



UNIVERSITA' DEGLI STUDI DI CATANIA

Ph.D. in Biotechnology
Curricula Molecular Biotechnology
(XXXVII Cycle)

Daniela Luciano

***VDAC1P8* as a Novel Regulatory Element in
Leukemogenesis: Integrative Genomic, Epigenetic, and
Functional Insights**

Ph.D. thesis

Supervisor: Prof. Francesca Guarino
Coordinator: Prof. Vito Nicola De Pinto

Academic Year 2024-2025

Sommario

I mitocondri sono organelli essenziali per le cellule eucariotiche, poiché supportano l'enorme richiesta energetica necessaria per mantenere l'omeostasi cellulare. La comunicazione tra il citosol e i mitocondri è mediata dalle porine mitocondriali, note anche come Voltage-Dependent Anion Selective Channels (VDAC), localizzate nella membrana mitocondriale esterna di tutti gli eucarioti. Le proteine VDAC consentono lo scambio di metaboliti e partecipano a numerosi processi cellulari grazie alla loro interazione con specifici enzimi citosolici e fattori apoptotici. Nei mammiferi sono presenti tre diverse isoforme: *VDAC1*, *VDAC2* e *VDAC3*. Pur condividendo un'alta omologia di sequenza, queste isoforme presentano funzioni e caratteristiche distinte.

In particolare, *VDAC1* è frequentemente sovraespresso nei tumori, ed uno degli aspetti più studiati è la sua interazione con l'esochinasi, un enzima cruciale nel metabolismo del glucosio. L'associazione stabile tra VDAC1 ed esochinasi è un elemento chiave dell'effetto Warburg, un fenomeno caratteristico delle cellule tumorali, in cui il metabolismo energetico è deviato dalla fosforilazione ossidativa mitocondriale verso la glicolisi anaerobica, anche in presenza di ossigeno. Questa alterazione consente alle cellule tumorali di soddisfare le proprie esigenze metaboliche per una rapida proliferazione e di ridurre la produzione di specie reattive dell'ossigeno (ROS), minimizzando il danno ossidativo.

L'interazione tra VDAC1 ed esochinasi non solo favorisce la glicolisi, ma impedisce anche a VDAC1 di legarsi a proteine pro-apoptotiche come Bax e Bak. Queste proteine sono essenziali per l'induzione dell'apoptosi. In questo contesto, VDAC1 funge da modulatore centrale che collega il metabolismo energetico alla sopravvivenza cellulare, rendendolo un bersaglio particolarmente interessante per strategie terapeutiche mirate.

Sebbene le proteine della famiglia VDAC, ed in particolare l'isoforma VDAC1, sono state ben caratterizzate nella loro funzione di canale e si conosce in alcuni casi il loro ruolo all'interno della cellula, poco si conosce a riguardo dei meccanismi di regolazione dell'espressione delle tre diverse isoforme. Solo recentemente sono stati delineati alcuni aspetti che riguardano la regolazione, dei meccanismi trascrizionali dei tre geni, riconducibili all'azione di alcuni fattori di trascrizione e del meccanismo epigenetico della metilazione grazie sia ad un approccio bioinformatico che sperimentale.

In questo contesto abbiamo cominciato ad esplorare anche possibili meccanismi di regolazione genica post-trascrizionali mediati dagli pseudogeni data l'elevata numerosità identificata per le isoforme *VDAC*.

Infatti gli pseudogeni stanno guadagnando sempre più attenzione nei processi tumorigenici. Tradizionalmente considerati frammenti genici privi di funzione, è stato dimostrato che alcuni pseudogeni regolano attivamente l'espressione genica, spesso agendo come ceRNA (RNA endogeni competitivi), sequestrando microRNA condivisi con i geni parentali e modulandone l'espressione. Questo meccanismo di regolazione, simile a quello osservato per i lncRNA, rappresenta un potenziale strumento di controllo dell'espressione genica ed è sempre più riconosciuto in molte patologie, in particolare nei tumori.

Poco o nulla si sa riguardo agli pseudogeni *VDAC*. Poiché i fattori e le circostanze che regolano l'espressione dei geni *VDAC* sono ancora in fase di studio, comprendere se vi sia un coinvolgimento o un ruolo per i loro pseudogeni potrebbe fornire una chiave importante per svelare la rete regolatoria che governa l'espressione delle isoforme *VDAC*.

Con il lavoro svolto in questa tesi di dottorato è stata eseguita un'analisi *in-silico* e *in vitro* per caratterizzare gli pseudogeni funzionali del gene parentale *VDAC1* e i potenziali meccanismi regolatori che potrebbero spiegare il ruolo biologico di *VDAC1* in condizioni fisiologiche e patologiche. Incrociando i dati di più database genomici e trascrittomici, è stata osservata una correlazione tra le variazioni dei livelli di espressione del gene *VDAC1* e di alcuni pseudogeni, in particolare lo pseudogene *VDACIP8*, che ha indirizzato la nostra attenzione su alcuni tumori, la leucemia mieloide acuta (AML) e il cancro al colon retto (COAD).

Per esempio, l'analisi comparativa *in-silico* tra il gene *VDAC1* e il suo pseudogene, *VDACIP8*, combinata con dati sperimentali ottenuti da modelli cellulari di AML, ha rivelato che la sovraespressione specifica di *VDACIP8* è associata alla downregulation del gene parentale *VDAC1*. Questi risultati indicano che *VDACIP8* potrebbe rappresentare un potenziale biomarcatore per AML e aprono nuove possibilità per lo studio dei meccanismi regolatori di *VDACIP8* e del suo possibile ruolo nella patogenesi dell'AML.

Abstract

Mitochondria are essential organelles for eukaryotic cells as they support the high energy demand required to maintain cellular homeostasis. Communication between the cytosol and mitochondria is mediated by mitochondrial porins, also known as Voltage-Dependent Anion Selective Channels (VDACs), which are located in the outer mitochondrial membrane of all eukaryotes. VDAC proteins facilitate the exchange of metabolites and participate in numerous cellular pathways through their interaction with specific cytosolic enzymes and apoptotic factors. In mammals, there are three distinct isoforms: VDAC1, VDAC2, and VDAC3. Despite their high sequence homology, these isoforms have different functions and characteristics.

In particular, VDAC1 is frequently overexpressed in tumors, and one of the most studied aspects is its interaction with hexokinase, a crucial enzyme in glucose metabolism. The stable association between VDAC1 and hexokinase is a key element in the Warburg effect, a phenomenon characteristic of cancer cells, where energy metabolism shifts from mitochondrial oxidative phosphorylation to anaerobic glycolysis, even in the presence of oxygen. This alteration allows cancer cells to meet their metabolic demands for rapid proliferation and to reduce the production of reactive oxygen species (ROS), thereby minimizing oxidative damage.

The interaction between VDAC1 and hexokinase not only supports glycolysis but also prevents VDAC1 from binding to pro-apoptotic proteins such as Bax and Bak. These proteins are essential for initiating apoptosis as they facilitate the release of apoptotic factors, such as cytochrome c, from the mitochondria into the cytoplasm. When VDAC1 is bound to hexokinase, its ability to interact with Bax and Bak is inhibited, contributing to apoptosis resistance, another hallmark of cancer cells. In this context, VDAC1 acts as a central modulator linking energy metabolism to cell survival, making it an attractive target for therapeutic strategies.

At the same time, pseudogenes are gaining increasing attention in tumorigenic processes. Traditionally considered non-functional genomic fragments, some pseudogenes have been found to actively regulate gene expression, often acting as ceRNAs (competitive endogenous RNAs) by sequestering microRNAs (miRNAs) shared with their parental genes and modulating their expression. This regulatory

mechanism, similar to that observed for lncRNAs, represents a potential tool for controlling gene expression and is increasingly recognized in many diseases, particularly cancer.

Little is known about the pseudogenes of *VDAC*. Since the factors and circumstances regulating *VDAC* gene expression are still under investigation, understanding whether their pseudogenes are involved or play a role could provide important insights into the regulatory network controlling the expression of *VDAC* isoforms.

To address this gap, this doctoral thesis conducts an in-silico and in vitro analysis of the physiological and clinical implications of the processed pseudogenes of human *VDAC* genes. By leveraging multiple genomic and transcriptomic databases, we observed that some *VDAC* pseudogenes are transcribed in specific tissues and pathological states. Our experimental data revealed a connection between the pseudogene *VDACIP8* and acute myeloid leukemia (AML) and in colorectum cancer (CRC).

A comparative in-silico analysis of the *VDAC1* gene and its pseudogene, *VDACIP8*, combined with experimental data obtained from AML cellular models, revealed that the specific overexpression of *VDACIP8* is associated with the downregulation of the parental *VDAC1* gene. These findings suggest that *VDACIP8* could represent a potential biomarker for AML and open new opportunities to study the regulatory mechanisms of *VDACIP8* and its possible role in the pathogenesis of AML.

Summary

1. Introduction	8
<i>1.1 Pseudogenes</i>	<i>8</i>
<i>1.2 Pseudogenes and Tumor</i>	<i>15</i>
<i>1.2.1 Pseudogenes in Leukemia</i>	<i>17</i>
<i>1.2.2 Pseudogenes in Colorectum Cancer</i>	<i>18</i>
<i>1.3 The role of VDAC mitochondrial genes in human physiology and pathology</i>	<i>19</i>
2. Aim of the PhD thesis	27
3. Material and Methods	28
<i>Section 1</i>	<i>28</i>
<i>3.1 Genomic data analysis</i>	<i>28</i>
<i>3.2 Transcriptomic data analysis</i>	<i>28</i>
<i>3.3 Sequence alignment analysis</i>	<i>29</i>
<i>3.4 Prediction analysis of transcription factor binding sites (TFBSs)</i>	<i>29</i>
<i>3.5 DNA methylation bioinformatic analysis</i>	<i>29</i>
<i>3.6 MethylCap-seq analysis</i>	<i>30</i>
<i>3.7 Cell cultures and treatment</i>	<i>31</i>
<i>3.8 Real-time amplification</i>	<i>32</i>
<i>Section 2</i>	<i>33</i>
<i>3.9 TCGA Data Analysis</i>	<i>33</i>
<i>3.9.1 RNA-seq Data Acquisition</i>	<i>33</i>
<i>3.9.2 Preprocessing and Filtering</i>	<i>33</i>
<i>3.9.3 Differential Expression Analysis</i>	<i>33</i>
<i>3.9.4 Identification of Differentially Expressed Genes (DEGs)</i>	<i>34</i>

3.10 RNA-seq Analysis of VDAC1 and VDAC1P8 in HL60/HTC116 Cell Lines	34
3.10.1 Cell culture and treatment	34
3.10.2 Real-time amplification	34
3.10.3 RNA Extraction and Library Preparation	35
3.10.4 Preprocessing and Alignment	35
3.10.5 RNA Sequencing	35
3.10.6 Gene Quantification and Filtering	36
3.10.7 Differential Expression Analysis	36
3.11 Systems Biology Approach	36
4. Results and Discussions	37
Section 1	37
4.1 Genomic and chromatin context of human VDACs pseudogenes	37
4.2 Expression analysis of VDAC genes and their pseudogenes in normal tissues and their tumor counterparts	40
4.3 VDAC1 and Pseudogenes in AML: Expression and Homology Insights	42
4.4 Phylogenetic Conservation and Functional Insight into the VDAC1P8 in Primates	44
4.5 VDAC1 and VDAC1P8 Expression and Epigenetic Regulation Across Human Development	46
4.6 Common Transcription Factor Binding Sites in VDAC1 and Its Pseudogenes	48
4.7 Epigenetic Regulation of VDAC1P8 Expression in AML	50
4.8 Exploring the Impact of VDAC1 KO on VDAC1P8 Levels in a Leukemic Context	53
Section 2	55
4.9 Comparative analyses between AML and CRC profiles from TCGA dataset	55
4.10 RNA-seq Analysis of VDAC1 and VDAC1P8 in Normal Tissue and HL60/HTC116	58
4.11 Modulation of VDAC1P8 pseudogene expression in HCT116	59
4.12 Epigenetic regulation of VDAC1 gene and VDAC1P8 expression in HTC116	61

5. Conclusions	62
References	66
Paper 1 enclosed to the present Ph.D. thesis	78
Acknowledgements	79

1. Introduction

1.1 Pseudogenes

The term "pseudogene" was first introduced by Jacq et al. in 1977 when they discovered a copy of the 5S rRNA gene in *Xenopus laevis*. This copy exhibited a 5'-end truncation and 14 base-pair mismatches, rendering it non-functional. Their study provided the first insight into pseudogenes, identifying a group of untranscribed genomic sequences that were homologous to the 5S rRNA gene in *Xenopus laevis*, thus establishing the concept of pseudogenes as non-functional relatives of functional genes. Since then, pseudogenes have been identified as being widely present in the genomes of most organisms, ranging from prokaryotes to eukaryotes (Hardison et al. 1979, Proudfoot NJ and Maniatis 1980).

The human genome contains approximately 11,000 pseudogenes, which is more than half the number of protein-coding genes. It has also been observed that pseudogenes outnumber protein-coding genes in the human genome. Their widespread presence suggests that pseudogenes may play an important role in fundamental physiological processes and in the progression of diseases. Traditionally, pseudogenes were viewed as "junk DNA" or "genomic fossils" (Proudfoot N. 1980, Zhang et al. 2003) because, for many years, the consensus was that gene duplication drives evolution by generating new genes (Chen et al. 2013, Ohno 1970). The scientific community considered pseudogenes for a long time as evolutionarily failed sequences. In fact, they are formed through duplications of functional genes or retrotranspositions, and over time, these copies can accumulate mutations, such as frameshifts, deletions, the premature presence of stop codons, or they may be relocated to inactive heterochromatic regions. As stated above, pseudogenes represent a reservoir for gene evolution and/or protein diversity. Most duplicated regions are either deleted from the genome or form disabling mutants of functional paralogs, known as pseudogenes (Innan et al. 2010, Lynch et al. 2000).

A pseudogene is a DNA sequence that resembles a gene but has been mutated into an inactive form over the course of evolution. However, recent advancements in next-generation sequencing and noncoding RNA research have uncovered multilayered functions of pseudogenic DNA, RNA, and proteins. This demonstrates how evolution does not always follow a linear path, but can find unexpected uses for elements previously considered "failed." Pseudogenes play important roles in transcriptional and post-transcriptional regulation and have the potential to evolve into novel genes, serving as a reservoir for gene renewal. Therefore, in the term "pseudogene," 'pseudo' implies sequence variations compared to the parental gene, not necessarily indicating a lack of function. Although sequence mutations render them 'pseudo-' relative to parental genes, many pseudogenes perform real and indispensable functions in physiological and pathological processes (Xiao-Jie et al. 2014).

