are estimated to use glycolysis around 200 times more than their normal counterpart cells (Akram 2013). This feature that cancer glycolysis is much more active than that of healthy tissue is called the Warburg effect. In 1930, Otto Warburg described this phenomenon and proposed that the mitochondria of cancer cells are dysfunctional and that this could be the main reason for cancer abnormalities, rather than the increase in proliferation (Akram 2013, Zheng 2012). Recent investigations have demonstrated that the enzymes in OXPHOS are functional in most tumors and cancer cell lines do not carry mutations that disengage their performance (Vander et al. 2009, Zheng 2012).

The increase in glycolytic activity by cancer cells has been shown to be regulated by oncogenes, the inactivation of tumor suppressors, mutations in the mitochondrial DNA, and the rapid proliferation of cancer cells. In solid tumors, generation of new vessels (angiogenesis) is limited due to the fast tumor growth and therefore cells suffer from hypoxic conditions (Zheng 2012). It is also known that genes involved in glycolysis are overexpressed in many cancers. In 70% of human cancers there is an overexpression of glycolysis genes especially in prostate, brain, and lymph node cancers where almost all of the mentioned pathway genes are highly expressed (Altenberg & Greulich 2004). The genes that are overexpressed in most cancer types are the ones for encoding glyceraldehyde-3-phosphate dehydrogenase, enolase 1, and pyruvate kinase (Altenberg & Greulich 2004). On the other hand, many of the glycolytic enzymes have pro-cancer non-glycolytic functions, an example being hexokinase and phosphoglucose isomerase that have antiapoptotic functions; phosphoglucose isomerase together with glyceraldehyde-3-phosphate dehydrogenase have a positive role in survival pathways; phosphofructokinase 1 and triose phosphate isomerase activate the cell cycle and aldolase promotes epithelial to mesenchymal transition (Lincet & Icard 2015).

It is important to highlight that aerobic glycolysis (the type of glycolysis that occurs both in normoxic or hypoxic conditions) acts together with OXPHOS, cooperating and keeping a balance to fulfill the energy demand of the cell. The metabolism of cancer cells differs markedly from the metabolism of normal tissues. Cancer cells have a higher level of plasticity and are more heterogeneous, being able to adapt themselves to

adverse conditions and different media giving them a survival advantage in comparison with normal cells.

1.1.2 The Contribution of Mitochondria to Cellular Respiration

Mitochondria are organelles responsible for energy production of a cell, and are involved in cell signaling and regulation of the cell cycle and growth as well as apoptosis (Yang et al. 2016). It has been shown that the mitochondria of cancer cells are different in function and structure as compared to normal cells (Yang et al. 2016). In healthy cells under normal O_2 environments, between 70% and 90% of the energy is supplied by the mitochondrial OXPHOS. Under hypoxic conditions, that most cancerous cells are exposed to, ATP is mainly produced via glycolysis and the pentose phosphate pathways (Zheng 2012, Yang et al. 2016). Mutations in mitochondrial DNA has been reported to play an important role for the transformation of normal cells into cancerous ones, as well as cancer progression and drug resistance that lead to an increase of cellular glycolytic activity (Rogalinska 2016).

Cellular respiration consists of a series of biochemical reactions taking place within the cells to produce energy as ATP from glucose, fatty acids, and amino acids. Respiration starts in the cytosol with the production of pyruvate from glucose in glycolysis. In low oxygen microenvironments, pyruvate will be reduced to lactic acid. Under normoxic conditions, pyruvate will be oxidized into acetyl-CoA plus one molecule of CO_2 and one molecule of NADH in the mitochondria by the pyruvate dehydrogenase complex. Acetyl-CoA in combination with oxaloacetate in the tricarboxylic acid (TCA) cycle will be further oxidized to CO_2 at the same time that NAD is reduced to NADH. The NADH molecules enter the electron transport chain to synthesize more ATP.

In cancer cells within tumors metabolism greatly varies. It has been demonstrated that not all cancer types follow a Warburg effect, and they show great heterogeneity within the same tumor (Hensley et al. 2016). In recent years it has been shown that some cancers use principally the OXPHOS to obtain their energy (Corbet & Feron 2017, Zheng 2012). The substrate necessary for the TCA cycle is provided from different sources such as amino acids and fatty acids, and it is possible to simultaneously observe elevated levels of fermentation and mitochondrial respiration (Hensley et al. 2016). During the last decades several studies have suggested that cancer cells rely on glycolysis to obtain ATP due to damaged and nonfunctional OXPHOS (Warburg effect)

and Otto Warburg to some extend misinterpreted his observations (Koppenol et al. 2011).

Moreover, nowadays it is known that in some cancers mitochondrial metabolism is necessary for tumorigenesis and to support tumor progression (Corbet & Feron 2017). Mutations of the TCA cycle enzymes lead to a formation of oncometabolites (TCA intermediaries, like citrate) that are responsible for the inhibition of the tumor suppressor p53, the stabilization of hypoxia-inducible factor, the activation of cell growth pathways, glycolysis, and reactive oxygen species (ROS) formation (Sajnani et al. 2017). The coordinated redox balance existing in normal tissues is out of control in cancer cells. Mitochondrial DNA mutations can elevate ROS production and high ROS levels induce genetic instability responsible for mutations (Yang et al. 2016, Okon & Zou 2015, Kang et al. 2015). In normal cells the production of ROS is proportional to the activity of the electron transport chain, and as mentioned earlier, is controlled by an antioxidant system that keeps the necessary amount of ROS high enough to function in protein regulation and signaling (Okon & Zou 2015, Kang et al. 2015). The elevated ROS in cancer cells has multiple functions apart from the induction of mitochondrial DNA mutations. The impact of ROS in cancer will depend on *i*) the levels of ROS, *ii*) the type of tumor and tissue, *iii*) the disease stage, and *iv*) the type/duration of treatment (Okon & Zou 2015). Elevated ROS levels have been implicated in cancer initiation, progression, and drug resistance (Okon & Zou 2015). Among ROS functions, elevated ROS accumulation has been attributed to the stabilization of the hypoxia-inducible factor 1A creating a pseudohypoxic environment under normoxic conditions, the hyperactivation of oncogenes to promote cancer development, and tumorigenesis mediated by K-ras, cancer cell proliferation and survival (Yang et al. 2016).

1.1.3 Cellular Proliferation

Cellular proliferation is the process by which cells grow and divide resulting in an increasing number of cells. The proliferation rate in adults is in equilibrium with cell loss by cell differentiation and cell death. A key characteristic of tumors is the uncontrolled nonregulated proliferation rate by which the signals that control a normal cell division behavior are no longer properly working. The uncontrolled proliferation of cells is considered one of the first symptoms in cancer development. A continued noninvasive proliferation of cells without spreading into distant tissues is observed in benign tumors

(Cooper & Hausman 2000). Malignant tumors (cancers) are the ones that apart of present abnormal and uncontrolled proliferations are also capable of invading the surrounding tissue and migrating to distant sites (Cooper & Hausman 2000). Several biomarkers connected to cell proliferation are widely used in the clinical context to characterize human cancers as to tumor grade, therapy responses, and prognosis (Wood et al. 2015).

Most of these upregulated proteins (biomarkers) used for therapeutic application are needed for efficient uptake of nutrients for producing the necessary biomass for new cells in fast-proliferating tumor cells. Evidence demonstrates that several signaling pathways that regulate proliferation do also regulate cell metabolism in cancer cells regarding the uptake and incorporation of nutrients for ATP production (Vander Heiden et al. 2009, Agathocleous & Harris 2013). Some of the pathways involved in regulating proliferation and metabolism are the mTOR, RAS, MYC, PI3K, and Hedgehog pathways (Agathocleous & Harris 2013). Changes in metabolism observed in cancer cells (Warburg effect and disruption of the redox balance, TCA cycle enzymes and lipid and amino acid metabolism) lead to utilization of metabolites for energy-producing pathways and biosynthetic anabolism for cell growth and proliferation (Agathocleous & Harris 2013). This behavior can also be observed in normal cells that display a Warburg effect mode in the production of ATP when stimulated to proliferate (Brand & Hermfisse 1997). This is not surprising if one takes into consideration that glycolysis serves as a starting point for pathways leading to the synthesis of nucleic acids, amino acids, and polysaccharides, all of which needed as building blocks in cell proliferation (Toyoda & Saitoh 2015).

1.1.4 Membrane Capacitance

Capacitance is defined as the ability to store electrical energy. Cells in their conductive electrolyte solution media (water plus ions) act as capacitors being able to store electricity (Mosgaard et al. 2015). A capacitor is formed by two conductive plates separated by a dielectric (an insulating material). Cell membranes act as an insulator (nonconductive material) and the inner and outer space of the cell as conductive plates with free ions that can transport electrical charges (Niebur 2008). A completely sealed membrane or insulator does not allow the passage of current from one side to the other. Artificial membranes built of pure phospholipid bilayers are quite good insulators. They

are highly impermeable to polar solutes, meaning that there are no ions that can transport charges across the hydrophobic central core of the bilayer (Niebur 2008). Nevertheless, natural membranes in cells are made up not only of phospholipids but also of different ion transporters and transmembrane proteins. The transmembrane ion transporter proteins that act as ion channels need, in order to cross the polar solutes through the membrane without a solvent, to build special interactions with the ions, and organize themselves to be selective and efficient in the transport (Fyles 2007). The composition of the lipid bilayer and the characteristics of the ion transporters give the cells different electrical properties. Depending on the cellular function, ion transporters differ in their distribution and quantity in different cell lines (Jones 2012).

Cell membranes themselves can also act as capacitors (Mosgaard et al. 2015). In this model, the hydrophobic central core of the phospholipid bilayer is the nonconductive material.

Capacitors can be charged connecting a cathode to one of the conductive plates and the anode to the second conductive plate producing a flow of electron (current). Cells can be charged applying a current across the membrane. The two conductive plates or in the case of a cell, the lipid bilayer, on one side, will be positively charged and, on the other side, negatively charged. For example, during the nerve impulse, neuron membranes are transiently charged due to the flow of current (charged molecules of Na^+ and K^+) through the ion channels (Mosgaard et al. 2015). Depending on the structure and composition of the cell membranes, different amounts of electrical charges can be stored. The charge stored by the capacitor is measured in Farads (F).

Cell membranes differ due to the particular lipid concentration, the number of lipid-rich microdomains, fatty acid composition and cell surface proteins, their content and degree of modification like by glycosylation (Christiansen et al. 2014). The plasma membrane is a complex structure and in charge not only for maintaining particular environments inside and outside of the cell (due to its function as a barrier and ion transporter), it also is a key component in cell signaling giving feedback between the extracellular and intracellular media (exocytosis and endocytosis, protein secretion, membrane receptors, etc). It is also implicated in compartmentalization and organization of the inside of organelles forming an endomembrane system with the Golgi apparatus, endoplasmic reticulum, nuclear membrane, and lysosomes (O'Connor et al. 2010) and for cell interactions, being responsible for cell-cell binding and cell-matrix contact in cell tissue organization. The plasma membrane's complexity in cells is

so important that it has been referred to as the capacitor for energy and metabolism (Ray et al. 2016). The different structures that form and provide the complexity of the plasma membrane are attributed to play specific rolls in the electrical properties of cells and tissues (Eisenberg 1980). Therefore, one can postulate that the insulating properties of the membranes, and in consequence their capacitance abilities, are unique characteristics for each cell line.

1.2 Public Data Repositories

A public data repository is an online database service used for long-term storage, preservation, archiving, managing, sharing, and accessing research data (Uzwyshyn 2016). The aim of a public repository is to provide benefit by sharing and improving knowledge allowing faster research and discoveries (Uzwyshyn 2016). "*The ability to build upon the findings of others is one of the cornerstones of science*" (Joung et al. 2015). To do research using the findings of others it is necessary that this information is made available and this information must be easily and permanently accessible and trustworthy. Therefore, data repositories must have the resources to secure the long-term preservation of the data, update the data, and provide a user-friendly platform to upload, download, and search the data, and to provide stable and resolvable identifiers and above all, the necessary technology for the data curation (Lemberger 2015).

Since the development of new sequencing technologies and microarrays, huge amounts of genome data have been accumulated in public archives. Every year more and more big data are generated that give information about genome evolution, gene regulation, cancer genetics and their effects on patients, drugs targets, etc. The explosion of data started in 2009 and in 2012 the amount of data produced was 4 times that produced in 2011 (Marx 2013). By the end of 2015, the European Bioinformatics Institute (EMBL-EBI), one of the largest data repositories for biology, had stored 75 petabytes (1 petabyte is 10^{15} bytes) of data and backups of genes, proteins, and small molecules (Cook et al. 2016); being in 2013 just 20 petabytes (Marx 2013). The data generated is continuously increasing nowadays. With the low price of high-throughput technology almost every small laboratory can be a big data generator and of course thanks to the online repositories a big data user (Marx 2013). The data curation process and the maintaining and development of the tools necessary to deal with the data

explosion require that new types of professionals and carriers evolve; data scientist, data archivist, data visualization experts, data curators, among others, will be more and more required (Uzwyshyn 2016).

Knowledge of gene expression has dramatic increased since the utilization of high-density immobilized oligonucleotide probes in arrays (Illumina, Affymetrix) (Rung & Brazma 2013) and the RNA sequencing (RNA-Seq) technique. Due to the requirements of being able to deal with and handle huge volumes of computational data, bioinformatics has become a central and important area of biology (Rung & Brazma 2013). Microarray analyses require expertise and computational resources based on the development of new tools and methods. Biologists benefit from such environments as Bioconductor (Huber et al. 2015) provided by the bioinformatics community to take advantage of the gene expression data generated and freely available in public repositories for microarray data like the Gene Expression Omnibus (GEO) of the U.S. National Center for Biotechnology Information (NCBI) or the ArrayExpress of EBI (Rung & Brazma 2013). To date, GEO contains more than 2 million samples freely available with raw data that every researcher in the world can work with (GEO, NCBI accessed 12 April 2017). A sample is an array measurement, for example, the gene expression of the complete genome of a determined cell line using one of the Affymetrix platforms. Other public repositories like The Genomics of Drug Sensitivity in Cancer Project (GDSC) funded by the Wellcome Trust Sanger Institute or The Cancer Imaging Archive supported by the NIH (U.S. National Cancer Institute) are focused on the improvement of cancer treatment. The GDSC is a freely available resource with curated data for reuse by the academic and medical communities. GDSC is performing a large-scale drug screen with compounds that include cytotoxic chemotherapeutics and targeted therapeutics from commercial sources, academic collaborators, and from the biotech and pharmaceutical industries. The drugs are tested in more than 1000 characterized human cancer cell lines and an incorporation of genomic analyses to identify drug response biomarkers and correlate them with the drug sensitivity patterns are in continued progress (Yang et al. 2012).

There is an enormous scientific and economic cost saving by the general access to previously published data. It allows the reanalysis of existing data to address different questions and the possibility to integrate and combine different data types giving enormous perspectives for leading research forward.

1.3 Correlations in Cellular Biology

Correlation is a statistical method used to evaluate the presence or not of a linear relationship between two variables and is measured by the correlation coefficient (r) (Mukaka 2012). The correlation coefficient gives the strength of the association. The calculated r is a dimensionless quantity between –1 and 1, with –1 a perfect negative correlation, 1 a perfect positive correlation, and 0 indicating the absence of a relationship among the two variables. A negative r indicates that the two variables are indirectly related, meaning that when the value of one variable goes up, the value of the other one goes down. On the other hand, a positive r shows that the variables are directly related: when the value of one variable goes up, the value of the second variable behaves the same (Fenton & Neil 2012, Mukaka 2012, McDonald 2009). In statistical terms, any other form of relation that does not follow linearity is not a correlation (Mukaka 2012). The calculation of a correlation or a linear regression needs two measurement variables (the ones that will be tested in order to find or not a linear relation) and a nominal variable. The nominal variable is not used directly for the calculation but is necessary to attribute the common measurements to the two measurement variables putting them together, for example a patient ID, a gene name, a cell line, etc. (McDonald 2009).

There are mainly two ways to calculate the correlation coefficients: the Pearson's product moment correlation coefficient and the Spearman's rank correlation coefficient. Pearson correlation is the most common statistical test used to evaluate the presence of a linear relationship between two variables and is the one used when the variables are normal distributed (data distributed in a bell-shape curve). Otherwise, Spearman's rank correlation is used when the data show skewness (non-normal distributed data with the presence of an asymmetry in the distribution), or one of the variables is ordinal (categorical) and is robust when outliers are present (Mukaka 2012, McDonald 2009). However, controversial information is found about the better performance test in non-normal distributed data. While some authors indicate that Pearson's formula is inadequate when one or both variables do not distribute normal (Mukaka 2012) others point out that Pearson's formula is sufficiently robust to deal with non-normal data and that should not be worried about when using regression or correlation coefficients (McDonald 2009). As stated by Chok (2010): "*Pearson's correlation coefficient could have significant advantages for continuous non-normal data, which does not have obvious*

outliers. Thus, the shape of the distribution should not be a sole reason for not using the Pearson product moment correlation coefficient". Moreover, Bishara & Hittner (2012) tested 12 different methods in non-normal distributed data to calculate correlations and concluded that Pearson's formula performs better after data transformation than non-parametric statistics like Spearman's. Important to highlight that when outliers are present, type I and II errors appear with the Pearson algorithm but there are ways to solve that. An increase in the robustness of Pearson in non-normal distributed data with outliers is done excluding the outlier or transforming the data to a normal shape before running Pearson correlations (Bishara & Hittner 2012, McDonald 2009). A transformation can be done using log-transformed data, natural log-transformed data, transformation to rank scores, square-root transformation, etc. (Bishara & Hittner 2012). Existing infinitive ways to transform the data to perform a properly statistical analysis of the data leaves them to follow a normal distribution (McDonald 2009).

In Pearson's product moment correlation, the null hypothesis (H_0) is that no relation between the two variables is present, meaning that the r coefficient is equal to 0, therefore a positive or negative r indicates a direct or inverse relation, respectively. The significance of the correlation analysis is calculated using the p-value. H_0 is rejected when the p-value is less than 0.05. A significant level (p-value) of 0.05 indicates the risk of finding a correlation when no relation exists, in percentages this means that there is a 5% probability of getting a relation when no relation actually exists (Fenton & Neil 2012). Rejecting the null hypothesis when it actually is true is called type-I error or significance level (false positive or to find a relation when no relation exists). A type-II error is to accept H_0 when it is, in fact, false (false negative or to not find an existing relation).

One important and widespread use of correlation coefficients in cellular biology is the use of Pearson's correlations on fluorescence microscopy as a statistic to quantify colocalization. Karl Pearson developed the correlation coefficient in 1896 and almost 100 years later Manders introduced its use in microscopy (Dunn et al. 2011). For evaluating the colocalization of two or more different fluorescence labels, almost all image analysis software uses Pearson coefficients to correlate the overlap of the pixels measuring the covariance in the signal level of the images (Dunn et al. 2011). Pearson correlation has been used also for searching relations between gene expression and cancer (Győrffy et al. 2016, Liu et al. 2014). Győrffy et al. used Pearson correlation analysis to search correlations between gene expression, breast cancer prognosis, and

the DNA methylation of genes. They found 48 genes that were significantly associated with breast cancer prognosis ($p < 0.05$) and 32 of these prognostic genes showed a direct correlation between its expression and its DNA methylation confirming that epigenetic aberrations like DNA methylations are associated with breast cancer features (Fleisher et al. 2014). Liu et al. (2014) performed gene expression analyses of hepatocellular carcinoma tissues and found that miR-122 levels correlated with glycolytic genes and that pyruvate kinase showed a highly significant negative correlation with the miRNA ($p < 0.0001$). Furthermore, Pearson correlation analysis has been used to search transcription factor target genes. The expression levels of a given transcription factor is correlated with the expression of candidate target genes. The relations obtained are corroborated by biochemical and functional analysis (Mansson et al. 2004). Moreover, gene expression levels also have been correlated using Pearson correlation on chemosensitivity in order to find candidate genes associated with a sensitivity of anticancer drugs in cancer cell lines (Nakamura et al. 2009). Other important uses of Pearson correlations are in attempts to find improvement in drug treatment of cancer, like in the prediction of anticancer drug responses using information of cell line similarity network data and drug similarity network data from the Cancer Cell Line Encyclopedia and Cancer Genome Project public databases (Zhang et al. 2015).

The aim of this study has been to perform a large number of assays in different human cell lines to characterize the inherent electrical and metabolic properties. Data on cellular features were used to search for correlations between gene expression of the entire genome and drug sensitivity of the cells. In addition to confirming some previously well-known relations, we found a sizeable amount of new relationships between cellular features and new candidate genes that may involve in cancer metabolism. Moreover, the sensitivity of cancer cells to widespread drugs was associated with certain cellular properties, giving further insights for possible improvements in cancer drug treatment.

2. Experimental Methods

2.1 Cellular Properties Acquisition

2.1.1 Cell Cultures

Human cell lines were cultivated under standard conditions in Dulbecco's modified eagle medium (DMEM) (PAA Laboratories) supplemented with 10% FBS (v/v) (PAA Laboratories) and 1% PenStrep (10,000 units/ml penicillin and 10,000 units/ml streptomycin) (Gibco Invitrogen) in a 5% CO_2 incubator at 37°C with a 95% humidity. When confluence arrived cells were split under sterile conditions. The detachment of the cells from the flask was done by using a trypsin substitute TrypLE Express (Thermo Fisher), afterwards washed with Dulbecco's phosphate buffered saline (DPBS) (Thermo Fisher) and recultivated, seeded or directly used for further analysis. The cell lines used in this study are shown in Table 1.

Table 1. List of cell lines used in this work

Cell line	Tissue	Disease	Cell type
MCF7	Breast	Adenocarcinoma	Epithelial
MDA-MB-231	Breast	Adenocarcinoma	Epithelial
HCT 116	Colon	Carcinoma	Epithelial
HT-29	Colon	Carcinoma	Epithelial
LS174T	Colon	Adenocarcinoma	Epithelial
AsPC-1	Pancreas	Adenocarcinoma	*
BxPC-3	Pancreas	Adenocarcinoma	Epithelial
JoPaca-1	Pancreas	Carcinoma	*
MIA PaCa-2	Pancreas	Carcinoma	Epithelial
Panc-1	Pancreas	Carcinoma	Epithelial
LNCaP	Prostate	Adenocarcinoma	Epithelial
PC-3	Prostate	Adenocarcinoma	Epithelial
DU 145	Prostate	Carcinoma	*
HaCaT	Skin	No malignancy	Keratinocyte
MDA-MB-435	Skin	Melanoma	Melanocyte
HeLa	Cervix	Adenocarcinoma	Epithelial
Hep G2	Liver	Carcinoma	Epithelial
1321N1	Central nervous system	Astrocytoma	Glial

*Not specified

2.1.2 Online Monitoring of Membrane Capacitance and Cellular Metabolism

Bionas Discovery 2500 instrument (Bionas GmbH) was used to measure the glycolytic activity, respiration, and the capacitance of cells seeded in sensor chips (Bionas Discovery SC1000) in real time (Fig. 1).

The glycolytic and respiratory activities are indirectly measured as pH changes and oxygen depletion in the media, respectively. The capacitance is measured directly by the sensor. To estimate glycolysis, respiration, and to measure capacitance each Bionas sensor chips contains five pH ISFET sensors (ion-sensitive field-effect transistors) (Bergveld 1981), two Clark-type electrode sensors (Clark et al. 1953) and one IDES sensor (Interdigitated Electrodes) (Ehret et al. 1997).

The **pH ISFET sensors measure the pH changes** by detecting ion concentrations in solution. The current between the reference electrode and the electrode in the solution changes according to the ion concentration in the media. Acidification in cancer cells is mainly due to lactic acid production at the end of the glycolysis.

The **Clark sensor measures the free oxygen** in the media. A fixed voltage is applied to the sensor, which contains a cathode and an anode. In the platinum cathode, the oxygen in the media is reduced while silver is oxidized in the anode generating a current. The current that is produced is proportional to the oxygen reduced.

The **membrane capacitance of the cells is measured by the IDES sensor**. Alterations in the capacitance value of the cells are the product of the changes in cellular morphology, adhesion/confluence, cell-cell contacts and cell-matrix binding. The IDES sensor contains an interdigitated palladium electrode pair in which an alternating current is applied. The flow of the current through the two electrodes is interrupted by the cells that adhere to the sensor surface, and such interruption is recorded.

The Bionas 2500 system allows reading of six independent chips. The influx of new media and an outflux of the used media are done in the fluid head of "biomodules" (Fig. 1). In each chip a different cell line can be seeded. The perfusion system supplies the cells with a constant flow of medium (running medium) and enables to do noninvasive, label-free measurements. The Bionas running medium consists of ultrapure water (Werner Reinstwassersysteme) with 1mM HEPES, 1 g/l glucose, 10

mg/l phenol red, 8.22 g/l DMEM powder, 1% PenStrept, 0.1% FCS and 2 mM L-glutamine. The pH was adjusted to 7.4 and the running medium was sterilized by filtration with a 0.2 µm hydrophobic filter (Millipore).

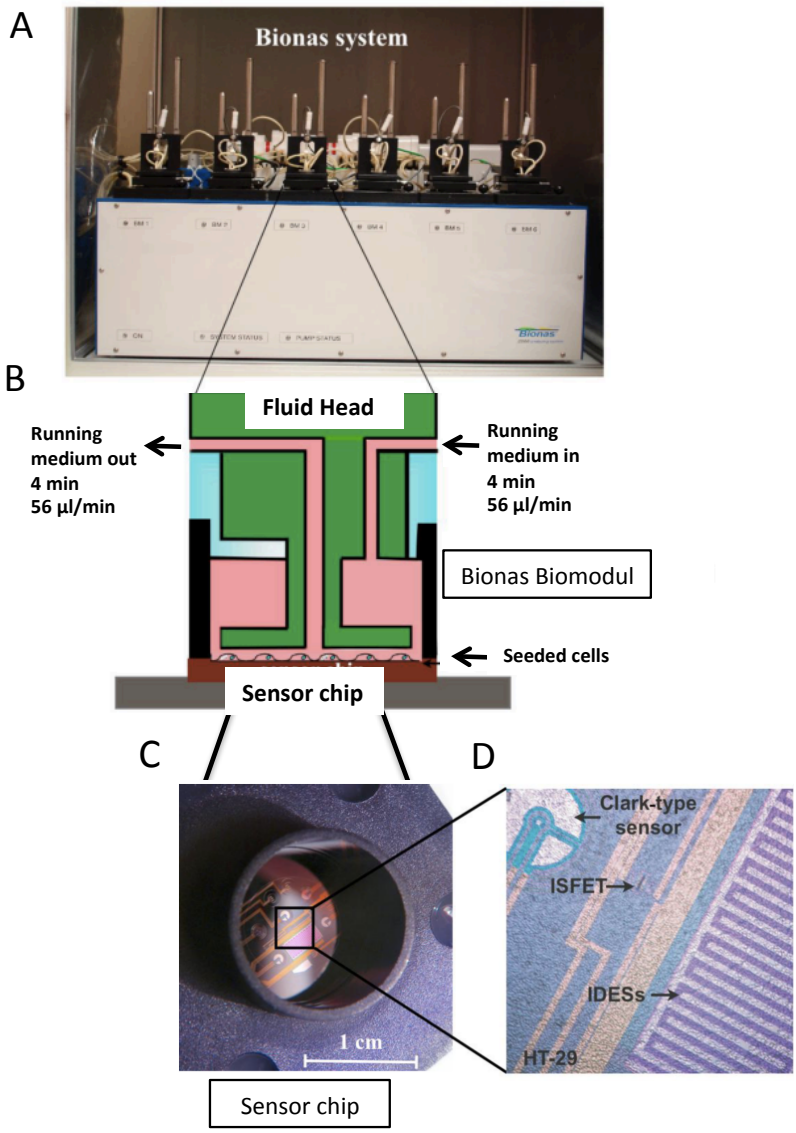


Fig. 1. Real-time monitoring of cell properties with the Bionas Discovery 2500 instrument and the Bionas Discovery SC1000 chip. A. The main body of the instrument with its six biomoduls and the tubes for perfusion. The pump system is not shown. **B.** Schematic view of a single biomodul. The chip with the cell monolayer is located at the bottom, with the fluid-head laying on top of the chip. During the measurement the system performs cycles of 4 min "go" and 4 min "stop" steps. In the 4 min "go" step, the cells are continually fed with fresh running medium. Running medium is injected into the chips through the fluid head, with a flow rate of 56 µl/min, going out continuously. In the second step of the cycle, the constant perfusion stops, leaving the cells with 6 µl of running medium for 4 min. **C.** Sensor chip. Each sensor chip contains five pH ISFET sensors (ion-sensitive field-effect transistors) for pH measurements, two Clark-type electrode sensors to measure oxygen concentration in the medium and one IDES sensor (interdigitated electrodes) for cellular capacitance measurements. **D.** Amplified view of a chip. A monolayer of HT-29 cells grown over the ISFET, Clark type and IDES sensors. Modified from Alborzinia et al. 2011.

2.1.2.1 Sensor Chip Seeding

Confluent cell cultures were used to seed ~200,000 cells per chip in 450 µl DMEM medium supplemented with 10% FCS and 1% PenStrept. Before seeding, the chips were sterilized with ethanol 70% for 20 min and washed with PBS three times. Chips with cells were incubated overnight with 5% CO_2 at 37°C and 95% humidity in a cell culture incubator.

2.1.2.2 Measurements

Before the cell-loaded sensor chips were placed in the system and the measurement program started, the Bionas tubes were disinfected with ethanol 70%. The disinfection program consists in sterilization with a continuous ethanol 70% flow of 12 min, followed by 4 min of rinsing with PBS. Finally, the tubes were filled with running medium. The measurement program performs a cycle of 4 min "go" and 4 min "stop" with running medium. In the 4 min "go" step the system injects running medium with a constant flow of 56 µl/min and in the "stop" step leaves the cells with ~6 µl of medium. At that point, the cells will consume oxygen and acidify the media. At the end of each cycle a measurement of the acidification of the media (as voltage in time µV/s), the cellular oxygen consumption (as current in time pA/s), and the cellular capacitance in nF are taken. To prevent contamination, a cleaning program was run at the end of each experiment. This program consists of 30 min constant flow of ethanol 70%. Finally, the cleaning program removes all the liquid from the tubes by pumping air.

2.1.2.3 Data Collection and Analysis

Data for the glycolytic activity estimation, cellular respiration, and cellular capacitance were obtained from the Bionas analyzer data file. Selected data contain measurements of human cell lines, cultivated under standard conditions (in DMEM supplemented with 10% FCS and 1% PenStrept in a 5% CO_2 incubator at 37°C) run with a 4 min "go" and 4 min "stop" cycle, and for which at least 4 independent experiments were performed. Eighteen cell lines met the quality requirements and the total data

generated by the sensors from ~1,300 chips was used for the analysis. The analyzer data files contain sensor measurements at a time point in which the cells showed stable cellular metabolism. This status was achieved around 4 h from the start of the measurement program. The analyzer data files were collected with the *R programming language for statistical computing and graphics* (Team, R. Core 2015) versions 2.7.0 and 3.2.1 (Supplements 5.1.1).

2.1.3 End Point Assays to Measure Cellular Metabolism

2.1.3.1 Adenosine Triphosphate (ATP) Production

The amount of extracellular ATP in the different cell lines was measured using ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer) and read in a Tecan ULTRA plate-reader. The ATPlite 1step estimates the amount of ATP based on the firefly (*Photinus pyralis*) luciferase activity. Light emitted by the ATP + luciferase + D-luciferin reaction is proportional to the ATP concentration.

The measurements were done according to the manufacturer's instructions. Briefly, cells were seeded in 96-well plates and ~10,000 cells per well were used. Plates were incubated under standard conditions for 24 h. Before adding the ATP reagent the plates were taken out from the incubator and kept at room temperature for 30 min. To each well, 100 µl of DMEM (without FCS) plus 100 µl of the ATP reagent was added. The cell suspension was shook for 2 min in the dark at 700 rpm in a thermo-mixer, and the luminescence was measured in a Tecan Safire microplate reader. Per cell line, three technical replicates and at least three independent measurements were carried out.

2.1.3.2 Mitochondrial Mass Content

The mitochondrial mass content of the cells in the different cell lines was estimated using MitoTracker Green = benzoxazolium, 2-[3-[5,6-dichloro-1,3-bis[[4-(chloromethyl) phenyl]methyl]-1,3-dihydro-2*H*-benzimidazol-2-ylidene]-1-propenyl]-3-methyl chloride 201860-17-5 (Thermo Fisher). MitoTracker diffuses into the cells and labels active mitochondria. Cells were trypsinized for 15 to 20 min until getting single

cells. The trypsinization was stopped with DMEM and ~250,000 cells were used to estimate the mitochondrial mass. The cells were centrifuged for 3 min at 1,600 rpm (Heraeus Multifuge Thermo Electronic) and washed two times with DPBS. Prewarmed (37°C) MitoTracker Green at a final concentration of 7.5 nM in DMEM (without FCS) was used to resuspend the cell pellet. The cell suspension was incubated by shaking for 30 min in the dark, repelleted and resuspended in prewarmed (37°C) DMEM without FCS. Mitochondrial mass was measured by flow cytometry using a FACSCalibur instrument. Excitation and emission settings were 490 nm and 516 nm, respectively (FL1 filter). At least three independent measurements were done per cell line.

2.1.3.3 Reactive Oxygen Species (ROS) Accumulation

ROS accumulation in cells was measured using dihydroethidium (DHE) (Sigma-Aldrich). DHE permeates cell membranes and freely diffuses to the cytosol where it exhibits blue fluorescence, until oxidized by superoxide when it turns into a product showing red fluorescence, and is retained by the cells. Cells were detached from the culture flask by trypsinization for 10 min. The reaction was stopped with DMEM. Cells were pelleted and resuspended in 500 µl DMEM (without FCS) containing 7.5 µM DHE. The cell suspension was incubated for 15 min in the dark with shaking and repelleted and later resuspended in DMEM without FCS. Accumulation of ROS was measured using the FACSCalibur flow cytometer (Becton Dickinson). The excitation was set at 488 nm and emission at 564–606 nm (FL2 filter). At least three independent measurements were done per cell line.

2.1.3.4 Proliferation Assays

To estimate the proliferation rate of the different cell lines, data from nontreated cells obtained from at least three independent measurements was used. Values obtained represent the mean of three technical replicates. All the data collected comes from experiments employing standard culture conditions. Confluent cells were plated in 96-well plates, 3,000 cells per well, incubated in DMEM medium for 72 h at 37° C in an incubator set with 5% CO_2 and 95% humidity.

The proliferation rate of each cell line was determined using two different approaches; MTT = 3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide reagent (Sigma-Aldrich) and SRB = sulforhodamine B reagent (Sigma-Aldrich). MTT penetrates viable eukaryotic cells. Cells with a mitochondrial activity reduce MTT producing a purple-colored product. The SRB assay measures the total protein content. The dye is incorporated into fixed cells and the amount of incorporated dye is proportional to the amount of cells (Vichai & Kanyawim 2006). The specific protocols are shown below:

MTT: The medium of the cells in the 96-well plates was replaced with 100 µl DMEM 1% FCS containing 0.5 mg/ml MTT reagent per well. Plates were then incubated for 3 h at 37°C in a 5% CO_2 incubator. The absorbance was measured in a Tecan Safire (Tecan) microplate reader at 595 nm adding 200 µl per well of DMSO (Sigma-Aldrich).

SRB: Cells were fixed with 10% (w/v) cold trichloroacetic acid (Sigma-Aldrich) for 1 h at 4°C. The trichloroacetic acid was removed washing the plates carefully with water two times. The excess of water was removed with a paper towel by tapping the plate several times. The rest of water in the plates was evaporated in an oven at 56°C. Dry cells were stained with SRB for 30 min at room temperature by adding 100 µl of 0.057% (w/v) SRB in 1% acetic acid (Sigma-Aldrich) in each well. The excess of dye was removed by washing three times with 1% (v/v) acetic acid (200 µl per well). The plates were dried in an oven at 56°C. The boundary between the SRB and the proteins was dissolved in 10 mM Tris (pH 10.5) adding 200 µl per well and shaking them for 10 min. The optical density determination was done at 510 nm in a Tecan Safire microplate reader.

2.2 Hierarchical Clustering

To group the cell lines based on metabolic features or gene expression, clusters were drawn using *R programming language* packages plyr version 1.8.4 library plyr (Wickham 2011) and rafalib version 1.0.0 library rafalib (Irizarry & Love 2016) The cluster method used was the hierarchical clustering with the hclust function, Euclidean distance and method complete (Supplements 5.1.6).

2.3 Gene Expression Analyses

Raw data (.CEL files) from gene expressions obtained using Affymetrix Human Genome U133 Plus 2.0 Array were obtained from GEO, NCBI (De Schutter et al. 2013, Barretina et al. 2012). The according public accession numbers and cell lines are listed in Table 2. All the analyzed data comes from cell lines cultivated in DMEM supplemented with 10% FCS in a 5% CO_2 incubator at 37°C. Data for HaCaT and HeLa lines were available in triplicate; a mean of the calculated expression value was used.

Table 2. Cell lines and gene accession numbers

Cell line	Gene accession number
MCF7	GSM887291
MDA-MB-231	GSM887295
HCT 116	GSM887062
HT-29	GSM887141
AsPC-1	GSM886870
BxPC-3	GSM886896
MIA PaCa-2	GSM887320
PANC-1	GSM887501
LNCaP	GSM887271
PC-3	GSM887506
DU 145	GSM886988
MDA-MB-435	GSM887298
Hep G2	GSM887079
1321N1	GSM886835
HeLa control a	GSM960275
HeLa control b	GSM960283
HeLa control c	GSM960291
HaCaT control a	GSM960278
HaCaT control b	GSM960286
HaCaT control c	GSM960294

Preprocessing of the .CEL files was done using the Robust Multi-array Average (RMA) normalization method (Irizarry et al. 2003). This method allows background subtraction, quantile normalization, and summarization (calculation of expression values). RMA was applied in all probe sets belonging to the 16 cell lines listed in Table 2. RMA calculations were done using Bioconductor (Huber et al. 2015), an open source software for bioinformatics that uses *R statistics programming language*. Bioconductor version 3.2 was run under *R* version 2.7.0. For the calculation, the package *affy* version 1.32.0 and the annotation package pd.ht.hg.u133.plus.pm were used (Gautier et al. 2004). The RMA function (rma()) was run with the default options (normalize=TRUE,

background=TRUE) (Supplements 5.1.2). Official gene symbols were downloaded from affymetrix.com and the *comma separated value* file HGU133.Plus2.na34.annot.csv was used.

2.3.1 Selection of Genes Related to Metabolism Pathways with Extreme Expression in the Affymetrix Array

The Entrez gene ID belonging to the glycolysis-gluconeogenesis, pentose phosphate, tricarboxylic acid cycle (TCA), and electron transport chain pathways were downloaded from Wikipathways (Kelder et al. 2011). The entire downloaded gene lists regards *Homo sapiens*. The specific Wikipathways used were: WP134 for pentose phosphate, WP534 for glycolysis-gluconeogenesis, WP78 for TCA, and WP111 for respiratory electron transport chain pathways. The Entrez gene identifiers acquired from Wikipathways were uploaded in DAVID 6.7 (Huang et al. 2009). DAVID is a free online bioinformatics resource database for annotation and gene list visualization. Using the gene ID conversion tool from DAVID, the official gene symbols were obtained. A total of seven genes were selected for further analysis of the pentose phosphate pathway, 48 genes from glycolysis-gluconeogenesis, and 116 genes of the TCA plus electron transport chain (Tables 3, 4, and 5).

Table 3. Entrez gene identifiers of genes belonging to the pentose phospate pathway WP134 from Wikipathways. Official gene symbols and gene names were obtained with DAVID 6.7 online database.

Wikipathways			DAVID		
Identifier	**Database**	**From**	**To**	**Species**	**David Gene Name**
2539	Entrez Gene	2539	G6PD	*Homo sapiens*	Glucose-6-phosphate Dehydrogenase(G6PD)
5226	Entrez Gene	5226	PGD	*Homo sapiens*	Phosphogluconate dehydrogenase (PGD)
6120	Entrez Gene	6120	RPE	*Homo sapiens*	Ribulose-5-phosphate-3-epimerase (RPE)
6888	Entrez Gene	6888	TALDO1	*Homo sapiens*	Transaldolase 1 (TALDO1)
7086	Entrez Gene	7086	TKT	*Homo sapiens*	Transketolase (TKT)
22934	Entrez Gene	22934	RPIA	*Homo sapiens*	Ribose 5-phosphate isomerase A (RPIA)
25796	Entrez Gene	25796	PGLS	*Homo sapiens*	6-Phosphogluconolactonase (PGLS)

Table 4. Entrez gene identifiers of genes belonging to the glycolysis-gluconeogenesis pathway WP534 from Wikipathways. Official gene symbols and gene names were obtained with DAVID 6.7 online database.

Wikipathways		DAVID			
Identifier	**Database**	**From**	**To**	**Species**	**DAVID gene name**
226	Entrez Gene	226	ALDOA	*Homo sapiens*	Aldolase, fructose-bisphosphate A(ALDOA)
229	Entrez Gene	229	ALDOB	*Homo sapiens*	Aldolase, fructose-bisphosphate B(ALDOB)
230	Entrez Gene	230	ALDOC	*Homo sapiens*	aldolase, fructose-bisphosphate C(ALDOC)
1737	Entrez Gene	1737	DLAT	*Homo sapiens*	dihydrolipoamide S-acetyltransferase(DLAT)
1738	Entrez Gene	1738	DLD	*Homo sapiens*	dihydrolipoamide dehydrogenase(DLD)
2023	Entrez Gene	2023	ENO1	*Homo sapiens*	enolase 1(ENO1)
2026	Entrez Gene	2026	ENO2	*Homo sapiens*	enolase 2(ENO2)
2027	Entrez Gene	2027	ENO3	*Homo sapiens*	enolase 3(ENO3)
2203	Entrez Gene	2203	FBP1	*Homo sapiens*	fructose-bisphosphatase 1(FBP1)
2538	Entrez Gene	2538	G6PC	*Homo sapiens*	glucose-6-phosphatase catalytic subunit(G6PC)
2645	Entrez Gene	2645	GCK	*Homo sapiens*	glucokinase(GCK)
2805	Entrez Gene	2805	GOT1	*Homo sapiens*	glutamic-oxaloacetic transaminase 1(GOT1)
2806	Entrez Gene	2806	GOT2	*Homo sapiens*	glutamic-oxaloacetic transaminase 2(GOT2)
2821	Entrez Gene	2821	GPI	*Homo sapiens*	glucose-6-phosphate isomerase(GPI)
3098	Entrez Gene	3098	HK1	*Homo sapiens*	hexokinase 1(HK1)
3099	Entrez Gene	3099	HK2	*Homo sapiens*	hexokinase 2(HK2)
3101	Entrez Gene	3101	HK3	*Homo sapiens*	hexokinase 3(HK3)
3939	Entrez Gene	3939	LDHA	*Homo sapiens*	lactate dehydrogenase A(LDHA)
3945	Entrez Gene	3945	LDHB	*Homo sapiens*	lactate dehydrogenase B(LDHB)
3948	Entrez Gene	3948	ldhc	*Homo sapiens*	lactate dehydrogenase C(LDHC)
4190	Entrez Gene	4190	MDH1	*Homo sapiens*	malate dehydrogenase 1(MDH1)
4191	Entrez Gene	4191	MDH2	*Homo sapiens*	malate dehydrogenase 2(MDH2)
5091	Entrez Gene	5091	PC	*Homo sapiens*	pyruvate carboxylase(PC)
5105	Entrez Gene	5105	PCK1	*Homo sapiens*	phosphoenolpyruvate carboxykinase 1(PCK1)
5160	Entrez Gene	5160	PDHA1	*Homo sapiens*	pyruvate dehydrogenase (lipoamide) alpha 1(PDHA1)
5161	Entrez Gene	5161	PDHA2	*Homo sapiens*	pyruvate dehydrogenase (lipoamide) alpha 2(PDHA2)
5162	Entrez Gene	5162	PDHB	*Homo sapiens*	pyruvate dehydrogenase (lipoamide) beta(PDHB)
5211	Entrez Gene	5211	PFKL	*Homo sapiens*	phosphofructokinase, liver type(PFKL)
5213	Entrez Gene	5213	PFKM	*Homo sapiens*	phosphofructokinase, muscle(PFKM)
5214	Entrez Gene	5214	PFKP	*Homo sapiens*	phosphofructokinase, platelet(PFKP)
5223	Entrez Gene	5223	PGAM1	*Homo sapiens*	phosphoglycerate mutase 1(PGAM1)
5224	Entrez Gene	5224	PGAM2	*Homo sapiens*	phosphoglycerate mutase 2(PGAM2)
5230	Entrez Gene	5230	PGK1	*Homo sapiens*	phosphoglycerate kinase 1(PGK1)
5232	Entrez Gene	5232	PGK2	*Homo sapiens*	phosphoglycerate kinase 2(PGK2)
5313	Entrez Gene	5313	PKLR	*Homo sapiens*	pyruvate kinase, liver and RBC(PKLR)
5315	Entrez Gene	5315	PKM	*Homo sapiens*	pyruvate kinase, muscle(PKM)
6513	Entrez Gene	6513	SLC2A1	*Homo sapiens*	solute carrier family 2 member

					1(SLC2A1)
6514	Entrez Gene	6514	SLC2A2	*Homo sapiens*	solute carrier family 2 member 2(SLC2A2)
6515	Entrez Gene	6515	SLC2A3	*Homo sapiens*	solute carrier family 2 member 3(SLC2A3)
6517	Entrez Gene	6517	SLC2A4	*Homo sapiens*	solute carrier family 2 member 4(SLC2A4)
6518	Entrez Gene	6518	slc2a5	*Homo sapiens*	solute carrier family 2 member 5(SLC2A5)
7167	Entrez Gene	7167	TPI1	*Homo sapiens*	triosephosphate isomerase 1(TPI1)
8050	Entrez Gene	8050	PDHX	*Homo sapiens*	pyruvate dehydrogenase complex component X (PDHX)
8789	Entrez Gene	8789	FBP2	*Homo sapiens*	Fructose-bisphosphatase 2(FBP2)
25874	Entrez Gene	25874	MPC2	*Homo sapiens*	Mitochondrial pyruvate carrier 2 (MPC2)
26330	Entrez Gene	26330	GAPDHS	*Homo sapiens*	Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic (GAPDHS)
51660	Entrez Gene	51660	MPC1	*Homo sapiens*	Mitochondrial pyruvate carrier 1 (MPC1)
92483	Entrez Gene	92483	LDHAL6B	*Homo sapiens*	Lactate dehydrogenase A like 6B (LDHAL6B)

Table 5. Entrez gene identifiers of genes belonging to the tricarboxylic acid cycle and electron transport chain, pathways WP78 and WP111 from Wikipathways. Official gene symbols and gene names were obtained with DAVID 6.7 online database.

Wikipathways				**DAVID**
Identifier	**Database**	**From**	**To**	**DAVID gene name**
50	Entrez Gene	50	ACO2	aconitase 2 (ACO2)
291	Entrez Gene	291	SLC25A4	solute carrier family 25 member 4 (SLC25A4)
292	Entrez Gene	292	SLC25A5	solute carrier family 25 member 5 (SLC25A5)
293	Entrez Gene	293	SLC25A6	solute carrier family 25 member 6 (SLC25A6)
498	Entrez Gene	498	ATP5A1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle (ATP5A1)
506	Entrez Gene	506	ATP5B	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide (ATP5B)
509	Entrez Gene	509	ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1 (ATP5C1)
513	Entrez Gene	513	ATP5D	ATP synthase, H+ transporting, mitochondrial F1 complex, delta subunit (ATP5D)
514	Entrez Gene	514	ATP5E	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit (ATP5E)
515	Entrez Gene	515	ATP5F1	ATP synthase, H+ transporting, mitochondrial Fo complex subunit B1 (ATP5F1)
516	Entrez Gene	516	ATP5G1	ATP synthase, H+ transporting, mitochondrial Fo complex subunit C1 (subunit 9) (ATP5G1)
517	Entrez Gene	517	ATP5G2	ATP synthase, H+ transporting, mitochondrial Fo complex subunit C2 (subunit 9) (ATP5G2)

518	Entrez Gene	518	ATP5G3	ATP synthase, H+ transporting, mitochondrial Fo complex subunit C3 (subunit 9) (ATP5G3)
521	Entrez Gene	521	ATP5I	ATP synthase, H+ transporting, mitochondrial Fo complex subunit E (ATP5I)
522	Entrez Gene	522	ATP5J	ATP synthase, H+ transporting, mitochondrial Fo complex subunit F6 (ATP5J)
539	Entrez Gene	539	ATP5O	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (ATP5O)
1327	Entrez Gene	1327	Cox4I1	cytochrome c oxidase subunit 4I1 (COX4I1)
1329	Entrez Gene	1329	Cox5b	cytochrome c oxidase subunit 5B (COX5B)
1337	Entrez Gene	1337	Cox6a1	cytochrome c oxidase subunit 6A1 (COX6A1)
1339	Entrez Gene	1339	COX6A2	cytochrome c oxidase subunit 6A2 (COX6A2)
1340	Entrez Gene	1340	Cox6b1	cytochrome c oxidase subunit 6B1 (COX6B1)
1345	Entrez Gene	1345	COX6C	cytochrome c oxidase subunit 6C (COX6C)
1346	Entrez Gene	1346	Cox7a1	cytochrome c oxidase subunit 7A1 (COX7A1)
1347	Entrez Gene	1347	COX7A2	cytochrome c oxidase subunit 7A2 (COX7A2)
1349	Entrez Gene	1349	COX7B	cytochrome c oxidase subunit 7B (COX7B)
1350	Entrez Gene	1350	COX7C	cytochrome c oxidase subunit 7C (COX7C)
1351	Entrez Gene	1351	COX8A	cytochrome c oxidase subunit 8A (COX8A)
1353	Entrez Gene	1353	cox11	COX11, cytochrome c oxidase copper chaperone (COX11)
1355	Entrez Gene	1355	Cox15	COX15, cytochrome c oxidase assembly homolog (COX15)
1431	Entrez Gene	1431	CS	citrate synthase (CS)
1738	Entrez Gene	1738	DLD	dihydrolipoamide dehydrogenase (DLD)
1743	Entrez Gene	1743	DLST	dihydrolipoamide S-succinyltransferase (DLST)
2271	Entrez Gene	2271	FH	fumarate hydratase (FH)
3418	Entrez Gene	3418	IDH2	isocitrate dehydrogenase (NADP(+)) 2, mitochondrial (IDH2)
3419	Entrez Gene	3419	IDH3A	isocitrate dehydrogenase 3 (NAD(+)) alpha (IDH3A)
3420	Entrez Gene	3420	IDH3B	isocitrate dehydrogenase 3 (NAD(+)) beta (IDH3B)
3421	Entrez Gene	3421	IDH3G	isocitrate dehydrogenase 3 (NAD(+)) gamma (IDH3G)
4191	Entrez Gene	4191	MDH2	malate dehydrogenase 2 (MDH2)
4508	Entrez Gene	4508	ATP6	ATP synthase F0 subunit 6 (ATP6)
4509	Entrez Gene	4509	ATP8	ATP synthase F0 subunit 8 (ATP8)
4512	Entrez Gene	4512	COX1	cytochrome c oxidase subunit I (COX1)
4513	Entrez Gene	4513	COX2	cytochrome c oxidase subunit II (COX2)
4514	Entrez Gene	4514	COX3	cytochrome c oxidase III (COX3)
4519	Entrez Gene	4519	CYTB	cytochrome b (CYTB)
4535	Entrez Gene	4535	ND1	NADH dehydrogenase, subunit 1 (complex I) (ND1)
4536	Entrez Gene	4536	ND2	MTND2 (ND2)
4537	Entrez Gene	4537	ND3	NADH dehydrogenase, subunit 3 (complex I) (ND3)
4538	Entrez Gene	4538	ND4	NADH dehydrogenase, subunit 4 (complex I) (ND4)
4539	Entrez Gene	4539	ND4L	NADH dehydrogenase, subunit 4L (complex I) (ND4L)

4540	Entrez Gene	4540	ND5	NADH dehydrogenase, subunit 5 (complex I) (ND5)
4541	Entrez Gene	4541	ND6	NADH dehydrogenase, subunit 6 (complex I) (ND6)
4694	Entrez Gene	4694	NDUFA1	NADH:ubiquinone oxidoreductase subunit A1 (NDUFA1)
4695	Entrez Gene	4695	NDUFA2	NADH:ubiquinone oxidoreductase subunit A2 (NDUFA2)
4696	Entrez Gene	4696	NDUFA3	NADH:ubiquinone oxidoreductase subunit A3 (NDUFA3)
4697	Entrez Gene	4697	NDUFA4	NDUFA4, mitochondrial complex associated (NDUFA4)
4698	Entrez Gene	4698	NDUFA5	NADH:ubiquinone oxidoreductase subunit A5 (NDUFA5)
4700	Entrez Gene	4700	NDUFA6	NADH:ubiquinone oxidoreductase subunit A6 (NDUFA6)
4701	Entrez Gene	4701	NDUFA7	NADH:ubiquinone oxidoreductase subunit A7 (NDUFA7)
4702	Entrez Gene	4702	NDUFA8	NADH:ubiquinone oxidoreductase subunit A8 (NDUFA8)
4704	Entrez Gene	4704	NDUFA9	NADH:ubiquinone oxidoreductase subunit A9 (NDUFA9)
4705	Entrez Gene	4705	NDUFA10	NADH:ubiquinone oxidoreductase subunit A10 (NDUFA10)
4706	Entrez Gene	4706	NDUFAB1	NADH:ubiquinone oxidoreductase subunit AB1 (NDUFAB1)
4707	Entrez Gene	4707	NDUFB1	NADH:ubiquinone oxidoreductase subunit B1 (NDUFB1)
4708	Entrez Gene	4708	NDUFB2	NADH:ubiquinone oxidoreductase subunit B2 (NDUFB2)
4709	Entrez Gene	4709	NDUFB3	NADH:ubiquinone oxidoreductase subunit B3 (NDUFB3)
4710	Entrez Gene	4710	NDUFB4	NADH:ubiquinone oxidoreductase subunit B4 (NDUFB4)
4711	Entrez Gene	4711	NDUFB5	NADH:ubiquinone oxidoreductase subunit B5 (NDUFB5)
4712	Entrez Gene	4712	NDUFB6	NADH:ubiquinone oxidoreductase subunit B6 (NDUFB6)
4713	Entrez Gene	4713	NDUFB7	NADH:ubiquinone oxidoreductase subunit B7 (NDUFB7)
4714	Entrez Gene	4714	NDUFB8	NADH:ubiquinone oxidoreductase subunit B8 (NDUFB8)
4715	Entrez Gene	4715	NDUFB9	NADH:ubiquinone oxidoreductase subunit B9 (NDUFB9)
4716	Entrez Gene	4716	NDUFB10	NADH:ubiquinone oxidoreductase subunit B10 (NDUFB10)
4717	Entrez Gene	4717	NDUFC1	NADH:ubiquinone oxidoreductase subunit C1 (NDUFC1)
4718	Entrez Gene	4718	NDUFC2	NADH:ubiquinone oxidoreductase subunit C2 (NDUFC2)
4719	Entrez Gene	4719	NDUFS1	NADH:ubiquinone oxidoreductase core subunit S1 (NDUFS1)
4720	Entrez Gene	4720	NDUFS2	NADH:ubiquinone oxidoreductase core subunit S2 (NDUFS2)

4722	Entrez Gene	4722	NDUFS3	NADH:ubiquinone oxidoreductase core subunit S3 (NDUFS3)
4723	Entrez Gene	4723	NDUFV1	NADH:ubiquinone oxidoreductase core subunit V1 (NDUFV1)
4724	Entrez Gene	4724	NDUFS4	NADH:ubiquinone oxidoreductase subunit S4 (NDUFS4)
4725	Entrez Gene	4725	NDUFS5	NADH:ubiquinone oxidoreductase subunit S5 (NDUFS5)
4726	Entrez Gene	4726	NDUFS6	NADH:ubiquinone oxidoreductase subunit S6 (NDUFS6)
4728	Entrez Gene	4728	NDUFS8	NADH:ubiquinone oxidoreductase core subunit S8 (NDUFS8)
4729	Entrez Gene	4729	NDUFV2	NADH:ubiquinone oxidoreductase core subunit V2 (NDUFV2)
4731	Entrez Gene	4731	NDUFV3	NADH:ubiquinone oxidoreductase subunit V3 (NDUFV3)
4967	Entrez Gene	4967	OGDH	oxoglutarate dehydrogenase (OGDH)
6341	Entrez Gene	6341	SCO1	SCO1, cytochrome c oxidase assembly protein (SCO1)
6389	Entrez Gene	6389	SDHA	succinate dehydrogenase complex flavoprotein subunit A (SDHA)
6390	Entrez Gene	6390	SDHB	succinate dehydrogenase complex iron sulfur subunit B (SDHB)
6391	Entrez Gene	6391	SDHC	succinate dehydrogenase complex subunit C (SDHC)
6392	Entrez Gene	6392	SDHD	succinate dehydrogenase complex subunit D (SDHD)
6834	Entrez Gene	6834	SURF1	SURF1, cytochrome c oxidase assembly factor (SURF1)
7350	Entrez Gene	7350	UCP1	uncoupling protein 1 (UCP1)
7351	Entrez Gene	7351	UCP2	uncoupling protein 2 (UCP2)
7352	Entrez Gene	7352	UCP3	uncoupling protein 3 (UCP3)
7381	Entrez Gene	7381	Uqcrb	ubiquinol-cytochrome c reductase binding protein (UQCRB)
7384	Entrez Gene	7384	UQCRC1	ubiquinol-cytochrome c reductase core protein I (UQCRC1)
7385	Entrez Gene	7385	Uqcrc2	ubiquinol-cytochrome c reductase core protein II (UQCRC2)
7386	Entrez Gene	7386	Uqcrfs1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1 (UQCRFS1)
7388	Entrez Gene	7388	Uqcrh	ubiquinol-cytochrome c reductase hinge protein (UQCRH)
8801	Entrez Gene	8801	SUCLG2	succinate-CoA ligase GDP-forming beta subunit (SUCLG2)
8802	Entrez Gene	8802	SUCLG1	succinate-CoA ligase alpha subunit (SUCLG1)
9016	Entrez Gene	9016	SLC25A14	solute carrier family 25 member 14 (SLC25A14)
9167	Entrez Gene	9167	COX7A2L	cytochrome c oxidase subunit 7A2 like (COX7A2L)
9377	Entrez Gene	9377	Cox5a	cytochrome c oxidase subunit 5A (COX5A)
9481	Entrez Gene	9481	SLC25A27	solute carrier family 25 member 27 (SLC25A27)
9551	Entrez Gene	9551	ATP5J2	ATP synthase, H+ transporting, mitochondrial Fo complex subunit F2 (ATP5J2)

10063	Entrez Gene	10063	COX17	COX17, cytochrome c oxidase copper chaperone (COX17)
10476	Entrez Gene	10476	ATP5H	ATP synthase, H+ transporting, mitochondrial Fo complex subunit D (ATP5H)
10632	Entrez Gene	10632	ATP5L	ATP synthase, H+ transporting, mitochondrial Fo complex subunit G (ATP5L)
10975	Entrez Gene	10975	UQCR11	ubiquinol-cytochrome c reductase, complex III subunit XI (UQCR11)
27089	Entrez Gene	27089	Uqcrq	ubiquinol-cytochrome c reductase complex III subunit VII (UQCRQ)
27109	Entrez Gene	27109	ATP5S	ATP synthase, H+ transporting, mitochondrial Fo complex subunit s (factor B) (ATP5S)
29796	Entrez Gene	29796	Uqcr10	ubiquinol-cytochrome c reductase, complex III subunit X (UQCR10)
55967	Entrez Gene	55967	NDUFA12	NADH:ubiquinone oxidoreductase subunit A12 (NDUFA12)
93974	Entrez Gene	93974	ATPIF1	ATPase inhibitory factor 1 (ATPIF1)
374291	Entrez Gene	374291	NDUFS7	NADH:ubiquinone oxidoreductase core subunit S7 (NDUFS7)

Using the official gene symbols from the selected pathways, the probe set IDs from the RMA normalized data were selected (Supplements 5.1.3). From the original table with 54,676 rows corresponding to the total probe set IDs in the Affymetrix U133 array, we extracted three small tables with 13 probe sets for the pentose phosphate pathway, 94 probe sets for the glycolysis-gluconeogenesis pathway, and 218 probe sets for TCA and electron transport chain.

The statistical method "Presence-Absence calls with Negative Probe sets" (PANP) was used to estimate the significance of the detection on Affymetrix U133 series microarrays (Warren et al. 2007). The probe sets in which an “Absence” value is given are the ones in which the expression level is below the threshold of detection, meaning that the expression level does not differ from zero significantly and that the fluorescence intensity generated in the sample occurred by unspecific hybridization.

The PANP function runs in the *R programming language* and uses Affymetrix-reporter probes with unknown hybridization partners to calculate the cutoffs and gives the p-values for each probe set. The *panp* package (Bioconductor) with the PANP function (panp()) was run with the default option for the p-values (0.01 and 0.02 as cutoffs, see Supplements 5.1.4). “P” (present) calls were given to the probe sets in which the p-value for the detection intensities was below 0.01; “M” (marginal) was given to the probe sets in which the p-value for the detection intensities was between 0.01 and 0.02;

and “A” (absence) was given to the probe sets in which the p-value for the detection intensities was above of 0.02.

In the Affymetrix U133 array the probe set IDs are named with a number followed by a letter suffix that indicates the probe set type:

• “_at” suffix: predicted to perfectly match the antisense strand of the gene of interest. Just one single transcript is included.

• “_a_at” suffix: at least one probe of the probe set cross-hybridized with transcripts from the same gene.

• “_s_at” suffix: at least one probe of the probe set cross-hybridized with transcript from the same gene or from homologous genes.

• “_x_at” suffix: is a mixed probe set that contains some probes that are identical or highly similar to other sequences.

A selection of one probe set per gene was done based on the probe set suffix and on the "Presence-Absence calls”. Independently of the "Presence-Absence calls”, all probe sets with a _x suffix were deleted. For glycolysis-gluconeogenesis together with the pentose phosphate pathways, 107 probe sets were reduced to 51 (51 genes). From the 218 probe sets belonging to the TCA together with the electron transport chain pathways, 104 probe sets remained (104 genes). Scale and center were applied to the expression values to normalize the expression of each gene in each cell line. To center the gene expression columns, the mean of the gene expression was calculated per each gene and then subtracted from the gene expression value in each cell line. These centered column values are divided by the column's standard deviation. A range to compare the expression values of each gene in the studied cell lines from −3 (lowly expressed genes) to 3 (highly expressed genes) was obtained. Scaling and centering was calculated using *R programming language* function biscale from the package softImpute version 1.4 library softImpute (Hastie et al. 2015) (Supplements 5.1.5). Genes that had high or low expression levels (extreme values after scale and center >1 or <–1) in at least 7 of the 16 cell lines were selected (Supplements 5.1.5). A set of 17 genes out of the 155 remained for correlation analysis with the metabolic assays.

2.3.2 Selection of Genes in the Total Probe Sets of the Affymetrix Array with Extreme Expression

Gene expression data of the total probe sets of the Affymetrix array normalized with the RMA function (Experimental Methods 2.3) were scaled and centered per probe set base in the expression value of the analyzed cell lines (Experimental Methods 2.3.1). In the complete data with 54,676 probe set IDs we selected the extreme expressed probe sets (values greater than 1 and lower than –1) per cell line (Supplements 5.1.5). A new range was done with the cell lines that presented the most extreme expressed probe set values. Probe sets were selected that were either highly or lowly expressed in at least 9 of the 16 cell lines. A total of 202 probe sets remained. The gene symbol of the 202 probe sets was added to search all the probe sets that target that gene symbol. A new list containing 481 probe sets was obtained. The probe sets with a _x suffix were deleted leaving 456 probe sets.

A final selection was done for genes with consistent information through their probe sets. Genes with just one probe set were discarded. There remained 36 probe sets corresponding to 17 genes for correlation analysis with the metabolic assays.

2.3.3 Real-Time Polymerase Chain Reaction (qPCR)

The gene expression of six genes belonging to the glycolysis-gluconeogenesis pathway (GOT1, HK1, HK2, LDHa, PFKM, SLC2A) was estimated by quantitative real-time PCR in MCF7, MDA-MB-231, HeLa, HCT 116, HT-29, Hep G2, ASPC-1, BxPC-3, MIA PaCa-2, PANC-1, PC-3, and DU 145 cell lines. The RNA for the qPCR came from cells cultivated under standard conditions. The total RNA from each cell line was isolated with RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. The quality of the RNA was examined by electrophoresis in agarose gels and quantified by spectrophotometry (260 nm/280 nm) in a NanoDrop 2000 (Thermo Scientific). A mass of 250 ng of total RNA for each cell line was transcribed into cDNA with the RevertAid Reverse Transcriptase (Thermo Fisher). For the synthesis, oligo-dT primers and random hexamer primers were used. The reverse transcription reaction program entailed incubation of 10 min at 25°C, a reverse transcription of 30 min at 48°C, and inactivation

of the reverse transcriptase for 5 min at 95°C. The amplification reactions contained 1 μl of cDNA, 1 μl SYBR Green Master (Qiagen), and 1 μl of the target primers (Eurofins Genomics). The primer sequences are listed in Table 6. The qPCRs were performed using a qTOWER (Analytik Jena) system with the following program: Taq polymerase activation (10 min at 95°C) and 40 amplification cycles (denaturation at 95°C for 15 sec and annealing/extension at 58°C for 1 min).

Table 6. List of the primer sequences used in real-time PCR for estimating the expression of some of the genes involves in the glycolysis-gluconeogenesis pathway

Target gene	Type of primer	Sequence
GOT1	Forward	5' CAC TAT CTG CCA ATC CTG 3'
	Reverse	5' CCT ACC CGC TTC TCC TTG AG 3'
HK1	Forward	5' GTT GCC AAC ATT CET AAG GTC C 3'
	Reverse	5' CAC TTG CAC CCG CAG AAT TCG 3'
HK2	Forward	5' GTG GCA CCC AGC TGT TTG AC 3'
	Reverse	5' CGA GAA GGT AAA ACC CAG TGG 3'
LDHa	Forward	5' GAA GGG AGA GAT GAT GGA TCT C 3'
	Reverse	5' CTT ATC TTC CAA GCC ACG TAG G 3'
PFKM	Forward	5' GGT GGA GAT CAC ATC AAG GAA G 3'
	Reverse	5' CTC GTT CCC GAA AGT CCT TGC 3'
SLC2A	Forward	5' CAG TTT GGC TAC AAC ACT GGA G 3'
	Reverse	5' GCA GGA TGC TCT CCC CAT AG 3'

2.4 Half Maximal Inhibitory Concentration (IC50) Estimations

2.4.1 IC_{50} Values from the Genomics of Drug Sensitivity in Cancer Project (GDSC)

The IC_{50}s of several drugs were obtained from the GDSC, part of the Wellcome Trust Sanger Institute (UK) and the Center for Molecular Therapeutics, Massachusetts General Hospital Cancer Center (USA) (Yang et al. 2012). For this study, the release 5 (June 2014) was used. IC_{50} values for 99 drugs in 11 out of the 18 cell lines that contain metabolism estimations were found (Table 7).

Table 7. IC_{50} values in μM obtained from the Genomics of Drug Sensitivity in Cancer project.

	IC_{50} μM										
Drug Name	MDA-MB-231	MCF7	HCT 116	HT-29	AsPC-1	BxPC-3	MIA PaCa-2	MDA-MB-435	PC-3	LNCaP	DU 145
681640	6.86	5.99	28.71	160.6	40.21	15.03	29.54	1.69	27.16	20.72	0.16
17-AAG	45.12	0.097	0.266	0.033	0.152	0.135	0.003	0.079	0.079	0.800	2.79
ABT-263	121.43	45.51	161.9	60.26	19.87	81.17	11.32	110.9	157.4	105.6	0.90
ABT-888	210.62	432.9	87.78	680.3	257.3	435.3	82.21	373.0	28.67	299.5	13.91
AG-014699	13.77	8.03	7.18	409	31.65	28.00	123.3	145.5	106.7	32.26	50.49
AICAR	2.401	1.894	2.555	17.723	1.327	1.376	3.429	2.480	300.7	8.310	6.767
AKT inhib. VIII	103.8	32.27	7.76	2.43	48.11	2.22	95.78	104.8	10.51	10.46	17.69
AMG-706	212.1	28.86	210.9	75.96	54.51	18.64	21.36	38.90	13.51	159.8	0.47
AP-24534	10.21	11.81	0.21	25.79	7.29	0.35	1.01	43.32	44.43	25.35	1.08
AS601245	12.31	1.50	231.5	4.27	317.8	1.86	0.98	28.42	11.62	73.99	12.02
ATRA	71.80	52.35	217.1	1.729	997	211.4	425.9	415.1	343.2	974	50.55
AUY922	0.091	0.003	0.152	0.044	2.54	0.043	0.015	0.218	0.132	0.082	0.013
Axitinib	18.65	46.55	103.9	29.30	30.02	74.39	1.83	37.16	21.80	34.85	4.13
AZD-2281	66.56	91.79	30.05	241.2	367.8	496.9	23.79	493.3	197.5	300.4	217.7
AZD6244	5.57	243.5	1.06	70.39	0.32	4.66	38.84	0.04	23.57	395.4	133.8
AZD6482	60.07	3.35	6.69	29.76	32.68	3.45	14.71	20.63	16.54	1.52	1.37
AZD6482	60.96	468.6	19.34	187.9	176.8	39.36	266.2	216.8	39.20	2.41	37.28
AZD7762	1.60	13.24	1.42	0.37	0.66	0.26	0.14	0.14	9.71	1.44	40.63
AZD8055	1.72	0.41	1.43	2.33	0.74	0.89	0.42	0.47	0.72	0.35	6.77
BAY 61-3606	547.9	1.15	3.98	2.81	432.9	2.86	5.40	169.9	418	2.18	7.97
Bexarotene	63.20	273.9	268.7	108.1	797.3	522.9	10.51	53.58	65.53	896.1	158.3
BIBW2992	28.28	16.47	4.41	55.24	19.65	25.01	25.21	44.13	21.11	15.42	0.45
Bicalutamide	106	39.30	59.95	88.40	74.47	68.05	77.87	50.48	10.28	41.29	98.21
BIRB 0796	206.4	768.9	324.7	28.19	427.9	263.6	14.53	734	152.8	431.7	3.94
Bleomycin	0.27	1.17	2.88	0.54	82.50	3.81	0.11	3.22	13.31	18.32	0.82
BMS-708163	339	149.8	18.80	118.8	60.85	110.4	192.4	61.96	78.38	454.4	163
BMS-754807	2.61	0.09	3.21	0.01	18.86	1.38	0.25	6.59	1.93	7.86	1.93
Bosutinib	47.62	135.4	2.23	4.61	14.56	62.63	1.05	4.49	102.6	147	67.74
Bryostatin 1	0.65	0.05	0.05	0.47	0.12	0.05	0.26	0.01	0.48	0.21	0.08
BX-795	8.24	35.83	4.99	38.55	52.72	10.01	2.26	5.22	4.17	30.60	429.1
Camptothecin	0.010	0.024	0.032	0.047	0.168	0.053	0.002	0.012	0.007	0.031	2.28
CCT007093	519.2	3.86	434.3	238	126.1	161.2	687.9	479.5	592.1	874.1	759.1
CCT018159	2.80	199.8	5.77	1.111	90.69	1.105	4.42	312.8	9.54	18.64	8.22
CEP-701	0.59	3.20	0.14	0.87	1.62	1.01	0.50	0.05	12.61	0.35	2.16
CHIR-99021	200.9	51.76	43.21	122.8	152.8	4.70	1.76	23.11	99.06	175.9	109.95
CI-1040	4.29	76.60	6.81	1.47	5.37	8.39	0.92	0.15	312.9	239.6	100.6
Cisplatin	31.69	56.64	118.4	96.55	20.95	8.25	1.48	166.4	49.66	515.4	11.12
Cytarabine	0.53	1.97	0.31	0.57	1.39	4.75	0.39	0.58	0.86	1.31	52.13
DMOG	367.9	15.360	658	846.2	3.924	639.1	139.1	1.859	417.6	3.206	824.5
Docetaxel	0.009	0.002	0.002	0.006	0.008	0.004	0.002	0.002	0.008	0.001	0.013
Doxorubicin	1.53	0.012	0.207	0.037	0.855	0.211	0.034	0.905	0.957	0.204	0.074
EHT 1864	260.8	14.76	40.91	700.6	95.64	464.0	58.12	297.3	216.4	14.35	363.3
Elesclomol	0.019	0.031	0.040	0.068	0.038	0.141	0.005	0.122	0.005	0.008	0.500
Embelin	21.82	8.35	28.52	7.52	55.66	6.41	4.70	11.01	14.81	21.92	8.76
Epothilone B	0.025	0.000	0.005	0.000	0.144	0.002	0.001	0.037	0.023	0.004	0.001
Etoposide	9.09	0.83	1.01	1.98	302.6	16.86	2.66	7.95	61.27	23.52	0.59
FH535	14.34	2.64	3.72	5.53	22.29	6.84	0.60	83.23	8.12	21.57	4.69
FTI-277	41.53	83.08	21.02	3.30	23.64	18.84	32.83	65.56	49.26	63.89	12.04

GDC-0449	65.44	897.4	129.1	1.515	38.47	878.3	29.67	763.8	609.9	541.8	18.32
GDC0941	16.71	0.90	30.56	1.47	6.75	6.08	0.90	3.31	5.92	0.80	151.5
Gefitinib	35.51	15.81	25.42	40.56	32.36	10.97	44.41	25.44	5.34	24.02	0.22
Gemcitabine	0.028	0.000	0.051	0.000	2.50	0.037	0.023	0.52	2.83	2.65	0.009
GSK-1904529A	8.19	29.27	51.92	33.69	48.54	35.82	59.71	49.83	7.31	12.90	18.99
GSK-650394	91.65	1.82	32.07	7.27	879.2	28.95	7.08	669.2	760.5	1.794	20.50
GW 441756	116.2	22.34	55.11	105.4	81.74	171.3	8.13	191.1	129.4	70.58	4.01
IPA-3	280.4	2.18	1.254	2.562	1.516	367.9	18.14	1.947	911.1	47.38	8.21
JNJ-26854165	4.56	17.19	14.62	22.51	36.26	14.46	12.83	59.18	36.03	1.21	15.78
JNK Inhib. VIII	378.6	279.6	684.7	562.9	516	936.5	57.23	238.1	42.28	586.7	19.87
JNK-9L	0.97	0.39	0.16	0.36	3.80	0.35	1.40	0.34	4.55	0.90	0.48
KU-55933	462.7	60.99	464.7	67.92	87.17	198.8	17.13	147.7	43.80	682.8	3.27
LAQ824	0.07	0.02	0.04	0.02	0.05	0.03	0.04	0.30	0.07	0.05	0.03
Lenalidomide	159.8	503	280.9	330.6	224.1	422.2	71.12	563.7	219	311.9	336.7
LFM-A13	81.98	464.4	703.1	865.8	344.7	42.52	600.8	979.3	583.9	178.5	445.2
Methotrexate	22.90	16.20	0.79	0.16	0.27	4.41	4.66	1.94	13.34	9.58	1.01
Midostaurin	0.28	0.33	0.24	0.98	9.86	0.49	20.74	15.85	4.67	35.34	0.32
Mitomycin C	3.24	0.01	1.02	0.04	2.90	0.77	0.28	1.34	4.07	0.45	0.01
MK-2206	177.5	2.86	137.2	9.55	0.22	40.57	1.77	23.41	11.58	0.61	0.65
Nilotinib	54.53	141.1	27.43	10.64	105.2	167.1	4.98	113.9	40.13	94.49	68.00
NSC-87877	1.911	657.8	441.1	205.6	1.559	1.169	542.7	392.3	815.5	1.640	1.713
NU-7441	121.5	80.22	212.8	17.62	30.81	100.5	5.93	13.03	8.27	17.52	98.49
Nutlin-3a	515.8	5.44	3.30	672.8	36.15	468.9	484.3	658.3	268.3	2.64	327.5
NVP-BEZ235	0.11	0.05	0.13	0.41	0.03	0.08	0.04	0.06	0.01	0.04	14.51
Obatoclax Mesylate	0.80	0.00	0.09	0.04	7.49	0.07	0.23	0.46	1.75	2.30	0.47
OSI-906	40.99	5.15	74.82	0.24	40.46	52.74	8.86	75.53	131.7	38.56	70.31
OSU-03012	4.16	1.06	9.30	3.59	249.1	6.36	18.81	7.37	8.25	31.05	2.31
PAC-1	236.5	35.31	97.56	6.71	35.14	2.09	191.2	160.1	120.3	21.38	195.8
Pazopanib	108	28.71	541.2	737.9	112.4	111.3	5.60	87.92	212.7	12.36	14.76
PD-0325901	0.16	3.86	0.29	0.01	0.05	0.39	0.01	0.00	9.07	24.43	20.28
PD-0332991	3.82	2.71	5.10	0.86	0.25	6.27	1.86	32.98	1.35	13.80	3.46
PD-173074	43.96	11.62	44.99	168.5	144.4	135.8	47.70	92.23	112.2	154.4	1.58
PF-4708671	631	83.29	54.03	922.9	125.9	1.083	170.1	133.7	214.2	62.79	196.3
PF-562271	2.50	91.71	3.51	9.12	65.60	2.35	15.42	164	6.13	37.19	2.49
PLX4720	1.079	39.86	923.4	2.72	87.62	247.2	17.04	0.22	53.36	1.174	137
QS11	42.39	2.26	51.47	8.45	1.381	56.65	8.88	348.7	55.22	13.43	7.86
RDEA119	1.12	162.5	0.52	0.09	1.07	1.60	0.32	0.03	75.77	410.8	1.06
RO-3306	273.7	11.18	19.67	42.83	80.39	385.4	10.50	123.7	18.00	128.5	209.9
SB 216763	74.90	145.9	258.4	27.18	433.7	59.73	70.22	751.8	268.7	246	11.65
SB590885	29.60	115.4	50.88	2.13	7.14	502.4	142.4	0.08	196.2	321.2	61.53
Shikonin	1.02	0.43	0.39	0.64	0.62	0.33	0.28	1.49	0.86	41.59	0.19
SL 0101-1	654.5	437	161.8	1.091	79.55	388.7	531.5	476.7	239.3	1.131	100.3
Temsirolimus	0.68	0.57	3.41	18.00	1.15	0.96	0.10	0.46	0.02	0.18	0.56
Thapsigargin	0.001	0.000	0.001	0.005	1.14	0.007	0.003	0.014	0.008	0.032	0.009
Tipifarnib	5.83	0.17	4.88	0.55	72.39	4.33	0.32	580	11.29	11.46	8.66
TW 37	0.108	0.34	0.053	2.13	0.97	2.48	0.48	2.83	0.72	0.81	0.052
Vinblastine	0.030	0.032	0.007	0.050	0.013	0.005	0.004	0.009	0.024	0.008	0.099
Vinorelbine	0.037	0.005	0.003	0.003	1.36	0.004	0.006	0.021	0.104	0.018	0.005
Vorinostat	4.70	5.01	1.18	3.12	2.71	11.37	0.77	2.67	7.71	6.80	748.2
VX-702	141.9	193.7	78.45	337.1	178.6	145.5	40.09	160.2	147	109.1	0.15
ZM-447439	28.49	89.42	1.81	8.62	44.97	19.38	2.44	11.80	28.46	59.37	125.3

2.4.2 Drug IC_{50} Estimations

The IC_{50} values were determined for 17-AAG (Sigma-Aldrich), CCT007093 (Sigma-Aldrich), LFM-A13 (Sigma-Aldrich), PAC-1 (Sigma-Aldrich), IPA-3 (Sigma-Aldrich), AKT inhibitor VIII (Sigma-Aldrich), and BAY 61-3606 (Sigma-Aldrich) in the human cell lines MCF7, MDA-MB-231, LNCaP, PC-3, DU 145, HT-29, HCT 116, ASPC-1, MIA PaCa-2, and BxPC-3. Inhibition of the cell proliferation was monitored by means of the SRB assay (see Experimental Methods 2.1.3.4). Cells were incubated for 72 h with a serial dilution (up to 8 concentrations) of each drug (Table 8). Calculations were done using the Excel add-in ED50plus v1.0 development by Mario H. Vargas (2000). The software is a pharmacological analysis tool in the form of a Microsoft Excel worksheet, which allows creating and analyzing dose-response curves. The fit was done using sigmoid and linear regressions.

Table 8. Compounds and concentrations (μM) used for the IC_{50} calculations.

Dilution	17-AAG	CCT007093	LFM-A13	PAC-1	IPA-3	AKT inhibitor VIII	BAY 61-3606
1	10	200	400	200	400	100	200
2	5	100	200	100	200	50	100
3	2.5	50	100	50	100	25	50
4	1.25	25	50	25	50	12.5	25
5	0.625	12.5	25	12.5	25	6.25	12.5
6	0.313	6.250	12.500	6.250	12.500	3.125	6.250
7	0.156	3.125	6.250	3.125	6.250	1.563	3.125
8	0.078	1.563	3.125	1.563	3.125	0.781	1.563

2.5 Dependencies between Cellular Properties and Gene Expression or Drug IC50s

The linear correlation value (r coefficient) between the metabolic features of the cell lines and the gene expression or drugs' IC_{50}s were obtained with the Pearson algorithm using the *Hmisc* package (version 3.17-4) and the function rcorr (Harrell et al. 2015) (Supplements 5.1.7) in *R programming language*.

2.6 Graphics and Statistics

Plots and statistics were done using *R programming language* version 2.7.0. Barplot. boxplots and scatterplots were drawn using the ggplot2 package version 2.1.0 functions ggplot and qplot (Wickham 2009). For the linear correlation scatterplots the geom_smooth method="lm", se = FALSE was used. Heatmaps of the gene expression were constructed using the *pheatmap* package version 1.0.8 and RColorBrewer with the function colorRampPalette (green, black, red) (Kolde 2015) (Supplements 5.1.6).

3. Results

To get a better understanding of the properties of cancer cells, assessment of the membrane capacitance and a set of metabolic assays were performed on different cancer cell lines. These parameters included the direct measurement of cellular capacitance (cell membrane properties), respiratory activity (O_2 consumption), glycolytic activity (acidification), energy metabolism (ATP level), intracellular ROS accumulation, and mitochondrial mass and proliferation rates (MTT and SRB assays) (Table 9). The information obtained from the studied cell lines was compared with public gene expression data (Barretina et al. 2012, De Schutter et al. 2013) and drugs IC_{50}s (Yang et al. 2013) to search for correlations. The results revealed a set of candidate genes that could be involved in cell matrix adhesion or cell-cell interactions and cancer metabolism, as well as new relations between known drugs and cancer features.

Table 9. Mean value of the cell capacitance and metabolism features of cancer cell lines. Cellular capacitance, glycolytic and respiration activities were measured with the Bionas 2500 online monitoring system. Cells were seeded in chips containing IDES sensors for the estimation of the cellular capacitance. ISFET sensors were used to measure glycolytic activity and Clark electrodes for determining respiration. Lower capacitance indicates a stronger cell matrix adhesion and cell-cell contact interactions. Higher respiration and glycolysis are observed in cells with greater pA/s and µV/s values, respectively. The ATP level was measured with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). The luminescence is proportional to the ATP concentration in the cells. The mitochondrial mass content was determined using MitoTracker Green (Thermo Fisher). The fluorescence intensity is proportional to the mitochondrial mass of the cells. ROS accumulation in cells was measured using dihydroethidium (DHE). The fluorescence intensity produced is directly related to the ROS accumulation in the cells. The proliferation rate was estimated using the MTT and SRB assays. MTT reagent labels active mitochondria and its reduction produces a colored compound that can be measured by absorbance at 595 nm. The SRB dye is incorporated into fixed cells and the amount of incorporated dye is proportional to the amount of cells; the dye has an absorbance at 510 nm. SD = Standard deviation. At least three independent experiments with three technical replicates were used to calculate the means of each assay per cell line.

Cell Line/Assay	Cell Capacitance (nF)	SD (nF)	Respiration Activity (pA/s)	SD (pA/s)	Glycolysis Activity (µV/s)	SD (µV/s)	ATP (Luminescent counts)	SD (Luminescent counts)	Mitochondrial Mass (Fluorescence intensity)	SD (Fluorescence intensity)	ROS (Fluorescence intensity)	SD (Fluorescence intensity)	SRB (Absorbance 510 nm)	SD (Absorbance 510 nm)	MTT (Absorbance 595 nm)	SD (Absorbance 595 nm)
MDA-MB-231	23.88	3.35	1.26	0.25	29.20	5.84	1,054,269	47,684	69.0	9.59	126.1	21.39	1.72	0.32	0.61	0.12
MCF7	6.02	0.98	1.67	0.37	19.30	3.55	1,650,847	158,507	105.6	12.72	117.2	16.05	1.88	0.31	0.63	0.11
HCT 116	15.15	4.79	2.47	0.57	42.41	5.08	688,385	94,945	130.6	24.14	62.0	12.43	4.03	0.30	1.69	0.37
HT-29	12.64	2.43	1.25	0.25	33.53	7.71	424,044	109,680	121.4	27.14	33.4	3.84	3.61	0.39	1.17	0.12
LS174T	23.90	3.33	2.45	0.39	20.81	5.46	612,440	50,000	69.2	10.53	57.2	11.33	3.70	0.52	1.32	0.12
AsPC-1	20.77	2.47	1.77	0.31	22.10	4.60	1,270,124	124,264	90.9	16.61	67.8	5.55	2.00	0.27	0.78	0.05
JoPaca-1	23.87	3.50	0.78	0.21	16.96	1.71	1,507,901	231,978	128.1	9.17	77.1	5.00				
MIA PaCa-2	21.75	3.25	2.01	0.41	23.76	2.10	1,330,008	126,339	90.8	9.69	95.7	12.52	2.41	0.28	0.93	0.21
Panc-1	10.16	1.13	3.41	0.27	26.54	4.54	1,498,013	170,845	108.4	1.32	148.7	22.98	2.28	0.50	0.83	0.16
BxPC-3	7.14	0.40	1.53	0.11	16.57	3.11	928,904	165,540	112.4	14.75	103.7	13.04	1.93	0.34	0.76	0.03
LNCaP	31.99	2.64	3.10	0.47	35.42	6.80	1,842,348	15,369	89.9	6.09						
PC-3	19.71	2.47	1.73	0.12	22.86	1.95	1,150,409	201,682	119.7	5.57	115.7	8.43	2.28	0.43	0.55	0.30
DU 145	3.61	0.32	1.72	0.20	23.99	4.21			116.8	8.12	58.2	6.40				
HaCaT	6.42	1.77	0.97	0.14	13.22	11.21	1,070,569	266,976	62.6	4.79	68.6	6.10				
MDA-MB-435	12.96	0.70	1.63	0.34	30.38	8.59	475,630	14,342	109.9	6.74	58.6	0.74				
HeLa	27.47	1.78	2.32	0.26	38.36	5.72	461,821	85,376	156.4	8.14	46.4	2.76	2.54	0.43	1.81	0.05
Hep G2	9.24	0.97	1.55	0.23	25.98	5.79	563,945	142,399	96.7	14.45	85.7	6.03	2.46	0.30	0.44	0.23
1321N1	14.16	4.07	0.86	0.10	27.93	3.39	1,449,614	225,083	110.8	7.91	74.1	13.75				

3.1 Cellular Capacitance and Metabolism

Data of cellular capacitance and metabolism was obtained from cell lines belonging to breast (MDA-MB-231 and MCF7), colon (HCT116, HT-29, and LS174T), pancreas (AsPC-1, MIA PaCa-2, JoPaCa-1, Panc-1, and BxPC-3), prostate (LNCaP, PC-3, and DU145), skin (HaCaT and MDA-MB-435), cervix (HeLa), liver (HepG2), and central nervous system (1321N1) (Experimental Methods 2.1.1, Table 1). HaCaT corresponds to keratinocytes that come from normal tissue and is the only noncancerous cell line included in this study. For the measurements, all the cell lines were cultivated in DMEM media with 10% FCS and 1% PenStrep in a 5% CO_2 incubator at 37°C with 95% humidity (see Experimental Methods 2.1.1 for details). The data were obtained from online monitoring (real-time measurements) and end-point assays. Cellular capacitance, glycolysis, and respiratory activities were estimated in real time using the Bionas 2500 system (Experimental Methods 2.1.2). The Bionas 2500 instrument contains a perfusion system that allows feeding of the cells with a flow of culture media for 4 min “go” followed by 4 min “stop” protocol. In the "stop" phase, cells consume the nutrients in the media and at the same time the measurements are done (Experimental Methods 2.1.2.2). Energy metabolism (ATP level), mitochondrial mass, ROS accumulation and proliferation rates of the studied cell lines were determined using end-point assays (Experimental Methods 2.1.3, Table 9).

3.1.1 Cellular Capacitance

The cell capacitance was measured with the Bionas 2500 online monitoring system (Experimental Methods 2.1.2, Fig. 1). For each cell line equal numbers of cells were seeded in sensor chips with an IDES sensor (Experimental Methods 2.1.2.1). In the sensor chips, an alternating current is applied and interruption of electric current is recorded. Cells that are seeded in the chip block the flow of electrons between the two electrodes of the IDES sensor. Electricity does not affect the viability of the cells, neither their physical properties nor metabolism (Ehret et al. 1997). The chips in the Bionas 2500 perform a continuous determination of the cellular capacitance. The IDES sensor allows to detect changes in cellular morphology, confluence, cell-cell contact, and cell-

matrix contacts (Ehret et al. 1997). Lower capacitance (in the nano Faraday, nF, scale) indicates stronger cell-matrix adhesion and cell-cell physical interactions. The prostate carcinoma cell line DU145 and the breast adenocarcinoma cell line MCF7 showed the lowest capacitance values, displaying the greatest interaction between the cells and the chip surface. In contrast, the prostate adenocarcinoma LNCaP and the cervix adenocarcinoma HeLa cells showed the highest capacitance values (Fig. 2). For the calculation, three independent experiments were performed for MDA-MB-435 and LNCaP (cell lines with fewer replicas) and 97 independent experiments for MCF7 (cell line with most replicas). All of them with six technical replicates. The data comes from selected experiments in which the culture and experimental conditions were identical in all the cell lines (Experimental Methods 2.1.2.3).

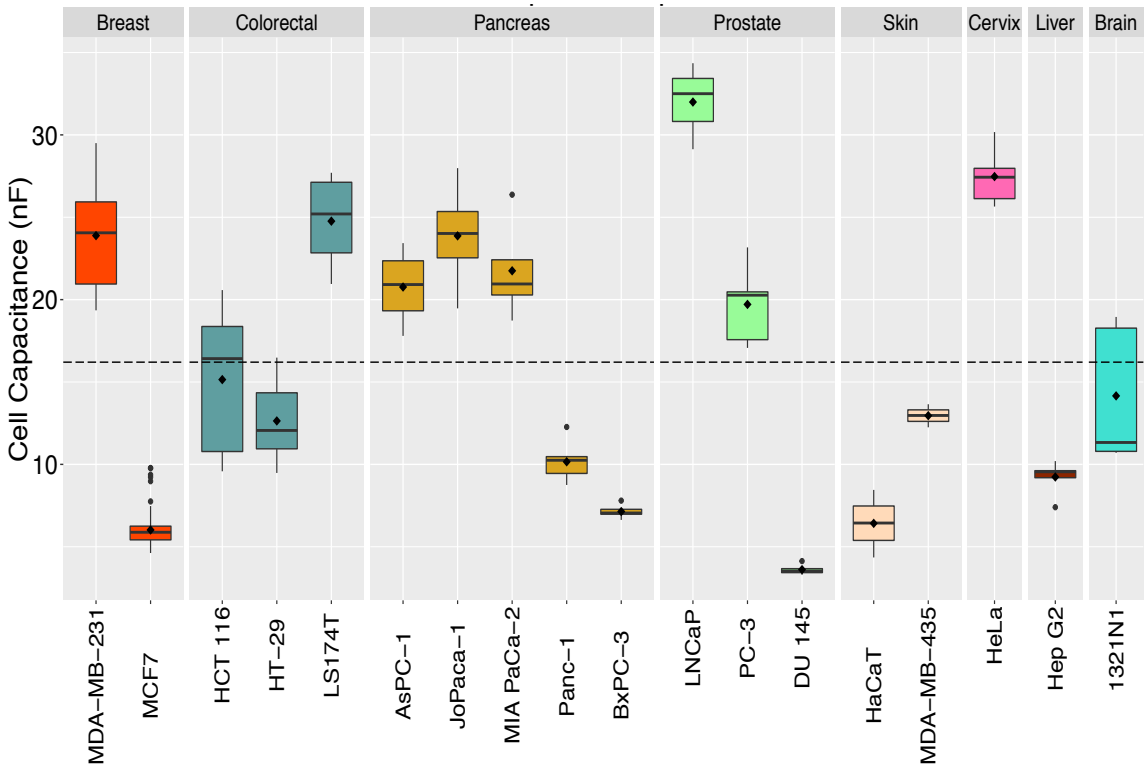


Fig. 2. Cellular capacitance per cell lines. The capacitance was measured on the Bionas 2500; cells are seeded in chips with IDES sensors. Equal numbers of cells per cell line were seeded in sensor chips. Higher capacitance values indicates low cell-cell and cell-matrix adhesion. Box-plot diagram for different cell lines. *Box* = 25th and 75th percentiles; *error bar* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value for all the analyzed cell lines; *black dots out of the box* = outliers. Cell lines are grouped by tissue type. At least three independent experiments were done per cell line and each one with six technical replicates.