Pseudogenes are DNA sequences that, while resembling functional genes, can no longer produce active proteins. They represent a crucial aspect of genomic evolution and can be classified into three main categories based on their origin and structure: unitary pseudogenes, duplicated pseudogenes, and retroposed pseudogenes (also known as processed pseudogenes) (Kovalenko et al. 2018, Pei et al. 2012). Unitary pseudogenes are formed as a result of spontaneous mutations that affect single-copy genes. Due to these mutations, the genes lose the ability to be transcribed or translated, becoming non-functional. Unlike other pseudogenes, unitary pseudogenes are not formed through gene duplication but rather from the degradation of an existing functional gene, often due to point mutations or frameshifts that compromise the gene's normal functioning. Despite losing their function, they retain the original intron-exon structure, which is a testimony to their ancient function. A key feature distinguishing unitary pseudogenes is the absence of a functional counterpart in the same species. In fact, these pseudogenes lack paralogs—no functional copies of the same gene exist within the genome.

However, they may still have orthologs in other species, indicating that these genes were once functional before being degraded over evolutionary time (Zhang et al. 2010) (**Figure 1a**). Duplicated pseudogenes, also known as unprocessed pseudogenes, arise through gene duplication events. During such events, which may occur imperfectly, a copy of the gene is duplicated but undergoes mutations that prevent it from functioning properly.

This process results in the loss of essential regulatory elements, such as promoters or enhancers, or the appearance of mutations that introduce premature stop codons or frameshifts. Consequently, the duplicated gene becomes non-functional and transforms into a pseudogene. Although they are no longer active, duplicated pseudogenes retain the intron-exon structure of their functional counterparts, demonstrating their common origin. These pseudogenes are often located near the functional genes from which they originate, for this reason may retain upstream regulatory elements of their parent genes, resulting in a stronger positive correlation with the parent gene and more active transcription than retroposed pseudogenes (Ma et al. 2021). From an evolutionary standpoint, duplicated pseudogenes serve as evidence of unequal crossing-over and other errors in recombination processes that can occur during DNA replication. Moreover, duplicated pseudogenes often retain upstream regulatory elements from their parent genes, which can result in a stronger correlation with the transcriptional activity of the parent gene compared to retroposed pseudogenes. This difference in genomic structure explains why duplicated pseudogenes may have more active transcription, as they often include a copy of the promoter but lack the 3' untranslated region (UTR). This absence limits their role as antisense regulators or miRNA decoys. Duplicated pseudogenes are also more conserved across species than retroposed pseudogenes, maintaining around 28.1% parent gene conservation, compared to just 3.4% in retroposed pseudogenes (Pei et al. 2012) (**Figure 1b**).

Unlike duplicated pseudogenes, retroposed pseudogenes, or processed pseudogenes, result from a process called retrotransposition. In this case, a mature mRNA, lacking introns, is reverse-transcribed into cDNA and reintegrated into the genome at a new location. This means that processed pseudogenes do not contain introns, as they originate from mRNA molecules that have already undergone splicing. A distinguishing feature of these pseudogenes is the presence of a poly(A) tail (Kaessmann et al. 2009, Tan et al. 2016), which during retrotranscription is used as primer, so it remains incorporated into the pseudogene. Additionally, processed pseudogenes are often flanked by short direct repeats (5-20 base pairs) resulting from the integration process. Dong et al. Identified a subgroup of processed pseudogenes that arise from the transcription of circular RNA (circRNA), which do not have the poly(A) tail and display an inverted order of introns compared to the original mRNA, reflecting a different transcription mechanism (Dong et al. 2016).

From an evolutionary perspective, pseudogenes offer an important window into genome dynamics and the evolutionary history of species. For instance, the numerous retroposed pseudogenes found in the human genome are the result of particularly intense retrotransposition activity that occurred around 40 million years ago in our primate ancestors (Pei et al. 2012) [19] (**Figure 1c**).

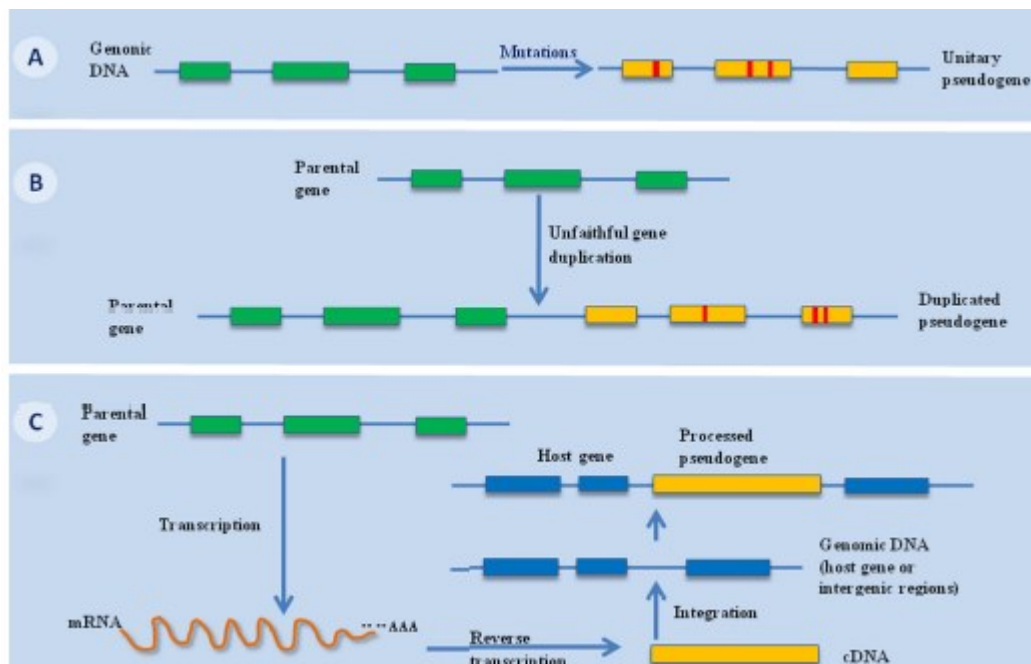


Figure 1a-c. The origin and classification of pseudogene. (a) Gene mutations that abrogate transcription and/or translation turn that ‘gene’ into ‘unitary pseudogene’. (b) Duplicated pseudogenes derive from unfaithful (uncompleted or mutated) gene duplications, with the original copy remaining fully functional (parental gene), whereas the duplicated copy becomes pseudogene. (c) Processed pseudogenes arise from mRNA that is reverse transcribed into cDNA and then integrated into another gene (host gene) or intergenic region. That is why they are often located distant from their parental gene and lack introns. Picture drawn from Xiao-Jie et al. 2014.

Today, over 14,700 pseudogenes have been annotated in the human genome, the majority of which belong to the category of retroposed pseudogenes. An emblematic example includes genes that encode essential proteins for cellular metabolism, such as *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) and ribosomal proteins. Due to their high expression levels, these genes are particularly prone to retrotransposition and, therefore, have numerous processed pseudogenes in the human genome. While pseudogenes are often considered non-functional sequences, they may actually play non-

negligible biological roles. Some studies suggest that, under certain circumstances, they can acquire new functions, a process known as neofunctionalization, or they may influence the gene expression of other functional genes. This ability to reuse or repurpose non-active genetic material makes pseudogenes a fascinating component of genomic complexity. In summary, pseudogenes reflect different mechanisms of genomic evolution and represent traces of ancient duplication, mutation, or retrotransposition events. Their widespread presence in the human genome, particularly among highly expressed genes like those coding for ribosomal proteins and metabolic enzymes, underscores the importance of studying them to better understand the evolution and dynamics of our genetic heritage.

For a long time, pseudogenes were thought to be non-functional. However, advances in analytical techniques have revealed that many pseudogenes are transcribed and play important regulatory roles. The functions and mechanisms of pseudogenes have been extensively described in other works (Chen et al. 2020, Cheetham et al. 2020). In summary, pseudogenes can influence gene expression through interactions at the levels of protein, RNA, and DNA.

Research has so far uncovered: (1) pseudogenes that produce truncated proteins similar to their parent genes, such as *SRGAP2C* and *Notch2NL* (Charrier et al. 2012, Suzuki et al. 2018, Fiddes et al. 2018), though once their protein-coding capacity is recognized, they are no longer classified as pseudogenes (Ma et al. 2021) (**Figure 2a**); (2) pseudogenes that are transcribed to generate antisense RNA sequences of their parent genes, like NOS pseudogenes (Korneev et al. 1999) (**Figure 2b**); (3) pseudogene transcripts that regulate parent gene expression by competitively binding with miRNAs, such as *PTENP1*, *BRAFPI*, and *KRASPI* (Poliseno et al. 2010, Karreth et al. 2015) (**Figure 2c**); (4) pseudogene transcripts that interact with RNA-binding proteins or transcription factors, thereby playing roles in transcriptional or post-transcriptional regulation, such as *HMGAI-p*, *Lethe*, and *RNA5SP141* (Chiefari et al. 2010, Rapicavoli et al. 2013, Chiang et al. 2018) (**Figure 2d**); and (5) pseudogene DNA regions that influence the expression of nearby genes by contributing to or regulating chromatin structure (Huang et al. 2017, Verges et al. 2017) (**Figure 2e**).

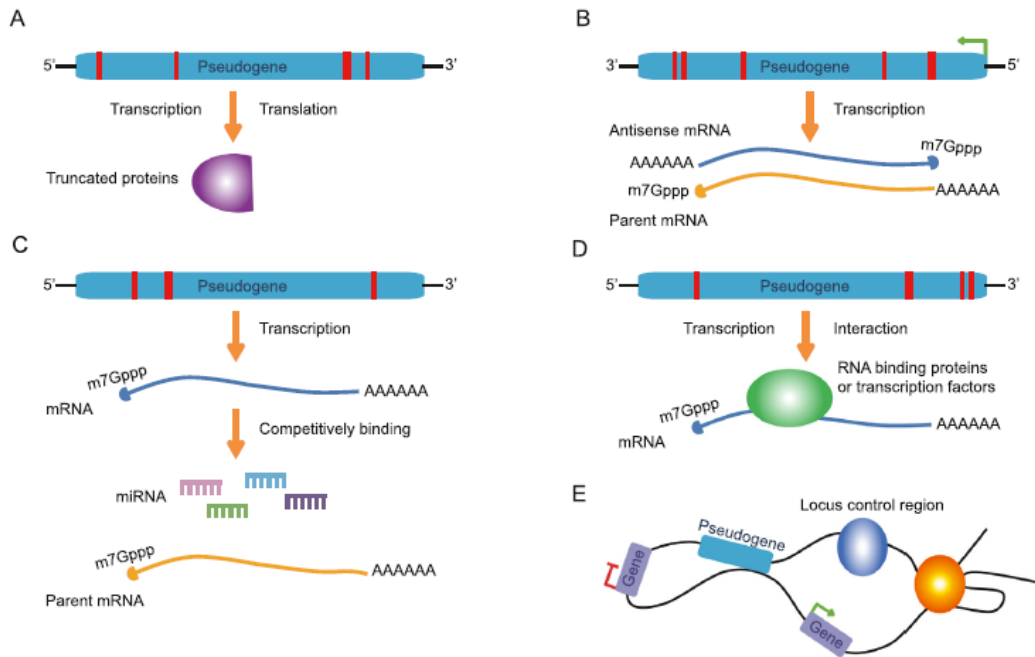


Figure 2a-e. Five models of pseudogene functioning. (a) *Truncated proteins encoded by pseudogenes.* (b) *Pseudogenes transcribe antisense transcripts of their parent gene.* (c) *Pseudogene transcripts competitively bind miRNA to perform regulatory functions.* (d) *Pseudogene transcripts play a regulatory role through interacting with RNA-binding proteins or transcription factors.* (e) *Pseudogenes regulate the expression of genes by affecting DNA folding.*

Given the discovery of functional pseudogenes and their widespread presence in the human genome, it is becoming clear that pseudogenes are not mere “junk” DNA, but rather serve as "gene repositories" that drive evolution. While pseudogenes may be non-functional when they first arise, they can be reactivated and acquire functionality over the course of evolution.

Pseudogene transcripts are classified as long noncoding RNAs (lncRNAs) and therefore share the general characteristics and functionalities of lncRNAs. In addition to these shared roles, pseudogenes can have unique functions based on the extent of sequence similarity to their parent genes. For instance, they can produce small interfering RNAs (siRNAs) that regulate the expression of their parent genes (Watanabe et al. 2008), or generate antisense RNA that pairs with parent gene RNA to form double-stranded RNA (dsRNA), influencing gene regulation (Poliseno et al. 2010).

Pseudogenes that encode truncated proteins may also interact with the full-length proteins of the parent genes, disrupting their function (Fiddes et al. Suzuki et al. 2018, Charrier et al 2012, Cheetham et al. 2020). As a result, pseudogenes may perform paralog-specific roles while acting as lncRNAs. However, unlike the considerable focus on understanding lncRNA functions, pseudogenes have not been thoroughly or systematically studied. They are often excluded from genome-wide analyses and functional screenings that use CRISPR or siRNA libraries (Chen et al. 2020). Therefore, in-depth and systematic research on the functions and mechanisms of pseudogenes is necessary, which would ultimately contribute to projects like the Encyclopedia of DNA Elements (ENCODE).

1.2 Pseudogenes and Tumor

Pseudogenes, once dismissed as non-functional "junk DNA," have emerged as important regulators in various biological processes, including cancer. These genetic elements are copies of functional genes that have lost their protein-coding ability due to mutations such as frameshifts or stop codons. However, recent research has shown that many pseudogenes retain regulatory functions, influencing cancer progression through complex interactions with RNA, DNA, and proteins.

One of the key mechanisms by which pseudogenes contribute to tumorigenesis is through the competing endogenous RNA (ceRNA) network (To et al. 2024). In this model, pseudogenes regulate the expression of their parent genes by acting as molecular sponges that sequester microRNAs (miRNAs), preventing them from targeting and degrading the mRNA of functional genes. For example, the *PTENP1* pseudogene regulates the tumor suppressor gene *PTEN* by competing for miRNAs that would otherwise bind to *PTEN* mRNA. This ceRNA activity results in increased *PTEN* expression, promoting tumor suppression. Dysregulation of this pathway has been implicated in cancers such as breast, prostate, and liver cancer (Poliseno et al., 2010).