3.1.2 Glycolytic Activity

The extracellular acidification rate was measured as a proxy for the cell's glycolytic activity (Experimental Methods Fig. 1). Equal numbers of cells per cell line were seeded in sensor chips with ISFET sensors (Experimental Methods 2.1.2.1). The ISFET sensor measures the pH changes in the media by detecting ion concentrations (Bergveld 1981). The chips are located in the Bionas 2500 for a continuous online monitoring of the acidification rate. The ISFET sensor allows to detect changes of media acidification due to lactic acid production by cells as a result of aerobic glycolysis, known to be activated in cancer cells (Warburg effect). Cell lines with higher acidification rate showed an elevated glycolysis pathway (Fig. 3). The colorectal carcinoma cell line HCT116 and the cervix adenocarcinoma HeLa cells showed the highest acidification rate and also displayed strong glycolytic metabolism in all the tested cell lines (Fig. 3). In contrast, the noncancerous skin keratinocyte HaCaT showed the lowest glycolytic activity (Fig. 3). The HaCaT and HT-29 cell lines show a large data dispersion, showing the largest distance between data in the 25th and 75th percentiles. Four and sixteen independent experiments were used to determine the acidification rates for the HaCaT and HT-29 cell lines, respectively (data not shown). All independent experiments for all cell lines are composed of six internal replicates. Selected data comes from experiments in which the culture and experimental conditions remained the same for all the cell lines (Experimental Methods 2.1.2.3).

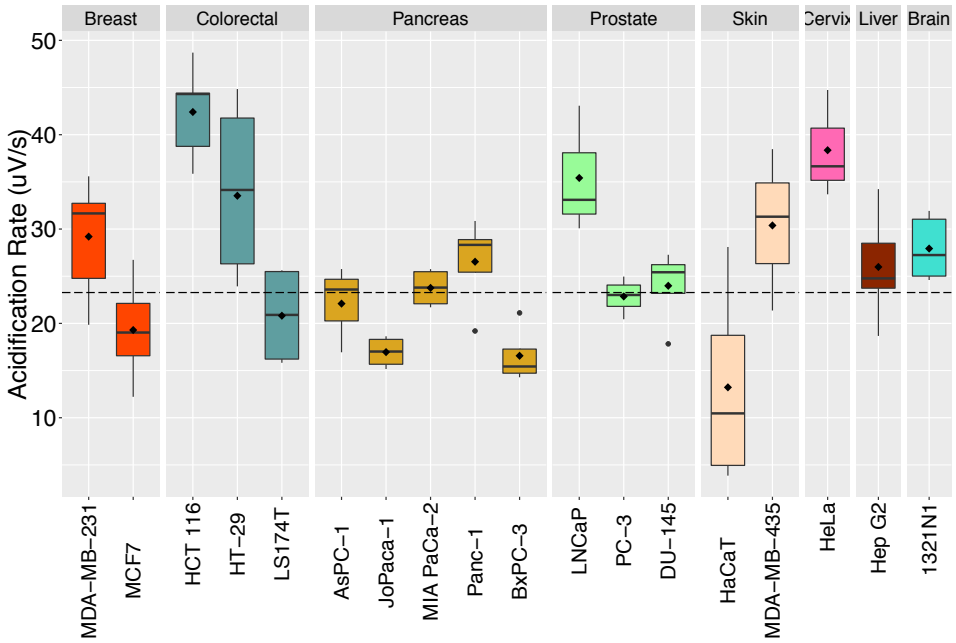


Fig. 3. Glycolytic activity per cell lines. The glycolytic activity was measured on the Bionas 2500 ; cells are seeded in chips with ISFET sensors. Equal numbers of cells per cell line were seeded in sensor chips. Higher acidification rate indicates stronger glycolytic activity. Box-plot diagram for different cell lines. *Box* = 25^{th} and 75^{th} percentiles; *error bar* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value for all the analyzed cell lines; *black dots out of the box* = outliers. Cell lines are grouped by tissue type. At least three independent experiments were done per cell line and each one with six technical replicates.

3.1.3 Respiration Activity

The mitochondrial activity of the cells was measured as oxygen consumption determined by Clark electrodes of the Bionas 2500 system in real time. Clark electrodes measure free oxygen in the culture medium (Clark et al. 1953). These electrodes are equipped with oxygen-permeable membranes to enable the free diffusion of oxygen (Clark et al. 1953). A fixed voltage is applied to the electrode, which allows oxygen reduction in the cathode, and oxidation of silver in the anode, thus generating an electric current. The amount of current produced is proportional to the oxygen concentration in the medium (Experimental Methods 2.1.2). The Bionas software calculates the difference between oxygen levels in the "go" cycle with those at the end of the "stop" cycle. Cell lines with a higher oxygen consumption showed a stronger respiratory activity (Fig. 4). The pancreatic cell lines Panc-1 (carcinoma) and JoPaCa-1 (carcinoma) show the highest and the lowest oxygen consumption in the tested cell lines, respectively. The prostate adenocarcinoma LNCaP also shows strong respiration activity, contrary to its tissue-type pairs (PC-3 and DU145), which rather show a respiration near the overall mean (Fig. 4). The skin cell line HaCaT, the brain astrocytoma 1321N1, and the pancreas JoPaCa-1 are found in the low oxygen-consumption range (Fig. 4).

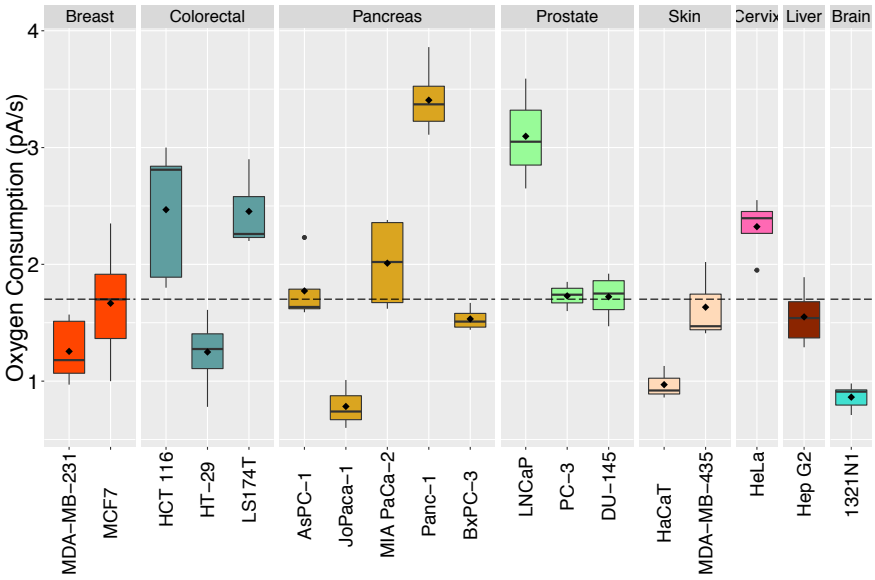


Fig. 4. Respiration activity per cell line. The oxygen consumption was measure in Bionas 2500; cells are seeded in chips with Clark electrodes. Equal numbers of cells per cell line were seeded in sensor chips. Higher oxygen consumption indicates stronger respiration rate. Box-plot diagram. *Box* = 25th and 75th percentiles; *error bar* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value between the analyzed cell lines; *black dots out of the box* = outliers. Cell lines are grouped by tissue type. At least three independent experiments were done per cell line and each one with six technical replicates.

3.1.4 Energy Metabolism (ATP Level)

The energy metabolism of the cells was measured as the amount of ATP produced by using firefly (*Photinus pyralis*) luciferase activity (see Experimental Methods 2.1.3.1). In this assay, the luminescence produced is proportional to the extracellular ATP concentration present in the cell suspension (Experimental Methods 2.1.3.1). At least three technical replicates and three independent measurements were done per cell line.

The lowest levels of energy metabolism activity are found in the colorectal cell lines (the carcinomas HCT116 and HT-29 and the adenocarcinoma LS174T). Also, one of the skin cell lines (the melanoma MDA-MB-435), the cervix adenocarcinoma HeLa and the liver carcinoma Hep G2) present low levels of energy metabolism activity. On the other hand, high levels of energy metabolism activity is found in four out of the five pancreatic cell lines (AsPC-1, JoPaCa-1, MIA PaCa-2, and Panc-1), with the remaining BxPC-3 having ATP levels corresponding to the mean of all the tested cell lines. The

breast adenocarcinoma MCF7 and the brain astrocytoma 131N1 are also among the cell lines with high levels of energy metabolism activity (Fig. 5).

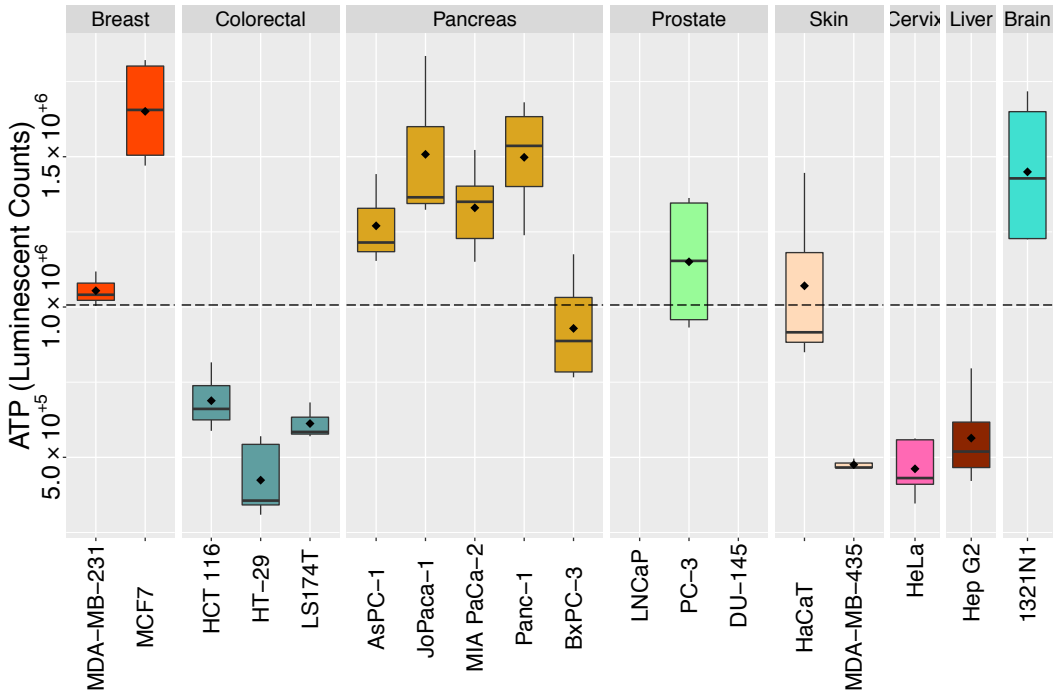


Fig. 5. Energy metabolism (ATP level) per cell line. The ATP production of the cell lines was measured with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). The luminescent counts are proportional to the ATP concentration. Box-plot diagram: *Box* = 25th and 75th percentiles; *error bars* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value between the analyzed cell lines; *black dots out of the box* = outliers. Cell lines are grouped by tissue type. At least three independent experiments were done per cell line and each one with three technical replicates.

3.1.5 Total Mitochondrial Mass

The mitochondrial mass of the different cell lines was estimated using MitoTracker Green compound that permeates cell membranes and labels active mitochondria (see Experimental Methods 2.1.3.2). At least three independent measurements with three technical replicates were carried out per cell line. High fluorescence intensity is proportional to the mitochondrial mass in the cells. The highest levels of mitochondrial mass were found in the cervix HeLa cell line. This cell line shows higher levels than all the other studied cell lines. It is followed in intensity by two of the three colorectal cell lines, the carcinomas HCT116 and HT-29. These two colorectal cell

lines present a high data dispersion, showing higher values of standard deviation and 25th-to-75th percentil distances. The cell lines with the lowest levels of active mitochondria are the breast adenocarcinoma MDA-MB-231, the colorectal adenocarcinoma LS174T, and noncancerous skin cell line HaCaT (Fig. 6). The breast, pancreatic, and prostate cell lines appear together around the mean of the total cell lines used for the analysis (Fig. 6).

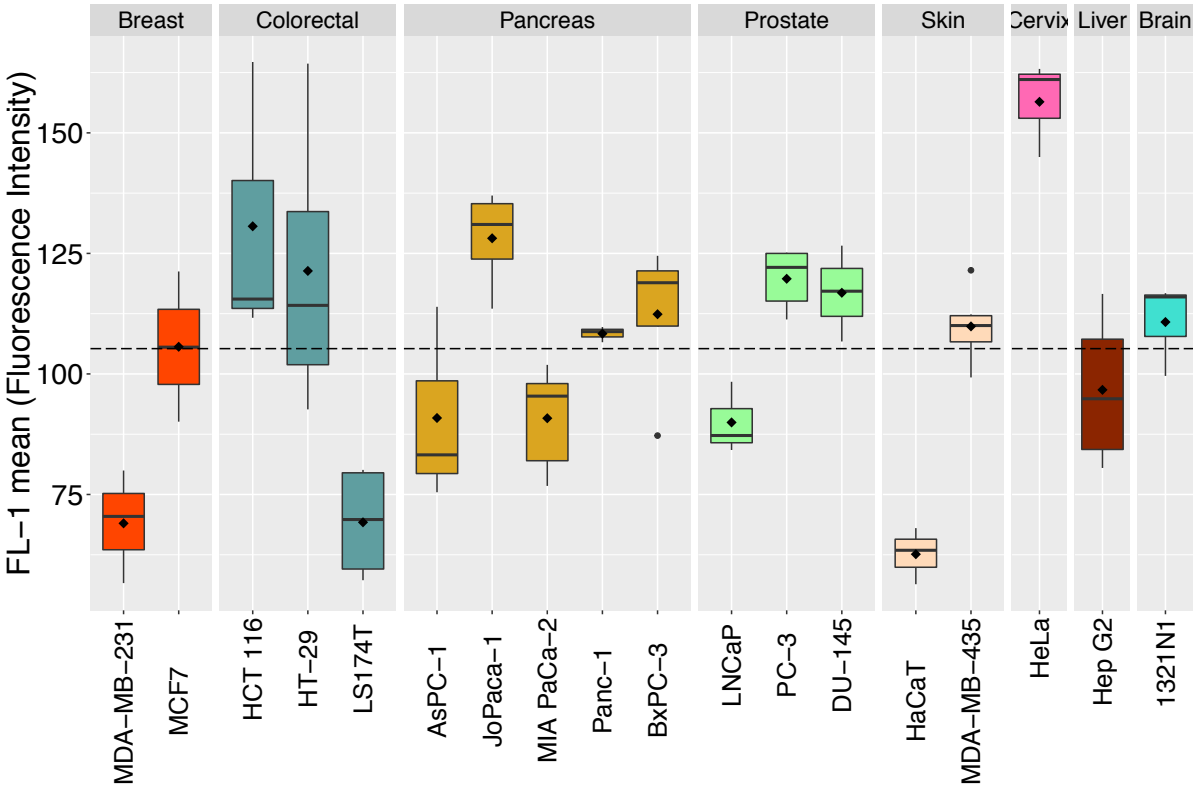


Fig. 6. Total mitochondrial mass per cell line. The mitochondrial mass content of the cells was determining using MitoTracker Green (Thermo Fisher). Cell lines with higher fluorescence intensity contain higher mitochondrial mass. Box-plot diagram: *Box* = 25th and 75th percentiles; *error bars* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value between the analyzed cell lines; *black dots out of the box* = outliers. Cell lines are grouped by tissue type. At least three independent experiments were done per cell line and each one with three technical replicates.

3.1.6 Reactive Oxygen Species (ROS) Accumulation

Reactive oxygen species (ROS) formation was measured using dihydroethidium (DHE) in the studied cell lines. DHE permeates into cell membranes and diffuses into the

cytosol, where it is oxidized by superoxide resulting in the emission of red fluorescence (Experimental Methods 2.1.3.3). At least three independent measurements with three technical replicates were done per cell line. The cell line with the highest levels of ROS is the pancreatic carcinoma Panc-1 followed by the breast adenocarcinoma MDA-MB-231 (Fig. 7). The lowest level of ROS accumulation is found in the colorectal HT-29 cell line. Its tissue type pairs, HCT116 and LS174T, also show low ROS, with its levels being under the mean for all cell lines tested (Fig. 7).

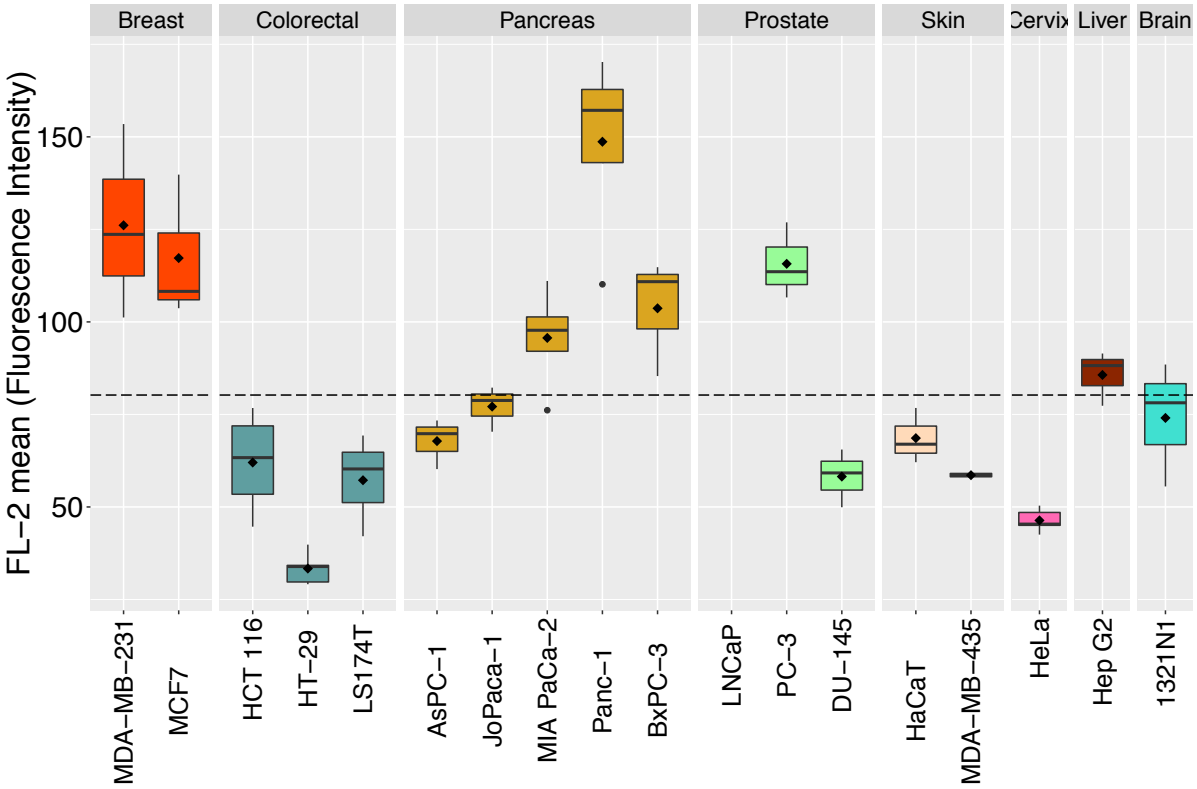


Fig. 7. Reactive oxygen species (ROS) formation per cell line. ROS accumulation in cells was measured using dihydroethidium (DHE). Cell lines with higher fluorescence intensity present a higher ROS formation. Box-plot diagram: *Box* = 25th and 75th percentiles; *error bars* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value between the analyzed cell lines; *black dots out of the box* = outliers. Cell lines are grouped by tissue type. At least three independent experiments were done per cell line and each one with three technical replicates.

3.1.7 Cell Proliferation Rate

The cell proliferation rate for different cell lines was estimated using two different approaches: *i*) metabolic activity measured by MTT assay and *ii*) the total

protein content of the cells using SRB assay. The MTT reagent diffuses into the cells and labels active mitochondria and therefore is an indirect measurement for cell proliferation. However, the SRB assay stains total proteins of fixed cells, after washing out the dead cells, which then is directly proportional to cell number and proliferation rate (Vichai & Kanyawim 2006) (Experimental Methods 2.1.3.4). For each cell line, equal numbers of cells were seeded and the cells were left growing for 72 h before performing the measurements. At least three independent measurements with three technical replicates were done per cell line and assay. The MTT and SRB assays proved to be reliable methods to compare proliferation rates in the different cell lines. Both yielded the same results for almost all tested cell lines (Figs. 8 and 9). Only HeLa and HepG2 cells showed contradictory information between the assays, being the MTT measurement in HeLa cells the one with the highest proliferation rate and HepG2 the one with the lowest. Besides, in SRB, both cell lines remain around the mean of all the tested cells (Figs. 8 and 9). In both assays, at the tissue-type level, the colorectal cell lines show the fastest proliferation and the breast cell lines the slowest proliferation rate (Figs. 8 and 9). In the three tissue types where the assays were tested, cell lines belonging to the same tissue type showed similar proliferation rates, i.e., "within cell-type". The prostate tissue type was only measured for the adenocarcinoma PC-3, therefore no "within cell-type" comparisons were possible (Figs. 8 and 9).

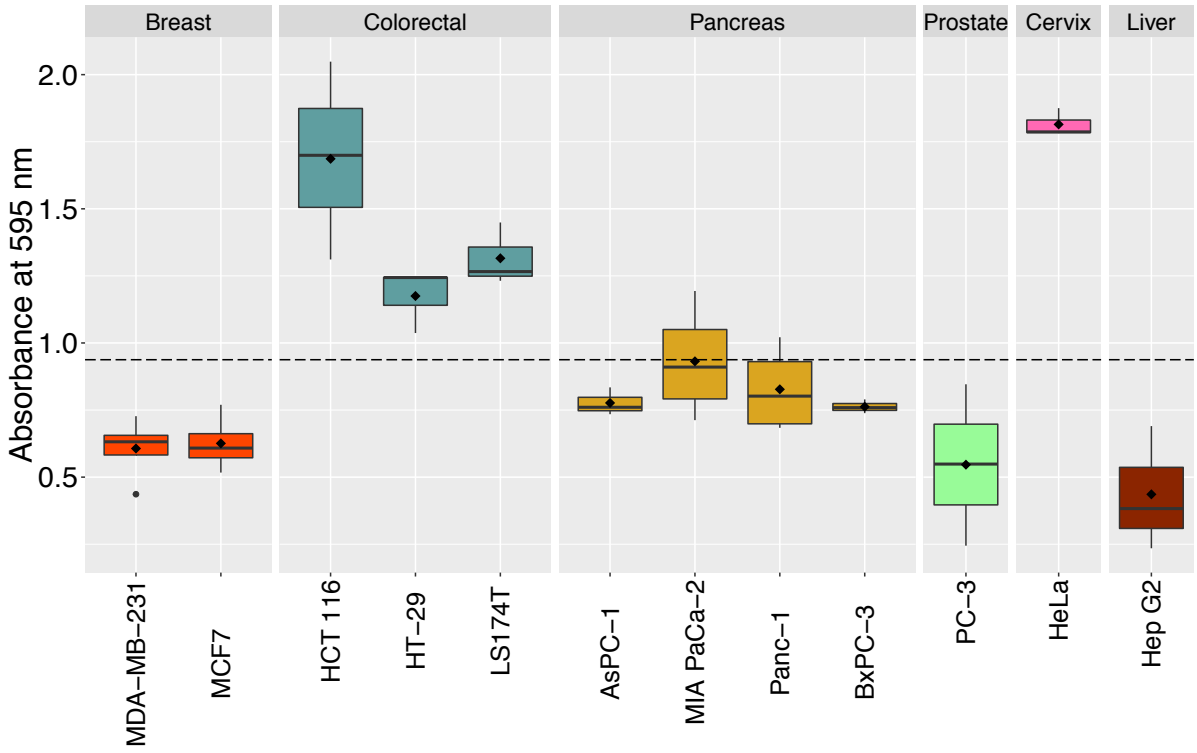


Fig. 8. Proliferation rate measured as metabolic activity (MTT assay) per cell line. The MTT reagent is reduced by active mitochondria producing a purple-colored product with an absorbance at 595 nm. Cells with higher absorbance present a higher metabolic activity that is used as an estimation of the proliferation rate. Equal low cell numbers per cell line were seeded in 96-well plates, the absorbance was measure after 72 h**.** Box-plot diagram: *Box* = 25th and 75th percentiles; *error bars* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value between the analyzed cell lines; *black dots out of the box* = outliers. Cell lines are grouped by tissue type. At least three independent experiments were done per cell line and each one with three technical replicates.

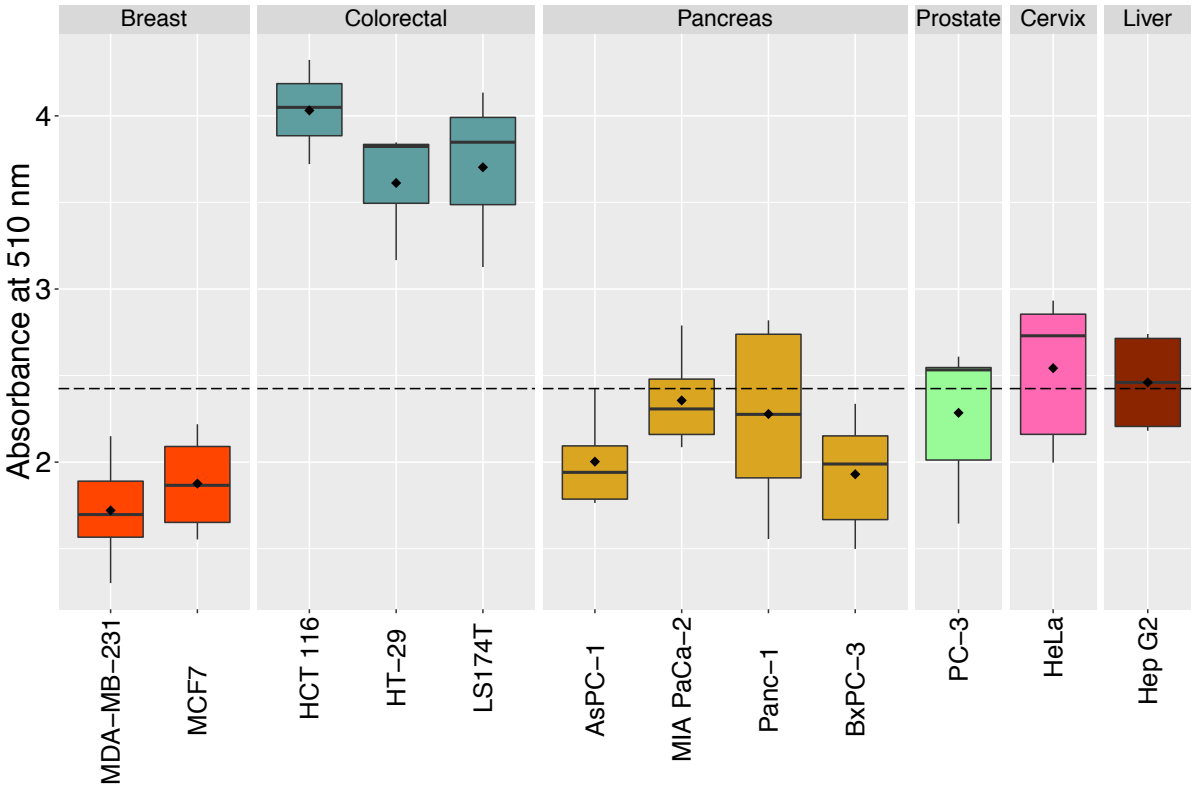


Fig. 9. Proliferation rate measured as total protein content (SRB assay) per cell line. In the SRB assay a dye is incorporated into fixed cells and the amount of incorporated dye is proportional to the amount of cells. Cells with higher absorbance have a higher protein content that is used as an estimation of the proliferation rate. Equal low cell numbers per cell line were seeded in 96-well plates, the absorbance was measured after 72 h. Box-plot diagram: *Box* = 25th and 75th percentiles; *error bars* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value between the analyzed cell lines; *black dots out of the box* = outliers. Cell lines are grouped by tissue type. At least three independent experiments were done per cell line and each one with three technical replicates.

3.2 Cellular Properties of the Cancer Cell Lines by Tissue Type. Summary Figure and Cluster Dendrograms

The basal metabolic profile and the cellular capacitance of the 18 cell lines belonging to different tissue types are shown in Fig. 10; this summarizes all the different features of the studied cell lines done before (Figs. 2 to 9).

The broad overview of the cellular properties in the studied cell lines allows comparisons of the cellular capacitance, the O_2 consumption (respiratory activity), media acidification (glycolytic activity), ATP levels (energy metabolism), intracellular ROS accumulation, mitochondrial mass, and proliferation rates (MTT and SRB assays) (Table 9) by cell lines and tissue type.

The two adenocarcinoma breast cancer cell lines differ in their capacitance and in their energy metabolism (ATP level) but present similar respiratory activity, ROS accumulation, and proliferation rates (Fig. 10). Hierarchical clustering performed on the data for the different assays (with measurements in each cell line) show the two adenocarcinoma breast cell lines in the same branch for the MTT proliferation assay and in proximal branches in the ROS formation and in the SRB proliferation assay (Fig. 11). However, respiration activity does not appear neither in the same or close branches, in fact, energy metabolism shows a closer distance between the two cell lines in the dendrogram (Fig. 11). By comparing the breast cancer cell lines with the other tissue types, it is possible to say that these cells present high ATP and ROS levels and low proliferation rates (Figs. 5, 7, 8, 9, and 10).

The two colorectal carcinoma cancer cell lines HCT116 and HT-29 show similar membrane capacitance, glycolysis activity (acidification), and mitochondrial mass, indicating discrepancies with their the adenocarcinoma LS174T (Fig. 10). Hierarchical clustering locates these two carcinoma cells in near branches and far away from the adenocarcinoma LS174T for cell capacitance, glycolysis, and mitochondrial mass assays (Fig. 11). The three cancer cell lines share similar levels in ATP, ROS, and proliferation rates (Fig. 10). Dendrograms of the mentioned assays confirm these observations, grouping the three cell lines in near branches for energy metabolism and ROS accumulation, and in the same branch for the proliferation assays (Fig. 11). Regarding the respiratory activity, the hierarchical clustering grouped LS174T and HCT 116 together in the same branch (Fig. 11). In contrast, for the breast and pancreas cancer cell

lines, cells of the colorectal tissue type present low ATP and ROS levels (Figs. 5, 7, and 10) and the highest proliferation rate among all the tissue types analyzed (Figs. 8, 9, and 10).

The pancreatic cells do not seem to show similarities between the type of disease (or type of tumor) and their cell properties, like the colorectal line does, with the two adenocarcinomas AsPC-1 and BxPC-3 or the three carcinomas MIA PaCa-2, JoPaCa-1, and Panc-1 having no similar features. The data for the O_2 consumption in pancreatic cells spreads in a wide range within this tissue type presenting the highest (Panc-1) and the lowest (JoPaCa-1) respiration activities in all tested cell lines (Fig. 10). Glycolysis activity, energy metabolism, and proliferation locate the pancreas cells roughly together, the measurements showing similar means (Fig. 10). The hierarchical cluster for the MTT proliferation assay locates the pancreas tissue type cells together in one group apart from the rest of the cell lines. Such a strong match is not displayed in the dendrogram for the SRB proliferation assay; clusters spread the pancreas cells in two of the three branches (Fig. 11). The dendrogram for energy metabolism shows the pancreas tissue type in close branches. Comparing the pancreas tissue type with breast and colorectal tissue types makes possible to observe that pancreas cells behave like breast cells showing relative high ATP and low proliferation. The ROS accumulation differs in the five pancreas cell lines (Figs. 5, 7, 8, 9, and 10).

The prostate cell lines PC-3 (adenocarcinoma) and DU145 (carcinoma) show almost equal respiration and glycolysis activities as well as mitochondrial mass and seems to differ from the adenocarcinoma LNCaP (Fig. 10). Hierarchical clustering for glycolysis and mitochondrial mass support the observations, showing the three cell lines in different branches but near the PC-3 to the DU 145. Dendrograms of the respiration activity present PC-3 and DU 145 in the same branch and far away of the LNCaP (Fig. 11). The highest and the lowest membrane capacitance of all the tested cell lines are present in the prostate cells being LNCaP that exhibits a weaker cell-cell contact and cell-matrix binding (high cell capacitance) and DU 145 with stronger cell adhesion and cell bindings (low cell capacitance).

The last tissue type group in which the data was collected was the skin cell line HaCaT, a noncancerous cell line, and the melanoma MDA-MB-435. These two cell lines were plotted together with the cervix HeLa, the liver HepG2, and the brain 1321N1. The mean value of the two skin cell lines seem to be similar for the cellular capacitance, O_2 consumption, acidification rate and ROS accumulation measurements (Fig. 10).

Hierarchical clusters build with the cell capacitance data show the two skin cell lines in different branches but together in one of the two main tails of the tree. In the glycolysis and respiration dendrograms, HaCaT and MDA-MB-435 are far away from each other; they cluster in separated branches. For the ROS accumulation, hierarchical clustering located the cells for the skin tissue type in different branches (Fig. 11).

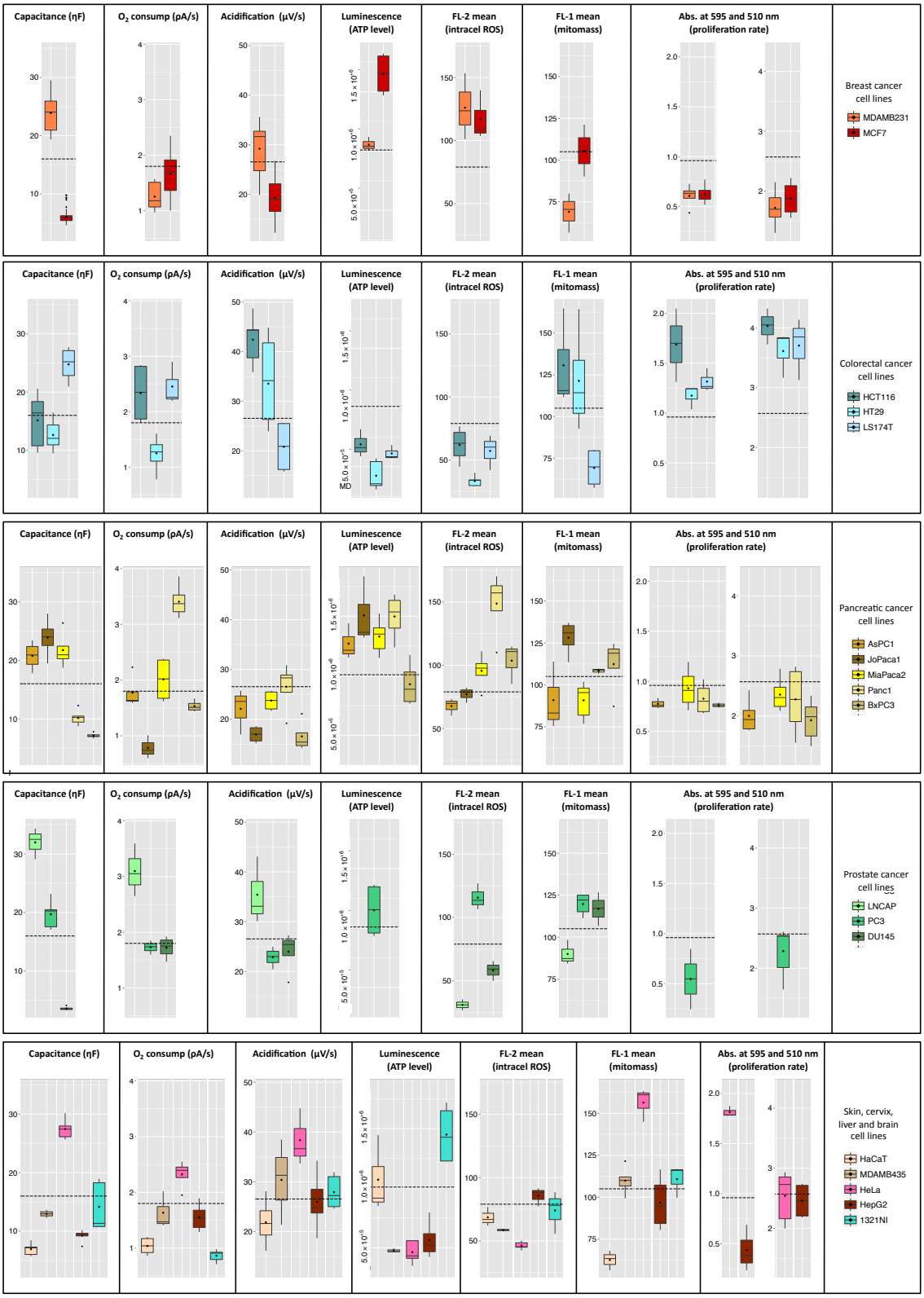


Fig. 10. Summary of cell capacitance and metabolic features of cell lines by tissue type. Cellular capacitance, glycolytic activity (acidification), and respiratory activity (O_2 consumption) were measured on the Bionas 2500 online monitoring system. Cells were seeded in chips containing IDES sensors for the estimation of the cellular capacitance, ISFET sensors to measure the acidification rate, and Clark-type electrodes for determining oxygen consumption. Lower capacitance indicates a stronger cell matrix adhesion and cell-cell contact interactions. A stronger acidification rate indicates a higher glycolysis level. Higher oxygen consumption indicates stronger respiration rate. The energy metabolism (ATP level) was measured with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). The luminescence is proportional to the ATP concentration in the cells. The mitochondrial mass content was determined using MitoTracker Green (Thermo Fisher). The fluorescence intensity is proportional to the mitochondrial mass of the cells. ROS accumulation in cells was measured using dihydroethidium (DHE). The fluorescence intensity produced is directly related to the ROS accumulation in the cells. The proliferation rate was estimated using MTT and SRB assays. MTT reagent labels active mitochondria and its reduction produces a colored compound that can be measured with absorbance at 595 nm. The SRB dye is incorporated into fixed cells and the amount of incorporated dye is proportional to the amount of cells, the dye has an absorbance at 510 nm. Box-plot diagram: *Box* = 25^{th} and 75^{th} percentiles; *error bars* = standard deviation; *black dot inside the box* = mean; *black line* = median; *dashed line* = mean value between the total analyzed cell lines; *black dots out of the box* = outliers. At least three independent experiments were done per cell line and each one with three technical replicates.

The cervix adenocarcinoma cell line HeLa shows a much higher mitochondrial mass than the rest of the studied cell lines (Figs. 6 and 10). Hierarchical clustering corroborates this observation localizing HeLa in the mitochondrial mass content dendrogram alone in a branch separate from the others. Looking deeper into the cellular property clusters, it is remarkable that in several of the metabolic assays HeLa cells appears closer to the colorectal cell lines, sharing with the three of them the same branch in the MTT proliferation assay dendrogram and being closer to the ones in the respiration activity and energy metabolism clusters. Moreover, in ROS formation dendrogram it shares the same branch with HT-29 and with HCT 116 in the glycolysis activity dendrogram (Fig. 11).

The liver carcinoma cell line Hep G2 seems to not share features with other cell lines except for cellular capacitance and glycolysis activity where, in the hierarchical clustering, it shares the same branches with pancreas carcinoma Panc-1 (Fig. 11).

The central nervous system astrocytoma 1321N1 cell line in the hierarchical clusters of the metabolic assay measurements appears near the pancreas cells JoPaCa-1 and Panc-1. The brain cell line localizes in the same branches with JoPaCa-1 regarding its respiratory activity and in the ROS accumulation dendrograms. Notice also that

1321N1 is the closest cell line to JoPaCa-1 and Panc-1 regarding its ATP level and appears near the branches of Panc-1 in glycolysis and mitochondrial mass (Fig. 11).

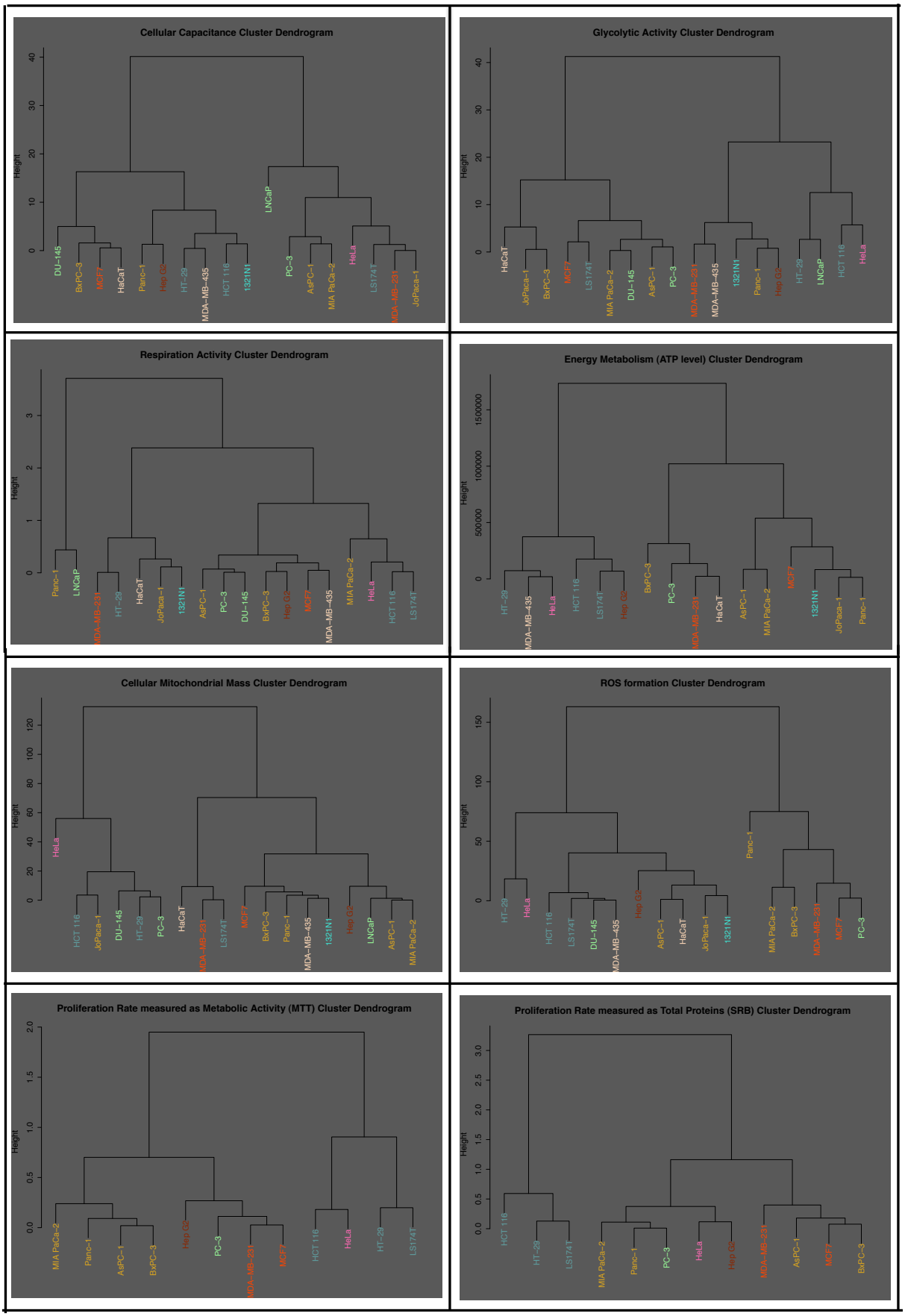


Fig. 11. Hierarchical clustering of cellular capacitance and metabolic features. Cell lines were grouped based on their electrical and metabolic properties. Cellular capacitance, glycolytic and respiratory activities were measured on the Bionas 2500 online monitoring system. Energy metabolism (ATP level) was measured with ATPlite 1step Luminescence ATP Detection Assay System (Perkin Elmer). The mitochondrial mass content was determined using MitoTracker Green (Thermo Fisher). ROS accumulation in cells was measured using DHE. The proliferation rate was estimated using MTT and SRB assays. The clustering calculations were done via *R programming language* and Euclidean distance approach. Cell lines belonging to the same tissue type share the same color. The height in the y-axis gives the values of the respective metabolism assay tested. The mean of at least three independent experiments per assay and per cell line was used for the cluster analysis.

3.3 Linear Relations between the Cellular Properties of the Studied Cell Lines

The data collected for the cellular properties of the studied cell lines (Figs. 2 to 10) was used to calculate a Pearson product moment correlation (see Experimental Methods 2.5) in order to search for linear relationships within the different cellular features. The Pearson correlation coefficient (r coefficient) indicates whether there is a linear relationship between two variables and how the two sets of data are related, i.e., whether a positive linear relation ($0 < r < 1$) or a negative linear relation ($-1 < r < 0$) exists, with r being the correlation coefficient. An r value equal to 1 indicates a perfect positive correlation between the two variables, and –1 is a perfect negative correlation. Zero indicates no linear correlation. Table 10 shows the r and its associated p-values obtained for the calculation of the Pearson product moment with the cellular properties data in the studied cell lines. The obtained significant (p-values < 0.05) relationships are plotted in Fig. 12.

The proliferation rate measured as metabolic activity by MTT and the proliferation rate measured as total protein content by SRB present a positive correlation ($r = 0.69$) (Fig. 12A). In the dependency graph it is possible to observe that the three colorectal cell lines have high proliferation rates for both measurements and localize near one another; the pancreatic cells present a low proliferation rate and localize together. The breast cancer cell lines exhibit an even lower proliferation rate.

A significant (p-value < 0.01) positive relation is also recorded for the glycolytic activity and the proliferation rate MTT assay ($r = 0.71$) (Fig. 12B), meaning that the cell lines that proliferate most, also possess a strong glycolytic activity. Similarly to the

proliferation rate assay dependency, colorectal cell lines display the highest acidification and proliferation rate values in the three tissue types tested, the breast tissue type cells show the lowest, and the pancreatic tissue type cell lines with moderate values. Cells belonging to the colorectal tissue type do not appear together in the plot due to their differences in the acidification rate. On the other hand, cells with a high proliferation rate accumulate less ROS (r = –0.68) (Figs. 12C and D). Notice that for this negative dependency the colorectal tissue type with its faster proliferation shows the lowest ROS formation. The opposite is the case in the breast cancer cell lines with high ROS and low proliferation.

It is remarkable that ROS accumulation negatively correlates with both MTT and SRB proliferation assays, having a significant (p-value < 0.05) inverse relationship with the metabolic activity as well as with the total protein content of the studied cell lines (Fig. 12C and D and Table 10). Moreover, ROS accumulation exhibits a significant (p-value < 0.01) positive relation with the energy metabolism (r = 0.64) (Fig. 12E). Pancreatic and breast tissue type cells show higher ROS and ATP levels in comparison with the colorectal ones which localize at the lower end in the correlation plot, showing low ROS accumulation and low energy metabolism (Fig. 12E). The last significant (p-value < 0.05) linear dependency found among the cellular properties of the studied cell lines is the negative relation between energy metabolism (ATP formation) and proliferation rate measured as total protein content in the SRB assay (r = –0.64) (Fig. 12F). Due to the positive dependency found between ATP level and ROS accumulation in cells, the negative relation between ATP and the SRB assay supports the dependency of ROS formation and proliferation rate in the studied cell lines.

Table 10. r coefficients and p values of the linear dependencies between the electrical and metabolic properties of the studies cell lines. A Pearson comparison between the different assays was used to search for correlations. The calculation was done using the mean of at least three independent experiments. Cellular capacitance, glycolytic and respiratory activities were measured on the Bionas 2500 online monitoring system. The ATP level was measured with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). The mitochondrial mass (mitomass) content was determined using MitoTracker Green (Thermo Fisher). ROS accumulation in cells was measured using DHE. The proliferation rate was estimated using MTT and SRB assays. The Pearson comparison calculation was done using *R programming language* package *Hmisc* function rcorr. In *yellow* the significant correlations (p -values < 0.05).

r coefficients	Capacitance	Respiration	Glycolysis	ATP	Mitomass	ROS	SRB
Capacitance	1						
Respiration	0.28	1					
Glycolysis	0.36	0.44	1				
ATP	0.14	0.11	-0.32	1			
Mitomass	-0.002	0.06	0.45	-0.21	1		
ROS	-0.11	0.24	-0.31	0.64	-0.24	1	
SRB	0.12	0.26	0.55	-0.64	0.21	-0.68	1.00
MTT	0.42	0.42	0.71	-0.55	0.53	-0.68	0.69

p -values	Capacitance	Respiration	Glycolysis	ATP	Mitomass	ROS	SRB
Respiration	0.2561						
Glycolysis	0.1387	0.0700					
ATP	0.5970	0.6650	0.2115				
Mitomass	0.9932	0.8096	0.0627	0.4180			
ROS	0.6655	0.3468	0.2221	0.0072	0.3476		
SRB	0.7092	0.4055	0.0630	0.0263	0.5220	0.0147	
MTT	0.1696	0.1765	0.0100	0.0654	0.0767	0.0154	0.0127

3.4 Cellular Properties and Affymetrix Gene Expression

Gene expression data was obtained from public repository files in GEO, NCBI (De Schutter et al. 2013, Barretina et al. 2012). Public accession numbers and cell lines are listed in Table 2. The data comes from experiments done in Affymetrix Human Genome U133 Plus 2.0 Arrays. To avoid introducing extra variability between the data obtained for the electrical and metabolic parameters with the Affymetrix arrays, all the gene expression data collected comes from cell lines cultivated in DMEM supplemented with 10% FCS in a 5% CO_2 incubator at 37°C, the same conditions set for the acquisition of the electrical and metabolism features. Gene expression data were obtained for 16 of the 18 cell lines used in this work. The cell lines were: MDA-MB-231 and MCF7 (breast tissue type), HCT 116 and HT-29 (colorectal tissue type), AsPC-1, MIA PaCa-2, Panc-1, and BxPC-3 (pancreas tissue type), LNCaP, PC-3, and DU 145 (prostate tissue type), HaCaT and MDA-MB-435 (skin tissue type), HeLa (cervix tissue type), Hep G2 (liver tissue type), and 1321N1 (central nervous system tissue type) (Experimental Methods, Table1). The raw data of the 16 cell lines collected from Affymetrix arrays was preprocessed using Bioconductor in *R programming language* and the Robust Multi-array Average (RMA) normalization method (Irizarry et al. 2003, Gautier et al. 2004) (Experimental Methods 2.3, Supplements 5.1.2).

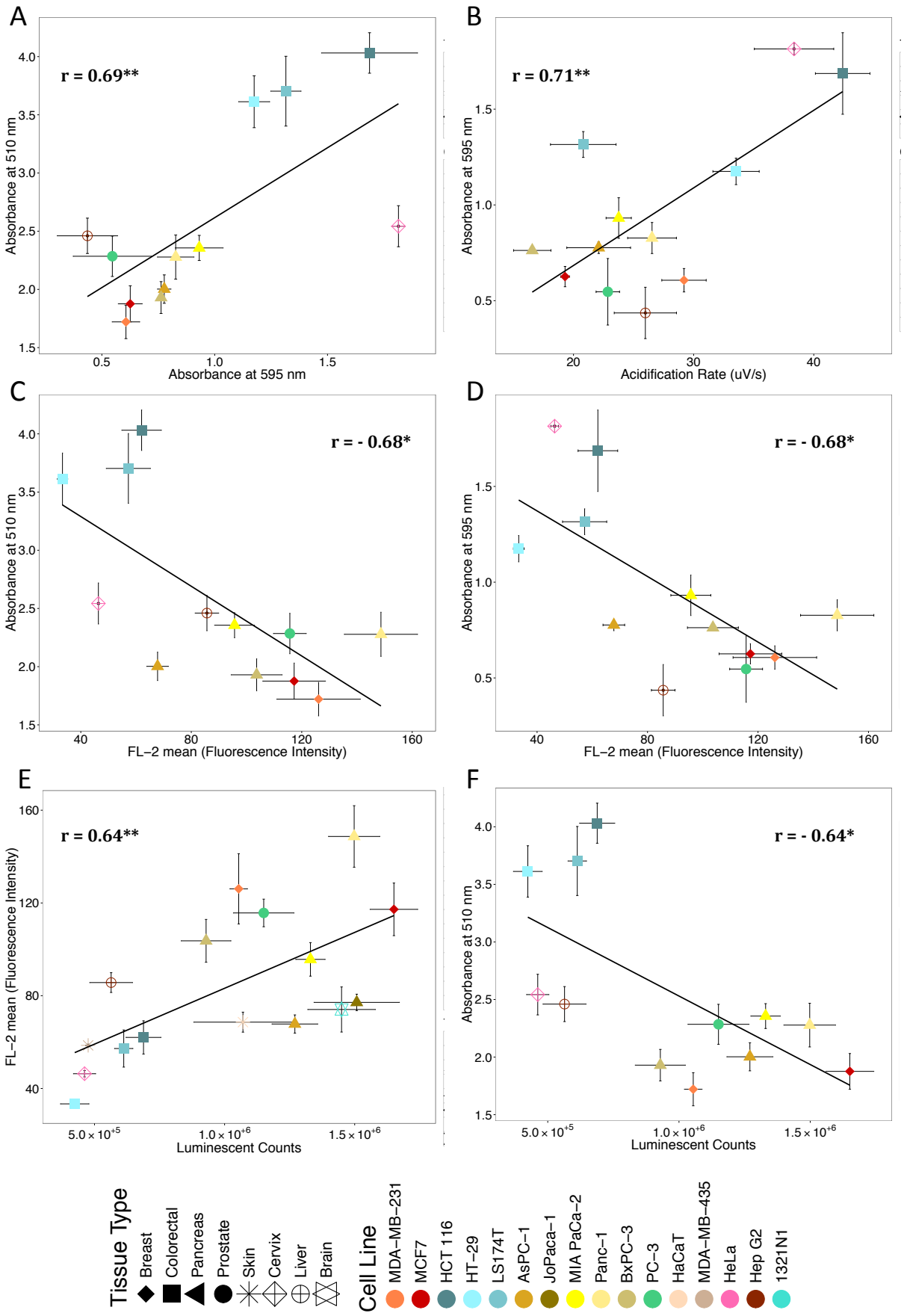

A
r = 0.69**
Absorbance at 510 nm
Absorbance at 595 nm
B
r = 0.71**
Absorbance at 595 nm
Acidification Rate (uV/s)
C
r = - 0.68*
Absorbance at 510 nm
FL−2 mean (Fluorescence Intensity)
D
r = - 0.68*
Absorbance at 595 nm
FL−2 mean (Fluorescence Intensity)
E
r = 0.64**
FL−2 mean (Fluorescence Intensity)
Luminescent Counts
F
r = - 0.64*
Absorbance at 510 nm
Luminescent Counts
Tissue Type
Breast
Colorectal
Pancreas
Prostate
Skin
Cervix
Liver
Brain
Cell Line
MDA−MB−231
MCF7
HCT 116
HT−29
LS174T
AsPC−1
JoPaca−1
MIA PaCa−2
Panc−1
BxPC−3
PC−3
HaCaT
MDA−MB−435
HeLa
Hep G2
1321N1

Fig. 12. Significant dependencies between the cellular features of the analyzed cell lines. Linear correlation coefficients between the cellular properties of the studied cell lines were obtained using the Pearson algorithm. Only the significant dependencies (p-values < 0.05) are shown. The proliferation rate was estimated using MTT and SRB assays. MTT reagent labels active mitochondria and its reduction produces a colored compound that can be measured by absorbance at 595 nm. The SRB dye is incorporated into fixed cells and the amount of incorporated dye is proportional to the amount of cells, the dye has an absorbance at 510 nm. The glycolytic activity (acidification rate) was measured on the Bionas 2500; cells were seeded in chips containing ISFET sensors. A stronger acidification rate indicates a higher glycolysis level. ROS accumulation in cells was measured using dihydroethidium (DHE). The fluorescence intensity produced is directly related to the ROS accumulation in cells. Energy metabolism (ATP level) was measured with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). Luminescence is proportional to ATP concentrations in cells. Pearson coefficients were calculated using the mean of at least three independent experiments per cell line and per metabolic assay. The calculations were done using *R programming language* package *Hmisc* and rcorr function. The shape of the *dots* gives the tissue types and the color indicates cell lines. The *black lines* represent the linear regressions. For each dependency the r correlation coefficient is shown; asterisks represent level of significance (p-values < 0.05* and p-values < 0.01**). *Error bars* = standard error (se). Linear correlations of: **A.** SRB and MTT proliferation assays. **B.** Glycolysis activity and MTT proliferation assay. **C.** ROS accumulation and SRB proliferation assay. **D.** ROS accumulation and MTT proliferation assay. **E.** Energy metabolism and ROS accumulation. **F.** Energy metabolism and SRB proliferation assay.

Entrez genes and gene symbols belonging to the metabolism-related pathways glycolysis, gluconeogenesis, pentose phosphate pathways, TCA, and electron transport chain were gathered using Wikipathways (Kelder et al. 2012) and DAVID (Huang et al. 2009)(Experimental Methods 2.3.1, Tables 3, 4, and 5 and Supplements 5.1.3). A selection of probe set IDs was done using the statistical method "Presence-Absence calls with Negative Probe sets" to estimate the significance of the detection on Affymetrix U133 series microarrays and to exclude the probe sets that could show an unspecific hybridization (Warren et al. 2007). The Presence-Absence call calculations were done using *R programming language* and Bioconductor (Experimental Methods 2.3.1 and Supplements 5.1.4). Additionally, all the probe sets with a _x suffix (mixed probe set that contains some probes that are identical or highly similar to other sequences) were deleted. For the glycolysis-gluconeogenesis together with the pentose phosphate pathways, the original 107 probe sets were reduced to 51 and just one probe set per gene was selected. From the 218 probe sets belonging to the TCA together with the electron transport chain pathways, 104 probe sets remained, with also just one probe set per gene. In order to search for possible relations with the metabolism properties of

the studied cell lines (Figs. 2 to 10) and their gene expression level in metabolism, the selected metabolism-related pathways were split in two; the nutrient-uptake related pathways (glycolysis-gluconeogenesis and pentose phosphate pathways) and the oxidative phosphorylation (OXPHOS) related pathways (TCA cycle and electron transport chain).

Table 11. Affymetric gene expression means of selected pathways by cell line. Affymetric Human Genome U133 Plus 2.0 Array gene expression raw data were obtained from GEO, NCBI. Public accession numbers of the cell lines are listed in Experimental Methods 2.3, Table 2. Preprocessing was done using Robust Multi-array Average (RMA) normalization method. The normalized expression mean of the genes belonging to the glycolysis-gluconeogenesys plus pentose phosphate pathway (*left side* of the table) and TCA cycle plus electron transport chain (*right side* of the table) were used to determine three categories: low, medium, and high total gene expression per cell line.

Glycolysis-gluconeogenesis and pentose phosphate pathways			**TCA and electron transport chain**		
Cell line	**Mean expression value**	**Expression level**	**Cell line**	**Mean expression value**	**Expression level**
Panc1	6.83	low	1321N1	10.06	low
DU145	7.00	low	AsPC1	10.14	low
HaCaT	7.00	low	HaCaT	10.17	low
MDAMB435S	7.02	low	MDAMB231	10.24	low
1321N1	7.07	low	HepG2	10.26	low
AsPC1	7.14	medium	HeLa	10.26	medium
HCT116	7.16	medium	Panc1	10.29	medium
PC3	7.20	medium	PC3	10.31	medium
BxPC3	7.21	medium	MDAMB435S	10.32	medium
MiaPaCa2	7.22	medium	DU145	10.33	medium
MCF7	7.28	medium	BxPC3	10.38	medium
HeLa	7.29	high	HCT116	10.39	high
LNCaP	7.34	high	MiaPaCa2	10.42	high
HT29	7.41	high	HT29	10.51	high
MDAMB231	7.42	high	MCF7	10.55	high
HepG2	7.66	high	LNCaP	10.66	high

The mean value of the total gene expression of the nutrient-uptake related pathways per cell line was used to determine three expressions levels (Table 11): low, medium, and high total gene expression. The same was done for the OXPHOS pathways. The cell lines were sorted by their expression level in the nutrient-uptake related pathways and OXPHOS related pathways and plotted against their cellular properties (data not shown). For all the generated plots, the one with the glycolysis-gluconeogenesis and pentose phosphate pathways against the cell capacitance (Fig. 2) suggests a positive correlation between the two mentioned variables in the studied cell lines (Fig. 13A).

Cells with high expression levels appear to have high cell-capacitance values and cells with low expression values a low capacitance. However, Pearson product moment correlation gives a nonsignificant (p-values > 0.05) and low dependency (r = 0.3) (Fig. 13A). Additionally, the plot generated with the expression level in the TCA cycle together with the electron transport chain versus respiration activity (O_2 consumption) (Fig. 4) seems also to show a positive correlation between the cell lines (Fig. 13B). Cells with low expression level in the OXPHOS related pathways appear to have a low O_2 consumption and the ones that show high expression of the genes related to the OXPHOS pathways a high respiration activity. Nevertheless, Pearson correlation calculation shows a nonsignificant (p-values > 0.05) moderately positive dependency (r = 0.44) (Fig. 13B).

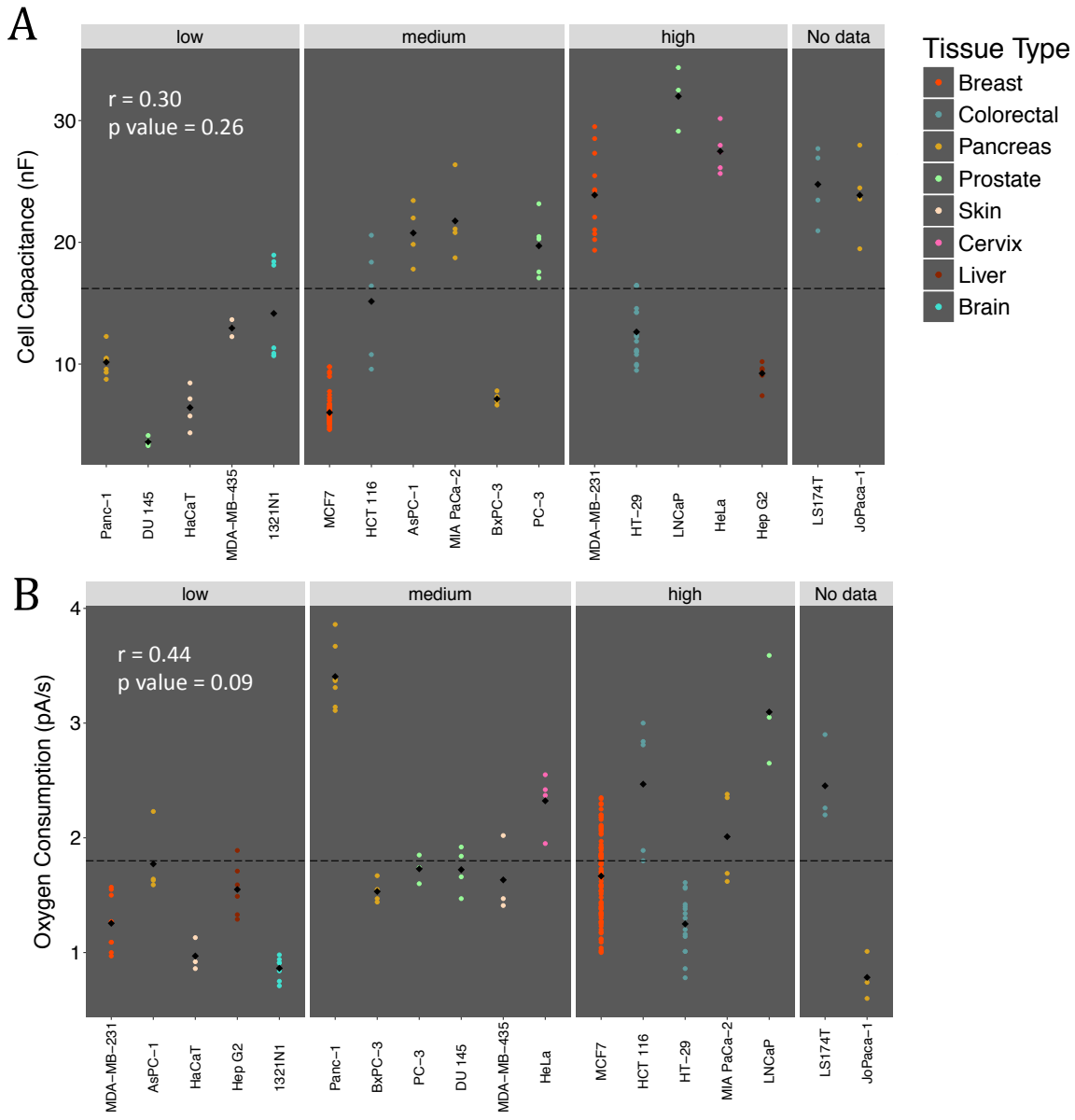
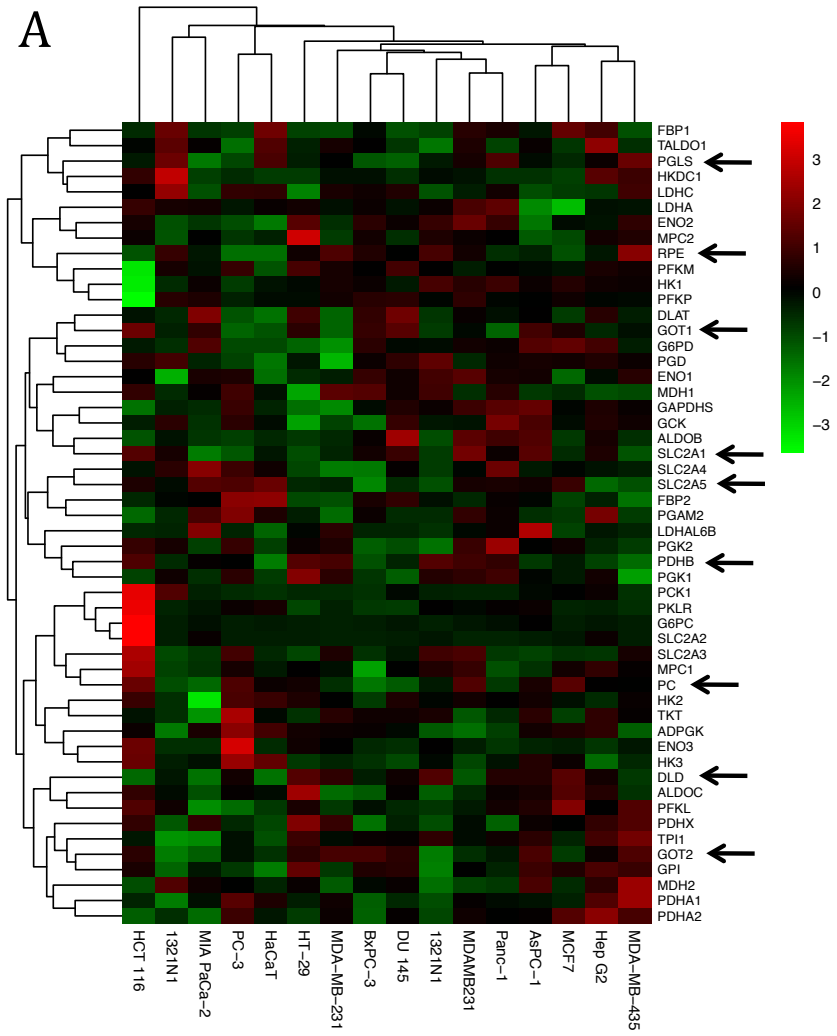


Fig. 13. Relation of cellular capacitance and respiration activity with Affymetrix gene expression level of selected metabolism pathways. The cell capacitance and the respiration activity were measured on the Bionas 2500 online monitoring system. Cells were seeded in chips containing IDES and Clark electrodes. Lower capacitance (nF values) indicates a stronger cell-matrix adhesion and cell-cell contact interactions. Higher oxygen consumption indicates a stronger respiration activity. Equal numbers of cells per cell line were seeded in the chips and at least three independent experiments were done per cell line, each one with six technical replicates. Affymetrix Human Genome U133 Plus 2.0 Array gene expression raw data were obtained from GEO, NCBI. Public accession numbers of the cell lines are listed in Experimental Methods, Table 2. Preprocessing was done using Robust Multi-array Average (RMA) normalization method. The normalized expression mean of genes belongs to the glycolysis-gluconeogenesis and pentose phosphate pathways together (**A**) and TCA cycle together with the electron transport chain (**B**) were used to determine three categories: low, medium, and high total gene expression per cell line. *Dashed line* indicates the mean between the analyzed cell lines. *Black dot* indicates the mean for the corresponding cell line. Linear correlation coefficients between the variables were obtained using Pearson algorithm. r coefficients and p-values for the dependencies are shown. Nonsignificant dependencies (p-value > 0.05) were found. No data shows the cell lines in which no Affymetrix gene expression is given. Linear correlation and RMA normalization were done using *R programming language* with the package *Hmisc* and rcorr function and Bioconductor package *affy.*

3.4.1 Selection of Genes Belonging to the Selected Metabolism Pathways

A selection of genes belonging to the glycolysis-gluconeogenesis pathway, pentose phosphate pathway, TCA cycle, and electron transport chain was obtained based on the expression values of the genes in the studied cell lines. For the selection, the expression of each gene was normalized across the cell lines. Scale and center was applied to normalize the gene expression relative to the set of cell lines used. A range to compare the expression values of each gene in the studied cell lines from −3 (lowly expressed genes) to 3 (highly expressed genes) was obtained (Figs. 14A and 15A) (Experimental Methods 2.3.1). Scaling and centering was calculated using *R programming language* package *softImpute* (Hastie et al. 2015) (Supplements 5.1.5). Normalized extreme expression values were the ones greater than 1 and lower than −1 (Supplements 5.1.5). A set of 17 genes out of the 155 remained for correlation analysis with the capacitance and metabolic assays (Figs. 14 and 15). In the metabolism-related glycolysis-gluconeogenesis and pentose phosphate pathways the selected genes were the ones that present, after scaling and being centered, the highest or the lowest expression levels in at least 7 of the 16 cell lines (Fig. 14B).

A



B

Cell Line/Gene	DLD	GOT1	GOT2	PC	PDHB	PGLS	RPE	SLC2A1	SLC2A5
MDA-MB-231	-1.15	-0.02	-0.58	1.28	1.04	0.44	0.38	1.78	0.40
MCF7	1.36	0.52	-0.75	1.37	-0.30	-0.46	-1.05	-0.54	0.93
HCT 116	0.33	1.38	0.70	-1.24	-0.42	-1.33	0.05	0.91	-0.53
HT-29	0.37	-0.46	0.25	0.11	-0.85	0.30	-0.21	0.52	-1.42
AsPC-1	-0.25	-0.40	-1.73	-1.00	-0.55	1.69	0.87	0.43	-0.07
MIA PaCa-2	-0.73	-0.15	1.24	0.09	-1.46	1.61	2.06	-1.10	-1.08
Panc-1	-1.58	0.80	-1.28	-1.40	0.19	-1.69	-0.25	-1.72	1.31
BxPC-3	0.61	1.07	1.19	0.45	-0.73	-0.06	-0.38	1.35	0.36
LNCaP	1.31	0.72	0.73	0.37	1.33	-0.33	0.35	-1.01	-0.49
PC-3	0.64	-1.39	-0.24	-0.70	0.85	1.22	-0.58	0.28	0.46
DU 145	-0.31	0.86	1.12	-1.61	-1.09	-1.17	0.59	0.38	-1.87
HaCaT	-1.59	-1.10	-0.65	0.28	-1.71	1.19	-1.53	-0.23	1.62
MDA-MB-435	0.76	-1.37	1.12	-0.58	1.11	0.06	1.22	-0.45	-0.35
HeLa	0.37	-1.39	-0.13	1.22	0.09	-0.94	-1.53	-1.16	1.24
Hep G2	-1.42	1.69	0.70	1.61	1.20	-0.28	-1.11	1.34	0.54
1321N1	1.28	-0.76	-1.71	-0.26	1.30	-0.26	1.13	-0.78	-1.06

Fig. 14. Selection of genes related to the glycolysis-gluconeogenesis and pentose phosphate pathways. Affymetrix Human Genome U133 Plus 2.0 Array gene expression raw data (.CEL files) were obtained from GEO, NCBI. Public accession numbers of the cell lines are listed in Experimental Methods 2.3, Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. The normalized expression was scaled and centered to compare the expression levels of each gene in the studied cell lines. A

range from –3 (lowly expressed genes in *green*) to 3 (highly expressed genes in *red*) was calculated. Genes with extreme expression values in at least 7 of the 16 cell lines were selected. Extreme expression values were the ones greater than 1 and lower than –1. **A.** Heatmap of the expression levels of genes belong to glycolysis-gluconeogenesis and pentose phosphate pathways. *Arrows* show the genes that present extreme expression after scaling and centering. **B**. Scale and center values of genes that present extreme expression in the analyzed cell lines. Scale-center and RMA normalizations were done using *R programming language* with the package *softImpute* function biscale and Bioconductor package *affy.*

From the 51 genes corresponding to the mentioned pathways, 9 were selected (Fig. 14). The proteins corresponding to the selected genes are:

1. DLD (dihydrolipoamide dehydrogenase) oxidizes dihydrolipoamide to lipoamide.
2. GOT1 (glutamic-oxaloacetic transaminase 1) located in the cytosol, catalyzes the reversible transamination of aspartate to oxaloacetate.
3. GOT2 (glutamic-oxaloacetic transaminase 2), located in mitochondria, catalyzes the reversible transamination of aspartate to oxaloacetate.
4. PC (pyruvate carboxylase), catalyzes the carboxylation of pyruvate to oxaloacetate.
5. PDHB (pyruvate dehydrogenase (lipoamide) beta) is one of the enzymes of the pyruvate dehydrogenase complex that catalyzes the conversion of pyruvate to acetyl-CoA.
6. PGLS (6-phosphogluconolactonase) of pentose phosphate pathway catalyzes the hydrolysis of 6-phosphogluconolactone to 6-phosphogluconate.
7. RPE (ribulose-5-phosphate-3-epimerase) of the pentose phosphate pathway catalyzes the reversible epimerization of D-ribulose 5-phosphate to D-xylulose 5-phosphate.
8. SLC2A1 (glucose transporter 1, solute carrier family 2 member 1) transports glucose across the plasmatic membrane.
9. SLC2A5 (solute carrier family 2 member 5) transports fructose into the intestinal lumen.

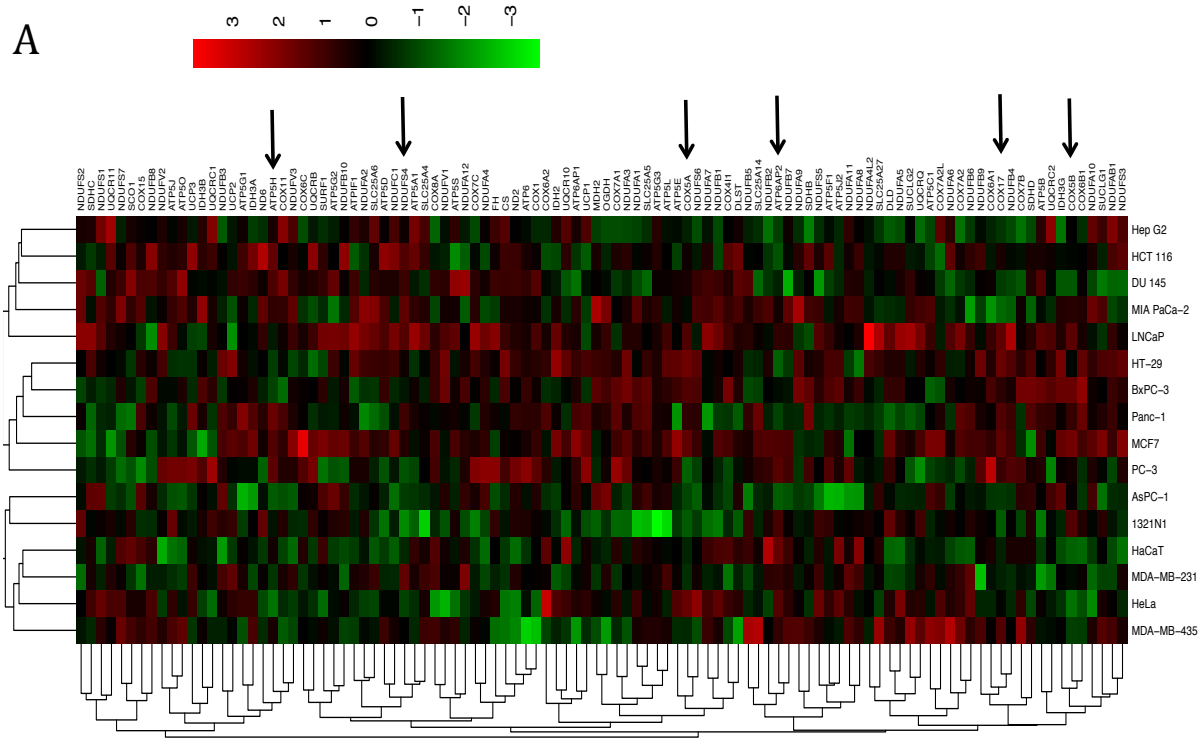
In the respiration-related pathways, the selected genes were the ones with higher and lower expression levels in at least 8 of the 16 cell lines (Fig. 15B). From the 104 genes corresponding to the TCA cycle and electron transport chain together, six were selected (Fig. 15). The proteins corresponding to the selected genes are:

10. ATP5H (mitochondrial Fo complex subunit D), one of the subunits of ATP synthase that catalyzes ATP synthesis.
11. ATP6AP2 (ATPase H^+ transporting accessory protein 2) is a proton-translocating ATPase.
12. COX17 (cytochrome c oxidase copper chaperone). Cytochrome c oxidase catalyzes the electron transfer from cytochrome c to oxygen.
13. COX5A (cytochrome c oxidase subunit 5a), a subunit of the cytochrome c oxidase complex.
14. COX5B (cytochrome c oxidase subunit 5b), a subunit of the cytochrome c oxidase complex.
15. NDUFS4 (ubiquinone oxidoreductase subunit S4), a subunit of complex I that removes electrons from NADH to forward them to the electron acceptor ubiquinone.

3.4.1.1 Dependencies between Cellular Properties of the Studied Cell Lines and Affymetrix Gene Expression of Pathways of Selected Genes

With the aim of finding undiscovered dependencies among genes and metabolism or cell capacitance, Pearson product moment correlation was used to search for novel relationships within the expression values of the previous selected genes and the cellular properties data. Linear correlations were calculated between the data of the studied cell lines for cellular capacitance, respiratory activity (O_2 consumption), glycolytic activity (acidification), energy metabolism (ATP level), intracellular ROS accumulation, mitochondrial mass, and proliferation rates (MTT and SRB assays) (Figs. 2 to 10) and the 15 selected metabolism-related genes (DLD, GOT1, GOT2, PC, PDHB, PGLS, RPE, SLC2A1, SLC2A5, ATP5H, ATP6AP2, COX17, COX5A, COX5B, and NDUFS4) (Figs. 14 and 15). The gene expression data of the 15 genes and their total probe sets in the Affymetrix Human Genome U133 Plus 2.0 Array were used for searching dependencies with the cellular features. Only those genes with consistent information across their probe sets were chosen as candidate genes.

A



B

Cell Line/ Gene symbol	ATP5H	ATP6AP2	COX17	COX5A	COX5B	NDUFS4
MDA-MB-231	-0.87	0.61	-0.62	-0.81	-0.29	1.28
MCF7	1.03	0.99	1.19	1.24	1.54	0.38
HCT 116	1.28	-1.33	0.26	0.12	-0.21	1.06
HT-29	-0.99	-1.51	1.4	1.51	-0.2	0.09
AsPC-1	-1.42	-1.47	-0.95	-1.16	-0.01	-1.00
MIA PaCa-2	0.53	0.14	-1.61	-0.31	0.47	1.23
Panc-1	1.38	-1.1	1.03	0.55	0.63	0.8
BxPC-3	-1.09	-0.3	0.49	1.07	1.41	-0.43
LNCaP	-0.51	0.52	1.24	1.00	1.21	1.11
PC-3	1.21	1.14	0.68	-1.53	0.16	-1.34
DU 145	1.02	-0.58	-1.00	0.25	-1.36	0.74
HaCaT	0.13	1.36	-0.61	-0.62	-1.52	-0.69
MDA-MB-435	0.23	0.59	1.03	-0.68	-1.02	-0.11
HeLa	-1.76	1.09	-0.19	1.41	-1.76	-1.11
Hep G2	0.25	-1.03	-0.88	-0.61	-0.08	0.11
1321N1	-0.41	0.88	-1.45	-1.43	1.04	-2.11

Fig. 15. Selection of genes related to TCA cycle and electron transport chain. Affymetrix Human Genome U133 Plus 2.0 Array gene expression raw data (.CEL files) were obtained from GEO, NCBI. Public accession numbers of the cell lines are listed in Experimental Methods 2.3, Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. The normalized expression was scaled and centered to compare the expression levels of each gene in the studied cell lines. A range from –3 (lowly expressed genes in *green*) to 3 (highly expressed genes in *red*) was calculated. Genes with extreme expression values in at least 8 of the 16 cell lines were selected. Extreme expression values were the ones greater than 1 and lower than –1. **A.** Heatmap of the expression levels of genes belonging to TCA cycle and electron transport chain. *Arrows* indicate the genes that present extreme expression after scaling and centering. **B**. Scale and center values of genes that present extreme expression in the analyzed cell lines. Scale-center and RMA normalizations were done using *R programming language* with the package *softImpute* function biscale and Bioconductor package *affy.*

PGLS and COX5B are the genes that shown significant correlation and consistent information across their probe sets with some of the metabolism data. The cellular capacitance did not show any relation with any of the selected metabolism genes (Table 12). PGLS (6-phosphogluconolactonase) is an enzyme that catalyzes the hydrolysis of 6-phosphogluconolactone to 6-phosphogluconate in the second reaction of the pentose phosphate pathway oxidative phase.

The PGLS gene presents four probe sets in the Affymetrix platform used, all of which show consistent information about the dependencies between the expression level and mitochondrial mass content, glycolytic activity, and proliferation rates measured as metabolic activity (MTT assay) and total protein content (SRB assay). The negative correlation observed with the mitochondrial mass is significant (p-values < 0.05) in three out of the four probe sets. The fourth probe set gave significance (Table 12). Additionally, the same three probe sets show also a negative correlation with glycolysis activity but just one of the probe set reached significance (p-value < 0.05) (Table 12). The significant dependency of PGLS with SRB assay is corroborated just for one out of the four probe sets, but all of them having a moderate negative correlation. Moreover, this is corroborated by according dependencies with the proliferation assay MTT, which presents a moderately negative correlation with the entire probe set of PGLS (Table 12). The negative dependencies found indicate that cell lines with high gene expression of PGLS present low mitochondrial mass content, low glycolytic activity, and low proliferation rates.

COX5B is a nuclear-coded subunit of the cytochrome c oxidase complex in the electron transport chain. COX5B contains two probe sets in the Affymetrix platform used for this study. One of the two probe sets shows a significant (p-value < 0.05) negative correlation with the proliferation assay MTT, which is in agreement with the moderate r coefficient of the second probe set. The calculated dependencies between the two probe sets and the SRB proliferation assay validated these relations (Table 12). The obtained correlations indicate that cell lines with high gene expression of COX5B present low proliferation rates.

Table 12. r coefficients and p-values of the significant dependencies between metabolism features and Affymetrix gene expression of candidate genes of the selected pathway. The mitochondrial mass content of the cells was determined using MitoTracker Green (Thermo Fisher). Cell lines with higher fluorescence intensity contain higher mitochondrial mass. The proliferation rate measured as metabolic activity (MTT assay) estimates the reduction of MTT

reagent by active mitochondria and can be measured by absorbance at 595 nm. Energy metabolism (ATP level) was estimated with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). Luminescence is proportional to ATP concentrations in the cells.
A set of genes were selected based on the expression values of genes belonging to glycolysis-gluconeogenesis, pentose phosphate, TCA cycle, and electron transport chain metabolic pathways. The gene expression raw data (.CEL files) were obtained from GEO, NCBI and correspond to Affymetrix Human Genome U133 Plus 2.0 Array. Public accession numbers of the cell lines are listed in Experimental Methods Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. The selection process included scale and center normalization of the gene expression in the studied cell lines. The genes with extreme expression values in at least 7 of the 16 cell lines were selected for Pearson comparisons with the metabolism assay. All the probe sets of the selected genes in the Affymetrix array are shown. The Pearson coefficients were calculated using the mean of at least three independent experiments per cell line and per metabolic assay. Pearson comparisons, scale-center, and RMA normalizations were obtained using *R programming language* with the packages *Hmisc* function rcorr, *softImpute* function biscale and Bioconductor package *affy.* In *yellow* the probe sets that present a significant dependence with the metabolic assays (p-value < 0.05).

r coefficients	Glycolysis	ATP	Mitomass	ROS	SRB	MTT
PGLS.1554316_at	0.00	0.34	-0.39	0.21	-0.17	-0.43
PGLS.218387_s_at	-0.51	0.17	-0.58	0.06	-0.41	-0.53
PGLS.218388_at	-0.47	0.08	-0.53	-0.03	-0.32	-0.42
PGLS.230699_at	-0.41	0.66	-0.68	0.55	-0.72	-0.51
COX5B.213735_s_at	-0.15	-0.28	-0.13	-0.29	-0.10	-0.27
COX5B.213736_at	-0.18	0.68	-0.13	0.59	-0.36	-0.60

p-values	Glycolysis	ATP	Mitomass	ROS	SRB	MTT
PGLS.1554316_at	0.9977	0.2142	0.1385	0.4488	0.6135	0.1850
PGLS.218387_s_at	0.0427	0.5457	0.0183	0.8218	0.2114	0.0923
PGLS.218388_at	0.0643	0.7776	0.0354	0.9149	0.3437	0.1995
PGLS.230699_at	0.1183	0.0072	0.0037	0.0330	0.0120	0.1061
COX5B.213735_s_at	0.5912	0.3119	0.6363	0.2893	0.7676	0.4292
COX5B.213736_at	0.5133	0.0052	0.6202	0.0213	0.2819	0.0494

Figure 16 shows the significant dependencies (p-values < 0.05) among three out of four probe sets of PGLS gene and the mitochondrial mass content. The entire set of probe sets belonging to the PGLS gene display a negative relationship between the expression of the gene and the mitochondrial mass content in the cell lines. The probe set 230699_at shows a strong correlation representing the highest r coefficient with mitochondrial mass. Although the probes in the probe-sets 230699_at and 218388_at interrogate just one transcript (_at suffix), the information differs for some cell lines from the others. Colorectal cells PC-3 and DU-145 show more similar expression values for the *PGLS* gene with the data of the 230699_at probe set than for the 218387_at or

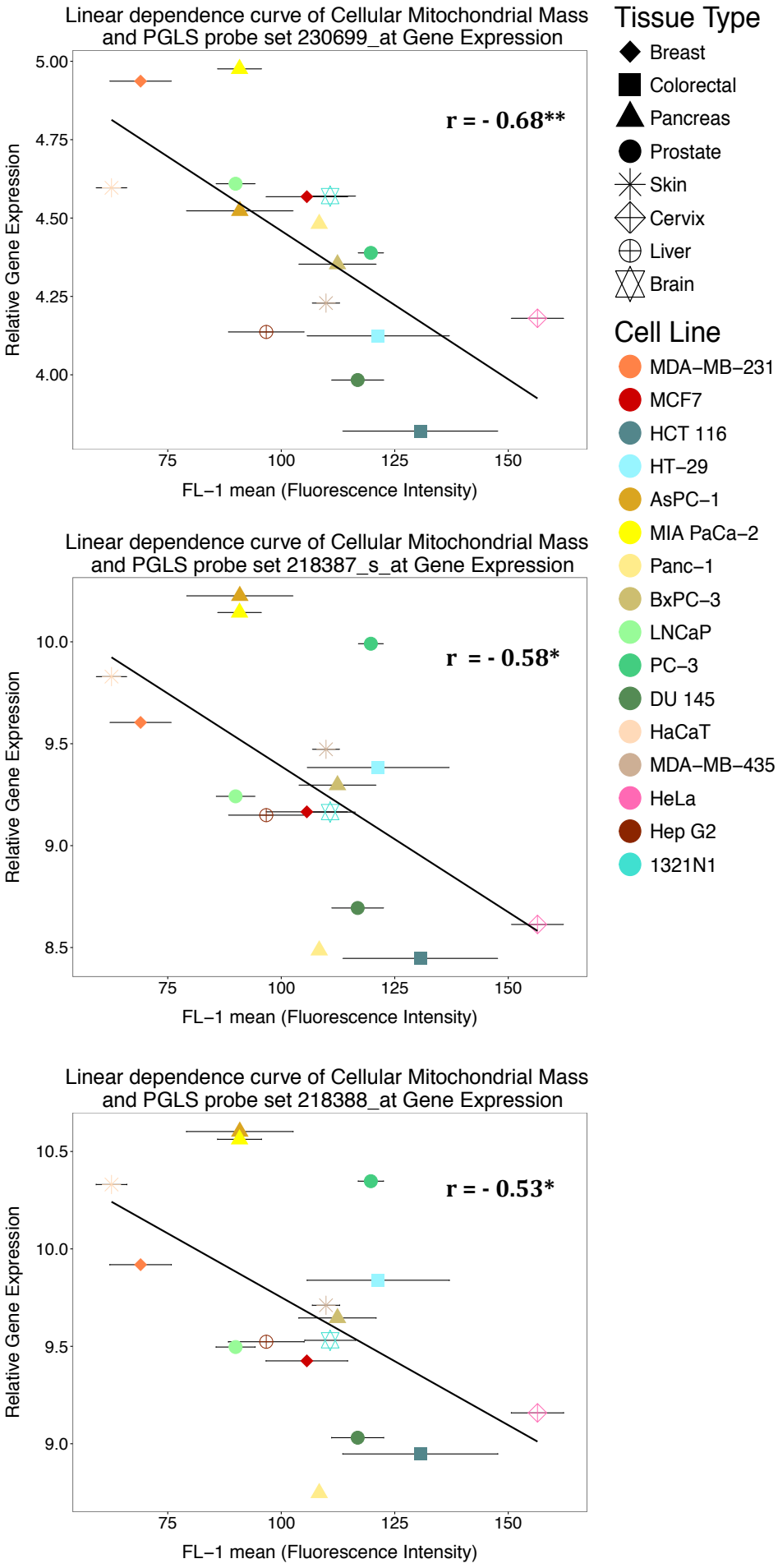

Linear dependence curve of Cellular Mitochondrial Mass and PGLS probe set 230699_at Gene Expression
r = - 0.68**
Relative Gene Expression
FL−1 mean (Fluorescence Intensity)
Linear dependence curve of Cellular Mitochondrial Mass and PGLS probe set 218387_s_at Gene Expression
r = - 0.58*
Relative Gene Expression
FL−1 mean (Fluorescence Intensity)
Linear dependence curve of Cellular Mitochondrial Mass and PGLS probe set 218388_at Gene Expression
r = - 0.53*
Relative Gene Expression
FL−1 mean (Fluorescence Intensity)
Tissue Type
Breast
Colorectal
Pancreas
Prostate
Skin
Cervix
Liver
Brain
Cell Line
MDA−MB−231
MCF7
HCT 116
HT−29
AsPC−1
MIA PaCa−2
Panc−1
BxPC−3
LNCaP
PC−3
DU 145
HaCaT
MDA−MB−435
HeLa
Hep G2
1321N1

Fig. 16. Significant dependencies between mitochondrial mass content and gene expression of candidate gene PGLS. Linear correlations of the gene expression values from probe sets belonging to the PGLS gene that presents significant dependencies (p-values < 0.05) with mitochondrial mass in the studied cell lines. The correlations were calculated using Pearson algorithm. Mitochondrial mass content of the cells was determined using MitoTracker Green (Thermo Fisher). Cell lines with higher fluorescence intensity contain higher mitochondrial mass. The Pearson coefficients were calculated using the mean of at least three independent experiments per cell line. The expression values were obtained from GEO, NCBI and correspond to Affymetrix Human Genome U133 Plus 2.0 Array. Public accession numbers of the cell lines are listed in Experimental Methods, Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. Pearson comparisons and RMA normalizations were obtained using *R programming language* with the packages *Hmisc* function rcorr and Bioconductor package *affy.* The shape of the *dots* gives the tissue types and the *color* the cell lines. The *blue lines* represent linear regressions. For each dependency the r correlation coefficient is shown; *asterisks* represent the level of significance (p-values < 0.05* and < 0.01**). *Error bars* = standard error (se).