In addition to acting as miRNA sponges, pseudogenes can influence cancer by producing truncated or aberrant proteins. *BRAFPI*, for example, has been identified in thyroid cancer, where it produces a shortened version of the BRAF protein. This truncated protein activates the MAPK signaling pathway, which drives cell proliferation and tumorigenesis. Similarly, *KRASPI* plays a role in enhancing the oncogenic activity of *KRAS* in pancreatic cancer, further demonstrating how pseudogene-derived proteins can mimic or interfere with the functions of their parent genes (Stasiak et al., 2021).

Pseudogenes can also regulate gene expression at the transcriptional level. For instance, *HMGAI-p*, a pseudogene of the *HMGAI* gene, interacts with RNA-binding proteins and transcription factors to regulate the stability of *HMGAI* mRNA. This interaction alters the course of tumorigenesis by modulating chromatin structure and transcriptional activity in cancers such as glioblastoma and leukemia (Ma et al. 2021).

Similarly, the *POU5F1B* pseudogene plays a critical role in gastric cancer by regulating the *POU5F1* gene, influencing cell proliferation and tumor development through chromatin remodeling mechanisms (Tutar et al. 2018). Recent studies have also uncovered that pseudogenes can be translated into small peptides that retain partial or altered functions of their parent proteins. For instance, *BRAFPI* in thyroid cancer generates pseudo-proteins that activate oncogenic pathways. This adds another layer of complexity to how pseudogenes contribute to cancer, as they not only modulate gene expression through RNA interactions but can also influence cellular pathways at the protein level (Stasiak et al., 2021).

Due to their diverse roles in cancer biology, pseudogenes have garnered attention as potential diagnostic and prognostic biomarkers. Some pseudogenes, such as *FTHIP3*, have been linked to drug resistance in breast cancer, where their overexpression is associated with resistance to chemotherapy agents like paclitaxel (Stasiak et al., 2021). Pseudogenes can also serve as biomarkers in liquid biopsies, as their transcripts are detectable in blood, urine, and other bodily fluids, providing a non-invasive method for cancer diagnosis and monitoring (Stasiak et al., 2021).

Overall, pseudogenes are emerging as critical players in the regulation of oncogenes and tumor suppressors, influencing cancer progression through multiple molecular mechanisms. Their potential as therapeutic targets and biomarkers makes them an exciting area of research in oncology.

1.2.1 Pseudogenes in Leukemia

Pseudogenes, once considered non-functional remnants of genomic evolution, are increasingly recognized for their regulatory roles in critical biological processes, including hematopoiesis and the leukemogenesis. Hematopoiesis, the process by which blood cells are formed, requires tight regulation of gene expression to ensure proper differentiation and function of hematopoietic stem and progenitor cells (HSPCs). Pseudogenes can influence this process by acting as ceRNAs, sequestering microRNAs (miRNAs) that would otherwise target mRNAs of functional genes. For example, the pseudogene HMGA1-p has been implicated in regulating its parent gene *HMGA1*, a key factor in stem cell proliferation, by binding miRNAs that suppress *HMGA1* expression. This regulatory mechanism is crucial for maintaining the balance between self-renewal and differentiation in HSPCs (Ma et al. 2021). Furthermore, pseudogenes are involved in chromatin remodeling and the modulation of transcription factors essential for hematopoiesis.

The *HBBP1* pseudogene, located in the β -globin locus, is particularly important for erythropoiesis. *HBBP1* is transcribed during erythroid differentiation, where it binds to RNA-binding proteins like HNRNPA1, stabilizing transcripts such as *TALI*, a master regulator of erythropoiesis (Luo et al. 2020). Disruption in the expression or function of such pseudogenes can lead to hematopoietic disorders such as different types of leukemia, underscoring their potential as key regulatory elements in normal and pathological hematopoiesis (Olsson et al. 2016). As our understanding of pseudogenes expands, their role as regulators in hematopoietic processes highlights their significance in both normal development and disease contexts, including hematologic malignancies (Huang et al. 2017). This growing body of evidence suggests that pseudogenes are not merely genomic relics but active participants in the regulation of hematopoiesis, contributing to the fine-tuning of gene networks that govern the formation and function of blood cells.

1.2.2 Pseudogenes in Colorectum Cancer

In colorectum cancer (CRC), pseudogenes have been implicated in tumorigenesis, progression, metastasis, and drug resistance. For example, the pseudogene *PTENP1*, a well-characterized ceRNA, modulates the tumor suppressor *PTEN* by sponging *PTEN*-targeting miRNAs, thereby stabilizing *PTEN* expression. Reduced *PTENP1* levels have been observed in CRC tissues, correlating with enhanced tumor cell proliferation and survival (Poliseno et al., 2010). Another example is *DUSP5P1*, a pseudogene of *DUSP5*, which has been shown to suppress CRC progression by inhibiting ERK signaling, a critical pathway in cancer cell proliferation (Zhang et al., 2020). Moreover, *PCA3*, although more studied in prostate cancer, has shown altered expression in CRC and potential as a biomarker for diagnosis and prognosis (Dong et al., 2018). The Pseudogene-*CTHRC1P1* was also found to regulate epithelial-to-mesenchymal transition (EMT) in CRC via the Wnt/ β -catenin pathway, indicating a functional role in metastasis (Liu et al., 2022).

Emerging transcriptomic analyses and high-throughput sequencing technologies have facilitated the identification of pseudogene signatures with prognostic and diagnostic value. Integrative studies suggest that certain pseudogene expression profiles are associated with CRC patient survival, tumor grade, and response to therapy (Li et al., 2020). Furthermore, some pseudogenes are involved in immune evasion by influencing immune checkpoint pathways and tumor microenvironment remodeling (Zhao et al., 2021), positioning them as potential targets for immunotherapeutic interventions.

Collectively, pseudogenes represent a novel class of non-coding RNAs that contribute to CRC pathobiology through intricate regulatory networks. Their functional versatility and disease-specific expression make them attractive candidates for biomarker development and therapeutic targeting. As the field of non-coding RNA biology advances, a deeper understanding of pseudogene-mediated mechanisms in CRC will likely lead to innovative diagnostic tools and personalized treatment strategies.

1.3 The role of *VDAC* mitochondrial genes in human physiology and pathology

The human VDACs (Voltage-Dependent Anion-selective Channels) gene family consists of three members, *VDAC1*, *VDAC2*, and *VDAC3*, encoding pore-forming proteins of the outer mitochondrial membrane (OMM) involved in the permeability of metabolites (ATP/ADP, NAD^+/NADH) and ion flux (Cl^- , K^+ , Na^+), the calcium homeostasis, the fatty acid and cholesterol transport, and the regulation of apoptotic signaling events (**Figure 3**) (Shoshan-Barmatz et al 2010, De Pinto et al. 2022).

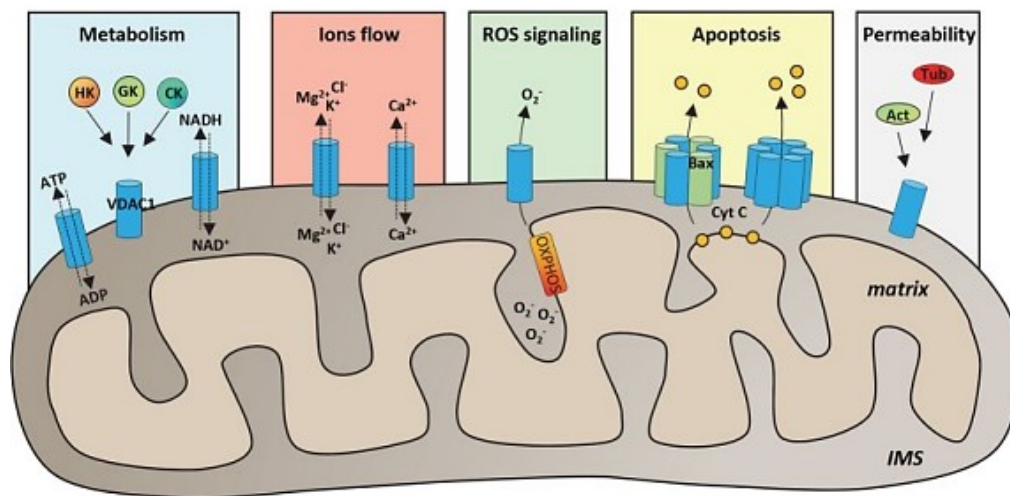


Figure 3. Schematic representation of VDAC functions. VDACs allow metabolite and ions exchanges across the OMM. VDACs mediate cellular energy production by transporting ATP/ADP, NADH, and Krebs cycle's intermediates. VDAC1 controls the flux of Mg^{2+} , Cl^- and K^+ ions across the OMM and participates in maintaining cytosolic Ca^{2+} levels in the physiological range. Furthermore, VDAC proteins are responsible for superoxide anion release to the cytosol. VDACs participate in numerous cellular pathways thanks to interacting with cytosolic enzymes and apoptotic factors. It has been shown that many cytoskeleton proteins, such as Actin (Act) or Tubulin (Tub), bind VDAC participating in the regulation of channel permeability. Picture from: Reina et al. 2017.

It has been long recognized their involvement in many pathologies, from tumors (Mazure et al. 2017) to neurodegenerative diseases like ALS (Israelson et al. 2017), Parkinson's disease (Rostovtseva et al. 2015), Alzheimer's disease (Reddy et al. 2013), type 2 diabetes (Zhang et al. 2019), encephalomyopathy (Huizing et al. 1997), but also in the design of potential therapeutic targets (Magri et al. 2018, Reina et al. 2017, Ben-Hail et al. 2016).

Their 3-D protein structure shows a transmembrane channel consisting of antiparallel 19 β -strands with an additional N-terminal region positioned inside the pore, containing elements of α -helix considered to play an essential role in protein stability and functionality of the open state of the channel underlying the molecular mechanism driving the gating of VDAC pores (Ben et al. 2016, De Pinto et al. 2008, Bayrhuber et al. 2008, Hiller et al. 2008, Ujwal et al. 2008).

The structure of VDAC2 is very similar to that of VDAC1, as shown by X-ray crystallography and solid-state NMR spectroscopy (Eddy et al. 2019, Schredelseker et al. 2014), while the structure of VDAC3 has not yet been obtained except through bioinformatic predictions suggesting an extensive similarity (Amodeo et al. 2014). Moreover, results achieved from functional experiments of VDAC channels reconstitution in artificial membranes have demonstrated that VDAC isoforms are able to form pores (De Pinto et al. 2008, Messina et al. 2011, Raghavan et al. 2012) showing that VDAC1 and VDAC2 have a similar conductance, ion selectivity and voltage dependence, while VDAC3 channel has revealed distinct features (Checchetto et al. 2014, Queralt-Martín et al. 2020, Reina et al. 2010, Saletti et al. 2018, Reina et al. 2020). Such variation have been recently addressed to the differences in number and distribution of cysteine residues since in humans, VDAC1, VDAC2, and VDAC3 have two, nine, and six cysteines, respectively (Pittalà et al. 2020). Indeed, cysteine residues are reactive amino acids that, through their redox-reactive thiol group (-SH), can undergo various oxidative post-translational modifications.

ROS generated by mitochondria regulate mitochondrial protein interactions and activity via oxidative modification of cysteine residues, it is possible that the different functional properties exhibited by the three isoforms may be associated with the cysteine content. This led to the recent hypothesis that the highest amount of cysteine content found in VDAC3 seems to correlate with the function of buffer the excess of mitochondrial ROS (Saletti et al. 2018, Pittalà et al. 2020).

Given the peculiar differences shown by the three isoforms, such as in the mitochondrial outer membrane localization (Neumann et al. 2010), the channel functionality and voltage dependence (Checchetto et al. 2014, Reina et al. 2020), the contribution of the N-terminal portion to cell viability and survival, and the cysteine content of sequence, it is highly presumable a more specialized role/function for each isoform in different biological contexts (Messina et al. 2012).

Analyzing the genome organization of *VDAC* isoforms in mammals, the genes are located in different chromosomes and share a similar exon organization, conserved among the whole eutherian group. Specifically, the genes have the same number of coding exons, characterized by the same size (Messina et al. 2012). However, *VDAC2* represents an exception since it contains an additional first exon encoding for the short pre-sequence of 11 amino acids, as shown in **Figure 4**.



Figure 4. Structure of genes coding for *VDAC* isoforms in humans. *VDAC1*, *VDAC2* and *VDAC3* show similar gene organization. The arrows show the Start codon. The ball and stick show the Stop codon position. The genes have the same number of coding exons with the same size (hg38 assembly). *VDAC2* represents an exception, since it contains an additional first exon encoding for the short pre-sequence of 11 amino acids. Picture modified from: Messina et al. 2012.

The three isoforms of mammalian *VDAC* are expressed ubiquitously, although at varying levels in different tissues, and they are considered functionally distinct, despite sharing a high sequence (about 70%) and structure similarity (Hinsch et al. 2004). Several evidences highlighted specific functions for each isoform, attributing to *VDAC1* a pro-apoptotic function (Anflous et al. 2001), while *VDAC2* an anti-apoptotic protein (Cheng et al. 2003).

Concerning *VDAC3*, it has been associated with sex tissue development and maintenance (Sampson et al. 2001, Cassarà et al. 2009). The *VDAC1* gene is the most abundant and best characterized isoform. Its location allows the control of metabolic cross-talk between mitochondria and cytosol, playing a crucial role in the regulation of mitochondrial metabolic and energetic function (**Figure 5a**).

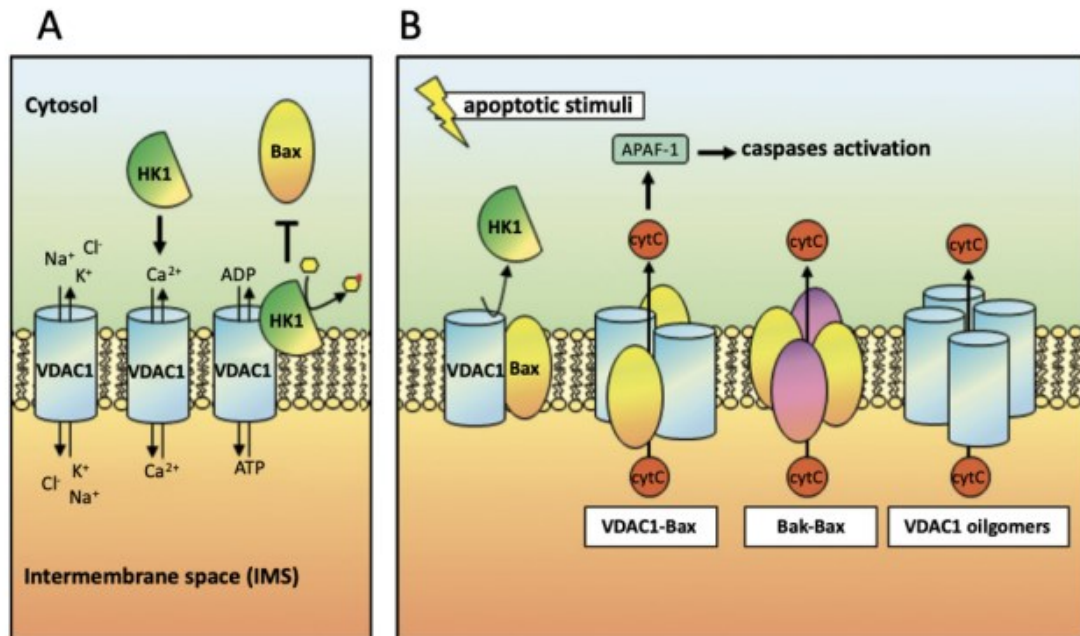


Figure 5a-b. Schematic representation of VDAC1 function in the cell and its involvement in apoptosis. (a) *VDAC1* mainly participates in the transport of ions and small hydrophilic molecules. Furthermore, *VDAC1* is the principal mitochondrial receptor of *HK1*. The binding *VDAC1*-*HK1* impairs the interaction with *Bax* (a pro-apoptotic factor). (b): *VDAC1* participation in apoptosis. In the presence of apoptotic stimuli, the detachment of *HK1* from *VDAC1* leads to the recruitment of *Bax* in the OMM and its interaction with *VDAC1*. The active *Bax* can also associate with *Bak*, forming hetero-oligomers which increase the permeability of the OMM. In Addition, *VDAC1* homo-oligomers participate in the cytochrome *c* (*cytC*) release to the cytosol. Picture modified from Magri et al. 2017.

VDAC1 deficiency indeed reduces mitochondria metabolism in *Vdac1*^{-/-} mice (Okada et al. 2004) confirming the importance of this protein as a carrier of metabolites across the OMM. *VDAC1* is widely considered a pro-apoptotic protein due to its binding to *Bax* (**Figure 5b**). Notably, this interaction influences channel activity and leads to the formation of *VDAC1*-*Bax* hetero-oligomers involved in the

release of Cytochrome C (Cyt c) (Vander et al. 1999). In addition, apoptotic stimuli can induce oligomerization of VDAC1, likewise participating in the passage of Cyt c to the cytosol. The two main isoforms of hexokinases (HK), known as HK1 and HK2, bind to VDAC1 with a different level of affinity to obtain direct access to mitochondrial ATP (Fiek et al. 1982).

Specifically, HKs compete with Bax for binding to VDAC1, by reducing the formation of VDAC1-Bax complexes. Apoptosis stimuli result in the detachment of HKs from VDAC1, increasing the propensity of the porin to bind Bax, thereby initiating the apoptotic process (Pastorino et al 2002, Abu-Hamad et al. 2008). Therefore, VDAC1 has become interesting from the oncology drug development (Magri et al. 2018).

Concerning VDAC2, several evidences have documented an anti-apoptotic function, aside transporting metabolites. Genetic studies have shown that *vdac2*^{-/-} mice are embryonic lethal, whereas both *vdac1*^{-/-} and *vdac3*^{-/-} mice are viable (Craigie et al. 2008). Furthermore, the effect of VDAC2 knockout on embryonic lethality cannot be compensated by overexpression of VDAC1 and VDAC3. VDAC2 is considered essential for cell survival, associated with an anti-apoptotic role because of the binding of VDAC2 with the pro-apoptotic protein Bak (Bcl-2 homologous antagonist/killer) in the OMM prevents its oligomerization, inhibiting Bak-dependent mitochondrial apoptosis (Sampson et al. 2001).

In this regard, VDAC2 plays an intriguing and controversial role showing to promote apoptosis interacting with BAX (BCL2 Associated X, Apoptosis Regulator), but not BAK, as recently found in acute myeloid leukemia (Chin et al. 2001). Another specific activity credited for VDAC2, is the regulation of the calcium homeostasis given the most affinity for Ca²⁺ ions (Subedi et al. 2011 Shimizu et al. 2015, Min et al 2012).

In cardiomyocytes, co-localization studies indicated an interaction between VDAC2 and ryanodine receptor type-2 (RyR2) at the sarcoplasmic reticulum (SR)– mitochondrial junctions in the sub-sarcolemmal regions. These junctions are indispensable for Ca²⁺ transfer from the sarcoplasmic reticulum to the mitochondria (Maurya et al. 2017, Prasad et al. 2015) . VDAC2 is also involved in steroidogenesis via interaction with steroidogenic acute respiratory protein (StAR) in the mitochondria-associated endoplasmic reticulum (ER) membrane (MAM) (Naghdi et al. 2016). Several other VDAC2-

interaction partners have been implicated in multiple cellular functions (Veresov et al.2014), summarized in **Figure 6**.

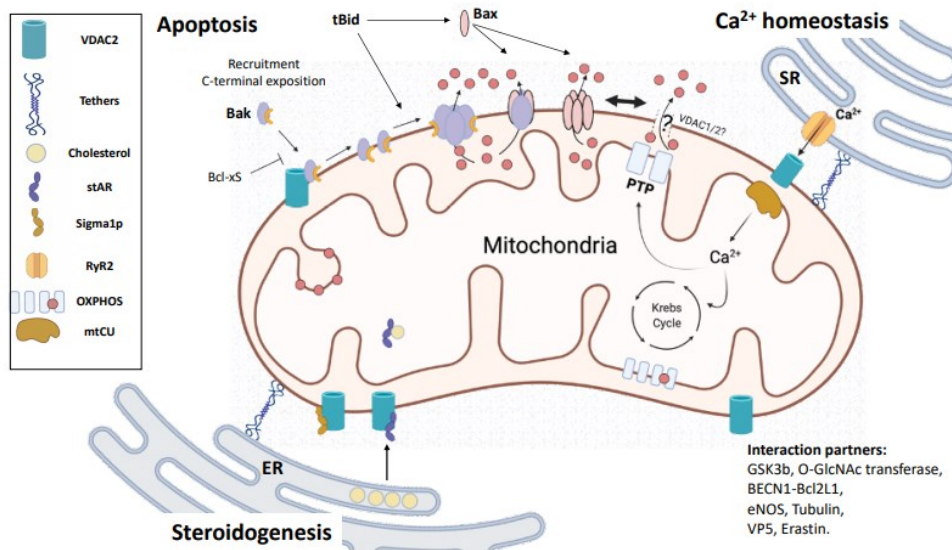


Figure 6. Schematic representation of VDAC2 function and interaction partners in the cell.

Apoptosis: VDAC2 plays a crucial role in apoptosis. In the healthy cells, VDAC2 sequesters Bak into the OMM in an inactive conformation, while Bax is present in the cytosolic fraction. When apoptosis is induced, truncated Bid form (tBid) translocates into OMM, where it displaces Bak from VDAC2 (tBid shows stronger affinity to VDAC2 as compared to Bak). Both Bax and Bak can homo- or hetero-oligomerize in the OMM to produce pores large enough for release of cytochrome c. Ca²⁺ homeostasis: In cardiac cells, VDAC2 plays a major role in the transfer of Ca²⁺ from SR to the mitochondria by interacting with RyR2. Steroidogenesis: Cholesterol is the substrate for steroid hormones, and it needs to be transported from ER to internal mitochondrial membrane for the first step of steroidogenesis. The transport of cholesterol in a hydrophilic environment is facilitated by stAR. Prasad et al., showed that stAR interacts with VDAC2, thus increasing the residence time of StAR at the MAM before it is transported across the mitochondrial membranes. Picture modified from: Veresov et al. 2014.

VDAC3 is the least known isoform. Unlike VDAC1 and VDAC2, which are ubiquitously expressed in all eukaryotes, VDAC3 exhibits a restricted organ distribution and it prevails in the cerebral cortex, liver, heart, testes, and spermatozoa (Sampson et al. 2001). This isoform seems to be essential in sperm mobility since VDAC3 knockout mice show an ultrastructural modification in the epididymal axoneme and in disassembly of cilia during the cell cycle by targeting Mps1 protein kinase to centrosomes

(Majumder et al. 2013). Literature available to date reveals that *VDAC3* considerably differs from isoforms 1 and 2. The complementation assay in *Saccharomyces cerevisiae* strain devoid of endogenous porin (Δ por1) has been extensively exploited for functional studies on VDAC proteins (Reina et al. 2016, De Pinto et al. 2010, Leggio et al. 2018, Magri et al. 2016). Δ por1 cells are unable to grow on glycerol-based media at elevated temperature (37°C), a condition where the yeast cells are mitochondrial-dependent for its survival. As reported, the heterologous expressions of mammalian *VDAC1* and *VDAC2* fully complement the growth defect associated with gene loss, while *VDAC3* only partially rescues the porin-less phenotype (De Pinto et al. 2010).

Initially, *VDAC3* was considered unable to form channels since its pore-forming activity has long been quite hard to be detected (Xu et al. 1999). Subsequently, it has been reported that under non-reducing conditions, VDAC3 inserts into artificial membranes as small and ungated channels with an average conductance of about 0.09 nS in 1 M KCl. Nevertheless, when refolded under highly reducing conditions or pre-treated with reducing agent (*e.g.*, DTT) (Conti Nibali et al. 2021, Okazaki et al. 2015), VDAC3 forms typical voltage-dependent pores displaying similar channel behavior with VDAC1 and VDAC2.

As reported in Okazaki et al. 2015, replacement of specific cysteine residues in human VDAC3 (as well as the replacement of all cysteines to alanine in Reina et al. 2020) affects both electrophysiological features and its ability to complement the growth defect in Δ por1 yeast cells. Overall, these results suggest that the cysteine residues of VDAC3 are involved in modulating channel activity. Despite some significant differences among VDAC, a realistic framework of their identities and the functional relationship within the cell has not yet provided. Many queries are still to be investigated. One of the unsolved questions is whether a functional correlation exists between the different isoforms and the abundance of cell-specific isoforms. In this sense, the activation of an individual isoform of *VDAC* gene, or the expression of more isoforms is the manifestation of cell type-and condition-specific regulatory mechanisms. The identification of a cell program in the expression of a predominant transcript isoform and specific alternatives could be useful to determine the signaling pathways of the cooperative relationship among VDAC isoforms and the functional activity of the cell type-specific transcript.

To this end, determining whether exists a mutual relationship or a gene coordination might also suggest the existence of a kind of isoform-specific protein–protein interactions leading to a tunable protein synthesis depending on specific cellular stimuli.

The assumption of a systems biology and genome approach may reveal some insights in their regulations showing some interesting insights in the basal and specific expression. A recent analysis of the promoter regions of each isoform has characterized the most represented TFs underlying some common mitochondrial-related mechanisms among which cell proliferation and differentiation, apoptosis, and metabolism regulation, while highlighted a set of unique TFs implicated in energy balance (*VDAC1*), in neurodevelopment (*VDA2*) and in germinal development (*VDAC3*) (Zinghirino et al. 2020, Zinghirino et al. 2021).

2. Aim of the PhD thesis

The first part of this Ph.D. thesis aims to analyse the pseudogenes of the *VDAC* family, with a particular focus on *VDACIP8*, using an *in-silico* approach and genomic and transcriptomic high-throughput databases. The objective was to identify and characterize *VDAC* pseudogenes, highlighting potential transcriptional or regulatory activities. The analysis revealed that *VDACIP8*, located in a genomic region enriched with active regulatory elements, is expressed in various normal tissues and shows elevated expression in AML. Furthermore, studies on its epigenetic status and the influence of methylation suggested that *VDACIP8* is regulated by epigenetic mechanisms, potentially serving as a biomarker in AML as reported in Pappalardo et al. 2023.

The second part of the thesis focuses on the experimental validation of the hypotheses derived from the *in-silico* analysis, with particular attention to the role of *VDACIP8* as a ceRNA that may modulate the expression of its parental gene, *VDAC1*. Through gene expression analysis in AML and CRC cell lines and the use of techniques such as treatment with demethylating agents and *VDAC1* silencing, it was demonstrated that *VDACIP8* is regulated by methylation and that there is an inverse relationship between its expression and that of *VDAC1*. These findings suggest an active role for *VDACIP8* in tumours and open new avenues for studying the regulatory mechanisms associated with *VDACIP8* and its pathological implications in AML and CRC.

To expand on these observations, we also analyzed the expression profiles of *VDAC1* and *VDACIP8* in colorectal cancer (CRC). Interestingly, *VDAC1* maintained relatively stable expression, consistent with its role as a housekeeping gene, whereas *VDACIP8* displayed no significantly increased expression in CRC. This pattern raises the possibility that *VDACIP8* contributes to tumorigenesis in a tissue-specific manner and warrants further investigation into its functional impact in solid tumors such as CRC.

3. Material and Methods

Section 1

The following methodology derived from the Paper 1 enclosed to the present Ph.D. thesis:

Pappalardo, X. G., Risiglione, P., Zinghirino, F., Ostuni, A., **Luciano, D.**, Bisaccia, F., De Pinto, V., Guarino, F., & Messina, A. (2023). *Human VDAC pseudogenes: an emerging role for VDACIP8 pseudogene in acute myeloid leukemia*. *Biological research*, 56(1), 33.

<https://doi.org/10.1186/s40659-023-00446-1>

3.1 Genomic data analysis

UCSC Genomic Browser (assembly GRCh38/hg38) was used to analyze the genomic structure of *VDAC* pseudogenes and gene regulatory data from hub tracks' selection. This included Pseudogene Annotation Set of GENCODE v.38lift37 (Ensemble 104), Eukaryotic Promoter Database (EPD) v.4–6, CpG islands track, Genotype-Tissue Expression (GTEx) RNA-seq v.8 (2019), ChIP-Seq data for RNA polymerase II, H3K4me3 and H3K4me1, and chromatin state segmentation by Hidden Markov Model from ENCODE/Broad project. GRCh37/hg19 assembly was used because it provided a more suitable representation of the genomic trait for data visualization and interpretation. The range of analysis of the genomic context of each pseudogene was set about 1000 kb upstream and downstream from the Refseq annotation. In addition, comparative genomics data were acquired from the Ensembl vertebrate genome browser (<https://www.ensembl.org/index.html>) to synteny search between human *VDAC* genes or their pseudogenes and more evolved primates. Also, the GeneCards—The Human Gene Database (<https://www.genecards.org/>) was queried to refine the genomic information related to *VDAC* pseudogenes.