218388_s_at. In these last two probe sets, Panc-1 shows a similar expression level to HCT 116, but a much higher one in the 230699_at probe set. Indeed, the two "_at" probe sets correlate with each other with an r coefficient of just 0.55. On the other hand, the 218388_at and 218387_s_at probe sets are highly correlated by an r coefficient of 0.99 (data not shown).

3.4.2 Relations between Cell Properties of the Studied Cell Lines and the Extremely Expressed Probe Sets in Affymetrix Arrays

The preprocessing (RMA normalized data, Experimental Methods 2.3) gene expression data of the total probe sets in the Affymetrix Human Genome U133 Plus 2.0 Array for the studied cell lines were normalized using scale and center standardization. The expression of each probe set was standardized across the cell lines. Scale and center was applied to normalize the probe set expression relative to the set of cell lines used. From the complete Affymetrix platform with more than 50,000 entries we selected the extreme expressed probe sets per cell line (Supplements 5.1.5). The choice was made for probe sets that were either highly or lowly expressed in at least 9 of the 16 cell lines. Using the gene symbol it was possible to collect the entire set of probe sets that join the

extreme expressed probe sets chosen. A final selection of genes with consistent information through their probe sets was done. That final selection includes the condition that no pair of the chosen probe sets could show an extreme expression in less than seven cell lines. Genes with just one probe set were discarded (Experimental Methods 2.3.1). With the remaining 36 probe sets, corresponding to 17 genes, linear relationships were searched with the metabolic properties of the studied cell lines and the probe set gene expression (Table 13). The linear dependencies were calculated using Pearson product moment correlation using *R programming language* package *Hmisc* and rcorr function.

Three probe sets showed a linear dependency with some of the metabolism assays. No probe sets showed a significant dependency with cellular capacitance (Table 13). The first probe set with a correlation to metabolism is the 1552502_s_at belonging to the RHBDL*2* (rhomboid-related protein 2) gene. RHBDL2 contains three probe sets in the Affymetrix array used, and the other two probe sets are consistent with 1552502_s_at (Fig. 17). It is remarkable that the only two probe sets of the IFI16 (gamma-interferon-inducible protein 16) gene (206332_s_at and 208965_s_at) are the second and third probe sets that present a correlation with metabolism (Fig. 18).

The cell lines with higher expression levels of RHBDL2 are the ones which have less glycolytic activity and mitochondrial mass (Fig. 17A). An r coefficient of –0.53 and –0.52 was calculated for the dependencies between 1552502_s_at probe set with glycolysis activity and mitochondrial mass content, respectively. The other two probe sets that interrogate the RHBDL2 gene, the 1554895_a_at and 1554897_s_at, corroborate these findings and present a highly significant positive relation (p-value < 0.0001) with 1552502_s_at (Fig. 17B). A much higher expression of the RHBDL2 gene is observable in the noncancerous skin cell line HaCaT, with the skin cell line pair MDA-MB-435 completely in the bottom of the plot with one of the lowest expression values. Moreover, is it not possible to observe a correlation between the tissue type of the cell lines and the gene expression level of RHBDL2. Glycolysis activity and mitochondrial mass content both present a nonsignificant (p-value > 0.05) moderately positive relation (Table 10).

Table 13. r coefficients and p values of the extreme expressed probe sets in Affymetrix array and the cell properties of the studied cell lines. In the total probe sets data from Affymetric Human Genome U133 Plus 2.0 Array the extreme expressed probe sets were selected for Pearson comparison with the electrical and metabolic features of the studies cell lines. The gene expression values were obtained from GEO, NCBI, and public accession numbers of the cell lines are listed in Experimental Methods, Table 2. Preprocessing of the .CEL files was done using

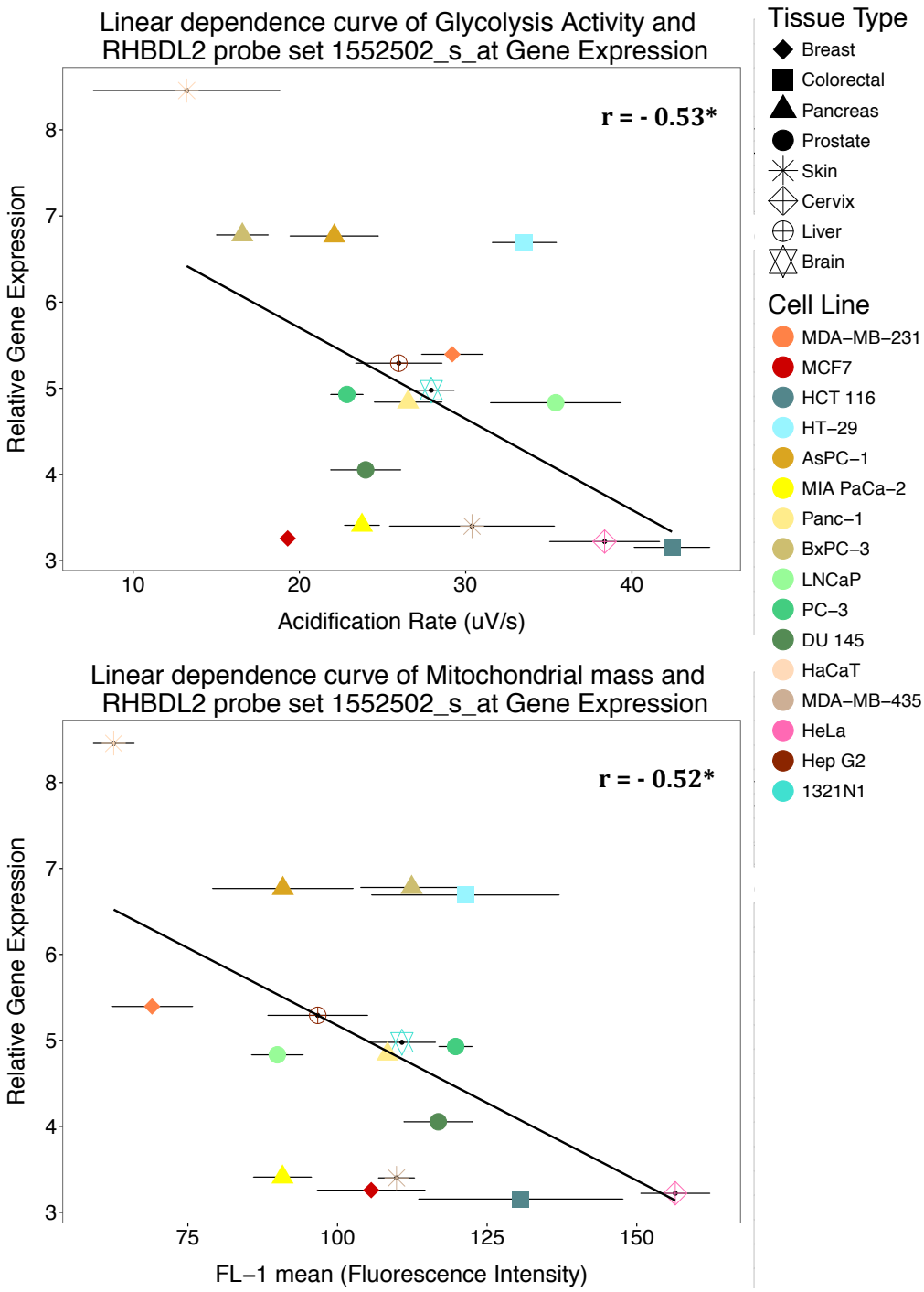
Robust Multi-array Average (RMA) normalization method. The normalized expressions per gene were scaled and centered to normalize them across the studied cell lines. A range to select the probe sets that were either highly or lowly expressed in at least 9 of the 16 cell lines was done. Cellular capacitance, glycolytic and respiratory activities were measure on the Bionas 2500 online monitoring system. The ATP level was measured with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). The mitochondrial mass (mitomass) content was determined using MitoTracker Green (Thermo Fisher). ROS accumulation in cells was measured using DHE. The proliferation rate was estimated using MTT and SRB assays. At least three independent experiments per assay and cell line were used for the calculations. Pearson comparisons, scale-center and RMA normalizations were obtained using *R programming language* with the packages *Hmisc* function rcorr, *softImpute* function biscale and Bioconductor package *affy*. In *yellow* the two probe sets of *IFI16* gene (206332_s_at and 208965_s_at) that show a significant linear dependency with respiration activity and one of the three probe sets of the *RHBDL2* gene that show a significant dependency with glycolysis and mitochondrial mass.

r coefficients	Capacitance	Respiration	Glycolysis	ATP	Mitomass	ROS	SRB	MTT
1555735_a_at	0.17	-0.09	0.03	0.04	0.00	0.32	-0.14	-0.47
201419_at	0.37	-0.16	0.07	-0.21	0.07	0.09	-0.17	0.14
201130_s_at	-0.06	0.09	-0.14	0.24	0.04	-0.09	0.08	-0.14
201131_s_at	-0.33	0.11	-0.35	0.27	-0.16	0.10	0.08	-0.30
201116_s_at	0.16	0.13	0.21	0.09	0.26	0.17	-0.12	-0.08
201117_s_at	0.22	0.19	0.23	0.13	0.28	0.16	-0.12	-0.01
203501_at	0.03	0.22	0.09	0.08	0.06	0.23	-0.23	-0.47
208454_s_at	-0.11	0.21	0.09	-0.06	0.22	0.20	-0.06	-0.31
217889_s_at	0.18	-0.40	0.00	-0.13	-0.11	-0.21	-0.22	0.14
222453_at	0.23	-0.27	-0.03	-0.09	-0.16	-0.24	-0.23	0.19
220318_at	0.00	-0.12	-0.27	0.23	-0.23	0.04	0.05	-0.25
223895_s_at	-0.20	0.03	-0.16	0.11	-0.02	0.04	0.26	-0.10
219121_s_at	-0.19	-0.01	-0.21	0.20	0.00	-0.08	0.24	-0.01
225846_at	-0.19	-0.07	-0.23	0.15	-0.01	-0.16	0.23	-0.01
225602_at	0.27	0.19	0.43	0.25	0.27	0.02	0.26	0.45
225604_s_at	0.18	0.04	0.34	0.16	0.33	-0.03	0.26	0.36
219976_at	-0.19	0.08	0.10	-0.10	0.17	-0.28	0.49	0.16
225792_at	-0.17	0.04	0.05	-0.16	0.20	-0.37	0.48	0.25
206332_s_at	-0.07	-0.59	-0.42	-0.12	-0.28	-0.10	-0.58	-0.23
208965_s_at	-0.08	-0.57	-0.42	-0.06	-0.30	-0.06	-0.58	-0.27
235301_at	0.16	-0.01	0.02	0.26	0.22	-0.09	-0.10	-0.05
244317_at	0.13	0.03	0.05	0.32	0.11	-0.07	-0.07	-0.14
218656_s_at	-0.40	-0.23	-0.40	0.05	-0.26	0.30	-0.42	-0.42
231411_at	-0.35	-0.09	-0.26	0.02	-0.24	0.39	-0.30	-0.28
233634_at	-0.21	-0.07	-0.13	0.02	-0.21	-0.25	0.42	0.04
239148_at	-0.33	-0.14	-0.26	-0.01	-0.10	-0.21	0.34	-0.03
239350_at	-0.27	-0.17	-0.19	-0.01	-0.16	-0.29	0.44	0.04
1555397_at	-0.37	-0.13	-0.33	0.10	-0.10	-0.26	0.06	-0.14
212338_at	-0.36	-0.22	-0.42	0.01	-0.17	-0.20	0.08	-0.20
205479_s_at	-0.16	0.02	-0.45	0.22	-0.29	0.47	-0.33	-0.29
211668_s_at	-0.19	0.03	-0.42	0.16	-0.23	0.49	-0.30	-0.39
1555579_s_at	0.23	-0.15	0.25	-0.15	0.02	-0.11	-0.20	-0.04
203329_at	0.33	-0.12	0.21	-0.02	0.04	-0.04	-0.22	0.11
1552502_s_at	-0.21	-0.42	-0.53	0.00	-0.52	-0.04	-0.19	-0.43
1554895_a_at	-0.22	-0.44	-0.42	-0.23	-0.39	-0.05	-0.15	-0.47
1554897_s_at	-0.03	-0.36	-0.41	0.14	-0.38	-0.01	-0.20	-0.25

205768_s_at	-0.26	0.03	0.10	-0.18	0.11	-0.30	0.53	0.07
205769_at	-0.20	0.01	0.12	-0.07	0.16	-0.36	0.60	0.28

p values	Capacitance	Respiration	Glycolysis	ATP	Mitomass	ROS	SRB	MTT
1555735_a_at	0.54	0.75	0.92	0.90	0.99	0.24	0.69	0.14
201419_at	0.16	0.55	0.79	0.46	0.79	0.75	0.62	0.68
201130_s_at	0.82	0.74	0.60	0.39	0.89	0.76	0.81	0.69
201131_s_at	0.21	0.69	0.18	0.32	0.55	0.72	0.81	0.36
201116_s_at	0.56	0.62	0.45	0.74	0.33	0.54	0.73	0.82
201117_s_at	0.42	0.47	0.39	0.66	0.30	0.56	0.72	0.98
203501_at	0.91	0.41	0.74	0.77	0.82	0.41	0.50	0.14
208454_s_at	0.67	0.43	0.75	0.83	0.40	0.49	0.86	0.35
217889_s_at	0.50	0.13	0.99	0.65	0.67	0.45	0.52	0.68
222453_at	0.39	0.31	0.92	0.76	0.54	0.40	0.49	0.57
220318_at	0.99	0.65	0.32	0.40	0.40	0.90	0.89	0.47
223895_s_at	0.46	0.93	0.54	0.69	0.94	0.89	0.44	0.78
219121_s_at	0.47	0.98	0.44	0.49	1.00	0.78	0.48	0.99
225846_at	0.49	0.81	0.39	0.59	0.96	0.58	0.49	0.98
225602_at	0.32	0.48	0.10	0.37	0.31	0.94	0.43	0.17
225604_s_at	0.51	0.87	0.19	0.56	0.21	0.92	0.44	0.28
219976_at	0.48	0.78	0.72	0.71	0.53	0.32	0.12	0.63
225792_at	0.53	0.87	0.85	0.58	0.45	0.17	0.14	0.46
206332_s_at	0.79	0.02	0.10	0.66	0.30	0.73	0.06	0.49
208965_s_at	0.77	0.02	0.10	0.84	0.27	0.84	0.06	0.43
235301_at	0.55	0.97	0.93	0.35	0.40	0.76	0.77	0.89
244317_at	0.62	0.90	0.86	0.25	0.68	0.81	0.84	0.68
218656_s_at	0.13	0.39	0.13	0.85	0.34	0.27	0.20	0.20
231411_at	0.18	0.75	0.33	0.95	0.37	0.15	0.37	0.41
233634_at	0.43	0.79	0.64	0.94	0.43	0.38	0.19	0.91
239148_at	0.21	0.61	0.34	0.98	0.71	0.45	0.31	0.94
239350_at	0.31	0.52	0.47	0.98	0.54	0.29	0.17	0.90
1555397_at	0.16	0.62	0.21	0.72	0.73	0.35	0.86	0.67
212338_at	0.17	0.42	0.10	0.98	0.52	0.46	0.82	0.55
205479_s_at	0.55	0.93	0.08	0.44	0.27	0.08	0.33	0.39
211668_s_at	0.48	0.90	0.11	0.58	0.39	0.06	0.37	0.24
1555579_s_at	0.40	0.57	0.36	0.60	0.94	0.69	0.55	0.90
203329_at	0.22	0.65	0.44	0.93	0.88	0.90	0.52	0.76
1552502_s_at	0.43	0.10	0.04	0.99	0.04	0.88	0.58	0.19
1554895_a_at	0.42	0.09	0.11	0.42	0.14	0.87	0.65	0.15
1554897_s_at	0.91	0.17	0.12	0.61	0.14	0.97	0.55	0.46
205768_s_at	0.33	0.90	0.71	0.53	0.70	0.27	0.09	0.84
205769_at	0.46	0.97	0.66	0.81	0.54	0.19	0.05	0.41

A



B

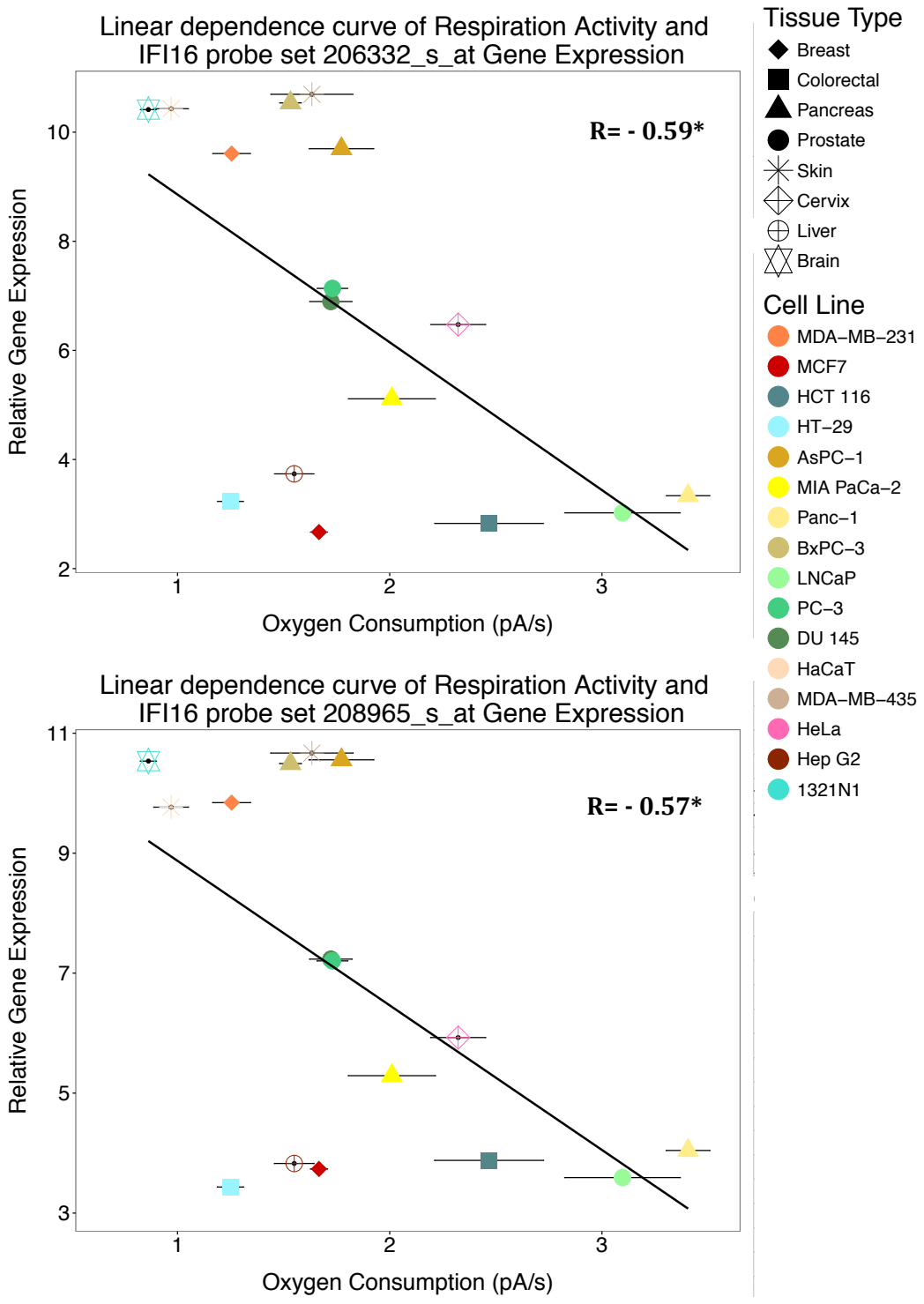
			r coefficients			p values		
nº of cell lines with extreme expression	Gen Symbol	Probe set ID	Glycolysis	Mitomass	1552502_s_at	Glycolysis	Mitomass	1552502_s_at
9	RHBDL2	1552502_s_at	-0.53	-0.52	1.00	0.04	0.04	
9	RHBDL2	1554897_s_at	-0.41	-0.38	0.88	0.12	0.14	8.00E-06
8	RHBDL2	1554895_a_at	-0.42	-0.39	0.85	0.11	0.14	6.30E-05

Fig. 17. Significant dependencies between glycolytic activity and mitochondrial mass with one of the probe sets of the RHBDL2 gene. From the total probe set data from Affymetrix Human Genome U133 Plus 2.0 Array the extreme expressed probe sets were selected for Pearson comparison with the cellular features of the studied cell lines. The glycolytic activity was measured on the Bionas 2500 containing chips with ISFET sensors. A stronger acidification rate indicates a higher glycolysis level. Equal numbers of cells per cell line were seeded in sensor chips. The mitochondrial mass content of the cells was determined using MitoTracker Green (Thermo Fisher). Cell lines with higher fluorescence intensity contain higher mitochondrial mass. The relationships were calculated using the mean of at least three independent experiments per cell line and per metabolic assay. The gene expression values were obtained from GEO, NCBI, and the public accession numbers of the cell lines are listed in Experimental Methods, Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. The normalized expressions per gene were scaled and centered to normalize them across the studied cell lines. A range to select the probe sets that were either highly or lowly expressed in at least 9 of the 16 cell lines was done. Pearson comparisons, scale-center and RMA normalizations were obtained using *R programming language* with the packages *Hmisc* function rcorr, *softImpute* function biscale and Bioconductor package *affy*. **A.** Linear dependency curves. The shape of the *dots* gives the tissue types and the *colors* the cell line. The *blue lines* represent the linear regressions. For each dependency, the r correlation coefficient is shown; asterisks represent the level of significance (p-values < 0.05*). *Error bars* = standard error (se). **B.** Number of cell lines with extreme expression of the probe sets, r coefficients, and p-values. The RHBDL2 gene has consistent information in the r coefficients in the three probe sets, but just one of them has a significant p score.

The description of IFI16 is "*encodes a member of the HIN-200 (hematopoietic interferon-inducible nuclear antigens with 200 amino acid repeats) family of cytokines. The encoded protein contains domains involved in DNA binding, transcriptional regulation, and protein-protein interactions. The protein localizes to the nucleoplasm and nucleoli, and interacts with p53 and retinoblastoma-1. It modulates p53 function, and inhibits cell growth in the Ras/Raf signaling pathway. Alternatively spliced transcript variants encoding different isoforms that have been found for this gene*" (NCBI/Gene summary, Gene ID: 3428). Our results shown that the two probe sets corresponding to IFI16 have a significant negative correlation (p-value < 0.02) with the respiration activity (Fig. 18), meaning that the cell lines that present a high expression of this gene respire less (less O_2 consumption). It is obvious that there is a group of cell lines that present much higher expression levels of IFI16, in contrast to a group with much lower expression of that gene (Fig. 18A). Among the highest expressing cells are the brain 1321N1, the two skin cell lines HaCaT and MDA-MB-435, the two adenocarcinoma pancreatic cell lines AsPC-1 and BxPC-3 and one of the two breast cancer cell lines the adenocarcinoma MDA-MB-231 (Fig. 18A). The carcinoma cell lines Hep G2 (liver) and HT-29, as well as HCT 116

(colon), MIA PaCa-2, and Panc-1 (pancreas) showed low expression levels of the IFI16 gene.

A



B

nº of cell lines with extreme expression	Gen Symbol	Probe set ID	R coefficient respiration	p value respiration	R coefficient 208965_s_at	p value 208965_s_at
10	IFI16	208965_s_at	-0.57	0.02	1.00	
9	IFI16	206332_s_at	-0.59	0.02	0.99	2.89E-13

Fig. 18. Significant dependencies between respiratory activity and the two probe sets of IFI16. In the total probe sets data from Affymetrix Human Genome U133 Plus 2.0 Array the extreme expressed probe sets were selected for Pearson comparison with the metabolic features of the studies cell lines. The respiratory activity of the cells was measured on the Bionas 2500 containing chips with Clark electrodes. Higher oxygen consumption indicates higher respiration rate. Equal numbers of cells per cell line were seeded in sensor chips. For the calculation the mean of at least three independent experiments per cell line with six technical replicates each one was used. The gene expression values were obtained from GEO, NCBI, and public accession numbers of the cell lines are listed in Experimental Methods, Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. The normalized expressions per gene were scaled and centered to normalize them across the studied cell lines. A range to select the probe sets that were either highly or lowly expressed in at least 9 of the 16 cell lines was done. The IFI16 gene has consistent information in the r coefficients and p-values in its two probe sets. Pearson comparisons, scale-center and RMA normalizations were obtained using *R programming language* with the packages *Hmisc* function rcorr, *softImpute* function biscale and Bioconductor package *affy*. **A.** Linear dependency curves. The shape of the *dots* shows the tissue types and the *color* the cell line. The *blue lines* represent linear regressions. For each dependency the r correlation coefficient is shown; *asterisks* represent the level of significance (p-values < 0.05*). *Error bars* = standard error (se). **B.** Number of cell lines with extreme expression of the probe sets, r coefficients, and p-values.

Besides, two of the seven cell lines with the lowest expression levels of IFI16 are the breast adenocarcinoma MCF7 and the prostate adenocarcinoma LS174T (Fig. 18A). The probe sets 206332_s_at and 208965_s_at are the only ones that interrogate the IFI16 gene and present an almost perfect positive correlation with each other, with an r coefficient of 0.99 (Fig. 18B).

3.4.3 Relationships between Cell Properties of the Studied Cell Lines and Affymetrix Gene Expression of the Probe Sets with High Significance

The expression values of the entire probe sets in the Affymetrix Human Genome U133 Plus 2.0 Array were used for Pearson product moment statistical tests to search for highly significant dependencies (p-values < 0.0001) of the cellular properties of the studied cell lines (Tables 14 and 15). The expression values of the pair probe sets from the ones that present highly significant relations with the cellular properties were collected and their dependencies recalculated (Table 16). From the 13 genes that present at least one probe set with a highly significant relation with the cellular features, four genes have consistent information across their set of probe sets (Table 16). These

are genes like *LARP4* for which one of the _at probe sets shows a moderate negative relation with the proliferation assay as metabolic activity MTT, and the others (_s_at, _a_at and _at) a positive one (Table 16). Even more drastic is the situation with the six probe sets that interrogate *ZNF333* gene. For this gene, one of the _a_at probe sets gave a highly significant positive dependency with the proliferation assay SRB (that measure the total protein content of the cells). Also, a positive but moderate dependency is present with the other _a_at. However, the four remaining probe sets, all of them with the _at suffix show moderated negative dependencies with the mentioned proliferation assay (Table 16). Moreover, from the 13 genes that present at least one probe set with a highly significant dependency with some of the cellular properties, all the ones that contain only one probe set were not taken into account.

Figure 19 plots the highly significant dependencies among the cellular properties and selected probe sets of genes with consistent information across their probe sets. All of these genes contain more than one probe set on the Affymetrix platform used. Notice that the entire set of probe sets belonging to the four candidate genes, present at least a significant (p-value < 0.05) relationship with the proliferation rate of the studied cell lines (Table 16).

The GCN1L1 gene (eIF-2-alpha kinase activator GCN1) with its probe set 216232_s_at shows a highly positive significant dependency (p-values < 0.0001) with the proliferation rate measured as total protein content in the SRB assay (r = 0.91) (Fig. 19A). UniProtKB/Swiss-Prot refers to GCN1L1 as follows: "*acts as a positive activator of the EIF2AK4/GCN2 protein kinase activity in response to amino acid starvation. Forms a complex with EIF2AK4/GCN2 on translating ribosomes; during this process, GCN1 seems to act as a chaperone to facilitate delivery of uncharged tRNAs that enter the A site of ribosomes to the tRNA-binding domain of EIF2AK4/GCN2, and hence stimulating EIF2AK4/GCN2 kinase activity. Participates in the repression of global protein synthesis and in gene-specific mRNA translation activation, such as the transcriptional activator ATF4, by promoting the EIF2AK4/GCN2-mediated phosphorylation of eukaryotic translation initiation factor 2 (eIF-2-alpha/EIF2S1) on 'Ser-52', and hence allowing ATF4-mediated reprogramming of amino acid biosynthetic gene expression to alleviate nutrient depletion*" (UniProtKB/Swiss-Prot accession number: Q92616). Furthermore, probe set 216232_s_at of GCN1L1 presents a significant (p-value < 0.05) dependency with the MTT proliferation assay (Table 15). The second probe sets that interrogate GCN1L1 in

the Affymetrix array corroborate the relations found giving a significant (p-value < 0.05 data not shown) correlation with both the proliferation assays, SRB and MTT (Table 16). The expression value of this gene measured according to the probe set 216232_s_at does not show much variation across the studied cell lines (Table 14). The cell lines with the highest expression are the colorectal HT-29 and HCT 116 (relative expression of 8.92 and 9.14, respectively) and the lowest are the skin cell lines HaCaT and MDA-MB-435 (relative expression of 8.04 and 7, respectively) (Table 14). Due to the missing data for the proliferation assays in the skin cell lines, the lowest values shown in the linear dependency curve in Fig. 19A are the ones for the breast cancer cell lines MDA-MB-231 and MCF7. These results indicate that the cell lines that proliferate the most present a high expression of the *GCN1L1* gene and this is also related to the tissue type from which the cells are derived.

Moreover, the proliferation rate (MTT/SRB assays) shows significant positive dependencies with HOXA7 (Homeobox 7) (Tables 15 and 16) for its two probe sets. NCBI describes HOXA7 as follows: "*In vertebrates, the genes encoding the class of transcription factors called homeobox genes are found in clusters named A, B, C, and D on four separate chromosomes. Expression of these proteins is spatially and temporally regulated during embryonic development. This gene is part of the A cluster on chromosome 7 and encodes a DNA-binding transcription factor which may regulate gene expression, morphogenesis, and differentiation. For example, the encoded protein represses the transcription of differentiation-specific genes during keratinocyte proliferation, but this repression is then overcome by differentiation signals*" (NCBI/Gene summary, Gene ID: 3204). The expression values of the two probe sets that interrogate HOXA7 showed a strong positive relation with the proliferation assay measured as metabolic activity in MTT, and is corroborated with a significant (p-values < 0.01, data not shown) positive relation found between the two probe sets and the proliferation assay SRB (Table 16). For this gene it is also possible to observe a significant (p-values < 0.01, data not shown) negative dependency between the two probe sets and ROS accumulation in the cells. As discussed in Sect. 3.3, the proliferation rate in the SRB and MTT assays presents a significant negative correlation (r = –0.68) (Table 10) with ROS accumulation in the cells. These findings indicate that cells with high proliferation rates also express the HOXA7 gene to a higher degree and show less ROS accumulation. The expression levels of HOXA7 are low for most of the cell lines; the two breast cancer cell lines and the four

pancreatic ones show equally low levels of the gene. In contrast, the two colorectal cancer cell lines HT-29 and HCT 116 present a much higher expression level of *HOXA7* (Fig. 19B). Even higher is the expression of *HOXA7* in HaCaT and HeLa cells (Table 14), being these two cell lines the ones with the stronger expression. The cell lines belonging to the prostate tissue type (LNCaP, PC-3, and DU 145) present a similarly low expression level as the pancreatic and breast tissue type cell lines (Table 14).

The third gene that presents a high dependency with the proliferation rate is INPP5B (inositol polyphosphate-5-phosphatase B). NCBI describes INPP5B as follows: "*This gene encodes a member of a family of inositol polyphosphate-5-phosphatases. These enzymes function in the regulation of calcium signaling by inactivating inositol phosphates. The encoded protein is localized to the cytosol and mitochondria, and associates with membranes through an isoprenyl modification near the C-terminus*" (NCBI/Gene summary, Gene ID: 3633). The highly sifnificant (p-value < 0.0001) positive dependency found for the probe set 213643_s_at of the INPP5B gene and the SRB assay is corroborated by the three probe sets that contain information for the gene, all of them having a significant (p-value < 0.01, data not shown) positive relation with the MTT proliferation assay (data not shown). The expression of the INPP5B gene behaves in the same way as the expression for HOXA7; most of the cell lines showed similar relative low expression levels with the two colorectal ones having together a higher expression (Fig. 19C). Interestingly, the lowest relative expressions are for the breast and pancreas cell lines, which are around 5.5, and the highest is for the brain 1321N1 with just 7.14. The colorectal HT-29 and HCT 116 present a relative expression of around 6.5 (Table 14). Cell lines with low expression levels of the INPP5B gene also present low proliferation rates.

Among the highly significant positive dependencies found between the set of candidate genes and the proliferation rate is LGALS8 (Galectin 8), the only one that shows a negative correlation with one of the proliferation assays and also with the cellular mitochondrial mass content (Table 16 and Fig. 19D). NCBI describes LGALS8 as follows: "*This gene encodes a member of the galectin family. Galectins are beta-galactoside-binding animal lectins with conserved carbohydrate recognition domains. The galectins have been implicated in many essential functions including development, differentiation, cell-cell adhesion, cell-matrix interaction, growth regulation, apoptosis,*

Table 14: Affymetric gene expression of the genes that shows the highly significant dependencies with the cellular properties of the studied cell lines. From more than 54,000 probe sets in which the R coefficient was calculated, the expression values of the ones with the highest significant correlations (p values < 0.0001***) are shown.

Gene Symbol	Probe Set ID	MDA-MB-231	MCF7	HCT 116	HT-29	AsPC-1	MIA PaCa-2	Panc-1	BxPC-3	LNCaP	PC-3	DU 145	MDA-MB-435	HaCaT	HeLa	Hep G2	1321N1
CCL19	210072_at	4.86	5.20	4.68	4.94	5.09	4.99	4.98	5.15	4.72	5.03	4.88	4.84	5.04	4.95	4.95	4.82
LGALS8	208935_s_at	8.35	8.11	6.72	7.22	8.07	7.81	8.30	7.32	9.17	6.47	7.62	6.54	9.65	4.92	8.19	5.82
LHX8	1569469_a_at	2.85	2.99	3.48	3.10	2.96	3.12	3.05	3.19	2.90	3.13	3.09	2.94	2.83	3.62	3.05	2.99
KLHL23 // / PHOSPHO2 -KLHL23	217505_at	3.13	3.18	2.95	2.84	2.81	3.03	3.20	3.11	3.07	2.97	3.04	2.85	2.85	2.82	3.04	2.94
GCN1L1	216232_s_at	8.19	8.27	9.14	8.92	8.42	8.65	8.63	8.57	8.38	8.68	8.45	7.00	8.04	8.66	8.44	8.55
INPP5B	213643_s_at	5.54	5.53	6.67	6.38	5.29	5.65	5.57	5.67	5.92	5.72	6.16	5.33	6.37	6.18	5.89	7.14
DUSP2	204794_at	4.29	4.59	7.42	6.04	4.02	6.11	6.60	5.09	6.24	3.72	3.91	4.98	4.79	8.39	3.69	3.88
GCC1	218912_at	7.40	7.62	8.05	7.92	7.38	7.78	7.69	7.67	7.62	7.63	7.19	7.39	8.30	8.43	7.43	8.01
HOXA7	206847_s_at	3.20	3.92	7.19	5.26	3.41	3.80	3.49	4.05	3.08	3.75	3.63	4.85	7.10	7.06	3.49	5.08
LARP4	212714_at	9.45	9.20	10.06	9.69	9.30	9.74	9.47	9.43	9.92	9.29	9.37	9.47	9.77	10.15	9.33	8.78
SLC39A10	225295_at	9.32	7.63	8.97	8.54	9.22	10.25	8.81	7.39	9.68	9.01	8.55	8.31	8.06	10.10	7.62	8.60
ZNF333	1563783_a_at	3.61	3.59	3.91	3.84	3.51	3.65	3.72	3.55	4.14	3.66	3.62	3.44	3.68	3.65	3.71	3.84
WDR86	243526_at	3.25	3.46	3.85	3.63	3.57	3.52	3.53	3.45	3.44	3.16	3.95	3.16	4.06	4.06	3.35	3.31

Table 15: p values of the highly significant dependencies between the cellular properties and the total probes sets from Affymetric Array. Linear correlation coefficients between the cellular features and the complete Affymetrix Human Genome U133 Plus 2.0 Array probe sets were calculated using Pearson algorithm. The gene expression values were obtained from GEO. NCBI. and public accession numbers of the cell lines are listed in Experimental Methods 2.3 Table 2. Pre-processing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. The p values of the highly significant dependencies (p values < 0.0001) found in the complete array are highlight. Pearson comparison and RMA normalization were obtained using R *programming language* with the packages *Hmisc* function rcorr and Bioconductor package *affy*.

Gene Symbol	Probe set ID	Capacitance	Respiration	Glycolysis	ATP	Mitomass	ROS	SRB	MTT
CCL19	210072_at	2.0E-01	4.6E-01	8.5E-05	6.5E-01	5.6E-01	3.4E-01	9.0E-01	7.4E-01
DUSP2	204794_at	1.4E-01	1.0E-02	8.1E-03	4.4E-01	5.0E-02	6.0E-01	8.0E-02	8.0E-05
GCC1	218912_at	5.8E-01	9.5E-01	4.1E-01	4.8E-01	2.7E-01	2.4E-01	6.7E-02	7.6E-05
GCN1L1	216232_s_at	6.8E-01	3.8E-01	3.3E-01	8.4E-01	1.5E-01	1.0E+00	9.9E-05	2.0E-02
HOXA7	206847_s_at	7.8E-01	5.6E-01	2.6E-01	5.3E-02	1.6E-01	8.9E-03	9.6E-03	9.2E-06
INPP5B	213643_s_at	7.0E-01	3.6E-01	2.2E-01	6.3E-01	3.0E-01	7.6E-02	6.0E-05	8.1E-03
KLHL23 /// PHOSPHO2-KLHL23	217505_at	4.9E-01	2.1E-01	4.2E-01	3.5E-02	4.0E-01	6.4E-05	1.5E-01	8.5E-02
LARP4	212714_at	1.1E-01	6.0E-02	4.3E-02	2.4E-01	4.2E-01	4.1E-01	2.7E-02	1.0E-05
LGALS8	208935_s_at	5.6E-01	7.4E-01	7.4E-02	8.1E-02	5.5E-05	8.4E-02	2.5E-01	1.0E-02
LHX8	1569469_a_at	5.7E-01	2.4E-01	3.4E-02	7.4E-02	2.7E-05	1.6E-01	7.4E-02	4.8E-04
SLC39A10	225295_at	5.1E-05	9.6E-02	4.4E-02	6.1E-01	6.5E-01	9.8E-01	7.3E-01	1.4E-01
WDR86	243526_at	5.9E-01	6.7E-01	7.4E-01	4.3E-01	4.0E-01	5.3E-02	8.8E-02	2.8E-05
ZNF333	1563783_a_at	1.7E-01	1.5E-01	3.4E-02	3.3E-01	8.4E-01	7.1E-01	6.2E-05	1.1E-01

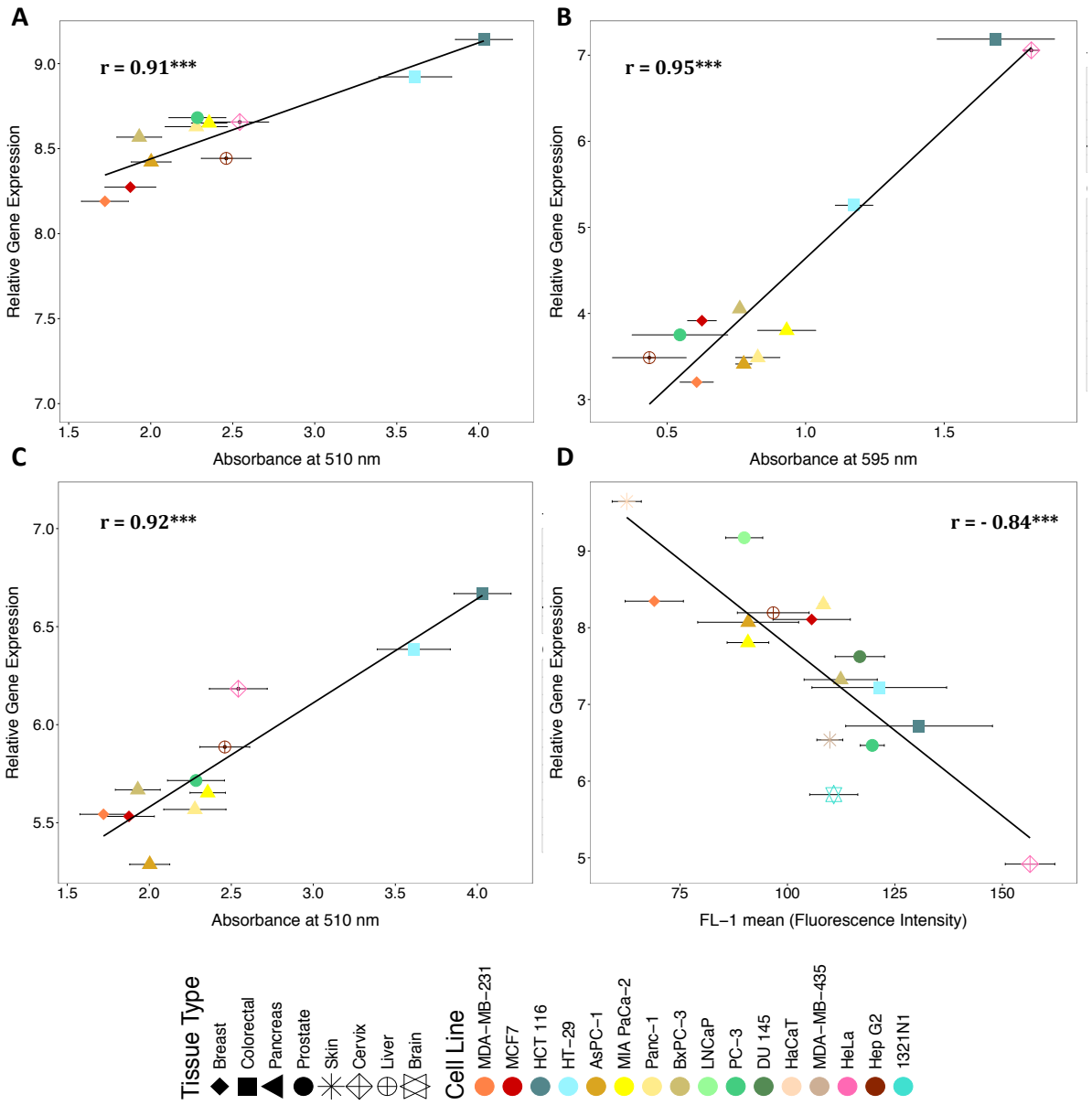


Fig. 19. Highly significant dependencies between the metabolic features of the analyzed cell lines and the total probe sets in Affymetrix Array. Linear correlation coefficients between the metabolic features and the complete Affymetrix Human Genome U133 Plus 2.0 Array probe sets were calculated using Pearson algorithm. Mitochondrial mass content was determined using MitoTracker Green (Thermo Fisher). The fluorescence intensity is proportional to the mitochondrial mass of the cells. The proliferation rate was estimated using MTT and SRB assays. MTT reagent labels active mitochondria and its reduction produces a colored compound that can be measured by absorbance at 595 nm. The SRB dye is incorporated into fixed cells and the amount of incorporated dye is proportional to the amount of cells; the dye absorbs at 510 nm. Gene expression values were obtained from GEO, NCBI, and public accession numbers of the cell lines are listed in Experimental Methods, Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. From more than 54,000 probe sets in which the r coefficient was calculated, the ones with the highest significant p-values (< 0.0001***) were selected and only those genes with consistent

information in all their corresponding probe sets were plotted. Pearson comparison and RMA normalization were obtained using *R programming language* with the packages *Hmisc* function rcorr and Bioconductor package *affy*. The shape of the *dots* gives the tissue types and the *color* the cell line. The *blue lines* represent linear regressions. For each independent dependency the r correlation coefficient is shown; *asterisks* represent the level of significance. *Error bars* = standard error (se). Linear correlations between: **A.** SRB proliferation assay and GCN1L1 probe set 216232_s_at. **B.** MTT proliferation assay and HOXA7 probe set 206847_s_at. **C.** SRB proliferation assay and INPP5B probe set 213643_s_at. **D.** Mitochondrial mass content and LGALS8 probe set 208935_s_at.