3.2 Transcriptomic data analysis

GEPIA (Gene Expression Profiling Interactive Analysis) web-based tool and GEPIA2 version were accessed in order to collect gene expression profiles of *VDAC* pseudogenes and their parental genes, in both tumor and healthy tissues [82, 83]. GEPIA provides data of TCGA [84] and GTEx (Genotype-

Tissue Expression project) dataset release v.8 (<https://www.gtexportal.org/home/>). Expression Atlas (www.ebi.ac.uk/gxa/home) [85] was also exploited for the analysis of adult and fetal human tissues in RIKEN functional annotation of the mammalian genome 5 (FANTOM5) project (<https://fantom.gsc.riken.jp/5/>) [86]. In addition, GWAS Catalog at <https://www.ebi.ac.uk/gwas/home> was queried for gene-linked phenotypes/functions in genome-wide association studies.

3.3 Sequence alignment analysis

Sequence alignment analysis was running at free software online (<http://multalin.toulouse.inra.fr/multalin/>). The percentage of homology between the examined sequences was obtained by BLAST website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The search for orthologous genes was performed using the BLAST/BAT tool at Ensembl.

3.4 Prediction analysis of transcription factor binding sites (TFBSs)

The prediction of potential transcription factors (TFs) and their transcription factor binding sites (TFBSs) on sequences associated with human *VDAC1* pseudogenes was conducted by hTFtarget database (<http://bioinfo.life.hust.edu.cn/hTFtarget#!/>) [87]. Each record of candidate TFs for query gene provides the chromosomal location of chromatin immunoprecipitation sequencing (ChIP-seq) peaks in different experimental conditions or cell lines. The association between candidate TF and target sequences is computed through a machine learning method and accumulated data of ChIP-seq peaks, epigenetic modification status of TFBSs and targets, co-regulation of TFs of the same targets, and co-association of TFs.

3.5 DNA methylation bioinformatic analysis

DNA methylation profiles of both regulatory sequences and gene bodies of *VDAC1* and *VDAC1P8* were investigated using data of single-base resolution methylomes (SRMs) retrieved from MethBank v.4.1 (Sep 2020) (<https://ngdc.cncb.ac.cn/methbank>) [88]. Since MethBank contains a large amount of SRM replicates both from healthy and pathological condition (prevalently lung and prostate carcinomas), our search strategy filtered SRM data for ‘human healthy tissues’ selecting epigenetic

change of guanine–cytosine ('GC') base pair found in the 'promoter' and 'gene body'. The average methylation data are calculated as β -value that reflects the methylation intensity at each CpG site. β -values of 0–1 (represented from 0 to 1) indicate signifying percent methylation, from 0 to 100%, respectively, for each CpG site. For illustration purpose, in case of some tissue replicates (adipose, adrenal gland, bladder, breast, lung, placenta, prostate, psoas) we averaged their β -values showing a single record for each sample. We also averaged samples belonging to the same apparatus clustering them into following systems: circulatory (aorta, carotid, right atrium, right and left ventricles, whole blood); digestive (pancreas, small and large intestine, stomach, colon, sigmoid colon, esophagus, esophagus squamous epithelium, adjacent esophageal tissue, gastric); lymphatic (spleen, thymus), nervous (anterior cingulate cortex, cerebral cortex (fetal), frontal lobe, hippocampus, nucleus accumbens, prefrontal cortex (Brodmann area 9), retina).

3.6 MethylCap-seq analysis

A full description of the MethylCap-seq procedure can be found in [89]. In summary, in our experiments 100 ng of genomic DNA from HL60 cells was subjected to enzymatic fragmentation, end repair/A-tailing and adapter ligation using Sure Select SureSelect XT HS2 DNA library preparation kit (Agilent). Adapter ligated DNA was subjected to methylation enrichment using MethylCap Kit (Diagenode) following the manufacturer's protocol, with some little modifications. The amounts of MethylCap protein and magnetic beads were decreased proportionally according to the recommended DNA to protein and beads ratio (0.2 μ g protein and 3 μ l beads per 100 ng DNA input). Single fraction elution with High Elution Buffer was applied. The eluted fraction was purified by Ipure kit V2 (Diagenode). Then, the methylation-enriched DNA libraries were amplified using Sure Select SureSelect XT HS2 DNA library preparation kit (amplification step). During amplification, a unique index from the primer was added to the sequencing adapter for each sample. The amplified libraries were purified using 1 $\times$ AMPure XP beads (Agilent). All final libraries were first quantified using D1000 Screen Tape Assay (Agilent). Libraries were then sequenced by NextSeq 2000 platform (Illumina; San Diego, CA, USA), 2 $\times$ 150 bp paired-end read.

FASTQ files were processed using a bioinformatic pipeline for data analysis described in Chemi et al. [90]. Briefly, FASTQ files were trimmed using Agilent AGeNT tool (<https://www.agilent.com/cs/library/usermanuals/public/G9983-90000.pdf>), they were aligned to the GRCh38 reference genome using bwa mem [91]. BAM files were deduplicated using Picard markduplicates (<https://gatk.broadinstitute.org/hc/en-us/articles/360037052812-MarkDuplicates-Picard>) followed by running Samtools [92] fixmate to assign mate quality scores. The QSEA R package [93] was used to analyze BAM files, with the use of a custom R package to extend QSEA, as described in Chemi et al. [90]. The entire genome was tiled into 300-bp non-overlapping windows, with the removal of windows lying within the encode exclusion list regions. Reads were then uniquely assigned into these almost nine millions bins according to their midpoint location. Reads were filtered by keeping read pairs where either end of the pair mapped with a mapping quality (MAPQ) score of at least 10. For paired reads, a fragment length between 50 and 1000 bp was required and a distance along the reference genome of at least 30 bp was required. Copy number variations were calculated from the non-enriched input sequencing for each sample, using HMMcopy with base parameters over 1-Mbp windows. Each sample was normalized for library size using TMM (trimmed mean of M values, part of QSEA) with a pooled reference sample. β -values (a scaled measure of methylation between 0 and 1) for each window in each sample were estimated within QSEA using the ‘blind calibration’ method; windows with insufficient reads to estimate a β -value were returned as NAs.

3.7 Cell cultures and treatment

The acute myeloid leukemia model cell lines (OCI AML2, OCI AML3, MOLM13, IMS M2) from American Type Culture Collection (Manassas, VA, USA) and HL60 (from DSMZ cell cultures, Braunschweig, German) were cultured in RPMI-1640 medium (Sigma Chemical Co.—St. Louis, MO, USA) supplemented with 10% fetal bovine serum, 2 mM L-Glutamine, 100 U/ml Penicillin and 100 μ g/ml Streptomycin (EuroClone, Pero, MI, Italy). The cells were grown at 37 °C in a humidified atmosphere containing 5% CO₂ and dissociated with 0,25% trypsin-0,02% EDTA solution. The HAP1 parental (C631) and knock-out for *VDAC1* (HZGHC005706c002) cell lines were purchased from Horizon Discovery (Waterbeach, UK). Cells were maintained as in [58], as was the monitoring of their

proliferation. For cell treatments, 5-Aza-2'-deoxycytidine (AzaD) was purchased from Sigma Chemical Co. (St. Louis, MO, USA) and resuspended in dimethyl sulfoxide (DMSO) at the concentration of 50 mg/mL, respectively. The cells were seeded at $0,1 \times 10^6$ /ml in 25 cm² culture flasks (NUNC, Thermo Scientific, Roskilde, Denmark) and treated with 5 μ M AzaD for 48 h.

3.8 Real-time amplification

Real-time amplification was performed in a Mastercycler EP Realplex (Eppendorf) in 96-well plates. cDNAs from leukemia cells line HL60 were used as template for the experiments. The reaction mixture was composed by 1 μ l cDNA, 0.2 μ M gene specific primers pairs for *VDAC1*, *VDAC1P8-201* and *GAPDH* (Additional file 12: Table S5) and 12.5 μ l of master mix (QuantiFast SYBR Green PCR kit, Qiagen). Analysis of relative expression level was performed using the housekeeping GAPDH gene as internal calibrator by the $\Delta\Delta C_t$ method. Three independent experiments were performed. Data were statistically analyzed by Student t-test. Values of * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ were taken as significant.

Section 2

All datasets, methodologies, and computational pipelines derive from original, unpublished work conducted as part of a thesis currently under submission. Experimental procedures were carried out in accordance with standard bioinformatics practices and are detailed below.

3.9 TCGA Data Analysis

3.9.1 RNA-seq Data Acquisition

RNA-seq data were obtained from TCGA via the GDC Data Portal (<https://portal.gdc.cancer.gov/>). Raw read count data, pre-processed using the STAR aligner, were downloaded in the `htseq.count` format. Only 50 primary tumor and 50 matched normal tissue samples were included in the analysis from a Caucasian cohort, equally gender distributed.

3.9.2 Preprocessing and Filtering

Individual `htseq.count` files were combined into a single raw count matrix (genes $\times$ samples). Genes with low expression (defined as having fewer than 10 counts in more than 50% of the samples) were filtered out to reduce background noise.

3.9.3 Differential Expression Analysis

Differential expression analysis was performed using the DESeq2 package (Love et al., 2014) in the R statistical environment (version 4.3.0). The count data were imported into a `DESeqDataSet` object, specifying a two-group experimental design: normal vs tumor samples. Normalization was performed using DESeq2's internal method based on size factors to correct for sequencing depth. The `DESeq()` function was used to estimate gene-wise dispersion and to fit the negative binomial model.

3.9.4 Identification of Differentially Expressed Genes (DEGs)

Genes were considered significantly differentially expressed if they met both criteria: adjusted p-value < 0.05 (Benjamini-Hochberg correction for false discovery rate) and absolute log2 fold change > 1.

3.10 RNA-seq Analysis of *VDAC1* and *VDAC1P8* in HL60/HTC116 Cell Lines

3.10.1 Cell culture and treatment

The adenocarcinoma colorectal cell lines HCT-116 were cultured in RPMI-1640 medium (Sigma Chemical Co.—St. Louis, MO, USA) supplemented with 10% fetal bovine serum, 2 mM L-Glutamine, 100 U/ml Penicillin and 100 ug/ml Streptomycin (EuroClone, Pero, MI, Italy). The cells are grown at 37°C in a humidified atmosphere containing 5% CO₂. Once the cells reached 75% confluence, they were divided into fractions and propagated or used for the experimental procedures at passages 10-20. The cells were grown at 37 °C in a humidified atmosphere containing 5% CO₂ and dissociated with 0,25% trypsin-0,02% EDTA solution. The cells were transiently transfected with 4ug of pCMS-MtDsRed carrying the encoding sequence of human *VDAC1* or with 10 uM siRNA directed against *VDAC1* mRNA using Lipofectamine 3000 (Life Technologies), as described in the manufacturer's instructions. The empty pCMS plasmid or scrambled siRNA were used as control.

3.10.2 Real-time amplification

After 48 h of incubation, total RNA was extracted using TRIzol® Reagent (ThermoFisher Scientific), according to the manufacturer's instructions. RNA concentration and purity were measured by a spectrophotometer, and 1 µg was used to synthesize cDNA by Applied Biosystems™ PowerUp™ SYBR™ Green Master Mix. Real-time amplification was performed in a Mastercycler EP Realplex (Eppendorf) in 96-well plates. The reaction mixture contained 1 µL cDNA, 0.3 µM gene-specific primers pairs (*hVDAC1*, *hVDAC1P8*, β -actin), and 5 µL of master mix (PowerUp™ SYBR™ Green Master Mix.). Three independent experiments of quantitative real-time were performed in triplicate for each sample. Thermocycling program consisted in a first activation at 95 °C for 15 min, followed by 40 cycles at 95 °C for 15 s, annealing at 57 °C for 15 s, extension at 68 °C for 15 s and a final step at 72 °C for 10 min. Analysis of relative expression level was performed by the $\Delta\Delta C_t$ method using

the housekeeping β -actin gene as an internal calibrator. Data were statistically analyzed by Student t-test. Values of * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ were taken as significant.

3.10.3 RNA Extraction and Library Preparation

Total RNA was extracted from cell lines using a standard phenol-chloroform protocol, followed by quality assessment using the Agilent TapeStation 4150. Only samples with RNA Integrity Number (RIN) > 7.0 were used for library preparation. RNA-seq libraries were prepared using the Twist Total RNA Library Prep Kit with Ribo-Zero rRNA Depletion (Twist Bioscience), according to the manufacturer's instructions. This protocol allows the removal of ribosomal RNA (rRNA) and enables the capture of both coding and non-coding transcripts. A total of 500 ng of input RNA was used per sample. After depletion and fragmentation, first- and second-strand cDNA synthesis, end repair, A-tailing, adapter ligation, and PCR amplification were performed. Final libraries were validated by Qubit fluorometric quantification and Bioanalyzer profiling.

3.10.4 Preprocessing and Alignment

Raw FASTQ files were subjected to quality control using FastQC. Adapter trimming and quality filtering were performed using the wrapper tool Trim Galore!. Reads were aligned to the GRCh38/hg38 human reference genome using STAR (Spliced Transcripts Alignment to a Reference) aligner (Dobin et al., 2013) with default parameters. Only uniquely mapped reads were retained.

3.10.5 RNA Sequencing

Sequencing was performed on an Illumina NovaSeq 6000 platform using paired-end 150 bp reads (PE150). Libraries were pooled equimolarly and sequenced to an average depth of 30 million reads per sample. Base calling and demultiplexing were carried out using Illumina bcl2fastq.

3.10.6 Gene Quantification and Filtering

Read counts per gene were obtained using featureCounts (Subread package), based on GENCODE v38 annotation. Genes with fewer than 10 counts in at least 50% of the samples were filtered out to improve the robustness of downstream analyses.

3.10.7 Differential Expression Analysis

Differential expression analysis was conducted using the DESeq2 package in R (v.4.3.0). The experimental design included three groups: one normal cell line and two tumor-derived cell lines. Count data were normalized using DESeq2's internal method (median-of-ratios), and dispersion estimates were calculated using empirical Bayes shrinkage. Differentially expressed genes (DEGs) were identified based on adjusted p-value < 0.05 and $|\log_2 \text{fold change}| > 1$.

3.11 Systems Biology Approach

The systems biology framework adopted in this study, particularly the integrative use of genomics and epigenetic analyses, was grounded in current literature and adhered to well-established experimental and computational protocols (Hasin et al., 2017; Cavalli & Heard, 2019). The interpretation of genomic data was conducted through a rigorous post-processing pipeline that followed widely accepted guidelines for methodological transparency and reproducibility in omics research (Bustin et al., 2009; Sandve et al., 2013). This pipeline included variant annotation, differential expression and methylation analyses, and multi-omics data integration to enable comprehensive biological interpretation. To contextualize the results within broader functional networks, downstream analyses included pathway enrichment and gene ontology (GO) analysis using curated databases such as the Gene Ontology Resource, KEGG, and Reactome. These steps enabled the identification of biologically relevant pathways and regulatory mechanisms, enhancing the interpretability and translational relevance of the findings.