Table 16. r coefficients of the highly significant dependencies between metabolism features and total probe sets from Affymetrix Array. Linear correlation coefficients between the cellular features and the complete Affymetrix Human Genome U133 Plus 2.0 Array probe sets were calculated using Pearson algorithm. Cellular capacitance, glycolytic activity (acidification), and respiratory activity (O_2 consumption) were measured on the Bionas 2500 online monitoring system. Cells were seeded in chips containing IDES sensors for the estimation of the cellular capacitance, ISFET sensors to measure the acidification rate, and Clark-type sensor for determination of oxygen consumption. Energy metabolism (ATP level) was measured with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). Mitochondrial mass content was determined by MitoTracker Green (Thermo Fisher). ROS accumulation in cells was measured using dihydroethidium (DHE). Gene expression values were obtained from GEO, NCBI, and public accession numbers of the cell lines are listed in Experimental Methods, Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. From more than 54,000 probe sets in which the r coefficient was calculated, the ones with the highest significant p-values (< 0.0001***) were selected. Genes with consistent information of the r coefficients in their total probe sets are highlighted. Pearson comparison and RMA normalization were obtained using *R programming language* with the packages *Hmisc* function rcorr and Bioconductor package *affy*.

Gene Symbol	Probe set ID	Capacitance	Respiration	Glycolysis	ATP	Mitomass	ROS	SRB	MTT
CCL19	**210072_at**	-0.39	-0.26	**-0.78**	0.13	-0.13	0.18	-0.70	-0.54
DUSP2	**204794_at**	0.38	0.62	0.64	-0.21	0.50	-0.14	0.55	**0.91**
GCC1	**218912_at**	0.15	-0.02	0.22	-0.20	0.29	-0.31	0.57	**0.92**
	243306_s_at	-0.22	-0.16	0.07	-0.17	0.19	-0.26	0.65	0.08
	243437_at	0.14	0.04	0.03	0.13	0.34	0.01	-0.02	0.04
GCN1L1	212139_at	0.14	0.30	0.22	0.15	0.31	0.02	0.54	0.73
	216232_s_at	0.11	0.24	0.26	0.06	0.37	0.00	**0.91**	0.68
HOXA7	**206847_s_at**	-0.08	-0.16	0.30	-0.51	0.37	-0.63	0.74	**0.95**
	235753_at	-0.05	-0.20	0.22	-0.49	0.27	-0.62	0.63	0.90
INPP5B	1563565_at	0.51	0.01	0.50	0.15	0.21	-0.17	0.29	0.71
	213643_s_at	-0.10	-0.25	0.33	-0.14	0.28	-0.46	**0.92**	0.75
	213804_at	-0.08	0.17	0.28	0.04	0.36	-0.27	0.57	0.73
KLHL23 /// PHOSPHO 2-KLHL23	213610_s_at	0.23	0.09	0.64	-0.29	0.24	-0.28	0.64	0.38
	217505_at	-0.19	0.33	-0.22	0.55	-0.22	**0.83**	-0.46	-0.54
LARP4	1555384_a_at	0.26	0.49	0.42	0.35	0.01	0.36	0.23	0.13

	212714_at	0.41	0.48	0.51	-0.32	0.22	-0.22	0.66	**0.95**
	214155_s_at	0.19	0.38	0.38	0.17	0.29	0.27	0.11	0.38
	238959_at	0.13	0.47	0.41	0.33	0.08	0.38	0.23	-0.13
	238960_s_at	0.14	0.32	0.45	0.25	0.09	0.21	0.34	0.00
LGALS8	208933_s_at	-0.20	0.05	-0.45	0.30	-0.76	0.38	-0.32	-0.77
	208934_s_at	-0.35	0.18	-0.41	0.40	-0.67	0.50	-0.22	-0.70
	208935_s_at	-0.16	0.09	-0.46	0.46	**-0.84**	0.45	-0.38	-0.73
	208936_x_at	-0.26	0.19	-0.45	0.39	-0.70	0.50	-0.31	-0.75
	210731_s_at	-0.24	0.01	-0.47	0.17	-0.73	0.35	-0.30	-0.75
	210732_s_at	-0.33	0.13	-0.41	0.38	-0.68	0.52	-0.23	-0.74
LHX8	**1569469_a_at**	0.15	0.31	0.53	-0.47	**0.85**	-0.37	0.56	0.87
SLC39A10	**225295_at**	**0.84**	0.43	0.51	0.14	0.12	0.01	0.12	0.48
WDR86	**243526_at**	-0.15	0.12	0.09	-0.22	0.23	-0.49	0.54	**0.93**
ZNF333	1552375_at	0.23	0.33	0.18	0.60	-0.20	0.50	-0.33	-0.41
	1559674_at	0.19	0.18	0.22	0.00	-0.10	0.21	-0.21	0.01
	1563783_a_at	0.36	0.38	0.53	0.27	0.05	0.10	**0.92**	0.51
	1569250_at	0.22	-0.11	0.23	0.29	-0.21	0.17	-0.10	-0.07
	1569251_a_at	0.40	0.12	0.30	0.39	-0.10	-0.01	0.13	0.27
	231369_at	-0.25	-0.11	0.00	0.10	-0.07	0.23	-0.02	-0.48

and RNA splicing. This gene is widely expressed in tumoral tissues and seems to be involved in integrin-like cell interactions" (NCBI/Gene summary, Gene ID: 3964). The probe set 208935_s_at shows a highly significant (p-value < 0.0001) and negative dependency with the mitochondrial mass content (Fig. 19D). The LGALS8 gene contains five probe sets more and in all of them a significant (p-value < 0.01 data not shown) negative dependency with the mitochondrial mass is found (Table 16). In the entire set of probe sets that interrogate the LGALS8 gene, a negative and significant (p-value < 0.01, data not shown) relation with the MTT proliferation assay is observable as well, but not for SRB. However, the dependencies found between the LGALS8 probe sets and SRB assay display the same tendency, a negative correlation (Table 16). The data obtained for the Pearson correlation coefficients indicate that the cell lines that express the higher level of the LGALS8 gene are the ones that present the lowest proliferation rate measured as metabolic activity and mitochondrial mass. As shown in Table 10, the mitochondrial mass content of the studied cell lines and the MTT proliferation assay do not show a significant dependency upon each other. However, a moderate positive correlation is present. Different to the expression pattern observed for the other three candidates genes, here, the colorectal cells HT-29 and HCT 116 are part of the ones that present the lowest expression levels (Fig. 19D). The relative expression of the LGALS8 gene for the breast and pancreas tissue type cells is high in comparison with the colorectal cell lines. The prostate cell lines differ in their expression levels and the same is observable for the

skin cell lines, MDA-MB-435 being one of the lowest and the noncancerous cell line HaCaT the highest in all the tested cell lines. The lowest relative expression is displayed by the cervix HeLa cell line (Fig. 19D).

3.4.4 Affymetrix Gene Expression and qPCR

The relative gene expression of the glycolysis-gluconeogenesis genes GOT1 (glutamic-oxaloacetic transaminase 1), HK1 (hexokinase 1), HK2 (hexokinase 2), LDHA (lactate dehydrogenase A), PFKM (phosphofructokinase-muscle), and SLC2A1 (solute carrier family 2 member 1) was estimated by qPCR. Real-time PCR was performed in the breast cancer cell lines MDA-MB-231 and MCF7, the colorectal cell lines HCT 116 and HT-29, the pancreatic cell lines AsPC-1, MIA PaCa-2, Panc-1, and BxPC-3, the prostate cell lines PC-3 and DU 145, the cervix HeLa, and the liver cell line Hep G2.

The RNA for the Reverse Transcriptase PCR comes from cells cultivated in DMEM medium at 37°C in 5% CO_2 incubator with 95% humidity. Primer sequences for the qPCR are listed in Table 6 (Experimental Methods 2.3.3). The relative gene expression was calculated as the mean of three independent experiments. Most of the amplified genes are highly expressed in the carcinoma cell lines Hep G2 and HCT 116 (Fig. 20). In contrast, HK1 and PFKM present the lowest expression values in all the tested cell lines for Hep G2. A high expression is also observed for LDHA in the adenocarcinomas AsPC-1 and PC-3 (Fig. 20).

Comparison was made between the results obtained from the qPCR gene expression estimations (Fig. 20) and the data from the Affymetrix Human Genome U133 Plus 2.0 Array obtained from public repositories files (Experimental Methods 2.3) (Fig. 21). The qPCR expression value of each of the six amplified genes was normalized across the cell lines. Scale and center were applied to normalize the gene expression relative to the set of cell lines used. A range of calculations was obtained to compare the expression values of each gene in the tested cell lines from –3 (lowly expressed genes) to 3 (highly expressed genes) (Experimental Methods 2.3.1). The same procedure was done regarding the gene expression of the same six genes obtained from the available public repositories (De Schutter et al. 2013, Barretina et al. 2012) (Experimental Methods 2.3). For the genes that contain more than one probe set in the Affymetrix, a mean of the probe sets was used. Probe sets with a _x suffix were deleted. Scaling and

centering were calculated using *R programming language* package *softImpute* (Hastie & Mazumder 2015) (Supplements 5.1.5).

Affymetrix arrays and qPCR standardized relative expression values of GOT1, HK1, HK2, LDHA, PFKM, and SLC2A1 in the 12 cell lines were plotted together in order to compare our results with the ones obtained by De Schutter et al. (2013) and Barretina et al. (2012) (Fig. 21). From a total of six glycolytic genes in the 12 cell lines, 53 out of the 72 comparisons show the same tendency, positive expression or negative expression, within the set of cell lines used. However, in most of the cases, they do not share the level of expression between the two methodologies. An exception is observed for GOT1 in Hep G2; HK1 in DU 145 and HeLa; HK2 in HT-29, HCT 116, Panc-1, and MCF7; PFKM in Panc-1; and SLC2A1 in MCF7, AsPC-1, and MIA PaCa-2 that showed relative similar expression values in relation to the cell lines used (Fig. 21). Major similarities are found in the prostate cell line PC-3 and the liver cell line Hep G2, where all the amplified genes follow the same Affymetrix patterns, being similarly either upregulated or downregulated (Fig. 21).

Fig. 20. Real-time polymerase chain reaction (qPCR) in key metabolic genes of the glycolysis-gluconeogenesis pathway. The RNA for the RT-PCR comes from cells cultivated in DMEM medium supplemented with 10% FBS (v/v) in a 5% CO_2 incubator at 37°C with a 95% humidity. Standard deviations come from three independent experiments with three internal replicates for each one. Primer sequences and amplification details are given in Experimental Methods 2.3.3.

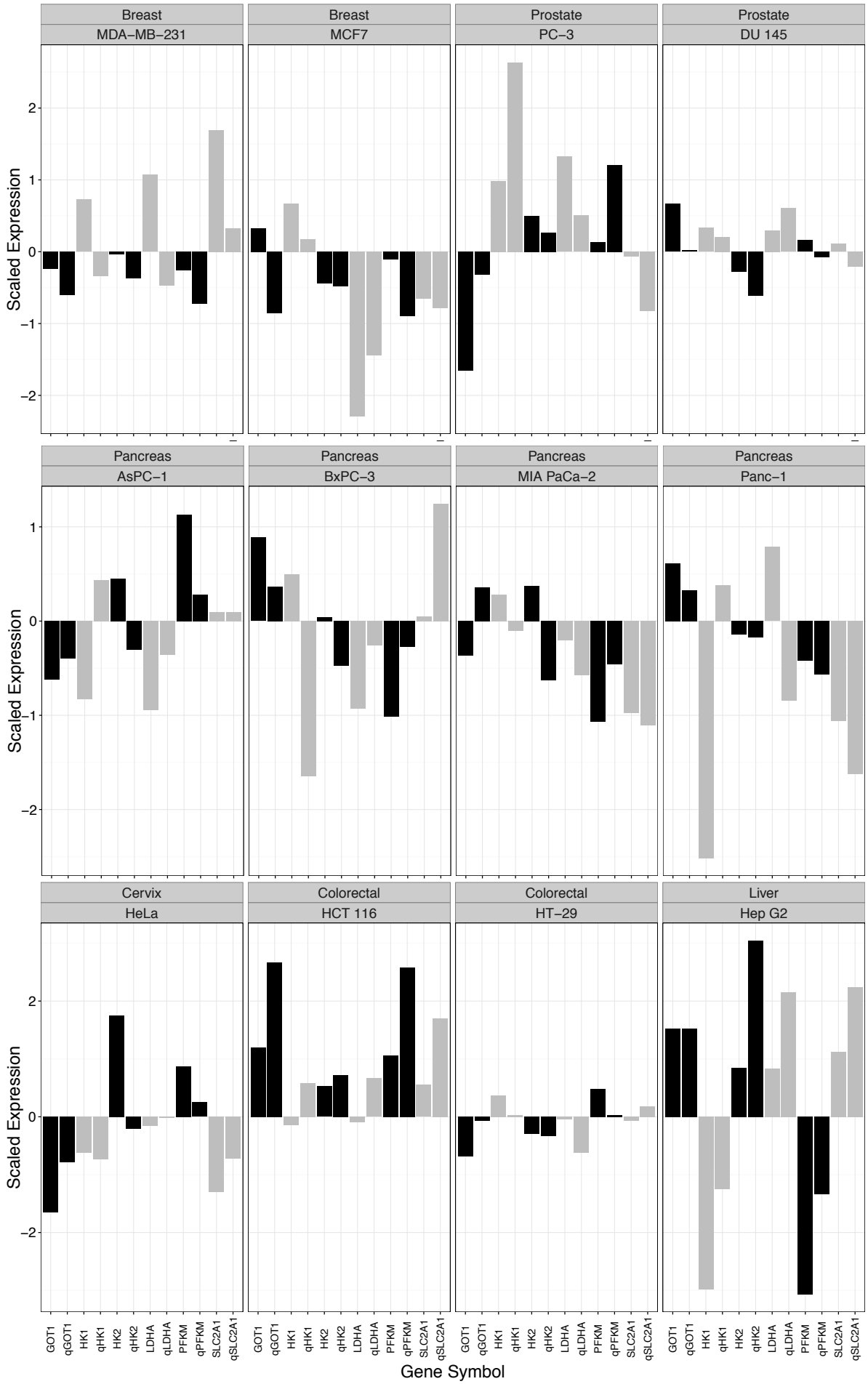

Breast
MDA−MB−231
Breast
MCF7
Prostate
PC−3
Prostate
DU 145
Pancreas
AsPC−1
Pancreas
BxPC−3
Pancreas
MIA PaCa−2
Pancreas
Panc−1
Cervix
HeLa
Colorectal
HCT 116
Colorectal
HT−29
Liver
Hep G2
Scaled Expression
2
1
0
−1
−2
GOT1
qGOT1
HK1
qHK1
HK2
qHK2
LDHA
qLDHA
PFKM
qPFKM
SLC2A1
qSLC2A1
Gene Symbol

Fig. 21. Affymetrix Gene expression and qPCR comparison. Affymetrix Gene expression raw data (.CEL files) were obtained from GEO, NCBI and correspond to the Human Genome U133 Plus 2.0 Array. Public accession numbers and cell lines are listed in Experimental Methods Table 2. Preprocessing of the .CEL files was done using Robust Multi-array Average (RMA) normalization method. Primer sequences and amplification details for the qPCR are given in Experimental Methods 2.3. The standard deviation of the qPCR experiments is not shown (for SD of qPCR see Fig. 19). Scale and center normalization were applied to the expression values of the Affymetrix and the qPCR data to normalize the expression of each gene in the set of studied cell lines. A range from –3 (lowly expressed genes in the lower part of the plot) to 3 (high expressed genes in the upper part of the plot) was obtained. Expression values that come from qPCR are denoted by a "q" followed by the corresponding gene symbol. Scale-center and RMA normalizations were obtained using *R programming language* package *softImpute* function biscale and Bioconductor package *affy*.

3.5 Cellular Properties and Drug IC50s

3.5.1 Relationships between Cell Properties and Drug IC50s of the Studied Cell Lines

To reach a better understanding of the impact of drugs in relation to cancer metabolism and cellular capacitance, measurements were obtained on the cellular properties of the studied cell lines (Figs. 2 to 10) and the log10 of the IC_{50} values for several drugs (Fig. 22 and Experimental Methods 2.4.1, Table 7) were used to search for relationships. Novel links between known drugs and cellular features were discovered. Pearson product moment correlations were used to find linear dependencies between the mentioned two variables. For the cellular properties, at least three independent experiments per assay and per cell line, on the average, were carried out. The IC_{50} values of nearly 100 compounds were obtained from the Genomics of Drug Sensitivity in Cancer project (Yang et al. 2013) (Fig. 22 and Experimental Methods 2.4.1, Table 7). Data for drug IC_{50}s were found for 11 out of the 18 cell lines used in this study. The data collection was done for the following cell lines: MDA-MB-231 and MCF7 (breast); HCT 116 and HT-29 (colon); AsPC-1, MIA PaCa-2, and BxPC-3 (pancreas); LNCaP, PC-3, and DU 145 (prostate); and MDA-MB-435 (skin) (Experimental Methods 1.1, Table 1). The r coefficients were acquired using *R programming language* packages *Hmisc* and the

rcorr function. Seven of the drugs showed a significant correlation ($p < 0.01$) with one of the cellular properties (Figs. 23–26).

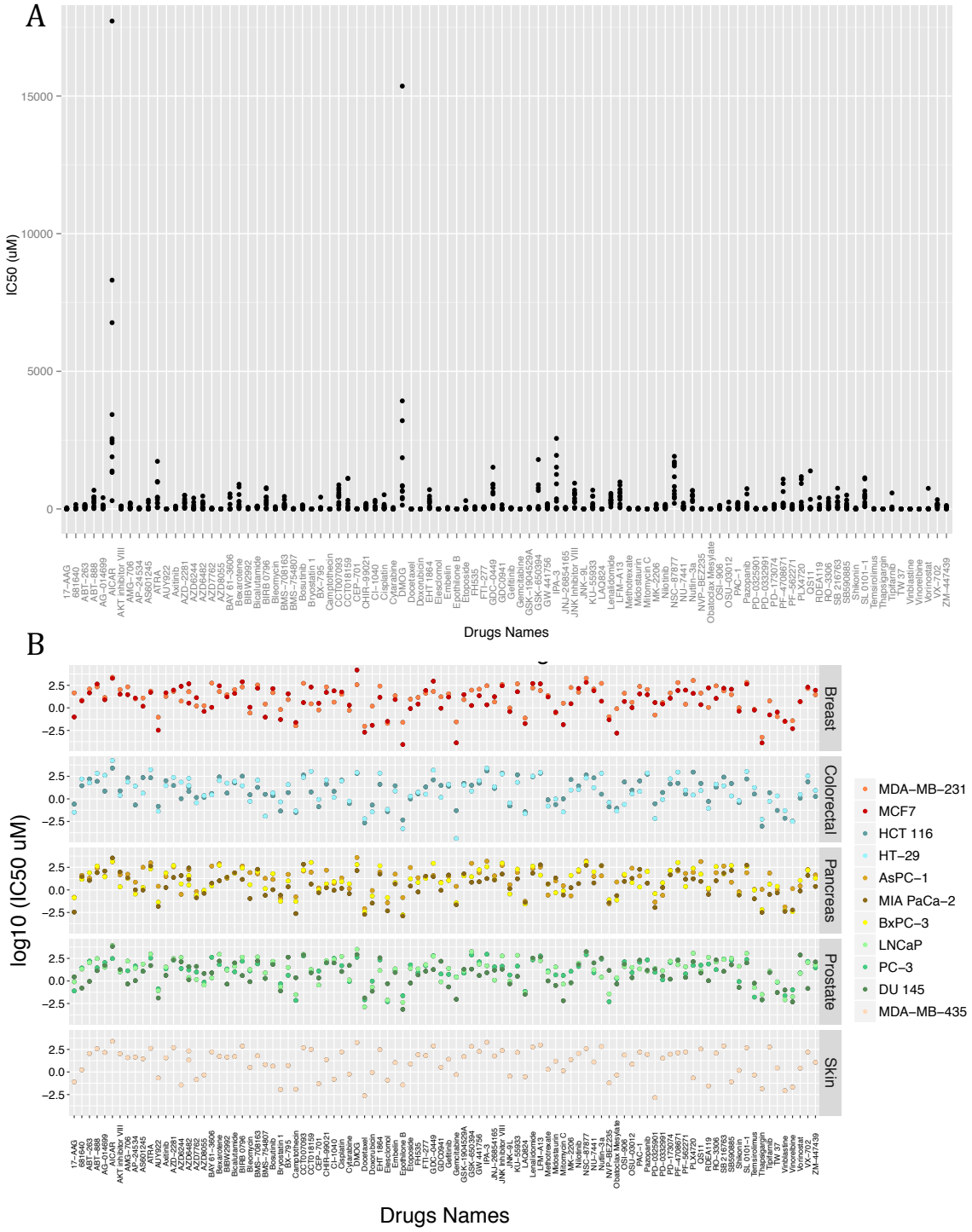


Fig. 22. Drug IC_{50}s of 99 compounds were obtained from the Genomics of Drug Sensitivity in Cancer project. Each *dot* corresponds to a cell line. Data on drug IC_{50}s for 11 of the 18 cell lines used in this study were obtained from the database. **A.** Drug IC_{50}s in μM. **B.** log10 of the drug IC_{50}s by tissue type.

Elesclomol is a copper chelator that induces growth arrest and apoptosis in cancer cell lines through the induction of oxidative stress and an increase in ROS production (Rae & Mairs 2017, Krishner et al. 2008). Preclinical studies showed an efficacy against acute myeloid leukemia (Hedley et al. 2016). Our results indicate a negative correlation of elesclomol with cellular capacitance, indicating stronger cell-matrix and cell-cell contacts of the prostate DU 145 and the pancreas cancer cell line BxPC-3 that may make them more resistant to the compound (Fig. 23).

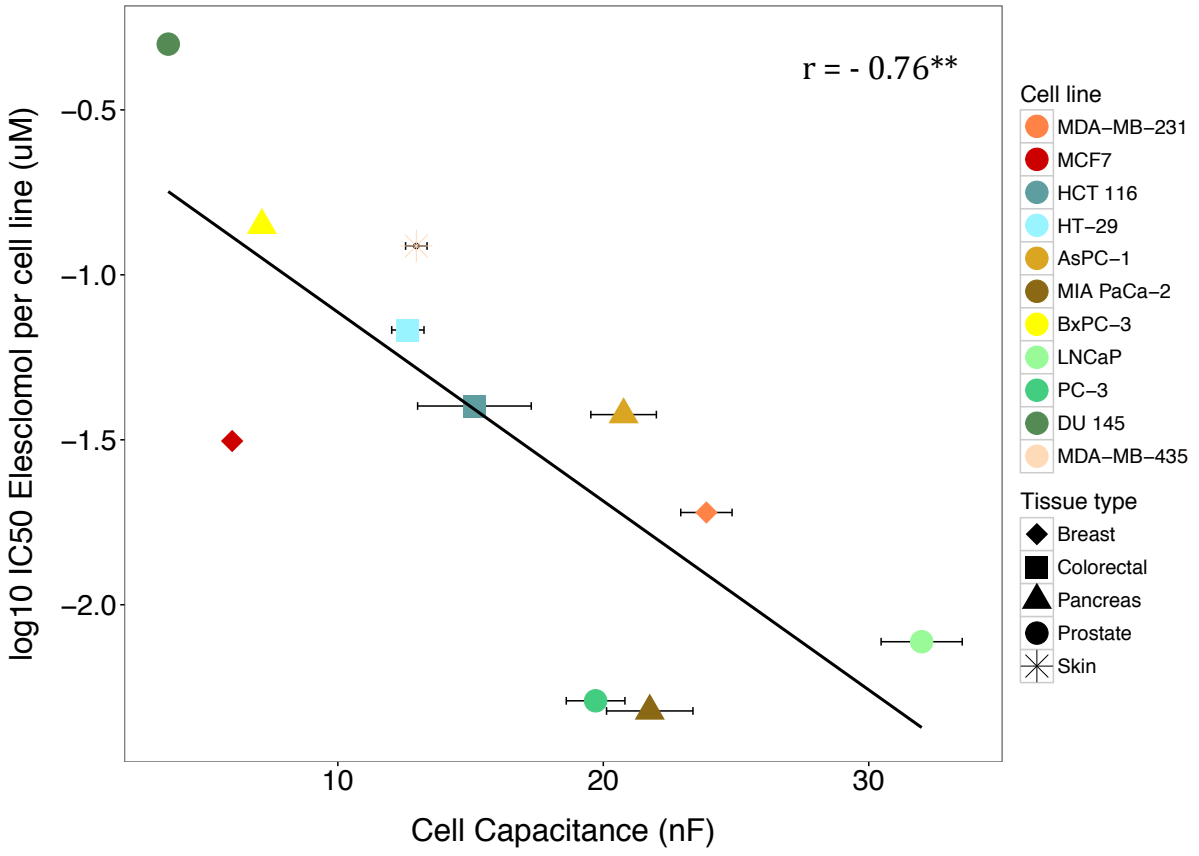


Fig. 23. Elesclomol - significant dependencies between cellular capacitance and the log10 IC_{50} of elesclomol. The cellular capacitance was measured with the Bionas 2500 online monitoring system. Cells were seeded in chips containing IDES sensors. Lower capacitance indicates a stronger cell matrix adhesion and cell-cell contact interactions. The mean value of at least three independent experiments per cell line was used for the Pearson correlation. The IC_{50} was obtained from the Genomics of Drug Sensitivity in Cancer project. The Pearson calculation was done using *R programming language* package *Hmisc* function rcorr. The shape of the *dots* gives the tissue types and the color the cell lines. The *blue line* represents the linear regression. The r correlation coefficient is shown and the *asterisks* represent the level of significance (p-values < 0.01**). *Error bars* = standard error (se).

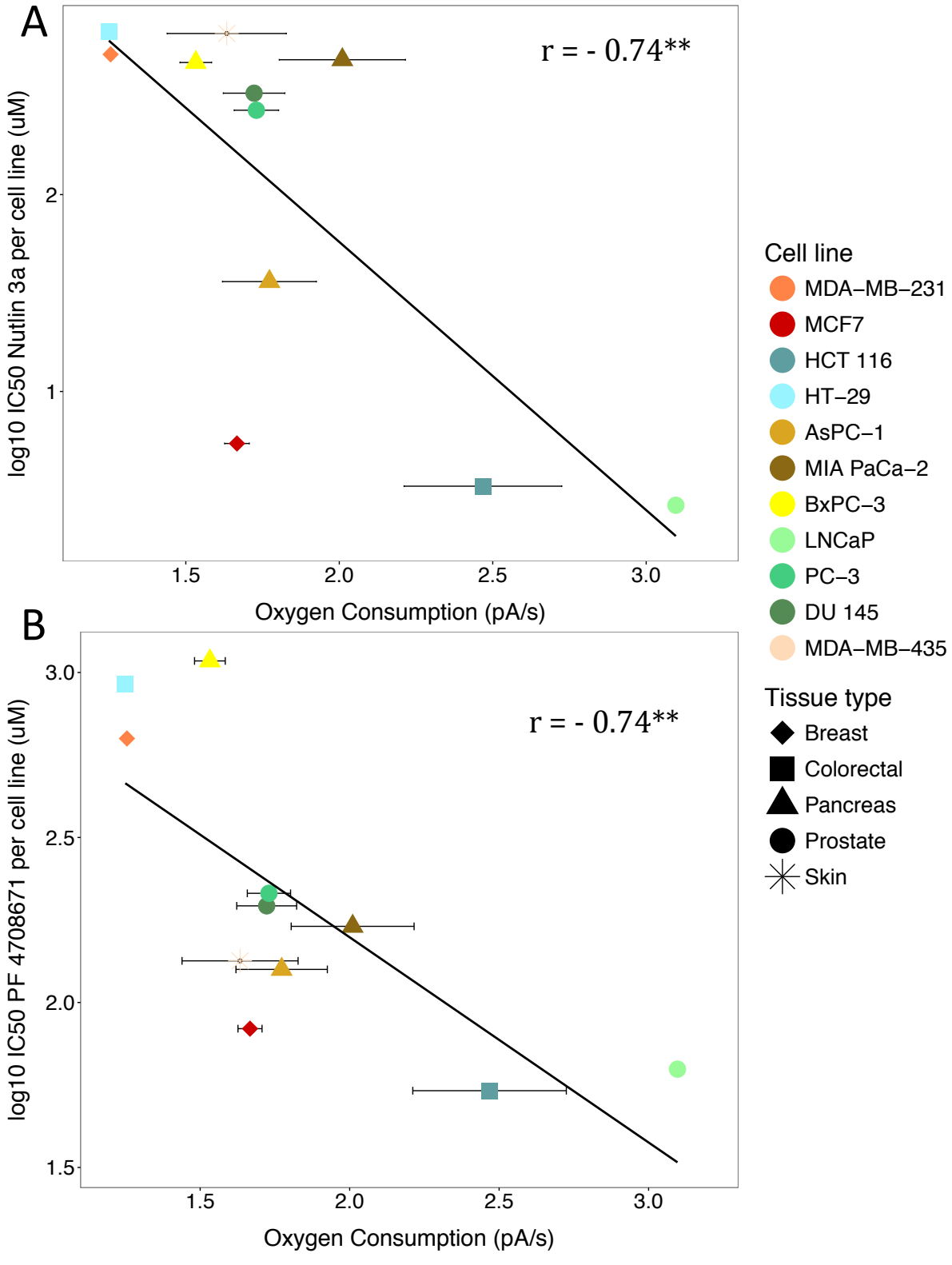


Fig. 24. Nutlin 3a and PF 4708671 – significant dependencies between respiration activity and the log10 IC_{50} of Nutlin 3a (A) and PF 4708671 (B). The respiration activity was measured with the Bionas 2500 online monitoring system. Cells were seeded in chips containing Clark electrodes. Higher oxygen consumption indicates stronger respiration rate. The mean value of at least three independent experiments per cell line was used for the Pearson

correlation. IC_{50} values were obtained from the Genomics of Drug Sensitivity in Cancer project. The Pearson calculations were done using *R programming language* package *Hmisc* function rcorr. The shape of the *dots* gives the tissue types and the *color* the cell lines. The *blue lines* represent the linear regressions. For each independent dependency the r correlation coefficient is shown; *asterisks* represent the level of significance (p-values < 0.01**). *Error bars* = standard error (se).

Nutlin 3a inhibits the interaction between mdm2 (a negative regulator of p53) and the tumor suppressor p53 (Kojima et al. 2006) and has been shown to have a negative correlation coefficient with the respiratory activity of the cancer cells (Fig. 24A). The cell lines with the lowest oxygen consumption, being the breast MDA-MB-231 and the colorectal HT-29, are the most resistant cells to Nutlin 3a (Fig. 24A).

PF 4708671 is an inhibitor of the p70 ribosomal S6 kinase 1 (p70S6K1) (Pearce et al. 2010). p70S6K1 is involved in the regulation of protein synthesis, proliferation, growth, and longevity and is activated by insulin and growth factors through PI3K and mTOR (Pearce et al. 2010). The cancer cell lines with high oxygen consumption rates are more sensitive to PF 4708671 as indicated by the negative correlation of the inhibiton of p70S6K1 with the respiratory activity (Fig. 24B).

EHT 1864 is a Rac1, Rac1b, Rac2, and Rac3 small family GTPase inhibitor (Onesto et al. 2008). The Rho family proteins are known to have an effect in the stimulation of proliferation, invasion, and metastasis of cancer cells (Onesto et al. 2008). Rac1 has been implicated in cell tumorogenesis transformation via Ras oncogenes (Kissil et al. 2007) and the four of them have been found to be overexpressed in cancer (Onesto et al. 2008). In this study EHT 1864 displayed an inverse relation with energy metabolism, the cells with low ATP formation (MDA-MB-435 and HT-29) being the more resistant ones towards the mentioned drug (Fig. 25A).

IPA3 is a small molecule inhibitor of p21-activated protein kinase 1 (PAK1) (Singhal and Kandel 2012). The analysis showed a negative correlation between energy metabolism of the cells and the log10 IC_{50} of IPA3, meaning that the cell lines that produce more ATP are more sensitive to this compound (Fig. 25B). The colorectal carcinoma cell lines HCT 116 and HT-29 displayed the lowest ATP levels and the highest IC_{50} values, while the adenocarcinoma MCF7 (breast) and LNCaP (prostate) with the highest ATP levels had the lowest IC_{50} values (Fig. 25B).

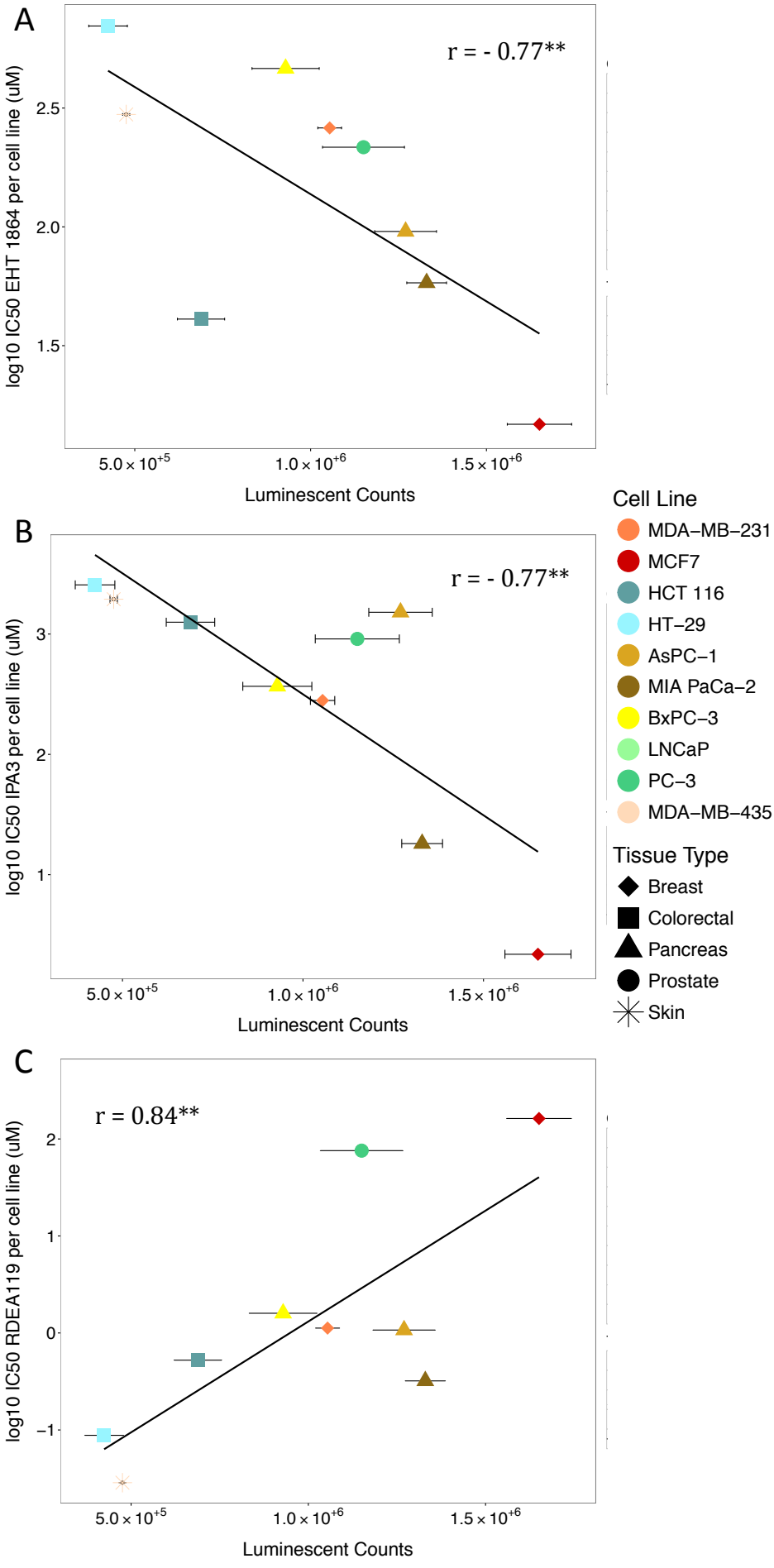

A
r = - 0.77**
log10 IC50 EHT 1864 per cell line (uM)
2.5
2.0
1.5
5.0 × 10+5
1.0 × 10+6
1.5 × 10+6
Luminescent Counts
B
r = - 0.77**
log10 IC50 IPA3 per cell line (uM)
3
2
1
5.0 × 10+5
1.0 × 10+6
1.5 × 10+6
Luminescent Counts
C
r = 0.84**
log10 IC50 RDEA119 per cell line (uM)
2
1
0
−1
5.0 × 10+5
1.0 × 10+6
1.5 × 10+6
Luminescent Counts
Cell Line
MDA–MB–231
MCF7
HCT 116
HT–29
AsPC–1
MIA PaCa–2
BxPC–3
LNCaP
PC–3
MDA–MB–435
Tissue Type
Breast
Colorectal
Pancreas
Prostate
Skin

Fig. 25. EHT 1864, IPA3, and RDEA119 – significant dependencies between energy metabolism and the log10 IC_{50} of EHT 1864 (A), IPA3 (B), and RDEA119 (C). The ATP level was estimated with ATPlite 1step Luminescence ATP Detection Assay System kit (Perkin Elmer). The luminescence is proportional to the ATP concentration in the cells. The mean value of at least three independent experiments per cell line was used for the Pearson correlation. IC_{50} values were obtained from the Genomics of Drug Sensitivity in Cancer project. The Pearson calculations were done using *R programming language* package *Hmisc* function rcorr. The shape of the *dots* gives the tissue types and the *color* the cell lines. The *blue lines* represent the linear regressions. For each independent dependency the r correlation coefficient is shown; *asterisks* represent the level of significance (p-values < 0.01**). *Error bars* = standard error (se).

RDEA119 is a highly selective allosteric inhibitor of MEK1/2 enzymes and has been shown to act in xenograft models of melanoma, colon and epidermal carcinomas, and pancreatic cancers (Iverson et al. 2009). Our results show a positive dependency between the cellular tolerance to RDEA119 and energy metabolism (Fig. 25C). Cell lines like the colorectal carcinoma HT-29 and the skin melanoma MDA-MB-435 with low ATP levels are the most sensitive to the mentioned drug (Fig. 25C).

Methotrexate is the most used drug in the treatment of rheumatoid arthritis (Phillips et al. 2003, Abolmaali et al. 2013, Bianchi et al. 2016) and was approved by the FDA in 1988 (Abolmaali et al. 2013). It is also used for the treatment of other diseases like psoriasis, multiple sclerosis, and the Crohn's disease. It successfully has been used for the treatment of breast cancer, acute lymphatic leukemia, osteogenic sarcoma, choriocarcinoma, lung cancer, bladder carcinoma, brain medulloblastoma, primary CNS lymphoma, and chronic myeloid leukemia (Abolmaali et al. 2013). Methotrexate is a folic acid analog that acts as an inhibitor of dihydrofolate reductase, an enzyme involved in purine and pyrimidine nucleotide synthesis. DNA replication and repairing are affected along with RNA synthesis resulting in an antiproliferative action that leads to cell death (Phillips et al. 2003, Abolmaali et al. 2013, Thomas et al. 2015). In our analysis methotrexate displayed a highly significant and strong positive correlation with ROS accumulation in cancer cells (Fig. 26). The cell lines that accumulate more ROS are the ones more resistant to the drug, like the breast MDA-MB-231 and MCF7 and the prostate LNCaP and PC-3 (Fig. 26).

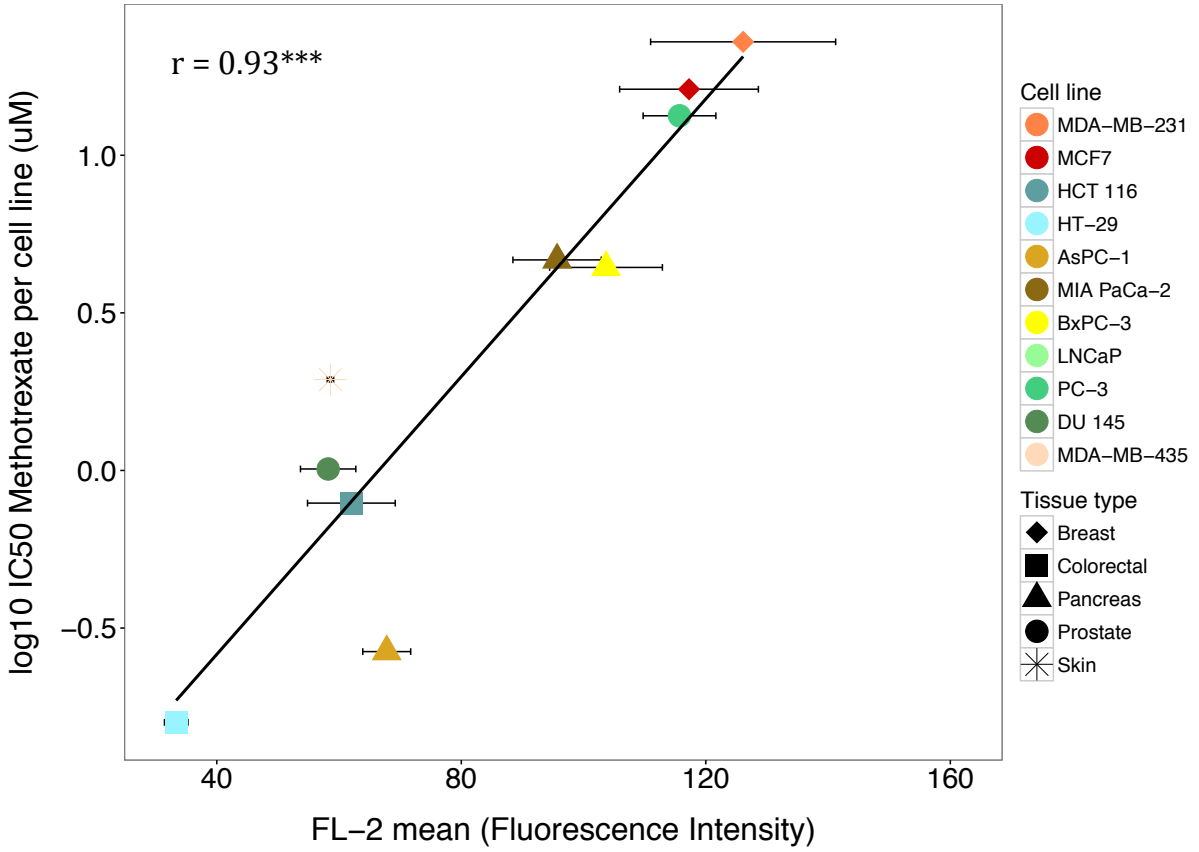


Fig. 26. Methotrexate – significant dependencies between ROS accumulation and the log10 IC_{50} of methotrexate. ROS accumulation in cells was measured using dihydroethidium (DHE). The mean value of at least three independent experiments per cell line was used for the Pearson correlation. IC_{50} values were obtained from the Genomics of Drug Sensitivity in Cancer project. The Pearson calculations were done using *R programming language* package *Hmisc* function rcorr. The shape of the *dots* gives the tissue types and the color the cell lines. The *blue lines* represent the linear regressions. For each independent dependency the r correlation coefficient is shown; *asterisks* represent the level of significance (p-values < 0.001***). *Error bars* = standard error (se).

3.5.2 Drug Sensitivity (IC_{50}) of the Cell Lines

The IC_{50} values of almost 100 drugs were collected from the Genomics of Drug Sensitivity in Cancer project (GDSC) (Yang et al. 2013) (Fig. 22 and Experimental Methods 2.4.1, Table 7). For seven of these compounds, the IC_{50} values were calculated for our cell lines. These drugs are 17-AAG, CCT007093, LFM-A13, PAC-1, IPA-3, AKT inhibitor VIII, and BAY 61-3606. The inhibition of cell proliferation was monitored by means of the SRB assay (Experimental Methods 2.1.3.4) in the human cell lines MCF7, MDA-MB-231, LNCaP, PC-3, DU 145, HT-29, HCT 116, ASPC-1, MIA PaCa-2, and BxPC-3. The serial dilution of each drug used for determining the IC_{50} was selected based on the

IC_{50} values obtained from the GDSC (Yang et al. 2013) (Experimental Methods 2.4.2, Table 8). The obtained absorbencies were fitted to sigmoid and linear regressions and the best fit was chosen (r coefficient near to –1). At least three independent measurements, with three technical replicates each, were done per drug and cell line. Fittings and IC_{50} calculations were performed using the Excel add-in ED50plus v1.0 developed by Mario H. Vargas (2000).

The results obtained from the GDSC project for the IC_{50} values of drugs (Yang et al. 2013) (Fig. 22 and Experimental Methods 2.4.1, Table 7) differ in most of the cases from the estimations done by us. Furthermore, it is possible to observe one order of magnitude difference between the two calculations for CCT007093 in MCF7, PAC1 in MDA-MB-231, AKT1/2 kinase inhibitor in HT-29, for BAY61-3606 in MCF7 and MIA PaCa-2, and for IPA3 in HCT 116 and HT-29 (Table 17). Similarly, the cases where the IC_{50} from the database and our estimations differed in less than 25% were the ones calculated for CCT007093 in HCT 116, HT-29 and MDA-MB-435, LFM-A13 in AsPC-1, PAC1 in AsPC-1, and BAY61-3606 in HCT 116 (Table 17)

Table 17. Database drug IC_{50}s compared with own estimations in our cell lines. Drug IC_{50} values were obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) project. From the total of drugs in the database, seven were used for our IC_{50} estimations. The IC_{50} calculations were done based on the inhibition of cell proliferation in ten cell lines. Cell proliferation was monitored by means of the SRB assay. Cells were incubated for 72 h with a serial dilution (up to eight concentrations) of each drug (Experimental Methods, Table 8). At least three independent measurements with three internal replicates were done per drug and cell line. The fit was adjusted to sigmoid and linear regressions. Based on the r coefficients, the best fit was selected. IC_{50} GDSC = IC_{50} Genomics of Drug Sensitivity in Cancer project; IC_{50} = our IC_{50} estimation; SD = standard deviation of our IC_{50} estimations; fit r coefficient = r coefficient between the estimation done by the data and the sigmoid or linear regressions to the fit model.

Drug	Cell line	IC_{50} GDSC (µM)	IC_{50} (µM)	SD (µM)	fit r coefficient
17-AAG	MDA-MB-231	45.12	6.14	1.04	-0.78
	LNCaP	0.80	1.29	0.53	-0.95
CCT007093	MCF7	3.86	158	51.03	-0.80
	HCT 116	434.3	435.3	227.5	-0.46
	HT-29	238	187.4	51.54	-0.55
	MIA PaCa-2	687.9	258.3	64.57	-0.61
	LNCaP	874.1	311.5	70.41	-0.58
	DU 145	759.1	406.1	46.31	-0.64
	MDA-MB-435	479.5	415.2	73.93	-0.72
LFM-A13	MDA-MB-231	81.9	348.9	34.63	-0.94
	MCF7	464.4	196.5	20.06	-0.95
	HCT 116	703.1	189.4	20.40	-0.96

	HT-29	865.8	189.3	22.21	-0.86
	AsPC-1	344.6	293.1	25.34	-0.94
	MIA PaCa-2	600.7	212.6	13.07	-0.94
	LNCaP	178.5	240.3	43.59	-0.91
	PC-3	583.9	219.2	34.46	-0.96
	DU145	445.2	231.2	18.59	-0.98
PAC1	MDA-MB-231	236.5	4.01	1.42	-0.81
	AsPC-1	35.14	43.67	25.21	-0.59
	MIA PaCa-2	191.2	106.3	23.56	-0.65
	MDA-MB-435	160.1	105.1	34.70	-0.68
AKT1/2 kinase inhibitor	MDA-MB-231	103.8	24.80	1.01	-0.92
	HCT 116	7.76	33.92	3.90	-0.82
	HT-29	2.43	44.96	12.50	-0.89
	MDA-MB-435	105	35.27	3.29	-0.84
BAY61-3606	MCF7	1.15	11.76	1.52	-0.88
	HCT 116	3.98	3.06	1.38	-0.94
	MIA PaCa-2	5.40	368	55.39	-0.65
	LNCaP	2.18	12.07	3.12	-0.74
	DU 145	7.97	26.18	6.11	-0.72
IPA3	HCT 116	1,254	37.58	10.38	-0.58
	HT-29	2,562	50.60	9.70	-0.59
	LNCap	47.38	77.12	13.21	-0.64

4. Discussion

4.1 Cellular Properties

The cellular features of 18 human cell lines were studied. The electrical (cellular capacitance) and metabolic properties (glycolysis and respiration activities, energy metabolism, mitochondrial mass content, levels of intracellular ROS and proliferation rates) were used to characterize the cell lines based on their tissue type source.

Although the histological subtype of each cell line was not in focus in this work, we could observe a difference between the colorectal carcinoma cells HCT 116 and HT-29 and their counterpart pair the adenocarcinoma LS174T for the cellular capacitance, glycolitic activity, and the mitochondrial mass content. The two carcinoma cells show stronger cell-matrix adhesion and cell-cell contact binding (low capacitance) and elevated glycolysis and mitochondrial mass content in comparison with the adenocarcinoma (Results, Fig. 10). Evidence points to differences among the molecular profile and metabolism between the histological subtype of tumors in some lung and cervix cancers (Meijer et al. 2012, Wright et al. 2013). In lung cancer, the carcinoma subtype displays a higher expression of glucose transporter 1 (GLUT1) in comparison with the adenocarcinoma subtype (Meijer et al. 2012). In colorectal cancer cell lines GLUT1 was found to be upregulated in cells with KRAS or BRAF mutations (Yun et al. 2009). The mutated cell lines present also higher glucose incorporation and higher glycolysis levels than the wild-type cells (Yun et al. 2009). In patients with cervix cancer, KRAS mutations were found only in adenocarcinomas subtype tumors and not in other carcinoma types (Wright et al. 2013). The limited number of colorectal cancer cell lines studied in this work impedes to draw a conclusion. However, the observation that carcinoma cells behave differently in glycolysis, and the varying amount of active mitochondria and different levels of capacitance in comparison with adenocarcinoma cells, may serve as a lead for attempting to improve colorectal cancer treatment.

On the other hand, cluster dendograms localize in the same branch or in closer ones the colorectal cell lines HCT116 and the LS174T for respiration activity, energy metabolism, and ROS formation (Results, Fig. 11). Both cell lines share the presence of KRAS mutations (Little et al. 2011, Lee et al. 2011) differing from HT-29 that carries a B-RAF^{V600E} mutation (Makrodouli et al. 2011).

B-RAF protein kinases regulate the MAP kinase/ERKs signaling pathway affecting cell division, differentiation, and secretion. Mutations in this pathway are associated with various cancers, inducing cellular proliferation (Cicenas et al. 2017). On the other hand, RAS genes are a family of small GTPase in which a mutation causes different malignancies, including lung, pancreatic, and colorectal carcinomas. Oncogenic forms of the RAS genes are prevalent in pancreatic carcinomas (>80%), colon carcinomas (40–50%), and lung carcinomas (30–50%) but are rare in human breast cancers (Jančík et al. 2010, Niemitz 2013). RAS, as B-RAF mutations, result in an increase in cellular proliferation (Cicenas et al. 2017).

Comparing the cell lines belonging to the colorectal tissue type with the cell lines belonging to others tissue types as studied in this work (breast, pancreas, prostate, and skin) reveals that the colorectal cells present the highest proliferation rate and the lowest energy metabolism and ROS accumulation (Results, Fig. 10). Furthermore, breast cancer cell lines, normally wt RAS (Niemitz 2013), show an inversed behavior with higher ATP and ROS levels and a low proliferation rate (Results, Fig. 10). Between the two breast cancer cell lines used in this study, MCF7 does not present mutations, neither in KRAS nor in B-RAF, in contrast to MDA-MB-231 that has mutations in both oncogenes (Forbes et al. 2016). The calculated Pearson correlation coefficient for energy metabolism, ROS accumulation, and proliferation rates corroborate these findings, indicating that cells with high ATP levels also have higher ROS levels ($r = 0.64^{**}$) (Results, Fig. 12E). Also, an inverse correlation is found between ROS and proliferation ($r = -0.68^{*}$) and ATP with proliferation ($r = -0.64^{*}$) (Results, Fig. 12C, D, and F).

Pearson correlation shows the glycolytic activity of the cells and the proliferation rate as a strongly positive related ($r = 0.71^{**}$) (Results, Fig. 12B). These correlations could indicate that cells that have more active glycolysis and proliferate more have less pyruvate to enter into the mitochondria for the TCA cycle and therefore less oxygen is consumed and less ROS is produced. Furthermore, glycolysis produces less ATP and a moderate nonsignificant negative relation is found between these two variables (Results 3.3, Table 10).

It has been shown that glycolysis contributes to the generation of biomass in providing the necessary building blocks for new cells; therefore, an increase in glycolysis is observed in proliferating cells (Lunt & Vander Heiden 2011). A positive correlation seems exist between glycolysis and cell proliferation. As was discussed in the

Introduction, some enzymes belonging to the glycolysis pathway display non-glycolytic functions crucial for cell proliferation (Lincet & Icard 2015).

Pearson correlation was applied in a study that could show a strong positive relation between cell proliferation and the glycolytic efficiency in cancer cells ($r = 0.82$, p-value = 0.04) (De Preter et al. 2016). In the same study, nonsignificant correlations were found between glycolysis and ATP levels and also not for the oxygen consumption with ATP nor with proliferation rate (De Preter et al. 2016).

4.2 Gene Expression and Metabolism

Data of gene expression from the public repository GEO, NCBI obtained using Affymetrix Human Genome U133 Plus 2.0 Array platform (De Schutter et al. 2013 and Barretina et al. 2012. Experimental Methods 2.3, Table 2) were used to search for linear relationships with the cellular properties of cancer cell lines. Pearson product moment correlations show that the expression levels of eight genes are correlated with the metabolism features of cancer cells. These eight genes provide consistent information through their corresponding probe sets. The metabolism-related candidate genes encode the following proteins: PGLS (6-phosphogluconolactonase), COX5B (cytochrome *c* oxidase subunit 5b), RHBDL2 (rhomboid-related protein 2), IFI16 (gamma-interferon-inducible protein 16), GCN1L1 (eIF-2-alpha kinase activator GCN1), HOXA7 (Homeobox 7), INPP5B (inositol polyphosphate-5-phosphatase B) and LGALS8 (galectin 8).

The *PGLS* gene is interrogated by four probe sets in the Affymetrix platform used for this study. Among all the candidate genes *PGLS* is the one that displays the highest number of relations with metabolism with all of the probe sets of this gene having inverse congruent correlation with the mitochondrial mass content, the glycolytic activity, and the proliferation rates measured as metabolic activity (MTT assay) and total protein content (SRB assay) (Results 3.4.1.1, Table 12 and Fig. 16). It is not surprising that PGLS appears to be related with glycolysis and cellular proliferation because PGLS is the second enzyme in the pentose phosphate pathway (PPP) and catalyzes the conversion of 6-phosphonoglucono-delta-lactone to 6-phosphogluconate. The PPP produces ribose and NADPH and is a metabolic pathway parallel and directly related to glycolysis (Cha, Jung & Koo 2017). Proliferating and cancer cells show an increase in glycolysis as well as in PPP (Jin et al. 2017, Cha, Jung & Koo 2017). Cancer

cells need pentose phosphate and NADPH for the synthesis of lipids and cell survival under stress, repairing ROS damage produced by the accelerated proliferation (Cha, Jung & Koo 2017, Jin et al. 2017). Lower rates of the PPP in cancer cells through siRNA against the first enzyme in the pathway, glucose-6-phosphate dehydrogenase, leads to a reduction in the proliferation rate of the cells (De Preter et al. 2016). In breast tissue, PGLS protein has been reported to be overexpressed in 98% of the tumors in comparison with normal tissue and thus has been identified as a breast cancer biomarker (Ou et al. 2008). Here, we present that, among cancer cells, the cell lines with high expression levels of PGLS present low glycolysis and low proliferation rates (Results 3.4.1.1, Table 12 and Fig. 16). No literature was found that could indicate a possible relation between PGLS and mitochondrial mass.

COX5B and its two probe sets in the Affymetrix platform show negative correlations with the proliferation assay MTT and that is in agreement with the moderate r coefficient between the two probe sets and the proliferation assay SRB (Results 3.4.1.1, Table 12). The obtained correlations indicate that cell lines with high gene expression of COX5B present low proliferation rates. COX5B is a nuclear encoded protein part of the cytochrome c oxidase complex and it has been shown that depletion of COX5B leads to a diminution of the cytochrome c oxidase activity, suggesting a regulatory activity of this gene (Galati et al. 2009). No much evidence has been shown about a possible relation between COX5B and cancer metabolism. In patients with breast cancer an upregulation of COX5B in comparison with the normal tissue is exhibited (Gao et al. 2015). Contrary to our findings, the same study reported that a downregulation of COX5B in the breast cancer cell lines MDA-MB-231, MDA-MB-468 and MCF-7 leads to an inhibition of cell proliferation. The same group has shown that cells with a COX5B knock-down also exert low migration activity, low ATP formation, and an increase in ROS production (Gao et al. 2015).

RHBDL2 contains three probe sets in the Affymetrix array used, and three of them show a middle to moderate negative correlation with glycolysis activity and mitochondrial mass content (Results 3.4.2, Fig. 17). The Rhomboids are a relatively new discovered family of intramembrane serine proteasess. RHBDL2 is one of the members of this family that has been implicated in wound healing of keratinocytes by the cleavage of thrombomodulin, a protein involved in cell adhesion and migration during wound healing (Etheridge et al. 2013, Cheng et al. 2014). In keratinocytes, an overexpression of RHBDL2 increases cell proliferation, reduces cell adhesion and supports the overcome

of anoikis (Cheng et al. 2014). The other two substrates that have been attributed to RHBDL2 in humans are the EGF ligand and the members of the EphrinB family (Etheridge et al. 2013). All of the three recognized substrates contain an EGF-like domain (Etheridge et al. 2013). RHBDL2 cleaves the EGF ligand inducing the activation of the epithelial growth factor receptor (EGFR) (Cheng et al. 2014). Limited information is available on the function of RHBDL2 and its relation to cancer is largely unknown. Here we show that cancer cells with high expression of RHBDL2 present low glycolysis activity and low mitochondrial mass suggesting a possible involvement of this protein in cancer metabolism (Results 3.4.2, Fig. 17).

The two probe sets of the IFI16 gene presented a negative correlation with the respiration activity at almost 0.6 (Results 3.4.2, Fig. 18). No information about a relation between cellular oxygen consumption and the IFI16 gene have been found. IFI16 is a transcriptional regulator that has a function in the inflammatory process (Shi et al. 2014, Ansari et al. 2015), in immune response (Ansari et al. 2015), in autophagy and cell survival under glucose depletion (Duan et al. 2011), and in cellular senescence associated with cell growth arrest (Alimirah et al. 2007). Its function in the inflammatory process and in immune responses are widely studied (Johnson et al. 2014, Ansari et al. 2015). IFI16 is a pathogen DNA sensor that acts by inducing the production of inflammatory cytokines (Ansari et al. 2015). The tumor suppressor p53 is one of the main regulators of IFI16 activity (Shi et al. 2014). In human diploid fibroblasts it was shown that the IFI16 protein is necessary for the activation of the ATM/AMPK/p53 pathway for autophagy upon glucose restriction and in low glucose microenviroments the expression of IFI16 is increased (Duan et al. 2011). Noncancerous prostate cells with an elevated expression of IFI16 protein show cellular senescence and cell growth arrest (Alimirah et al. 2007). In cancer prostate cell lines, IFI16 protein is not present or its amount is very low (Alimirah et al. 2007) being an advantage for the cancer cells exposed to an energetic stress (Duan et al. 2011). An overexpression of IFI16 in the prostate cancer cell lines LNCaP and DU 145 leads to an inhibition of proliferation (Alimirah et al. 2007). In contrast to the prostate cells, in liver cancer cells the IFI16 protein is expressed, but not in healthy adults hepatocytes (Shi et al. 2014). In our study the prostate cancer cell lines DU 145 and PC-3 showed a mid-to-high expression of the IFI16 gene in comparison with the other cell lines studied. On the other side, the prostate LNCaP and liver HepG2 showed low gene expressions (Results 3.4.2, Fig. 18). It is known that the expression of the IFI16 gene is well regulated at the transcriptional

and posttranscriptional levels (Shi et al. 2014) suggesting that the protein levels of IFI16 is not necessarily reflected by the gene expression levels.

The GCN1L1 gene with one of its two probe sets shows a significant, highly positive correlation with the proliferation rate measured by total protein content of the SRB assay ($r = 0.91$ and p-values < 0.0001) (Results 3.4.3, Fig. 19A). The second probe set that interrogates GCN1L1 also shows a significant dependency with SRB. The relation with the proliferation is corroborated by both probe sets that present also a significant relation with the MTT assay (Results 3.4.3, Table 15). It is well known that GCN1L1 in Saccharomyces cerevisiae growing in amino acid reduced media is involved in the activation of the eukaryotic initiation factor 2α that represses the synthesis of the GCN4 transcription factor which induces the expression of the genes necessary for amino acid biosynthesis (Kubota et al. 2001). Years later, it was found out in humans that GCN1L1 is one of the factors that is associated with the CDK8 complex (Knuesel et al. 2009). CDK8 is an oncogene that presents a kinase activity necessary for the regulation of β-catenin activity (Firestein et al. 2008). Colon cancer cells present high levels of CDK8 and Wnt/β-catenin hyperactivity (Firestein et al. 2008). In renal carcinogenesis, mutations were identified in members of the CDK8 complex including GCN1L1 that could participate in the dysregulation of the Wnt/β-catenin pathway inducing cancer transformation (Arai et al. 2014). It has also been well documented that hyperactivity of the Wnt/β-catenin pathway induces cell proliferation (MacDonald, Tamai & He 2009). We found that the gene expression of GCN1L1 is directly related to the proliferation of cancer cells and the cell lines that express more GCN1L1 show a high proliferation rate.

Both proliferation assays (MTT and SRB) show significantly positive dependencies on HOXA7 for its two probe sets. The expression values of the two probe sets that interrogate HOXA7 showed a strong positive relation with the proliferation assay measured as metabolic activity MTT ($r = 0.95$ and 0.90, p-values < 0.0001) (Results 3.4.3, Tables 15 and 16), and is corroborated with a significant (p-values < 0.01 data not shown) positive relation between the two probe sets and the proliferation assay SRB (Results 3.4.3, Table 16). For this gene it is also possible to observe a significant negative dependency between the two probe sets and ROS accumulation in cells (p-values < 0.01, data not shown). The proliferation rate SRB and MTT assays present a significant negative correlation ($r = -0.68$) (Results 3.3, Table 10) with ROS accumulation in cells.

HOX genes are transcriptional factors involved in morphogenesis, organogenesis, differentiation, cell migration, and proliferation (Zhang et al. 2013, Li et al. 2015). Several studies have shown a direct function of HOXA7 in the positive regulation of cellular proliferation in cancer cells. Knockdown of HOXA7 in human precursor B-cell leukemia MCF7, HepG2, and QGY-7703 cells reduces cell proliferation (Orlovsky et al. 2011, Zhang et al. 2013, Li et al. 2015). Our findings corroborate the relation of HOXA7 and proliferation indicating that cells with high proliferation rates also have a higher expression of the HOXA7 gene giving a novel insight on the function of this transcriptional factor and ROS accumulation in cancer cells.

The expression of the INPP5B gene presents a highly positive dependency with the SRB assay ($r = 0.92$, p-value < 0.0001) that is corroborated by the three probe sets that contain information on the gene, all of them having a significantly positive relation also with the proliferation assay MTT (p-value < 0.01 data not shown). A deficiency in the complementation of Ocrl's functionality by INPP5B in humans, and not in mouse, is responsible for the development of Lowe syndrome, an oculocerebrorenal X-linked genetic disorder (Bothwell et al. 2010). There is no evidence that links INPP5B with cancer. We show that cancer cell lines with low expression levels of the INPP5B gene also present low proliferation rates, being the first time that this gene is implicated in cancer metabolism.

LGALS8 shows a high negative correlation with the cellular mitochondrial mass content ($r = -0.84$, p-value < 0.0001) and also with the MTT proliferation assay ($r = -0.73$ p-value < 0.01) (Results 3.4.3, Table 16 and Fig. 19D). The LGALS8 gene contains five probe sets more and in all of them a significant (p-value < 0.01, data not shown) negative dependency with the mitochondrial mass and the MTT proliferation assay are observable (Results 3.4.3, Table 16). The dependencies found between the LGALS8 probe sets and the SRB assay display the same tendency, a negative correlation (Results 3.4.3, Table 16). LGALS8 is part of the galectin family of secreted proteins that bind glycoproteins of the cell surface and the extracellular matrix. Galectins are involved in cell-cell and cell-matrix interactions, cell-cell communication, migration, proliferation, apoptosis, and angiogenesis (Delgado et al. 2011, Metz et al. 2016). In glioblastoma U87 cells a reduction of LGALS8 by shRNA leads to a decrease in the proliferation rate and induces apoptosis (Metz et al. 2016). On the other hand, in colorectal cancer cell lines, LGALS8 was inversely correlated to tumor growth and migration rates (Nagy et al. 2002). In our study, the data obtained with the Pearson correlation coefficients indicates

that the cell lines that express a higher level of the LGALS8 gene are the ones that present the lowest proliferation rate. No evidence was found that relate LGALS8 activity and the mitochondrial mass content of the cells.

The estimation of changes in gene expression levels in many cases assumes that these changes will be propagated to protein expression and several studies take as a fact that the gene expression is informative in predicting protein levels (Kern et al. 2003). Regrettably, incongruences have been found in studies that calculate the correlation between gene expression and protein levels (Guo et al. 2008, Maier, Güell and Serrano 2009). In patients with acute myeloma leukemia, congruent results were obtained in correlating the mRNA of 39 specific genes and their protein abundances in samples from more than 100 patients (Kern et al. 2003). However, mRNA expression of freshly isolated human monocytes correlated with their protein expression levels shows huge variations in the different investigated genes and between different individuals. That part of the studied genes presents a significant positive correlation with protein expression (Maier, Güell & Serrano 2009). Nevertheless, an overall congruence is found indicating a general concordance between the abundance of mRNA and proteins (Maier, Güell & Serrano 2009). This indicates the necessity to corroborate the relation of the candidate genes found and cancer metabolism before drawing any conclusions from our findings.

4.3 Drug Sensitivity and Cellular Properties of Cancer Cell Lines

The IC_{50} values of 99 compounds were obtained from the Genomics of Drug Sensitivity in Cancer project (Yang et al. 2013) (Results 3.5.1, Fig. 22 and Experimental Methods 2.4.1, Table 7) and were used for Pearson correlation calculations with the cellular properties of cancer cell lines to search for linear relations between these two variables. Pearson product moments show that the sensitivity of seven of these drugs present a correlation with some of the cellular features of cancer cells. The compounds that correlate with the cellular properties are: elesclomol, Nutlin 3a, PF 4708671, EHT 1864, IPA3, RDEA119, and methotrexate.

Elesclomol is a copper chelator that transports the copper ions into mitochondria inducing ROS production and, as a consequence, growth arrest and apoptosis in cancer cell lines (Krishner et al. 2008, McCarty and Contreras 2014, Rae and Mairs 2017). A study done in patients with melanoma treated with elesclomol shows that the compound is more effective in the individuals with low lactate dehydrogenase (LDH) levels in the serum (McCarty and Contreras 2014). Patients with elevated LDH show cancer with a Warburg effect phenotype, meaning high glycolysis and less oxidative phosphorylations in mitochondria, which can be the reason for the observed low activity of the drug (McCarty and Contreras 2014). Our findings show that elesclomol is negatively correlated with the cellular capacitance ($r = -0.76$ p-value < 0.01), indicating that the cancer cell lines presenting stronger cell-matrix and cell-cell contacts are more resistant to the compound (Results 3.5.1, Fig. 22).

Nutlin 3a is an indirect activator of wild-type p53. The compound inhibits the interaction between MDM2, an ubiquitin ligase, and negative regulator of p53 and the tumor suppressor p53 (Kojima et al. 2006). p53 is a tumor suppressor involved in genomic stability and in the regulation of the cell cycle, senescence, and apoptosis (Lago et al. 2011). It has also been shown to be involved in the promotion of oxidative phosphorilation stimulating mitochondrial respiration and inhibiting glycolysis and that mutated p53 contributes to the Warburg effect phenotype in cancer (Lago et al. 2011). We show that cancer cells that are more sensitive to Nutlin 3a present a high respiration rate ($r = -0.74$, p-value < 0.01) (Results 3.5.1, Fig. 24A).

PF 4708671 is an inhibitor of the p70 ribosomal S6 kinase 1 (p70S6K1) (Pearce et al. 2010). P70S6K1 is involved in the regulation of protein synthesis, proliferation, growth, survival, motility, longevity, and chemotherapy drug resistance in cancer cells (Pearce et al. 2010, Qiu et al. 2016). Cells treated with PF 4708671 show cell cycle arrest and a reduction in the proliferation rate and invasion (Qiu et al. 2016). We show that cancer cell lines with high oxygen consumption are more sensitive to PF 4708671, indicating that the inhibitor of p70S6K1 has a negative effect on respiration activity ($r = -0.74$, p-value<0.01) (Results 3.5.1, Fig. 24B).

EHT 1864, as discussed in more detail in 3.5.1, is a Rac inhibitor (Onesto et al. 2008). Rac proteins are directly involved in proliferation, invasion, and metastasis in

cancer cells (Onesto et al. 2008). We show that EHT 1864 correlates negatively with energy metabolism ($r = -0.77$, p-value < 0.01) (Results 3.5.1, Fig. 25A).

IPA3 is a small molecule inhibitor of p21-activated protein kinase 1 (PAK1) (Singhal and Kandel 2012). PAKs promote tumor cell proliferation (Liu et al. 2016). In colon carcinoma and melanoma cell lines with mutations either in BRAF or in RAS (NRAS or KRAS) genes, treatment of cells with RAS mutations with IPA3 appears to be more effective (Singhal and Kandel 2012). RAS and specifically KRAS mutant cells show high nucleotide biosynthesis, glycolysis, and PPP activity (Lv et al. 2016). In the mouse model, mutant KRAS increased glucose and glutamine uptake and, on the other hand, cancer cells cultivated under low glucose conditions acquired KRAS mutations (Lv et al. 2016). Here we show that there is a negative correlation between energy metabolism in cells and the IC_{50} of IPA3 meaning that the cell lines that produce more ATP are more sensitive to this compound ($r = -0.77$, p-value < 0.01) (Results 3.5.1, Fig. 25B). This data provides further insight into treatment strategies of cancers with BRAF or RAS mutations using IPA3 and targeting metabolism.

RDEA119 is a highly selective allosteric inhibitor of MEK1/2 enzymes (Iverson et al. 2009). Crystallization of the MEK-RDEA119 complex indicates that the drug binds on the adjacent side of the Mg-ATP binding region interacting with ATP, preventing the binding and phosphorilation of the MEK substrate ERK (Iverson et al. 2009). MEKs are downstream of BRAF in the RAS-RAF-MEK-ERK pathway implicated in cellular proliferation (Iverson et al. 2009). RDEA119 that has shown to act in xenograft models of melanoma, colon and epidermal carcinomas, and pancreatic cancers (Iverson et al. 2009) is now in phase II clinical studies (Caunt et al. 2015). Our results show a highly positive relation between the cellular tolerance to RDEA119 and energy metabolism ($r = 0.84$, p-value < 0.01) (Results 3.5.1, Fig. 25C).

Methotrexate, as elaborated in Sect. 3.5.1, is an approved drug for use in several conditions including breast cancer, acute lymphatic leukemia, osteogenic sarcoma, choriocarcinoma, lung cancer, bladder carcinoma, brain medulloblastoma, primary central nervous system lymphoma, and chronic myeloid leukemia (Abolmaali et al. 2013) and is the most used drug in the treatment of rheumatoid arthritis (Phillips et al. 2003, Abolmaali et al. 2013, Bianchi et al. 2016). Methotrexate has been widely studied

in the context of its synthesis as an analog of folic acid. It is an inhibitor of dihydrofolate reductase, an enzyme that is necessary for purine and pyrimidine syntheses. Due to its effect of disrupting DNA synthesis, methotrexate has an antiproliferative effect (Phillips et al. 2003, Abolmaali et al. 2013, Thomas et al. 2015). Here we show that the IC_{50} of cells to methotrexate displays a highly significant and strongly positive correlation with ROS accumulation in cancer cells ($r = 0.93$, p-value < 0.001) (Results 3.5.1, Fig. 26).

In conclusion, the detailed metabolic and electrochemical characterization of several cancer cell lines presented in this work demonstrates a wide variability regarding metabolic activity and morphological properties of cancer cells. The data obtained provide an overview of various growth parameters including proliferation rates, energy metabolism, measured as glycolytic and respiration rates, formation of reactive oxygen species (ROS), mitochondrial mass, as well as cell-cell and cell-matrix interactions, which can be taken as a cell-specific footprint of the cancer cells analyzed. With the large data sets obtained for these complementary metabolic and cell growth parameters it became possible to calculate correlations between the analyzed parameters and further available gene expression and drug sensitivity data sets in a detailed computational analysis. Using a robust statistical analysis, based on Pearson correlation coefficient calculation, several significant correlations could be identified. The results support among others the relevance of HOXA7 as a potential drug target and link the efficacy of known and potential anticancer drugs with metabolic properties.

The results obtained confirm in particular the efficacy of drugs aimed at the control of cell metabolism and provides hints to specific metabolic markers and metabolism-associated expressed genes indicating that metabolic parameters of cancer cells could provide valuable additional information to optimize drug selection and adjust patient-specific therapy by combining drugs targeting both metabolism and proliferation.

5. Supplements

5.1 R Scripts

The data collection, calculations, statistics and plotting for this study were done using RStudio versions 0.98.1103 and 0.99.473. RStudio is an open-source integrated development environment used as an interface for R *programming language* for statistical computing and graphics (Team, R. Core. 2015). The R versions 2.7.0 and 3.2.1 were used.

5.1.1 Bionas Data Collection

```
path <- "/Users/admin/Documents/PhD/basal levels Results/human1"


## Get all the ANALYZER folders from the different cell line


folders1 <- list.files(path=path)

folders2 <- list.dirs(folders1)

file <- list.files(folders2, full.names=TRUE, pattern="*ANALYZER_DATA.dat",include.dirs =
TRUE)

write.csv(file, file="path.csv", row.names=TRUE)

out.file<-""

for(i in 1:length(file)){
  read <- readLines(file[i])
  out.file <- rbind(out.file, read)
}
write.csv(out.file, file="raw data all cell lines.csv", row.names=TRUE)

class(out.file) #matrix

read2<- out.file[,c(11,42,72)] # read just ides, o2 and isfet. Columns 11, 42 y 72
colnames(read2) <- c("IDES DATA in norming distance per BM (nF)", "ELECTRODES DATA in
norming distance per BM (pA/s)", "FET DATA in norming distance per BM (uV/s)")

write.csv(read2, file="raw data IDES,O2 and FET all cell lines.csv", row.names=TRUE)

### IDES

IDES <- read2[,1] #to have just the column with IDES data
View(IDES)
class(IDES) #character
```

```
IDES1 <- as.numeric(unlist(strsplit(gsub(",", ".", IDES), "\t"))) # write dots and splitt the experiments with comas

View(IDES1)
class(IDES1) #numeric

IDES2 <- matrix(IDES1, ncol=6, byrow=TRUE) # make a matrix with the 6 columns, one per each biomodul

write.csv(IDES2, file="raw data IDES all cell lines.csv", row.names=TRUE)

#### O2

O2 <- read2[,2] #just the column with O2 data
View(O2)

O2a <- as.numeric(unlist(strsplit(gsub(",", ".", O2), "\t")))
View(O2a)

O2b <- matrix(O2a, ncol=6, byrow=TRUE)

write.csv(O2b, file="raw data O2 all cell lines.csv", row.names=TRUE)

#### FET

FET <- read2[,3] #just the column with ISFET data
View(FET)

FET1 <- as.numeric(unlist(strsplit(gsub(",", ".", FET), "\t")))
View(FET1)

FET2 <- matrix(FET1, ncol=6, byrow=TRUE)

write.csv(FET2, file="raw data FET all cell lines.csv", row.names=TRUE)
```

5.1.2 Robust Multi-Array Average (RMA)

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

biocLite("affy") #The package contains functions for exploratory oligonucleotide array analysis
library(affy)

biocLite("pd.ht.hg.u133.plus.pm") #Annotation package for pd.ht.hg.u133.plus.pm built with pdInfoBuilder.
Data<-ReadAffy()

#Robust Multi-array Average preprocessing methodology. This strategy allows background subtraction, quantile normalization and summarization (via median-polish). The generated data are stored as ExpressionSet class in the 'eset' object

eset<-rma(Data).
```

```
write.exprs(eset, file="Data.txt", sep="\t") # Writes expression values to text file in working directory.

summarized <- rma(eset)
show(summarized)

HeLa<-rowMeans(hela, na.rm=TRUE)
View(HeLa)

write.table(HeLa, file="HeLamean.txt",sep = " ",row.names = FALSE, col.names = "Mean")

HaCaT<-rowMeans(Hacat, na.rm=TRUE)
View(HaCaT)

write.table(HaCaT, file="HaCaTmean.txt",sep = " ",row.names = FALSE, col.names = "Mean")
```

5.1.3 Pathways Gene Selection

```
ggenes <- HGU133plusna34plus16celllines # file with all probes sets and their gene symbol
colnames(ggenes)[3]<-"1321N1" #to have the column for 1321N1 without X (X1321N1)
View(ggenes)

### Genes for the Glycolysis/Gluconeogenesys, TCA cycle and Pentose phosphate pathways were downloaded from The Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7).

# SELECTION OF PROBES AGAINS GLYCOLYSIS GENES
gly.sub <- ggenes[ggenes$Gene.Symbol %in% c("ALDOA","ALDOB", "ALDOC", "DLAT", "DLD",
"ENO1", "ENO2", "ENO3", "FBP1", "G6PC", "GAPDHL6", "GCK", "GOT1", "GOT2", "GPI", "HK1",
"HK2", "HK3", "LDHB", "LDHC", "MDH1", "MDH2", "PC", "PCK1", "PDHA1", "PDHA2", "PDHB",
"PFKL", "PFKM", "PFKP","PGAM2","PGK1", "PGK2", "PKLR", "PKM2", "SLC2A1", "SLC2A2",
"SLC2A3","SLC2A4","SLC2A5","TPI1","PDHX","FBP2","MPC2","GAPDHS","MPC1","HKDC1","ADP
GK","LDHAL6B"), ]
gly.sub
write.csv(gly.sub, file="glycolysis genes.csv") #, sep = " ",row.names = TRUE, col.names = TRUE)

# SELECTION OF PROBES AGAINS TCA GENES
tca.sub <- ggenes[ggenes$Gene.Symbol %in%
c("AC02","SLC25A4","SLC25A5","SLC25A6","ATP5A1","ATP5B","ATP5C1","ATP5D","ATP5E","AT
P5F1","ATP5G1","ATP5G2","ATP5G3","ATP5I","ATP5J","ATP6AP1","ATP5O","COX4I1","COX5B","
COX6A1","COX6A2","COX6B1","COX6C","COX7A1","COX7A2","COX7B","COX7C","COX8A","COX11
","COX15","CS","DLD", "DLST","FH","IDH2","IDH3A","IDH3B","IDH3G","MDH2","COX3","ATP6",
"COX1","COX2","ATP6","CYTB","ND1","ND2","ND3","ND4","ND5","ND6","NDUFA1","NDUFA2",
"NDUFA3", "NDUFA4","NDUFA5","NDUFA6", "NDUFA7","NDUFA8","NDUFA9","NDUFA10",
"NDUFAB1","NDUFB1","NDUFB2","NDUFB3","NDUFB4","NDUFB5","NDUFB6","NDUFB7","NDUF
B8","NDUFB9","NDUFB10","NDUFC1","NDUFC2","NDUFS1","NDUFS2","NDUFS3","NDUFV1","ND
UFS4","NDUFS5","NDUFS6","NDUFS8","NDUFV2","NDUFV3","OGDH","SCO1","SDHA","SDHB","SD
HC","SDHD","SURF1","UCP1","UCP2","UCP3", "UQCRB","UQCRC1","UQCRC2","UQCRFSL1",
"UQCRHL","SUCLG2","SUCLG1","SLC25A14","COX7A2L","COX5A",  "SLC25A27",
"ATP5J2","COX17","ATP6AP2","ATP5H","ATP5L","UQCR11","UQCRQ","ATP5S","UQCR10","NDUF
A12",  "NDUFA4L2",  "ATPIF1","NDUFA11","NDUFS7"), ]
```

```
write.csv(tca.sub, file="TCA genes.csv")
tca.sub

# SELECTION OF PROBES AGAINS PENTOSE PHOSPHATE PATHWAY GENES
pent.sub <- ggenes[ggenes$Gene.Symbol %in% c("TALDO1", "RPE", "G6PD", "RPIA", "TKT",
"PGLS", "PGD"), ]
write.csv(pent.sub, file="PPhosp genes.csv")
phosgenes <- as.character(pent.sub$Probe.Set.ID)
View(phosgenes)
```

5.1.4 Present and Absence Calls

```
#Presence-Absence Calls on AffyMetrix HG-U133 Series Microarrays with panp package
biocLite("panp")
library(panp)

pa.calls() #run a function with no arguments to obtain a summary of usage information

#default p-value cutoffs of 0.01 and 0.02. So in this case, intensities above the intensity at the
0.01 cutoff will be called ”P” (present); intensities between the two cutoffs will be assigned an
”M” (marginal), and those below the intensity at the 0.02 p-value will get an ”A”(absent)

PA <- pa.calls(eset)
PAcalls <- PA$Pcalls #The presence/absence calls and p-values are returned as two matrices,

Pvalues <- PA$Pvals
write.table(PAcalls, file="PAcalls_rma.csv", sep=",", col.names=NA)
write.table(Pvalues, file="Pvalues_rma.csv", sep=",", col.names=NA)

#### present and absence calls for the probes belong to the pathway genes

tcagenes <- as.character(tcagenesnames[ ,1])
View(tcagenes)
tca.sub <- Pvalues_rma[Pvalues_rma$X %in% tcagenes, ]
write.csv(tca.sub, file="tcasubPvalues.csv")

glygenes <- as.character(glygenesnames[ ,1])
View(glygenes)
gly.sub <- Pvalues_rma[Pvalues_rma$X %in% glygenes, ]
write.csv(gly.sub, file="glysubPvalues.csv")

pentgenes <- Pvalues_rma[Pvalues_rma$X %in% phosgenes, ]
write.csv(pentsubPvalues, file="pentsubPvalues.csv")

glygenes <- as.character(glygenesnames[ ,1])
View(glygenes)
gly.sub <- PAcalls_rma[PAcalls_rma$X %in% glygenes, ]
write.csv(gly.sub, file="glysubPAcalls.csv")

tcagenes <- as.character(tcagenesnames[ ,1])
View(tcagenes)
tca.sub <- PAcalls_rma[PAcalls_rma$X %in% tcagenes, ]
```

```
write.csv(tca.sub, file="tcasubPAcalls.csv")

pentsubPAcalls <- PAcalls_rma[PAcalls_rma$X %in% phosgenes, ]
write.csv(pentsubPAcalls, file="pentsubPAcalls.csv")
# collection of data from the PP, gly and tca probes names selected

finalglyprobesdata <- as.character(glyprobesnamesselected[ ,2])
View(finalglyprobesdata)
gly.final <- Data[Data$X %in% finalglyprobesdata, ]
write.csv(gly.final, file="finalglyprobesdata.csv")

tca.final <- Data[Data$X %in% tcaprobesnamesselected$Probe.Set.ID, ]
write.csv(tca.final, file="finalTCAprobesdata.csv")

pp.final <- Data[Data$X %in% pentsubPAcalls$Probe.Set.ID, ]
write.csv(pp.final, file="finalPPprobesdata.csv")

### with A calls
glyprobesselectionwithAvalues <- as.character(glyprobesselectionwithAvalues[ ,2])

gly.final <- Data[Data$X %in% glyprobesselectionwithAvalues, ]
write.csv(gly.final, file="finalglyprobesdatawithAcalls.csv")

finaltcaprobesdata <- as.character(tcaprobesnamesselected[ ,1])
View(finaltcaprobesdata)
tca.final <- Data[Data$X %in% finaltcaprobesdata, ]
write.csv(tca.final, file="finaltcaprobesdata.csv")

pentfinal <- as.character(PPprobesselected[ ,1])
pentfinal <- Data[Data$X %in% pentfinal, ]
write.csv(pentfinal, file="pentfinal.csv")
```

5.1.5 Selection of the Highly and Lowest Express Probes

######## Scaled and center

```
#TCA + electron transport chain

datatca <- TCAprobesforscale
rnames <- datatca[,1]

# assign labels in column 1 to "rnames"
mat_datatca <- data.matrix(datatca[,2:ncol(datatca)]) # transform from column 2 into a matrix
rownames(mat_datatca) <- rnames

colnames(mat_datatca)[1]<-"1321N1"

scaledtca<-biScale(mat_datatca, row.center=TRUE, row.scale=TRUE, col.center=FALSE,
col.scale=FALSE, trace=TRUE)
View(scaledtca)

write.csv(scaledtca, file="scaledtca.csv", row.names=TRUE)
```

```
#Glycolysis

datag <- glyandpphosforscale
rnames <- datag[,1]

# assign labels in column 1 to "rnames"
mat_datag <- data.matrix(datag[,2:ncol(datag)]) # transform column 2-5 into a matrix
rownames(mat_datag) <- rnames

colnames(mat_datag)[1]<-"1321N1"

scaledgly<-biScale(mat_datag, row.center=TRUE, row.scale=TRUE, col.center=FALSE,
col.scale=FALSE, trace=TRUE)

write.csv(scaledgly, file="scaledgly.csv", row.names=TRUE)

#### PENTOSE PHOSPHATE ####

datapent <- penttoscale
rnames <- datapent[,1]

# assign labels in column 1 to "rnames"
mat_datapent <- data.matrix(datapent[,2:ncol(datapent)]) # transform column 2-n into a matrix
rownames(mat_datapent) <- rnames

colnames(mat_datapent)[1]<-"1321N1"

scaledpent<-biScale(mat_datapent, row.center=TRUE, row.scale=TRUE, col.center=FALSE,
col.scale=FALSE, trace=TRUE)

write.csv(scaledpent, file="scaledpent.csv", row.names=TRUE)

# USED FOR THE ANALYSIS OF THE 16 CELL LINES with gene expression data

### Krebs cycle pathway

select.range<-function (scaledtca, min, max, data)
{
 if (nargs() > 3){
   min.cond <- scaledtca > min
    max.cond <- scaledtca < max

   cond <- min.cond & max.cond

 }
 else cat("Usage: select.range(scaledtca,min,max,datavec)\n")
}
minmaxtca<-select.range(scaledtca,-1.0,1.0, data)

write.csv(minmaxtca, file="MinMaxTCA1.0.csv", row.names=TRUE)

### glycolysis/gluconeogenesys pathway

select.range<-function (scaledgly, min, max, data)
```

```
{
 if (nargs() > 3){
   min.cond <- scaledgly > min
   max.cond <- scaledgly < max

   cond <- min.cond & max.cond
   }
 else cat("Usage: select.range(scaledgly,min,max,datavec)\n")
}
minmaxgly<-select.range(scaledgly,-1.0,1.0, data)
minmaxgly

write.csv(minmaxgly, file="MinMaxglyandPP1.0.csv", row.names=TRUE)

### pentose phosphate pathway

select.range<-function (scaledpent, min, max, data)
{
 if (nargs() > 3){
   min.cond <- scaledpent > min
   max.cond <- scaledpent < max

   cond <- min.cond & max.cond

 }
 else cat("Usage: select.range(scaledpent,min,max,datavec)\n")
}
minmaxpent<-select.range(scaledpent,-1.5,1.5, data)

write.csv(minmaxpent, file="MinMaxPP.csv", row.names=TRUE)
```

5.1.6 Heatmap and Hierarchical Clustering

```
install.packages("pheatmap")
library(pheatmap)
install.packages("softImpute")
library(softImpute)
library("RColorBrewer")
display.brewer.all()

colfunc <- colorRampPalette(c("green", "black", "red"))

pheatmap (mat_datatca,scale="none",col=colfunc(150), border_color = NA,cellwidth = 15,
cellheight = 7,fontsize_row=6,fontsize_col=7, main = "Heatmap of genes involved in \nKrebs
Cycle and Electron Transport Chain", fontsize = 7, filename="tca7.pdf")

pheatmap (mat_datagly,scale="none",col=colfunc(150), border_color = NA,cellwidth = 15,
cellheight = 7,fontsize_row=6,fontsize_col=7, main = "Heatmap of genes involved in Glycolisys,\n
Gluconeogenesis and Pentose Phosphate Pathways", fontsize=7, filename="gly7.pdf")

##### Hierarchical Clustering
# Hierarchical Clustering. Euclidean distance. hclust method complete
```

```
install.packages("plyr")
library(plyr)

install.packages("rafalib") # pretty clusters
library(rafalib)

## Bionas Capacitance

mm <- ddply(labexperimentsdata, "Cell.Line", summarise, Im = mean(Bionas.Impedance,
na.rm=T)) #make a new tablewith the cell lines and the mean value of the impedance
mm # cluster used just one values per condition (mean)

rownames(mm) <- mm[,1]
rownames(mm)

Capacitance <- dist(mm, method = "euclidean") # distance matrix
Capacitance
fit <- hclust(Capacitance)
plot(fit)

colbars1 <-
c("orangered","orangered","cadetblue","cadetblue","cadetblue","goldenrod","goldenrod","golden
rod","goldenrod","goldenrod","palegreen","palegreen","palegreen","peachpuff","peachpuff","hot
pink","orangered4","turquoise")
op = par(bg = "grey35")
myplclust(fit, lab.col = colbars1, main = "Capacitance Cluster Dendrogram") # gives title

#### O2

mm <- ddply(labexperimentsdata, "Cell.Line", summarise, Oxigen.Consumption =
mean(Bionas.O2, na.rm=T)) #make a new tablewith the cell lines and the mean value of the
impedance
mm
rownames(mm) <- mm[,1]

Oxygen.Consumption <- dist(mm, method = "euclidean") # distance matrix
fit <- hclust(Oxygen.Consumption)
plot(fit)
op = par(bg = "grey35")
myplclust(fit, lab.col = colbars1, main = "Oxygen Consumption Cluster Dendrogram")

#### FET

mm <- ddply(labexperimentsdata, "Cell.Line", summarise, FET = mean(Bionas.Acid, na.rm=T))
#make a new tablewith the cell lines and the mean value of the impedance
mm
rownames(mm) <- mm[,1]

Acid.Lactic.Production <- dist(mm, method = "euclidean") # distance matrix
fit <- hclust(Acid.Lactic.Production)
plot(fit)

op = par(bg = "grey35")
myplclust(fit, lab.col = colbars1, main = "Acid Lactic Production Cluster Dendrogram")
```

```
#### ATP

colbars1 <-
c("orangered","orangered","cadetblue","cadetblue","cadetblue","goldenrod","goldenrod","golden
rod","goldenrod","goldenrod","palegreen","peachpuff","peachpuff","hotpink","orangered4","turq
uoise")
mm <- ddply(labexperimentsdata, "Cell.Line", summarise, ATP = mean(ATP, na.rm=T)) #make a
new tablewith the cell lines and the mean value of the impedance
mm

dfsub<-subset(mm,!(is.na(mm["ATP"]))) # take out nas. Cell lines in which no measurenment
was done
dfsub

rownames(dfsub) <- dfsub[,1]
Intracellular.ATP<- dist(dfsub, method = "euclidean") # distance matrix
fit <- hclust(Intracellular.ATP)
plot(fit)

op = par(bg = "grey35")
myplclust(fit, lab.col = colbars1, main = "Intracellular ATP Cluster Dendrogram")

#### Mitomass

mm <- ddply(labexperimentsdata, "Cell.Line", summarise, Mito = mean(Mitomass, na.rm=T))
#make a new tablewith the cell lines and the mean value of the impedance
mm
rownames(mm) <- mm[,1]

Cellular.Mitochondrial.Mass<- dist(mm, method = "euclidean") # distance matrix
fit <- hclust(Cellular.Mitochondrial.Mass)
plot(fit)

op = par(bg = "grey35")
myplclust(fit, lab.col = colbars1, main = "Cellular Mitochondrial Mass Cluster Dendrogram")

#### ROS

mm <- ddply(labexperimentsdata, "Cell.Line", summarise, ROS = mean(ROS, na.rm=T)) #make a
new tablewith the cell lines and the mean value of the impedance
mm
dfsub<-subset(mm,!(is.na(mm["ROS"]))) # take out nas. Cell lines in which no measurenment
was done
dfsub

rownames(dfsub) <- dfsub[,1]