4. Results and Discussions

Section 1

The following analyses and results derive from Paper 1:

Pappalardo, X. G., Risiglione, P., Zinghirino, F., Ostuni, A., **Luciano, D.**, Bisaccia, F., De Pinto, V., Guarino, F., & Messina, A. (2023). *Human VDAC pseudogenes: an emerging role for VDACP8 pseudogene in acute myeloid leukemia*. Biological research, 56(1), 33.
<https://doi.org/10.1186/s40659-023-00446-1>

4.1 Genomic and chromatin context of human VDACs pseudogenes

Using available bioinformatics resources, we identified pseudogene counterparts for all three *VDAC* isoforms within the human genome. We then characterized the chromatin landscape and transcriptional potential of the genomic loci harbouring these pseudogenes, focusing on features such as chromatin accessibility, enhancer strength (strong vs. weak), and elongation competency.

This integrative analysis allowed us to pinpoint genomic regions potentially permissive to transcriptional activation. Based on data retrieved from NCBI Entrez Gene, the UCSC Genome Browser, and the Ensembl database, we catalogued a total of 13 *VDAC1* pseudogenes, 4 *VDAC2* pseudogenes, and 1 *VDAC3* pseudogene. Regulatory elements proximal to these loci were identified using the UCSC Genome Browser (assembly GRCh38/hg38), and a summary of these findings is presented in **Table 1**.

Table 1. Features and chromatin status of genomic sequences containing *VDAC* pseudogenes, from UCSC Genome Browser GRCh38/ hg38.

Name	Entrez ID	Genomic coordinates			Regulatory build	
		Location	Size (bp)	Strand	5'-Flanking region	3'-Flanking region
(a)						
VDAC1P1	642585	chrX:80184923–80186706	1784	+	Heterochromatin	Insulator and weak enhancer
VDAC1P2	10064	chrX:49396238–49397962	1725	–	Heterochromatin	
VDAC1P3	100310838	chrX:91236801–91238513	1713	–	Weak enhancers, weak and active promoter (overlapping intronic region of <i>PCDH11X</i> gene)	
VDAC1P4	7418	chr1:180403931–180405937	2007	+	weak and strong enhancers, weak transcriptional elongation, weak promoter (overlapping intronic region of <i>ACBD6</i> gene)	
VDAC1P5	10187	chr12:55195669–55197407	1739	–	Heterochromatin	
VDAC1P6	359800	chrY:5074406–5076119	1714	–	Heterochromatin (overlapping intronic region of <i>PCDH11Y</i> gene)	
VDAC1P7	100310839	chr3:77365047–77366767	1721	–	Insulator	Heterochromatin (overlapping intronic region of <i>ROBO2</i> gene)
VDAC1P8 ^a	100310840	chr6:143811645–143813335	1691	+	Weak enhancer, weak transcriptional elongation (proximity to <i>PEX3</i> gene)	Weak and strong enhancer, weak transcriptional elongation (proximity to <i>FUCA2</i> gene)
VDAC1P9	391106	chr1:157693990–157694816	827	–	Heterochromatin	Weak enhancer, weak transcriptional elongation
VDAC1P10	643536	chr1:215549841–215551581	1741	+	Flanking heterochromatin and weak enhancers inside gene body	
VDAC1P11	100310841	chr9:97048814–97050837	2024	–	Weak transcriptional elongation, weak and strong enhancer (overlapping intronic region of <i>ZNF169</i> gene)	
VDAC1P12	100874289	chr13:34656914–34657447	534	–	Heterochromatin	
VDAC1P13	100420568	chr2:42690279–42691137	859	–	Weak enhancers, weak transcriptional elongation (overlapping intronic region of <i>KCNK3</i> gene)	
(b)						
VDAC2P1	54015	chr21:17466509–17467792	1284	–	Heterochromatin (overlapping <i>LINC00478</i> gene)	
VDAC2P2	643996	chr12:9188920–9190119	1200	–	Heterochromatin	
VDAC2P3	401959	chr1:118182962–118184349	1388	–	Insulator, weak enhancer, heterochromatin	Weak transcriptional elongation, polycomb repressed, heterochromatin
VDAC2P4	100420574	chr2:135554739–135555449	711	+	Heterochromatin, insulator, Polycomb repressor	
(c)						
VDAC3P1	341965	chr14:100430801–100431912	1112	+	Weak transcriptional elongation, heterochromatin, Polycomb repressed	Weak enhancer, heterochromatin, Polycomb repressed

The table shows in (a) information on *VDAC1* pseudogenes *VDAC1P1–13*, in (b) on *VDAC2* pseudogenes *VDAC2P1–4* and in (c) about the only one *VDAC3* pseudogene *VDAC3P1*. ^aGenomic position is referred to splicing variant with *ENST00000406025.2* ID at Ensembl database

Several pseudogenes—including *VDAC1P1*, *VDAC1P6*, *VDAC1P7*, *VDAC1P9*, and *VDAC1P10*—were found within heterochromatic regions characterized by low transcriptional activity and weak enhancer signatures. Others, such as *VDAC1P2*, *VDAC1P5*, *VDAC1P12*, all *VDAC2* pseudogenes, and the sole *VDAC3* pseudogene, were mapped to polycomb-repressed domains, suggesting a silenced chromatin context (see **Table 1**). A subset of pseudogenes—*VDAC1P3*, *VDAC1P4*, *VDAC1P11*, and *VDAC1P13*—reside in regions of moderate

transcriptional activity but overlap with introns of unrelated host genes, such as *PCDH11-X*, *ACBD6*, *ZNF169*, and *KCNG-3*.

These overlaps may predispose the pseudogene transcripts to degradation post-splicing, especially in the absence of protective regulatory mechanisms. Among these, *VDACIP8* emerged as particularly noteworthy due to its favourable genomic and chromatin context, which may indicate functional potential. Located on the opposite DNA strand between the 3' ends of *PEX3* and *FUCA2* (**Table 1**), *VDACIP8* lies in a relaxed chromatin environment, enriched with both strong and weak enhancers, and exhibits moderate elongation capacity according to ENCODE segmentation data.

4.2 Expression analysis of *VDAC* genes and their pseudogenes in normal tissues and their tumor counterparts

Using chromatin state data from the UCSC Genome Browser, we mapped several *VDAC* pseudogenes across the human genome and noted their localization within either active or repressed chromatin regions. To investigate their transcriptional relevance, we analyzed expression profiles in normal and tumor tissues using the GEPIA web server, which integrates TCGA and GTEx data for comparative expression analysis.

Among the pseudogenes examined, *VDACIP8* stood out due to its location in an active chromatin environment and its relatively widespread expression across multiple tissue types. In normal tissues, *VDACIP8* expression ranged between 1 and 9 TPM, with particularly elevated levels observed in the uterus, ovary, lung, rectum, colon, and thyroid. In contrast, most tumor tissues exhibited reduced expression of *VDACIP8* compared to their normal counterparts. A notable exception was AML, where *VDACIP8* was markedly upregulated, reaching 17.75 TPM, approximately 2.5-fold higher than in normal hematopoietic tissue.

Other pseudogenes, such as *VDACIP1*, displayed consistently low expression levels across both normal and tumor tissues, generally ranging from 0.01 to 0.9 TPM. *VDACIP2* was detected in a subset of tumor tissues (excluding AML), with values between 0.11 and 0.44 TPM, and was weakly expressed in normal tissues (0.01–0.26 TPM). In contrast, *VDACIP6*, located within a repressed chromatin region, exhibited minimal expression (0.01–0.07 TPM) in both normal and cancerous tissues such as the head and neck, kidney, and bone marrow, with a modest increase noted in colorectal cancers.

Additional GEPIA analysis revealed that pseudogenes like *VDAC1P3*, *VDAC1P4*, *VDAC1P9*, *VDAC1P13*, and *VDAC1P11* were expressed only in select tissues and at very low levels (0.01–0.1 TPM). No detectable expression was found for *VDAC1P5*, *VDAC1P7*, *VDAC1P10*, *VDAC1P12*, and *VDAC2P2–4*, consistent with their genomic positioning in transcriptionally repressive domains.

Interestingly, while the canonical *VDAC* genes are generally upregulated in most tumor types, AML presents an inverse pattern: *VDAC1* expression drops significantly, from 221.12 TPM in normal tissues to 80.01 TPM in AML samples. In parallel, *VDAC1P8* expression increases to approximately 17.5 TPM, mirroring the 2.5-fold elevation observed in earlier analyses. This inverse expression pattern between *VDAC1* and its pseudogene *VDAC1P8* in AML raises the possibility of a functional interplay between the two, suggesting a potential regulatory or compensatory role for *VDAC1P8* in leukemogenesis. These findings underscore the need for further experimental studies to clarify the biological significance of the *VDAC1–VDAC1P8* gene/pseudogene pair in AML pathophysiology.

4.3 *VDAC1* and Pseudogenes in AML: Expression and Homology Insights

The decision to investigate the *VDAC1* pseudogenes in greater detail stemmed from a notable observation: while *VDAC1* expression is significantly downregulated in AML relative to normal tissues, several of its pseudogenes—*VDACIP8*, *VDACIP1*, and *VDACIP11*—are conversely upregulated in AML. This inverse expression pattern prompted a closer examination of the sequence relationships between *VDAC1* and its pseudogene counterparts. To this end, we performed a multi-sequence alignment analysis, comparing each *VDAC1* pseudogene transcript to the full-length *VDAC1* transcript. The alignment revealed a high degree of sequence homology across several pseudogenes. Notably, *VDACIP1* exhibited 97% sequence identity with the *VDAC1* open reading frame (ORF), indicating a nearly identical sequence. Furthermore, certain pseudogenes, including *VDACIP10*, *VDACIP12*, and *VDACIP13*, were found to contain unique sequence extensions beyond the *VDAC1* ORF, suggesting divergence at their 3' or 5' ends. Due to its unique complexity, *VDACIP8* was excluded from this multi-alignment analysis.

Unlike most *VDAC1* pseudogenes, which typically consist of a single processed transcript, *VDACIP8* harbors 14 annotated non-coding splice variants, according to GENCODE v39 and Ensembl release 105, corroborated by all other queried databases. These variants are classified into multiple transcript biotypes: eight as intron-retaining, five as processed transcripts, and one as a transcribed processed pseudogene. Detailed analysis of these splice variants revealed that the *VDACIP8-201* transcript, composed of a single exon, shares 89% sequence identity with the *VDAC1* ORF.

In contrast, transcripts *VDACIP8-210* and *VDACIP8-205*, which contain five and four exons, respectively, show partial overlap: only the first exon aligns with *VDACIP8-201* and demonstrates 87.53% and 89.28% homology with the *VDACI* ORF. The remaining exons in these transcripts do not correspond to any *VDACI* sequence. Similarly, other splice variants of *VDACIP8* contain exonic regions with no homology to the parental gene, making them unlikely to contribute directly to *VDACI*-related functions.

4.4 Phylogenetic Conservation and Functional Insight into the *VDAC1P8* in Primates

Analyzing gene orthology and synteny plays a critical role in uncovering the evolutionary conservation of genes and understanding their potential physiological and pathological functions across species (Ido et al., 2012; Glover et al., 2019). Orthologous genes, derived from a common ancestor, often retain similar functions, providing valuable insights into conserved biological roles. In parallel, synteny analysis, which examines the conservation of gene order within chromosomal regions across species, can further strengthen functional interpretations by highlighting genomic regions that have remained evolutionarily stable—often suggesting essential biological functions (Nehrt et al., 2011).

In this study, we explored the evolutionary conservation of *VDAC* pseudogenes in primates, drawing on published literature and data from the NCBI Entrez Gene database, which reports numerous predicted pseudogenes in *Mus musculus*, including 13 for mVDAC1, and three each for mVDAC2 and mVDAC3. Previous findings by Ido et al. (2012, 2014) indicated the presence of syntenic *VDAC1* pseudogene clusters in the genomes of mice and rats. However, they also noted a lack of syntenic correspondence between rodent and human *VDAC1* pseudogenes, suggesting a degree of evolutionary divergence between these species.

Given the sparse information on *VDAC1* pseudogenes in primates, we focused on identifying potential orthologs of the 13 human *VDAC1* pseudogenes in other primates. Using the Ensembl BLASTN tool, we conducted homology searches across nine species—Marmoset, Macaque, Drill, Baboon, Gibbon, Orangutan, Gorilla, Bonobo, and Chimpanzee. Homologous sequences were found across all examined genomes, representing a range of evolutionary lineages. However, many of these hits corresponded to unannotated genomic regions, overlapped with protein-coding genes, or intersected with non-coding RNAs of unknown function.

Furthermore, the high sequence similarity among the human pseudogenes complicated species-specific assignment of orthologs, as multiple human pseudogenes matched the same primate loci. Due to these ambiguities—and considering the notable overexpression of *VDACIP8* in AML—we narrowed our focus to *VDACIP8*. BLASTN results confirmed that *VDACIP8* is highly conserved across all analyzed primates, from New World monkeys to great apes. We then performed a synteny analysis using Ensembl's "Synteny" browser to assess whether the genomic loci of *VDACIP8* orthologs in primates are retained within the same syntenic blocks as in humans. The analysis revealed that *VDACIP8* orthologs are consistently positioned within conserved genomic regions, specifically in active chromatin segments corresponding to the human 6q24.2 locus, which also contains functionally annotated genes.

This conservation of both sequence and genomic context suggests that *VDACIP8* may hold a functionally relevant role, potentially subject to selective pressures that maintain its position in the genome. These findings underscore the importance of synteny and orthology in identifying functionally significant non-coding elements, and further support the hypothesis that *VDACIP8* may contribute to regulatory or disease-related processes, particularly in hematopoietic malignancies such as AML.

4.5 *VDAC1* and *VDAC1P8* Expression and Epigenetic Regulation Across Human Development

The expression profiles of the *VDAC1* gene and its pseudogenes, with particular emphasis on *VDAC1P8*, were initially examined using the GEPIA server. This analysis prompted a more detailed study of their expression patterns in normal adult and fetal human tissues. To this end, RNA-Seq CAGE data from the RIKEN FANTOM5 project, accessed through the Expression Atlas (www.ebi.ac.uk/gxa/home), were analyzed. The results confirmed that both *VDAC1* and *VDAC1P8* are expressed in a variety of healthy adult and fetal tissues. *VDAC1* shows relatively stable expression levels across both adult and fetal tissues, whereas *VDAC1P8* exhibits markedly higher expression in fetal tissues (**Figure 7b**). In adult tissues, the highest expression of *VDAC1P8* was detected in bone marrow, diencephalon, and heart (**Figure 7a**).

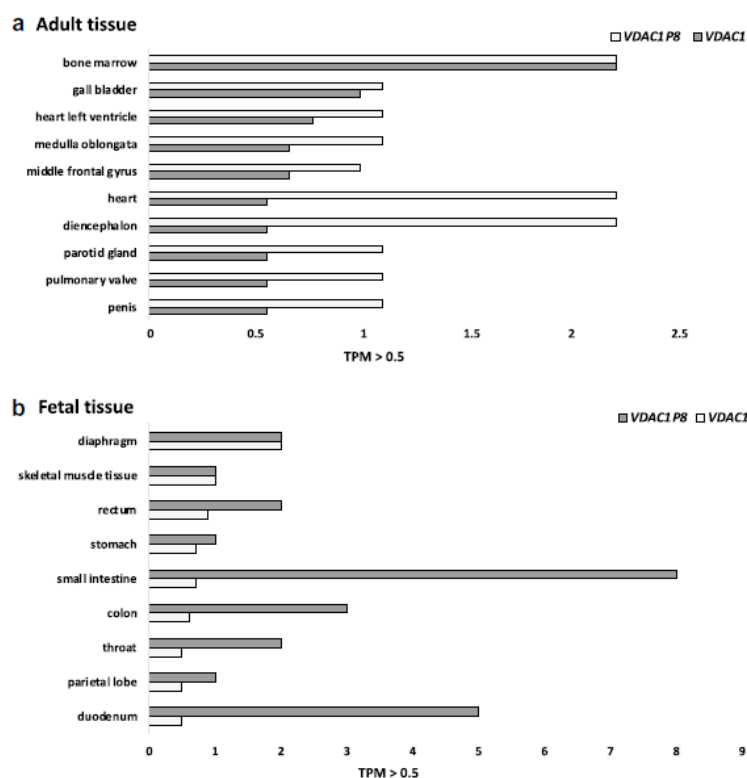


Figure 7a-b illustrates the tissue-specific transcript expression of *VDAC1* and *VDAC1P8* in adult (a) and fetal (b) samples, based on RNA-Seq CAGE data from the RIKEN FANTOM5 project (cut-off > 0.05 TPM). *TPM* stands for transcripts per kilobase million.

In contrast, fetal tissues displayed significant *VDACIP8* expression in the colon, duodenum, small intestine, and throat, with levels up to ten times greater than those of *VDACI* (**Figure 7b**).

These findings suggest that *VDACIP8* is predominantly expressed during fetal development, despite the general trend of pseudogene silencing during embryogenesis. Additionally, correlation analysis of *VDACI* and *VDACIP8* expression across tissues revealed an inverse relationship: tissues with low or no *VDACIP8* expression tend to have higher *VDACI* levels, and vice versa, except in adult bone marrow. Data regarding other *VDACI* pseudogenes were not available in the database.

Further investigation using MethBank explored the methylation status of the promoter and gene body regions of both *VDACI* and *VDACIP8*. As shown in **Fig. 4a and 4b**, methylation levels (β -values) were consistently higher in the promoter and gene body of *VDACIP8* compared to *VDACI* across samples from the central nervous, circulatory, digestive, and lymphatic systems.

Notably, the promoter of *VDACIP8* exhibited approximately tenfold higher methylation than that of *VDACI*. Comparable methylation patterns were observed in other *VDACI* pseudogenes implicated in AML. Although methylation data on *VDAC* genes remain limited, the CG methylation patterns in humans and mice indicate that *VDACI*'s promoter is less methylated than those of other isoforms, supporting its role as a stable housekeeping gene with consistent expression across tissues.

4.6 Common Transcription Factor Binding Sites in *VDAC1* and Its Pseudogenes

Expression analysis of *VDAC* pseudogenes showed that *VDACIP8*, *VDACIP11*, and *VDACIP4* are upregulated in AML, whereas *VDACIP1* and *VDACIP2* exhibit reduced expression compared to normal tissues. To validate these expression trends, we analyzed the transcription factor binding site (TFBS) profiles of *VDAC1* pseudogenes and compared them with those of the *VDAC1* gene. This was done using the hTFtarget database (<http://bioinfo.life.hust.edu.cn/hTFtarget#!/>), focusing on transcription factors relevant to blood and bone marrow. Additional factors—*BRD4*, *CBX3*, *POLR2A*, *RAD21*, and *STAG1*—have recently been identified as epigenetic regulators involved in the hematopoietic system's shift from normal to diseased states [38, 50–52]. In the regulatory region of *VDACIP11*, six TFBS linked to leukemogenesis (*JUND*, *CBX3*, *RCOR1*, *GATA1*, *JUN*, and *FOS*) were identified as potential regulators.

To assess whether *VDAC1* and the pseudogenes *VDACIP8* and *VDACIP11* share similar or distinct regulatory patterns associated with hematological tumorigenesis, we compared their TFBS distributions. The analysis revealed that 19 out of 24 TFBS found in *VDACIP8* and implicated in leukemogenesis are also potential regulators of *VDAC1*, whereas only 3 out of 8 such TFBS were present in *VDACIP11*. Notably, transcription factors *JUND*, *CBX3*, *RCOR1*, and *GATA1* were common regulators for *VDAC1* and both pseudogenes *VDACIP8* and *VDACIP11*. In contrast, *VDACIP1* was associated with *LMNB1*, a gene frequently overexpressed in cancers such as lung adenocarcinoma, breast cancer, and leukemia [53]. Together, these findings imply that *VDAC1* and pseudogenes *VDACIP8* and *VDACIP11* may be regulated by a shared transcriptional network involved in leukemia progression.

Interestingly, only *VDACIP8* and *VDACIP11*—both overexpressed in AML—harbor binding sites for transcription factors known to contribute to tumor development. Specifically, 24 TFBS were identified in *VDACIP8*'s regulatory region (**Table 2**), including transcription factors previously linked to leukemogenesis, such as *MYC*, *TAL1*, *JUND*, *BATF*, *FOS*, *CEBPB/D*, *GATA1/2/3*, *CTCF*, *ZNF384*, *RCOR1*, *SPI1*, *ETS1*, *ERG*, *FLI1*, *IRF4*, and *RUNX1* [34–49].

Table 2 List of transcription factors TFs for *VDAC1P8* detected in blood and bone marrow.

Class	Family	Subfamily	Name	Tissue		
				Bone marrow	Blood	
Basic helix-loop-helix factors (bHLH)	bHLH-ZIP	AP-4	MYC	Myc/Max factors	✓	X
	Tal-related	TAL-HEN	TAL1	Tal-related factors	X	✓
Basic leucine zipper factors (bZIP)	Jun-related	Jun	JUND	JunD proto-oncogene		
	B-ATF-related		BATF	B-ATF-related		
	Fos-related	Fos	FOS	Fos proto-oncogene	✓	X
	CEBP-related	CEBP	CEBPB	CCAAT/enhancer binding protein beta		
			CEBPD	CCAAT/enhancer binding protein delta		
Other C4 zinc finger-type factors	GATA-related	GATA double	GATA1	GATA binding protein 1		
			GATA2	GATA binding protein 2	X	✓
			GATA3	GATA binding protein 3		
C2H2 zinc finger factors	More than 3 adjacent zinc fingers	CTCF-like	CTCF	CCCTC-binding factor	✓	X
		ZNF362-like	ZNF384	Zinc Finger Protein 384	X	✓
Tryptophan cluster factors	Myb-SANT	NCoR-like	RCOR1	REST corepressor 1	✓	X
	ETS-related	SPI-like	SPI1	Spi-1 proto-oncogene	✓	
		ETS-like	ETS1	Ets-protoncogene 1	X	✓
			ERG	Ets-related factors		
			FLI1	Fli-1 protoncogene		
Runt domain factors	Runt-related		IRF4	Interferon-regulatory factors		
			RUNX1	Runt domain factors	✓	
Type		Name		Tissue		
				Bone marrow	Blood	
Other TFs	Epigenetic reader		BRD4	Bromodomain containing 4	✓	
	Heterochromatin protein 1 HP1 family		CBX3	Chromobox 3	✓	X
	Synthesizing messenger RNA		POLR2A	RNA polymerase II subunit A	✓	
	CTCF-transcriptional coactivators		RAD21	RAD21 cohesin complex component	✓	
			STAG1	Stromal antigen 1	X	✓

4.7 Epigenetic Regulation of *VDACIP8* Expression in AML

Following data obtained from public repositories on *VDACIP8* and its parental gene, we conducted a detailed in vitro analysis of the expression levels of the *VDAC1/VDACIP8* pair in various AML cell lines, including HL60, OCI-AML/2, OCI-AML/3, MOLM13, and IMS-M2. These cell lines represent different AML subtypes, such as myelocytic, monocytic, and megakaryocytic. Experimental validation confirmed the expression of both *VDAC1* and *VDACIP8* across all AML cell lines examined, supporting the in-silico data. However, *VDAC1* consistently exhibited higher expression levels than its pseudogene *VDACIP8*. The identity of these genetic products was verified through sequencing of the amplified products (**Figure 8a**).

To further explore this, we assessed the effect of treatment with 5-aza-2'-deoxycytidine (AzaD), a demethylating agent, on *VDAC1* and *VDACIP8* expression in HL60 cells. Results indicated that while *VDAC1* expression remained stable post-treatment, *VDACIP8* expression increased by approximately 1.8-fold compared to untreated control cells (**Figure 8b**). This finding suggests that *VDACIP8* expression may be regulated epigenetically through DNA methylation, unlike *VDAC1*.

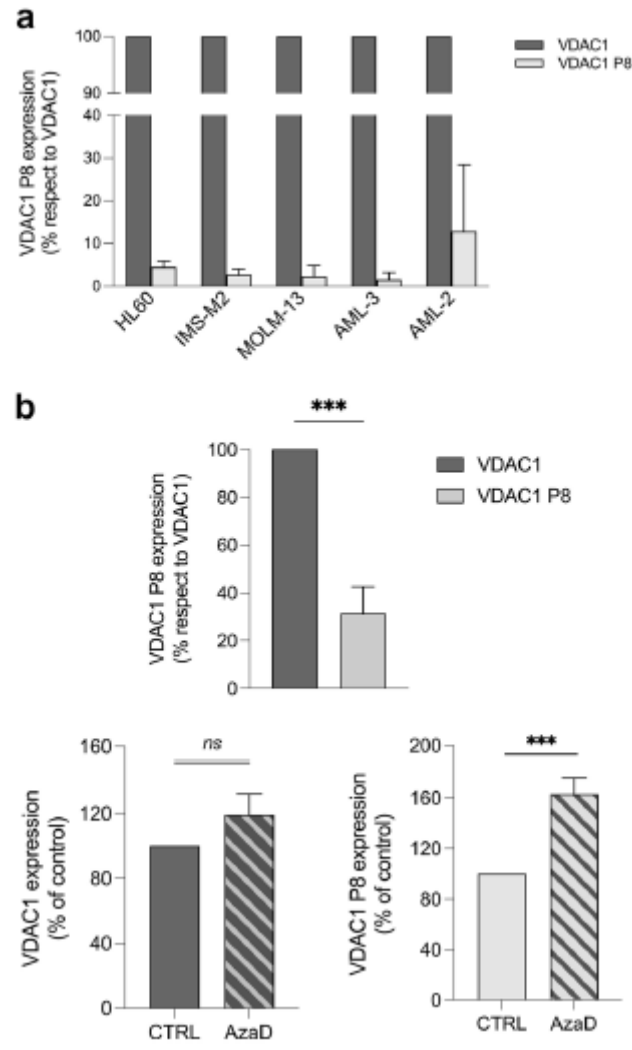


Figure 8a-b. Expression of *VDAC1* gene and *VDAC1P8* pseudogene in HL60 AML cell line. In (a), the expression of *VDAC1* gene and *VDAC1P8* pseudogene was evaluated in different models of leukemia cell line (HL60, IMS-M2, MOLM-13, AML-3, AML-2) by real-time PCR relative quantification. In (b), the expression of *VDAC1* and *VDAC1P8* both assessed in HL60 by real-time PCR relative quantification were also measured following 48 h of AzaD treatment. The variation of mRNA in treated samples were expressed relatively to control samples, after normalization with the housekeeping gene *GAPDH*, by the $\Delta\Delta C_t$ method. *VDAC1P8* expression is reported as % respect to *VDAC1* indicating with a value of 100 the level of *VDAC1*. Data were statistically analyzed by T-test and a value of $P < 0.05$ was considered significant.

Supporting this hypothesis, a methylome analysis was conducted using MethylCap-seq to evaluate potential changes in the methylation patterns of *VDAC1* and *VDACIP8*'s promoter and gene body regions. The analysis revealed no significant methylation changes in the regulatory regions of the *VDAC1* gene after AzaD treatment. However, *VDACIP8* regulatory regions, particularly a region referred to as “promoter 3,” exhibited a clear reduction in methylation in treated cells compared to controls.

Additionally, the entire gene body sequence of *VDACIP8* appeared demethylated in AzaD-treated HL60 cells, contrasting with the methylation status in untreated cells. These findings suggest that *VDACIP8* expression in AML cells is subject to epigenetic regulation, specifically through DNA methylation, while *VDAC1* expression seems unaffected by this mechanism in AML cells. Modulating *VDACIP8* expression via demethylation indicates that this pseudogene could play a role in the molecular mechanisms underlying AML, potentially acting as a biomarker or therapeutic target.

In the context of AML, it has been proposed that the dysregulation of the *VDAC1/VDACIP8* pair may arise from the epigenetic control of *VDACIP8*, which could indirectly influence the expression of the parental gene *VDAC1*. It is possible that *VDACIP8* acts as a regulatory element through non-coding RNA mechanisms (as also suggested in the text), modulating *VDAC1* expression via competition with miRNAs or other regulatory molecules

4.8 Exploring the Impact of *VDAC1* KO on *VDAC1P8* Levels in a Leukemic Context

We performed an experiment designed to explore the potential regulatory link between *VDAC1* and its pseudogene *VDAC1P8* in the context of AML. To this end, two cell lines derived from HAP1 were used: a wild-type line (HAP1 WT) and a *VDAC1* knock-out line (HAP1 Δ *VDAC1*), generated through CRISPR/Cas9 gene editing. The HAP1 cell line was chosen as a model because, being nearly haploid, it facilitates the observation of expression differences without the complications of multiple chromosome copies, although it does not represent a perfect model of AML.