Reactive.Oxygen.Species<- dist(dfsub, method = "euclidean") # distance matrix
fit <- hclust(Reactive.Oxygen.Species)
plot(fit)

op = par(bg = "grey35")
myplclust(fit, lab.col = colbars1, main = "Reactive Oxygen Species Cluster Dendrogram")
```

```
#### MTT

colbars1 <-
c("orangered","orangered","cadetblue","cadetblue","cadetblue","goldenrod","goldenrod","golden
rod","goldenrod","palegreen","hotpink","orangered4")

mm <- ddply(labexperimentsdata, "Cell.Line", summarise, MTT = mean(MTT, na.rm=T)) #make
a new tablewith the cell lines and the mean value of the impedance
mm
dfsub<-subset(mm,!(is.na(mm["MTT"]))) # take out nas. Cell lines in which no measurenment
was done
dfsub

rownames(dfsub) <- dfsub[,1]

MTT<- dist(dfsub, method = "euclidean") # distance matrix
fit <- hclust(MTT)
plot(fit)

op = par(bg = "grey35")
myplclust(fit, lab.col = colbars1, main = "MTT Cluster Dendrogram")

#### SRB

mm <- ddply(labexperimentsdata, "Cell.Line", summarise, SRB = mean(SRB, na.rm=T)) #make a
new tablewith the cell lines and the mean value of the impedance
mm
dfsub<-subset(mm,!(is.na(mm["SRB"]))) # take out nas. Cell lines in which no measurenment
was done
dfsub

rownames(dfsub) <- dfsub[,1]

SRB<- dist(dfsub, method = "euclidean") # distance matrix
fit <- hclust(SRB)
fit
plot(fit)

op = par(bg = "grey35")
myplclust(fit, lab.col = colbars1, main = "SRB Cluster Dendrogram")

############## SCALEING AND CLUSTERING BY COLUMNS THE GENE EXPRESSION DATA
(WILL BE THE TOTAL GENE EXPRESSION PER CELL LINE) ####################

#glycolysis/gluconeogenesys

scaledbycolgly<-biScale(mat_datag, row.center=FALSE, row.scale=FALSE, col.center=TRUE,
col.scale=TRUE, trace=TRUE)

write.csv(scaledbycolgly, file="scaledbycolgly.csv", row.names=TRUE)

rnames <- ggenetocluster[,1]
ggenetocluster <- data.matrix(ggenetocluster[,2:ncol(ggenetocluster)])

rownames(ggenetocluster) <- rnames
```

```
View(ggenetocluster)

# Ward Hierarchical Clustering
d <- dist(ggenetocluster, method = "euclidean") # distance matrix
fit <- hclust(d, method="ward")
plot(fit) # display dendogram
groups <- cutree(fit, k=2) # cut tree into 2 clusters
# draw dendogram with red borders around the 2 clusters
rect.hclust(fit, k=2, border="red")

# krebs cycle plus electron transport chain

scaledbycoltca<-biScale(mat_datatca, row.center=FALSE, row.scale=FALSE, col.center=TRUE,
col.scale=TRUE, trace=TRUE)

write.csv(scaledbycoltca, file="scaledbycoltca.csv", row.names=TRUE)

rnames <- tcagenetocluster[,1]
# assign labels in column 1 to "rnames"
tcagenetocluster <- data.matrix(tcagenetocluster[,2:ncol(tcagenetocluster)])

rownames(tcagenetocluster) <- rnames

# Ward Hierarchical Clustering
d <- dist(tcagenetocluster, method = "euclidean") # distance matrix
fit <- hclust(d, method="ward")
plot(fit) # display dendogram
groups <- cutree(fit, k=2) # cut tree into 2 clusters
# draw dendogram with red borders around the 2 clusters
rect.hclust(fit, k=2, border="red")
```

5.1.7 Pearson Correlation Coefficients Estimation

```
probes <- as.character(probes.metabolic.assays.and.drugs.IC50.for.pearson[ ,1])
View(probes)
probes.sub <- All.probes.set.16.cell.lines[All.probes.set.16.cell.lines$Probe.Set.ID %in% probes, ]
write.csv(probes.sub, file="probes.sub.csv")

########## PEARSON CORRELATION BETWEEN extreme expresses probes sets and
metabolic assays ######

install.packages("Hmisc")
library(Hmisc)

mat <- data.matrix(DRUGSPEARSON[,2:ncol(DRUGSPEARSON)]) # transform from column 2
into a matrix
rownames(mat) <- DRUGSPEARSON[,1]

correlation <-rcorr(mat, type="pearson")
correlation
str(correlation) # shows the 3 components of the function;r, n and p
```

```
df.correlation.r=data.frame(correlation$r) #convert the r part to a data frame

write.csv(df.correlation.r,"correlationmatrixDRUGSPEARSON.csv") # the data frame can be
exported

df.correlation.P=data.frame(correlation$P) #convert the p part to a data frame

write.csv(df.correlation.P,"pvaluescorrelationmatrixDRUGSPEARSON.csv") # the data frame can
be exported

rownames(pvaluescorrelationmatrixDRUGSPEARSON)<-
pvaluescorrelationmatrixDRUGSPEARSON[,1]
rownames(pvaluescorrelationmatrixDRUGSPEARSON)

# selection of the p values < 0.05

select.range<-function (pvaluescorrelationmatrixDRUGSPEARSON, min, max, data)
{
 if (nargs() > 3){
   min.cond <- pvaluescorrelationmatrixDRUGSPEARSON > min
   max.cond <- pvaluescorrelationmatrixDRUGSPEARSON< max

   cond <- min.cond & max.cond

 }
 else cat("Usage: select.range(pvaluescorrelationmatrixDRUGSPEARSON,min,max,datavec)\n")
}
minmaxtca<-select.range(pvaluescorrelationmatrixDRUGSPEARSON,0,0.05, data)

write.csv(minmaxtca, file="MinMaxpvalues0.05DRUGSPEARSON.csv", row.names=TRUE)

########## PEARSON CORRELATION BETWEEN all probes sets and metabolic assays
###### search p < 0.0001 in all the probes

mat <-
data.matrix(probes.sets.for.pearson5001to10000[,2:ncol(probes.sets.for.pearson5001to10000)
]) # transform from column 2 into a matrix
rownames(mat) <- probes.sets.for.pearson5001to10000[,1]
correlation <-rcorr(mat, type="pearson")
correlation
str(correlation) # shows the 3 components of the function;r, n and p
df.correlation.r=data.frame(correlation$r) #convert the r part to a data frame
write.csv(df.correlation.r,"correlationmatrixprobe1to5000PEARSON.csv") # the data frame can
be exported

df.correlation.P=data.frame(correlation$P) #convert the p part to a data frame
write.csv(df.correlation.P,"pvaluescorrelationmatrixprobe1to5000PEARSON.csv") # the data
frame can be exported

PP <- pvaluescorrelationmatrixprobe1to5000PEARSON

max <- subset(PP, PP[,2] < 0.0001)
max <- subset(PP, PP[,6] < 0.0001)
```

```
mat <- data.matrix(forpearsonprobes5001to10000[,2:ncol(forpearsonprobes5001to10000)]) # transform from column 2 into a matrix
rownames(mat) <- forpearsonprobes5001to10000[,1]
rownames(mat)

correlation <-rcorr(mat, type="pearson")
correlation
str(correlation) # shows the 3 components of the function;r, n and p
df.correlation.r=data.frame(correlation$r) #convert the r part to a data frame
write.csv(df.correlation.r,"correlationmatrixprobe5001to10000PEARSON.csv") # the data frame can be exported

df.correlation.P=data.frame(correlation$P) #convert the p part to a data frame
write.csv(df.correlation.P,"pvaluescorrelationmatrixprobe5001to10000PEARSON.csv") # the data frame can be exported

PP <- df.correlation.P

max1 <- subset(PP, PP[,1] < 0.0001)
max2<- subset(PP, PP[,2] < 0.0001)
max3 <- subset(PP, PP[,3] < 0.0001)
max4 <- subset(PP, PP[,4] < 0.0001)
max5 <- subset(PP, PP[,5] < 0.0001)
max6 <- subset(PP, PP[,6] < 0.0001)
max7 <- subset(PP, PP[,7] < 0.0001)
max8 <- subset(PP, PP[,8] < 0.0001)

mat <- data.matrix(allprobes10001to20000[,2:ncol(allprobes10001to20000)]) # transform from column 2 into a matrix
rownames(mat) <- allprobes10001to20000[,1]
rownames(mat)

correlation <-rcorr(mat, type="pearson")
correlation
str(correlation) # shows the 3 components of the function;r, n and p
df.correlation.r=data.frame(correlation$r) #convert the r part to a data frame
write.csv(df.correlation.r,"correlationmatrixprobe10001to20000PEARSON.csv") # the data frame can be exported

df.correlation.P=data.frame(correlation$P) #convert the p part to a data frame
write.csv(df.correlation.P,"pvaluescorrelationmatrixprobe10001to20000PEARSON.csv") # the data frame can be exported

PP <- df.correlation.P

max1 <- subset(PP, PP[,1] < 0.0001)
max2<- subset(PP, PP[,2] < 0.0001)
max3 <- subset(PP, PP[,3] < 0.0001)
max4 <- subset(PP, PP[,4] < 0.0001)
max5 <- subset(PP, PP[,5] < 0.0001)
max6 <- subset(PP, PP[,6] < 0.0001)
max7 <- subset(PP, PP[,7] < 0.0001)
max8 <- subset(PP, PP[,8] < 0.0001)
```

```
mat <- data.matrix(allprobespearson20001to30000[,2:ncol(allprobespearson20001to30000)])
# transform from column 2 into a matrix
rownames(mat) <- allprobespearson20001to30000[,1]
rownames(mat)

correlation <-rcorr(mat, type="pearson")

str(correlation) # shows the 3 components of the function;r, n and p
df.correlation.r=data.frame(correlation$r) #convert the r part to a data frame
write.csv(df.correlation.r,"correlationmatrixprobe20001to30000.csv") # the data frame can be
exported

df.correlation.P=data.frame(correlation$P) #convert the p part to a data frame
write.csv(df.correlation.P,"pvaluescorrelationmatrixprobe20001to30000.csv") # the data frame
can be exported

PP <- df.correlation.P

max1 <- subset(PP, PP[,1] < 0.0001)
max2<- subset(PP, PP[,2] < 0.0001)
max3 <- subset(PP, PP[,3] < 0.0001)
max4 <- subset(PP, PP[,4] < 0.0001)
max5 <- subset(PP, PP[,5] < 0.0001)
max6 <- subset(PP, PP[,6] < 0.0001)
max7 <- subset(PP, PP[,7] < 0.0001)
max8 <- subset(PP, PP[,8] < 0.0001)

mat <- data.matrix(allprobespearson30001to40000[,2:ncol(allprobespearson30001to40000)])
# transform from column 2 into a matrix
rownames(mat) <- allprobespearson30001to40000[,1]
rownames(mat)

correlation <-rcorr(mat, type="pearson")

str(correlation) # shows the 3 components of the function;r, n and p
df.correlation.r=data.frame(correlation$r) #convert the r part to a data frame
write.csv(df.correlation.r,"correlationmatrixprobe30001to40000.csv") # the data frame can be
exported

df.correlation.P=data.frame(correlation$P) #convert the p part to a data frame
write.csv(df.correlation.P,"pvaluescorrelationmatrixprobe30001to40000.csv") # the data frame
can be exported

PP <- df.correlation.P

max1 <- subset(PP, PP[,1] < 0.0001)
max2<- subset(PP, PP[,2] < 0.0001)
max3 <- subset(PP, PP[,3] < 0.0001)
max4 <- subset(PP, PP[,4] < 0.0001)
max5 <- subset(PP, PP[,5] < 0.0001)
max6 <- subset(PP, PP[,6] < 0.0001)
max7 <- subset(PP, PP[,7] < 0.0001)
max8 <- subset(PP, PP[,8] < 0.0001)
```

```
mat <- data.matrix(allprobespearson30001toend[,2:ncol(allprobespearson30001toend)]) #
transform from column 2 into a matrix
rownames(mat) <- allprobespearson30001toend[,1]
rownames(mat)

correlation <-rcorr(mat, type="pearson")
correlation
str(correlation) # shows the 3 components of the function;r, n and p
df.correlation.r=data.frame(correlation$r) #convert the r part to a data frame
write.csv(df.correlation.r,"correlationmatrixprobe30001toend.csv") # the data frame can be
exported

df.correlation.P=data.frame(correlation$P) #convert the p part to a data frame
write.csv(df.correlation.P,"pvaluescorrelationmatrixprobe30001toend.csv") # the data frame
can be exported

PP <- df.correlation.P

max1 <- subset(PP, PP[,1] < 0.0001)
max2<- subset(PP, PP[,2] < 0.0001)
max3 <- subset(PP, PP[,3] < 0.0001)
max4 <- subset(PP, PP[,4] < 0.0001)
max5 <- subset(PP, PP[,5] < 0.0001)
max6 <- subset(PP, PP[,6] < 0.0001)
max7 <- subset(PP, PP[,7] < 0.0001)
max8 <- subset(PP, PP[,8] < 0.0001)

to5000plusgenename <-
merge(correlationmatrixprobe1to5000PEARSON,genesymbol,by="Probe.Set.ID")

colnames(correlationmatrixprobe1to5000PEARSON)
rownames(correlationmatrixprobe1to5000PEARSON)

colnames (correlationmatrixprobe1to5000PEARSON) [1]<- "Probe.Set.ID"

class(correlationmatrixprobe1to5000PEARSON) #dataframe
class(genesymbol)#dataframe

a<- sub("X","",colnames(correlationmatrixprobe30001toend)) #column names without the X as
first letter
class(a) #character
a<-data.frame(a)
colnames(a) <- "Probe.Set.ID"
a <- a[-c(1), ] #to eliminate a row

to5000plusgenename = data.frame(a, correlationmatrixprobe30001toend) # to bind dataframe
a with dataframe correlationmatrixprobe30001toend
to5000plusgenename <- to5000plusgenename[-c(1), ] #to eliminate a row

to5000plusgenename <- merge(a,genesymbol,by="Probe.Set.ID")
colnames(to5000plusgenename)[1]<-"Probe.Set.ID"

rcorrplusgenename40001toend <- merge(genesymbol,to5000plusgenename,by="Probe.Set.ID")

rcorrplusgenename40001toend <- rcorrplusgenename40001toend[,c(1,2,5,6,7,8,9,10,11,12)]
```

```
write.csv(rcorrplusgenename40001toend, file="rcorrplusgenename40001toend.xls",
row.names=TRUE)


#### probes 1 to 5000
a<- sub("X","",colnames(correlationmatrixprobe1to5000PEARSON)) #column names without
the X as first letter
a<-data.frame(a) ### give me just the first column with the probes sets names without the X
colnames(a) <- "Probe.Set.ID"
a <- a[-c(1), ] #to eliminate a row
a<-data.frame(a)
colnames(a) <- "Probe.Set.ID"

correlationmatrixprobe1to5000PEARSON = data.frame(a,
correlationmatrixprobe1to5000PEARSON)

rcorrplusgenename1to5000 <-
merge(genesymbol,correlationmatrixprobe1to5000PEARSON,by="Probe.Set.ID")
rcorrplusgenename1to5000 <- rcorrplusgenename1to5000[,c(1,2,4,5,6,7,8,9,10,11)]
rcorrplusgenename1to5000<- rcorrplusgenename1to5000[,c(2,1,3,4,5,6,7,8,9,10)]
write.csv(rcorrplusgenename1to5000, file="rcorrplusgenename1to5000.xls",
row.names=TRUE)

#### probes 5001 to 10000
a<- sub("X","",colnames(correlationmatrixprobe5001to10000PEARSON)) #column names
without the X as first letter
a<-data.frame(a) ### give me just the first column with the probes sets names without the X
a <- a[-c(1), ] #to eliminate a row
a<-data.frame(a)
colnames(a) <- "Probe.Set.ID"

correlationmatrixprobe5001to10000PEARSON = data.frame(a,
correlationmatrixprobe5001to10000PEARSON)

rcorrplusgenename5001to10000 <-
merge(genesymbol,correlationmatrixprobe5001to10000PEARSON,by="Probe.Set.ID")
rcorrplusgenename5001to10000<- rcorrplusgenename5001to10000[,c(1,2,4,5,6,7,8,9,10,11)]
rcorrplusgenename5001to10000<- rcorrplusgenename5001to10000[,c(2,1,3,4,5,6,7,8,9,10)]
write.csv(rcorrplusgenename5001to10000, file="rcorrplusgenename5001to10000.xls",
row.names=TRUE)

#### probes 10001 to 20000
a<- sub("X","",colnames(correlationmatrixprobe10001to20000PEARSON)) #column names
without the X as first letter
a<-data.frame(a) ### give me just the first column with the probes sets names without the X
a <- a[-c(1), ] #to eliminate a row
a<-data.frame(a)
colnames(a) <- "Probe.Set.ID"

correlationmatrixprobe10001to20000PEARSON = data.frame(a,
correlationmatrixprobe10001to20000PEARSON)

rcorrplusgenename10001to20000 <-
merge(genesymbol,correlationmatrixprobe10001to20000PEARSON,by="Probe.Set.ID")
```

```
rcorrplusgenename10001to20000<-
rcorrplusgenename10001to20000[,c(1,2,4,5,6,7,8,9,10,11)]
rcorrplusgenename10001to20000<- rcorrplusgenename10001to20000[,c(2,1,3,4,5,6,7,8,9,10)]
write.csv(rcorrplusgenename10001to20000, file="rcorrplusgenename10001to20000.xls",
row.names=TRUE)

#### probes 20001 to 30000
a<- sub("X","",colnames(df.correlation.r)) #column names without the X as first letter
a<-data.frame(a) ### give me just the first column with the probes sets names without the X
a <- a[-c(1), ] #to eliminate a row
a<-data.frame(a)
colnames(a) <- "Probe.Set.ID"

correlationmatrixprobe20001to30000 = data.frame(a, df.correlation.r)

rcorrplusgenename20001to30000 <-
merge(genesymbol,correlationmatrixprobe20001to30000,by="Probe.Set.ID")
rcorrplusgenename20001to30000<- rcorrplusgenename20001to30000[,c(1,2,3,4,5,6,7,8,9,10)]

rcorrplusgenename20001to30000<- rcorrplusgenename20001to30000[,c(2,1,3,4,5,6,7,8,9,10)]
write.csv(rcorrplusgenename20001to30000, file="rcorrplusgenename20001to30000.xls",
row.names=TRUE)

#### probes 30001 to 40000
a<- sub("X","",colnames(correlationmatrixprobe30001to40000)) #column names without the X
as first letter
a<-data.frame(a) ### give me just the first column with the probes sets names without the X
a <- a[-c(1), ] #to eliminate a row
a<-data.frame(a)
colnames(a) <- "Probe.Set.ID"

correlationmatrixprobe30001to40000 = data.frame(a, correlationmatrixprobe30001to40000)

rcorrplusgenename30001to40000 <-
merge(genesymbol,correlationmatrixprobe30001to40000,by="Probe.Set.ID")
rcorrplusgenename30001to40000<-
rcorrplusgenename30001to40000[,c(1,2,4,5,6,7,8,9,10,11)]
rcorrplusgenename30001to40000<- rcorrplusgenename30001to40000[,c(2,1,3,4,5,6,7,8,9,10)]
write.csv(rcorrplusgenename30001to40000, file="rcorrplusgenename30001to40000.xls",
row.names=TRUE)

pvalues10001to20000 <- pvaluescorrelationmatrixprobe10001to20000PEARSON[,c(1,2,3,
4,5,6,7,8,9)]
write.csv(pvalues10001to20000, file="pvalues10001to20000.xls", row.names=TRUE)

pvalues20001to30000 <- df.correlation.P[,c(1,2,3, 4,5,6,7,8)]
write.csv(pvalues20001to30000, file="pvalues20001to30000.xls", row.names=TRUE)
pvalues30001to40000 <- pvaluescorrelationmatrixprobe30001to40000[,c(1,2,3, 4,5,6,7,8,9)]
write.csv(pvalues30001to40000, file="pvalues30001to40000.xls", row.names=TRUE)

PP <- pvalues10001to20000
class(pvalues10001to20000)
class(PP)

max1 <- subset(PP, PP[,1] < 0.0001)
```

```
max2<- subset(PP, PP[,2] < 0.0001)
max3 <- subset(PP, PP[,3] < 0.0001)
max4 <- subset(PP, PP[,4] < 0.0001)
max5 <- subset(PP, PP[,5] < 0.0001)
max6 <- subset(PP, PP[,6] < 0.0001)
max7 <- subset(PP, PP[,7] < 0.0001)
max8 <- subset(PP, PP[,8] < 0.0001)

pvaluescorrelationmatrixprobe5001to10000 <-
pvaluescorrelationmatrixprobe5001to10000PEARSON[,c(1,2,3, 4,5,6,7,8,9)]
```

6. References

Abolmaali, S. S., Tamaddon, A. M., & Dinarvand, R. (2013). A review of therapeutic challenges and achievements of methotrexate delivery systems for treatment of cancer and rheumatoid arthritis. *Cancer Chemotherapy and Pharmacology*, *71*(5), 1115-1130.

Agathocleous, M., & Harris, W. A. (2013). Metabolism in physiological cell proliferation and differentiation. *Trends in Cell Biology*, *23*(10), 484-492.

Akram, M. (2013). Mini-review on glycolysis and cancer. *Journal of Cancer Education*, *28*(3), 454.

Alborzinia, H., Can, S., Holenya, P., Scholl, C., Lederer, E., Kitanovic, I., & Wölfl, S. (2011). Real-time monitoring of cisplatin-induced cell death. *PLoS One*, *6*(5), e19714.

Alimirah, F., Chen, J., Davis, F. J., & Choubey, D. (2007). IFI16 in human prostate cancer. *Molecular Cancer Research*, *5*(3), 251-259.

Altenberg, B. A., & Greulich, K. O. (2004). Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics*, *84*(6), 1014-1020.

Ansari, M. A., Dutta, S., Veettil, M. V., Dutta, D., Iqbal, J., Kumar, B., ... & Chandran, B. (2015). Herpesvirus genome recognition induced acetylation of nuclear IFI16 is essential for its cytoplasmic translocation, inflammasome and IFN-β responses. *PLoS pathogens*, *11*(7), e1005019.

Arai, E., Sakamoto, H., Ichikawa, H., Totsuka, H., Chiku, S., Gotoh, M., ... & Fujimoto, H. (2014). Multilayer-omics analysis of renal cell carcinoma, including the whole exome, methylome and transcriptome. *International journal of cancer*, *135*(6), 1330-1342.

Bar-Even, A., Flamholz, A., Noor, E., & Milo, R. (2012). Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nature chemical biology*, *8*(6), 509-517.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., ... & Reddy, A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, *483*(7391), 603-607.

Berg, J. M., Tymoczko, J. L., & Stryer, L. (2002). Glycolysis and Gluconeogenesis, Chapter 16 in Biochemistry. 5th Edition**,** W.H. Freeman, New York

Bergveld, P. (1981). The operation of an ISFET as an electronic device. *Sensors and Actuators*, *1*, 17-29.

Bianchi, G., Caporali, R., Todoerti, M., & Mattana, P. (2016). Methotrexate and rheumatoid arthritis: current evidence regarding subcutaneous versus oral routes of administration. *Advances in therapy*, *33*(3), 369-378.

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological methods*, *17*(3), 399.

Bothwell, S. P., Farber, L. W., Hoagland, A., & Nussbaum, R. L. (2010). Species-specific difference in expression and splice-site choice in Inpp5b, an inositol polyphosphate 5-phosphatase paralogous to the enzyme deficient in Lowe Syndrome. *Mammalian genome*, *21*(9-10), 458-466.

Brand, K. A., & Hermfisse, U. (1997). Aerobic glycolysis by proliferating cells: a protective strategy against reactive oxygen species. *The FASEB journal*, *11*(5), 388-395.

Caunt, C. J., Sale, M. J., Smith, P. D., & Cook, S. J. (2015). MEK1 and MEK2 inhibitors and cancer therapy: the long and winding road. *Nature reviews Cancer*, *15*(10), 577-592.

Cha, Y. J., Jung, W. H., & Koo, J. S. (2017). Differential Site-Based Expression of Pentose Phosphate Pathway-Related Proteins among Breast Cancer Metastases. *Disease markers*, *2017*.

Cheng, T. L., Lai, C. H., Jiang, S. J., Hung, J. H., Liu, S. K., Chang, B. I., ... & Wu, H. L. (2014). RHBDL2 is a critical membrane protease for anoikis resistance in human malignant epithelial cells. *The Scientific World Journal*, *2014*.

Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data* (Doctoral dissertation, University of Pittsburgh).

Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L., & Packer, N. H. (2014). Cell surface protein glycosylation in cancer. *Proteomics*, *14*(4-5), 525-546.

Cicenas, J., Tamosaitis, L., Kvederaviciute, K., Tarvydas, R., Staniute, G., Kalyan, K., ... & Valius, M. (2017). KRAS, NRAS and BRAF mutations in colorectal cancer and melanoma. *Medical Oncology*, *34*(2), 26.

Clark, L. C., Wolf, R., Granger, D., & Taylor, Z. (1953). Continuous recording of blood oxygen tensions by polarography. *Journal of applied physiology*, *6*(3), 189-193.

Cook, Charles E., et al. "The European Bioinformatics Institute in 2016: data growth and integration." Nucleic acids research 44.D1 (2016): D20-D26.

Cooper, G. M., & Hausman, R. E. (2000). A molecular approach. *The cell. 2nd ed. Sunderland, MA: Sinauer Associates*.

Corbet, C., & Feron, O. (2017). Cancer cell metabolism and mitochondria: nutrient plasticity for TCA cycle fueling. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*.

De Preter, G., Neveu, M. A., Danhier, P., Brisson, L., Payen, V. L., Porporato, P. E., ... & Gallez, B. (2016). Inhibition of the pentose phosphate pathway by dichloroacetate unravels a missing link between aerobic glycolysis and cancer cell proliferation. *Oncotarget*, *7*(3), 2910.

De Schutter, T., Andrei, G., Topalis, D., Naesens, L., & Snoeck, R. (2013). Cidofovir selectivity is based on the different response of normal and cancer cells to DNA damage. *BMC medical genomics*, *6*(1), 18.

Delgado, V. M. C., Nugnes, L. G., Colombo, L. L., Troncoso, M. F., Fernández, M. M., Malchiodi, E. L., ... & Wolfenstein-Todel, C. (2011). Modulation of endothelial cell migration and angiogenesis: a novel function for the "tandem-repeat" lectin galectin-8. *The FASEB Journal*, *25*(1), 242-254.

Duan, X., Ponomareva, L., Veeranki, S., & Choubey, D. (2011). IFI16 induction by glucose restriction in human fibroblasts contributes to autophagy through activation of the ATM/AMPK/p53 pathway. *PLoS One*, *6*(5), e19532.

Dunn, K. W., Kamocka, M. M., & McDonald, J. H. (2011). A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology*, *300*(4), C723-C742.

Ehret, R., Baumann, W., Brischwein, M., Schwinde, A., Stegbauer, K., & Wolf, B. (1997). Monitoring of cellular behaviour by impedance measurements on interdigitated electrode structures. *Biosensors and Bioelectronics*, *12*(1), 29-41.

Eisenberg, R. S., & Mathias, R. T. (1980). Structural analysis of electrical properties of cells and tissues. *CRC Crit Rev Bioeng*, *4*(3), 203-32.

Etheridge, S. L., Brooke, M. A., Kelsell, D. P., & Blaydon, D. C. (2013). Rhomboid proteins: a role in keratinocyte proliferation and cancer. *Cell and tissue research*, *351*(2), 301-307.

Fenton, N., & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian networks*. CRC Press, Boca Raton, Florida.

Firestein, R., Bass, A. J., Kim, S. Y., Dunn, I. F., Silver, S. J., Guney, I., ... & Chheda, M. G. (2008). CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature*, *455*(7212), 547-551.

Fleischer, T., Frigessi, A., Johnson, K. C., Edvardsen, H., Touleimat, N., Klajic, J., ... & Helland, Å. (2014). Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome biology*, *15*(8), 435.

Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., ... & Stefancsik, R. (2016). COSMIC: somatic cancer genetics at high-resolution. *Nucleic acids research*, *45*(D1), D777-D783.

Fyles, T. M. (2007). Synthetic ion channels in bilayer membranes. *Chemical Society Reviews*, *36*(2), 335-347.

Galati, D., Srinivasan, S., Raza, H., Prabu, S. K., Hardy, M., Chandran, K., ... & Avadhani, N. G. (2009). Role of nuclear-encoded subunit Vb in the assembly and stability of cytochrome c oxidase complex: implications in mitochondrial dysfunction and ROS production. *Biochemical Journal*, *420*(3), 439-449.

Gao, S. P., Sun, H. F., Jiang, H. L., Li, L. D., Hu, X., Xu, X. E., & Jin, W. (2015). Loss of COX5B inhibits proliferation and promotes senescence via mitochondrial dysfunction in breast cancer. *Oncotarget*, *6*(41), 43363.

Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, *20*(3), 307-315.

Guo, Y., Xiao, P., Lei, S., Deng, F., Xiao, G. G., Liu, Y., ... & Jiang, H. (2008). How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta biochimica et biophysica Sinica*, *40*(5), 426-436.

Győrffy, B., Bottai, G., Fleischer, T., Munkácsy, G., Budczies, J., Paladini, L., ... & Santarpia, L. (2016). Aberrant DNA methylation impacts gene expression and prognosis in breast cancer subtypes. *International journal of cancer*, *138*(1), 87-97.

Harrell, F. (2015). Hmisc: Harrell Miscellaneous. R package version 3.17-1.

Hastie, T., Mazumder, R., Lee, J. D., & Zadeh, R. (2015). Matrix completion and low-rank SVD via fast alternating least squares. *Journal of Machine Learning Research*, *16*, 3367-3402.

Hedley, D., Shamas-Din, A., Chow, S., Sanfelice, D., Schuh, A. C., Brandwein, J. M., ... & Schimmer, A. D. (2016). A phase I study of elesclomol sodium in patients with acute myeloid leukemia. *Leukemia & lymphoma*, *57*(10), 2437-2440.

Hensley, C. T., Faubert, B., Yuan, Q., Lev-Cohain, N., Jin, E., Kim, J., ... & Wodzak, M. (2016). Metabolic heterogeneity in human lung tumors. *Cell*, *164*(4), 681-694.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, *4*(1), 44-57.

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., ... & Gottardo, R. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature methods*, *12*(2), 115-121.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, *4*(2), 249-264.

Irizarry, R.A., & Love, M.I (2015). Convenience Functions for Routine Data Exploration. Package "Rafalib". CRAN repository R project.

Iverson, C., Larson, G., Lai, C., Yeh, L. T., Dadson, C., Weingarten, P., ... & Hamatake, R. (2009). RDEA119/BAY 869766: a potent, selective, allosteric inhibitor of MEK1/2 for the treatment of cancer. *Cancer research*, *69*(17), 6839-6847.

Jančík, S., Drábek, J., Radzioch, D., & Hajdúch, M. (2010). Clinical relevance of KRAS in human cancers. *BioMed Research International*, *2010*.

Jin, K., Li, L., Sun, X., Xu, Q., Song, S., Shen, Y., & Deng, X. (2017). Mycoepoxydiene suppresses HeLa cell growth by inhibiting glycolysis and the pentose phosphate pathway. *Applied microbiology and biotechnology*, *101*(10), 4201-4213.

Johnson, K. E., Bottero, V., Flaherty, S., Dutta, S., Singh, V. V., & Chandran, B. (2014). IFI16 restricts HSV-1 replication by accumulating on the hsv-1 genome, repressing HSV-1 gene expression, and directly or indirectly modulating histone modifications. *PLoS pathogens*, *10*(11), e1004503.

Jones, C. J. (Ed.). (2012). *Epithelia: Advances in cell physiology and cell culture*. Springer Science & Business Media, California

Joung, J. K., Voytas, D. F., & Kamens, J. (2015). Accelerating research through reagent repositories: the genome editing example. *Genome biology*, *16*(1), 255.

Kang, S. W., Lee, S., & Lee, E. K. (2015). ROS and energy metabolism in cancer cells: alliance for fast growth. *Archives of Pharmacal Research*, *38*(3), 338-345.

Kelder, T., van Iersel, M. P., Hanspers, K., Kutmon, M., Conklin, B. R., Evelo, C. T., & Pico, A. R. (2011). WikiPathways: building research communities on biological pathways. *Nucleic acids research*, *40*(D1), D1301-D1307.

Kern, W., Kohlmann, A., Wuchter, C., Schnittger, S., Schoch, C., Mergenthaler, S., ... & Haferlach, T. (2003). Correlation of protein expression and gene expression in acute leukemia. *Cytometry Part B: Clinical Cytometry*, *55*(1), 29-36.

Kirshner, J. R., He, S., Balasubramanyam, V., Kepros, J., Yang, C. Y., Zhang, M., ... & Bertin, J. (2008). Elesclomol induces cancer cell apoptosis through oxidative stress. *Molecular cancer therapeutics*, *7*(8), 2319-2327.

Kissil, J. L., Walmsley, M. J., Hanlon, L., Haigis, K. M., Kim, C. F. B., Sweet-Cordero, A., ... & Jacks, T. (2007). Requirement for Rac1 in a K-ras–Induced Lung Cancer in the Mouse. *Cancer research*, *67*(17), 8089-8094.

Knuesel, M. T., Meyer, K. D., Donner, A. J., Espinosa, J. M., & Taatjes, D. J. (2009). The human CDK8 subcomplex is a histone kinase that requires Med12 for activity and can function independently of mediator. *Molecular and cellular biology*, *29*(3), 650-661.

Kojima, K., Konopleva, M., McQueen, T., O'Brien, S., Plunkett, W., & Andreeff, M. (2006). Mdm2 inhibitor Nutlin-3a induces p53-mediated apoptosis by transcription-dependent and transcription-independent mechanisms and may overcome Atm-mediated resistance to fludarabine in chronic lymphocytic leukemia. *Blood*, *108*(3), 993-1000.

Kolde, R. (2015). pheatmap: Pretty heatmaps. R package version 1.0. 8.

Koppenol, W. H., Bounds, P. L., & Dang, C. V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism. *Nature Reviews Cancer*, *11*(5), 325-337.

Kubota, H., Ota, K., Sakaki, Y., & Ito, T. (2001). Budding yeast GCN1 binds the GI domain to activate the eIF2α kinase GCN2. *Journal of Biological Chemistry*, *276*(20), 17591-17596.

Lago, C. U., Sung, H. J., Ma, W., Wang, P. Y., & Hwang, P. M. (2011). p53, aerobic metabolism, and cancer. *Antioxidants & redox signaling*, *15*(6), 1739-1748.

Lee, J., Lee, I., Han, B., Park, J. O., Jang, J., Park, C., & Kang, W. K. (2011). Effect of simvastatin on cetuximab resistance in human colorectal cancer with KRAS mutations. *JNCI: Journal of the National Cancer Institute*, *103*(8), 674-688.

Lemberger, T. (2015). Image data in need of a home.

Li, Y., Yang, X. H., Fang, S. J., Qin, C. F., Sun, R. L., Liu, Z. Y., ... & Li, G. (2015). HOXA7 stimulates human hepatocellular carcinoma proliferation through cyclin E1/CDK2. *Oncology reports*, *33*(2), 990-996.

Lincet, H., & Icard, P. (2015). How do glycolytic enzymes favour cancer cell proliferation by nonmetabolic functions?. *Oncogene*, *34*(29), 3751-3759.

Little, A. S., Balmanno, K., Sale, M. J., Newman, S., Dry, J. R., Hampson, M., ... & Cook, S. J. (2011). Amplification of the driving oncogene, KRAS or BRAF, underpins acquired resistance to MEK1/2 inhibitors in colorectal cancer cells. *Sci. Signal.*, *4*(166), ra17-ra17.

Liu, A. M., Xu, Z., Shek, F. H., Wong, K. F., Lee, N. P., Poon, R. T., ... & Luk, J. M. (2014). miR-122 targets pyruvate kinase M2 and affects metabolism of hepatocellular carcinoma. *PloS one*, *9*(1), e86872.

Liu, Y., Wang, S., Dong, Q. Z., Jiang, G. Y., Han, Y., Wang, L., & Wang, E. H. (2016). The P21-activated kinase expression pattern is different in non-small cell lung cancer and affects lung cancer cell sensitivity to epidermal growth factor receptor tyrosine kinase inhibitors. *Medical Oncology*, *33*(3), 22.

Lunt, S. Y., & Vander Heiden, M. G. (2011). Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annual review of cell and developmental biology*, *27*, 441-464.

Lv, J., Wang, J., Chang, S., Liu, M., & Pang, X. (2015). The greedy nature of mutant RAS: a boon for drug discovery targeting cancer metabolism?. *Acta biochimica et biophysica Sinica*, *48*(1), 17-26.

Månsson, R., Tsapogas, P., Åkerlund, M., Lagergren, A., Gisler, R., & Sigvardsson, M. (2004). Pearson correlation analysis of microarray data allows for the identification of genetic targets for early B-cell factor. *Journal of Biological Chemistry*, *279*(17), 17905-17913.

MacDonald, B. T., Tamai, K., & He, X. (2009). Wnt/β-catenin signaling: components, mechanisms, and diseases. *Developmental cell*, *17*(1), 9-26.

Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS letters*, *583*(24), 3966-3973.

Makrodouli, E., Oikonomou, E., Koc, M., Andera, L., Sasazuki, T., Shirasawa, S., & Pintzas, A. (2011). BRAF and RAS oncogenes regulate Rho GTPase pathways to mediate migration and invasion properties in human colon cancer cells: a comparative study. *Molecular cancer*, *10*(1), 118.

Marx, V. (2013). Biology: The big challenges of big data. *Nature*, *498*(7453), 255-260.

McCarty, M. F., & Contreras, F. (2014). Increasing superoxide production and the labile iron pool in tumor cells may sensitize them to extracellular ascorbate. *Frontiers in oncology*, *4*.

McDonald, J. H. (2009). *Handbook of Biological Statistics* (Vol. 2, pp. 173-181). Baltimore, MD: Sparky House Publishing.

Meijer, T. W., Schuurbiers, O. C., Kaanders, J. H., Looijen-Salamon, M. G., de Geus-Oei, L. F., Verhagen, A. F., ... & Bussink, J. (2012). Differences in metabolism between adeno-and squamous cell non-small cell lung carcinomas: spatial distribution and prognostic value of GLUT1 and MCT4. *Lung cancer*, *76*(3), 316-323.

Metz, C., Döger, R., Riquelme, E., Cortés, P., Holmes, C., Shaughnessy, R., ... & Soza, A. (2016). Galectin-8 promotes migration and proliferation and prevents apoptosis in U87 glioblastoma cells. *Biological research*, *49*(1), 33.

Mosgaard, L. D., Zecchi, K. A., & Heimburg, T. (2015). Mechano-capacitive properties of polarized membranes. *Soft Matter*, *11*(40), 7899-7910.

Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3), 69-71.

Nagy, N., Bronckart, Y., Camby, I., Legendre, H., Lahm, H., Kaltner, H., ... & Zick, Y. (2002). Galectin-8 expression decreases in cancer compared with normal and dysplastic human colon tissue and acts significantly on human colon cancer cell migration as a suppressor. *Gut*, *50*(3), 392-401.

Nakamura, T., Nakatsu, N., Yoshida, Y., Yamazaki, K., Dan, S., Sadahiro, S., ... & Yamori, T. (2009). Identification of candidate genes determining chemosensitivity to anti-cancer drugs of gastric cancer cell lines. *Biological and Pharmaceutical Bulletin*, *32*(11), 1936-1939.

Niebur, E. (2008). Electrical properties of cell membranes. *Scholarpedia*, *3*(6), 7166.

Niemitz, E. (2013). Ras pathway activation in breast cancer. *Nature genetics*, *45*(11), 1273-1273.

O'Connor, C. M., Adams, J. U., & Fairman, J. (2010). Essentials of cell biology. *Cambridge, MA: NPG Education*, *1*.

Okon, I. S., & Zou, M. H. (2015). Mitochondrial ROS and cancer drug resistance: Implications for therapy. *Pharmacological research*, *100*, 170-174.

Onesto, C., Shutes, A., Picard, V., Schweighoffer, F., & Der, C. J. (2008). Characterization of EHT 1864, a novel small molecule inhibitor of Rac family small GTPases. *Methods in enzymology*, *439*, 111-129.

Orlovsky, K., Kalinkovich, A., Rozovskaia, T., Shezen, E., Itkin, T., Alder, H., ... & Buchberg, A. (2011). Down-regulation of homeobox genes MEIS1 and HOXA in MLL-rearranged acute leukemia impairs engraftment and reduces proliferation. *Proceedings of the National Academy of Sciences*, *108*(19), 7956-7961.

Ou, K., Yu, K., Kesuma, D., Hooi, M., Huang, N., Chen, W., ... & Soon, S. Y. (2008). Novel breast cancer biomarkers identified by integrative proteomic and gene expression mapping. *Journal of proteome research*, *7*(4), 1518-1528.

Pearce, L. R., Alton, G. R., Richter, D. T., Kath, J. C., Lingardo, L., Chapman, J., ... & Alessi, D. R. (2010). Characterization of PF-4708671, a novel and highly specific inhibitor of p70 ribosomal S6 kinase (S6K1). *Biochemical Journal*, *431*(2), 245-255.

Phillips, D. C., Woollard, K. J., & Griffiths, H. R. (2003). The anti-inflammatory actions of methotrexate are critically dependent upon the production of reactive oxygen species. *British journal of pharmacology*, *138*(3), 501-511.

Qiu, Z. X., Sun, R. F., Mo, X. M., & Li, W. M. (2016). The p70S6K specific inhibitor PF-4708671 impedes non-small cell lung cancer growth. *PloS one*, *11*(1), e0147185.

Rae, C., & Mairs, R. J. (2017). Evaluation of the radiosensitizing potency of chemotherapeutic agents in prostate cancer cells. *International journal of radiation biology*, *93*(2), 194-203.

Ray, S., Kassan, A., Busija, A. R., Rangamani, P., & Patel, H. H. (2016). The plasma membrane as a capacitor for energy and metabolism. *American Journal of Physiology-Cell Physiology*, *310*(3), C181-C192.

Rogalinska, M. (2016). The role of mitochondria in cancer induction, progression and changes in metabolism. *Mini reviews in medicinal chemistry*, *16*(7), 524-530.

Romano, A. H., & Conway, T. (1996). Evolution of carbohydrate metabolic pathways. *Research in microbiology*, *147*(6-7), 448-455.

Rung, J., & Brazma, A. (2013). Reuse of public genome-wide gene expression data. *Nature Reviews Genetics*, *14*(2), 89-99.

Sajnani, K., Islam, F., Smith, R. A., Gopalan, V., & Lam, A. K. Y. (2017). Genetic alterations in Krebs cycle and its impact on cancer pathogenesis. *Biochimie*, 135, 164–172

Shi, X., Liu, J., Liu, Q., & Li, M. (2014). IFI16 mis-localization can be a contributing factor to hepatocellular carcinoma progression. *Medical hypotheses*, *82*(3), 398-400.

Singhal, R., & Kandel, E. S. (2012). The response to PAK1 inhibitor IPA3 distinguishes between cancer cells with mutations in BRAF and Ras oncogenes. *Oncotarget*, *3*(7), 700.

Team, R. C. (2015). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria.

Thomas, S., Fisher, K. H., Snowden, J. A., Danson, S. J., Brown, S., & Zeidler, M. P. (2015). Methotrexate is a JAK/STAT pathway inhibitor. *PLoS One*, *10*(7), e0130078.

Toyoda, Y., & Saitoh, S. (2015). Adaptive regulation of glucose transport, glycolysis and respiration for cell proliferation. *Biomolecular concepts*, *6*(5-6), 423-430.

Uzwyshyn, R. (2016). Research Data Repositories: The What, When, Why, and How. *Computers Libraries 36.3*.

Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*, *324*(5930), 1029-1033.

Vichai, V., & Kirtikara, K. (2006). Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nature protocols*, *1*(3), 1112-1116.

Warren, P., Taylor, D., Martini, P. G., Jackson, J., & Bienkowska, J. (2007, October). PANP-a new method of gene detection on oligonucleotide expression arrays. In *Bioinformatics and Bioengineering, 2007. BIBE 2007. Proceedings of the 7th IEEE International Conference on* (pp. 108-115). IEEE.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1-29.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer, New York.

Wood, C. E., Hukkanen, R. R., Sura, R., Jacobson-Kram, D., Nolte, T., Odin, M., & Cohen, S. M. (2015). Scientific and Regulatory Policy Committee (SRPC) Review* Interpretation and Use of Cell Proliferation Data in Cancer Risk Assessment. *Toxicologic pathology*, *43*(6), 760-775.

Wright, A. A., Howitt, B. E., Myers, A. P., Dahlberg, S. E., Palescandolo, E., Hummelen, P., ... & Quick, C. M. (2013). Oncogenic mutations in cervical cancer. *Cancer*, *119*(21), 3776-3783.

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., ... & Ramaswamy, S. (2012). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, *41*(D1), D955-D961.

Yang, Y., Karakhanova, S., Hartwig, W., D'Haese, J. G., Philippov, P. P., Werner, J., & Bazhin, A. V. (2016). Mitochondria and mitochondrial ROS in cancer: novel targets for anticancer therapy. *Journal of cellular physiology*, *231*(12), 2570-2581.

Yun, J., Rago, C., Cheong, I., Pagliarini, R., Angenendt, P., Rajagopalan, H., ... & Diaz, L. A. (2009). Glucose deprivation contributes to the development of KRAS pathway mutations in tumor cells. *Science*, *325*(5947), 1555-1559.

Zhang, Y., Cheng, J. C., Huang, H. F., & Leung, P. C. (2013). Homeobox A7 stimulates breast cancer cell proliferation by up-regulating estrogen receptor-alpha. *Biochemical and biophysical research communications*, *440*(4), 652-657.

Zhang, N., Wang, H., Fang, Y., Wang, J., Zheng, X., & Liu, X. S. (2015). Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. *PLoS computational biology*, *11*(9), e1004498.

Zheng, J. (2012). Energy metabolism of cancer: Glycolysis versus oxidative phosphorylation. *Oncology letters*, *4*(6), 1151-1157.