In this experiment, the expression of *VDAC1P8* was compared between the two cell lines to understand whether the absence of *VDAC1* could influence the levels of its pseudogene. In HAP1 WT cells, where *VDAC1* is normally expressed, *VDAC1P8* levels were significantly lower, approximately ten times lower than those of *VDAC1*. However, in HAP1 Δ *VDAC1* cells, where *VDAC1* was silenced, *VDAC1P8* expression was notably increased, doubling compared to *VDAC1* levels observed in the same cells. These results suggest that in the absence of *VDAC1*, the system might compensate by increasing *VDAC1P8* expression, hypothesizing a negative feedback mechanism.

This finding supports the hypothesis that *VDAC1P8* may play a regulatory role in *VDAC1* expression, possibly acting as a competitive endogenous RNA (ceRNA) or through epigenetic control. The data are shown in **Figure 9a-b**, which illustrates the variations in *VDAC1* and *VDAC1P8* expression levels in the two HAP1 cell lines.

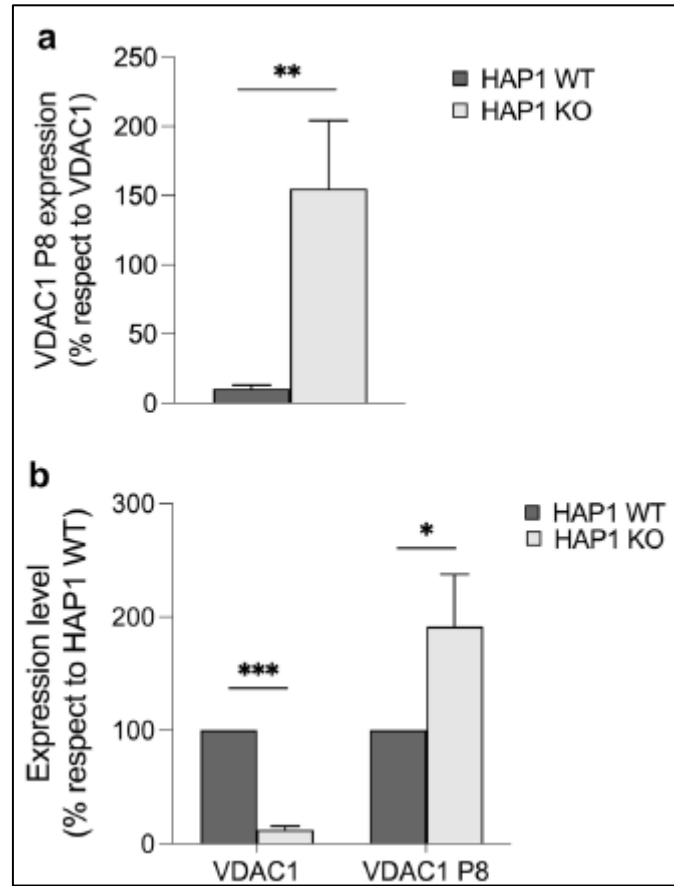


Figure 9a-b. Expression of *VDAC1* gene and *VDAC1P8* pseudogene in *VDAC1* KO leukemic cell lines. The level of *VDAC1* and *VDAC1P8* transcripts expression were evaluated in HAP1 WT and HAP1*VDAC1* cells by real-time PCR relative quantification. In (a) *VDAC1P8* transcript level was measured relatively to that of *VDAC1*. In (b) the expression level of both *VDAC1* and *VDAC1P8* in HAP1*VDAC1* cells was measured relatively to their relative amount found in HAP1 WT. After normalization with the housekeeping gene GAPDH, by the $\Delta\Delta C_t$ method. *VDAC1P8* expression is reported as % respect to *VDAC1* indicating with a value of 100 the level of *VDAC1*. Data were statistically analyzed by T-test and a value of $P < 0.05$ was considered significant.

Section 2

4.9 Comparative analyses between AML and CRC profiles from TCGA dataset

To further explore the transcriptional dynamics of *VDAC1* and its pseudogene *VDAC1P8*, we conducted a series of comparative expression analyses between AML and CRC profiles using data extracted from the TCGA dataset (<https://portal.gdc.cancer.gov/>). Our analyses were focused on a heterogeneous Caucasian cohort including 50 patients and 50 healthy controls, evenly distributed by sex. In the first experiment, we evaluated their expression profiles in AML samples relative to healthy controls (Figure 10).

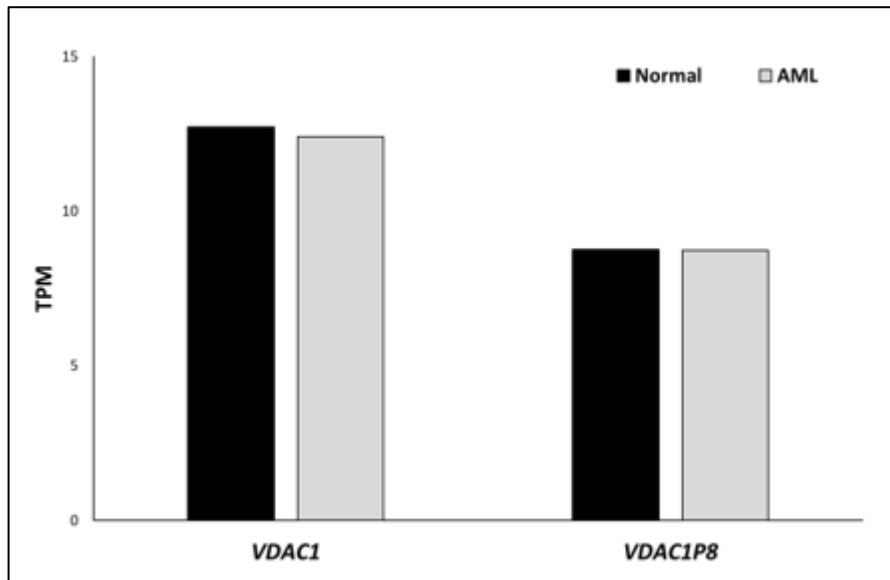


Figure 10. Gene expression levels of *VDAC1* and its pseudogene *VDAC1P8* in AML compared to normal samples (TCGA dataset). *VDAC1* expression remains stable between normal (12.40 TPM) and AML samples (12.72 TPM), while *VDAC1P8* also shows comparable levels (8.76 vs 8.74 TPM), suggesting no major transcriptional deregulation in AML. *TPM*= transcript per million.

The results showed that *VDAC1* expression remained relatively stable, with mean transcript levels of 12.40 TPM in normal samples and 12.72 TPM in AML samples. Similarly, *VDAC1P8* exhibited no significant changes, maintaining comparable levels in normal (8.76 TPM) and AML (8.74 TPM) tissues. These findings suggest an absence of major transcriptional deregulation for both genes in AML (**Figure 10**).

In the second experiment, we assessed CRC samples and observed a slight increase in *VDAC1* expression in CRC (13.98 TPM) compared to normal tissues (13.93 TPM), while *VDAC1P8* showed a minor decrease (6.46 vs. 6.55 TPM) (**Figure 11**). These subtle differences indicate minimal transcriptional variation in CRC.

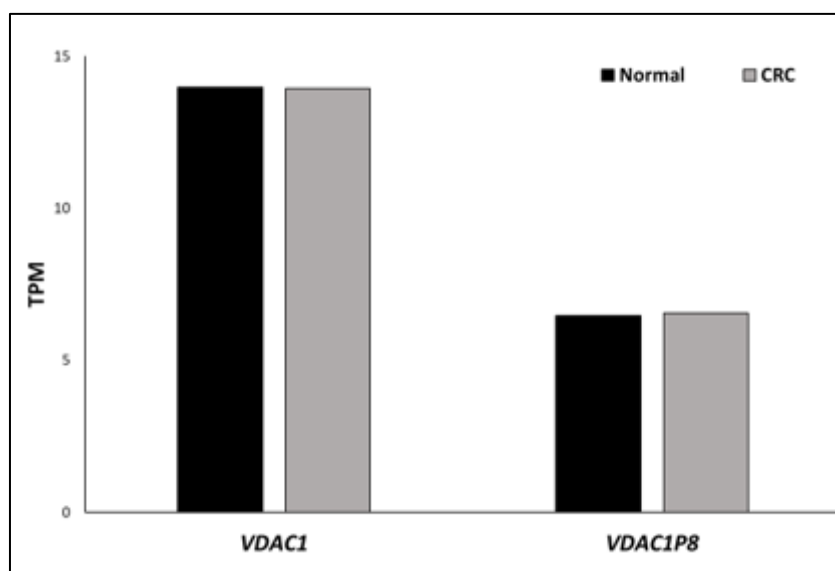


Figure 11. Gene expression levels of *VDAC1* and its pseudogene *VDAC1P8* in CRC compared to normal samples (TCGA dataset). *VDAC1* expression is slightly increased in CRC (13.98 TPM) compared to normal (13.93 TPM), while *VDAC1P8* shows a slight decrease (6.46 vs 6.55 TPM), indicating minimal transcriptional differences in CRC for these genes. TPM= transcript per million.

Finally, a direct comparison of AML and CRC expression profiles revealed that *VDAC1* levels were higher in CRC than in AML, whereas *VDAC1P8* expression was notably reduced in CRC compared to AML (6.46 vs. 8.74 TPM) (**Figure 12**). This contrast points to a potential tissue-specific regulation of the pseudogene, suggesting that *VDAC1P8* may be subject to distinct transcriptional control mechanisms depending on the cancer type.

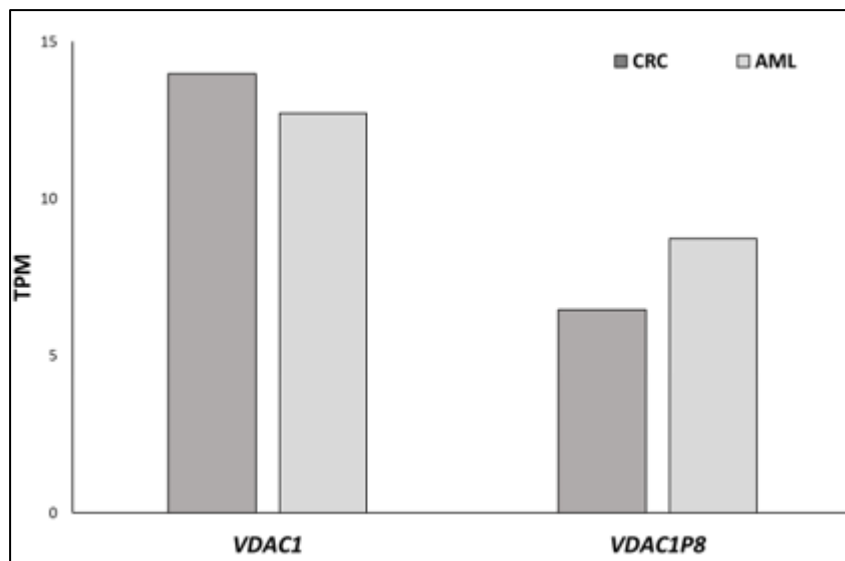


Figure 12. Comparative expression of *VDAC1* and *VDAC1P8* in AML and CRC samples from TCGA. *VDAC1* levels are higher in CRC (13.98 TPM) than in AML (12.72 TPM), whereas *VDAC1P8* expression is reduced in CRC (6.46 TPM) compared to AML (8.74 TPM), suggesting a possible tissue-specific regulation of the pseudogene. TPM= transcript per million

4.10 RNA-seq Analysis of *VDAC1* and *VDAC1P8* in Normal Tissue and HL60/HTC116

RNA-seq analysis was extended to include normal tissue and two representative cancer cell lines, HL60 as AML model, and HTC116 as CRC. *VDAC1* showed consistent expression across all samples, with transcript levels of 13.16 TPM in normal, 13.17 TPM in HL60, and 13.45 TPM in HTC116 cells (**Figure 13**). In contrast, *VDAC1P8* expression dropped significantly in both cancer cell lines—2.53 TPM in HL60 and 3.14 TPM in HTC116—compared to normal tissue (7.65 TPM), highlighting a consistent downregulation of the pseudogene in both hematopoietic and epithelial tumor contexts.

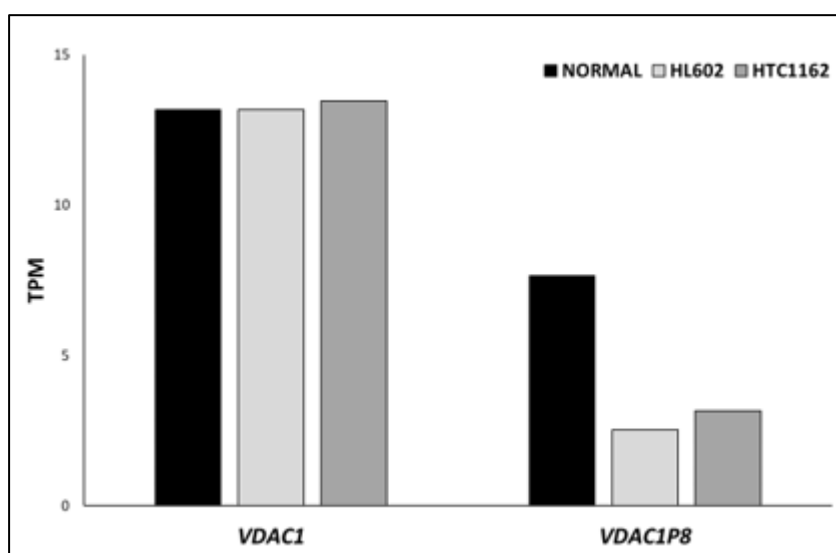


Figure 13. RNA-seq analysis of *VDAC1* and *VDAC1P8* in normal tissue and two cell lines (HL60, HTC116). *VDAC1* shows consistent expression across normal (13.16 TPM), HL60 (13.17 TPM), and HTC116 (13.45 TPM). In contrast, *VDAC1P8* expression drops significantly in HL60 (2.53 TPM) and HTC116 (3.14 TPM) compared to normal (7.65 TPM), indicating pseudogene downregulation in both hematopoietic and epithelial cancer cell lines. TPM= transcript per million.

4.11 Modulation of *VDACIP8* pseudogene expression in HCT116

Based on the information reported in the public databases we found that *VDAC1* mRNA is overexpressed in CRC while the expression of *VDACIP8* is among the lowest found in tumour. To understand whether *VDACIP8* and *VDAC1* mRNA expression are co-regulated by a controlled mechanism as we hypothesized (Pappalardo et al. 2023), the level of *VDAC1* and *VDACIP8* RNA was evaluated in HCT116 cell line of CRC. In this work, we silenced or overexpressed *VDAC1* mRNA and measured the relative amount of *VDACIP8* transcript. Oligonucleotide interfering with *VDAC1* transcript expression silenced the endogenous *VDAC1* reducing its level of 2 folds (Figure 14a). At the same time, *VDACIP8* RNA increased its expression of 1,5 folds in *VDAC1* silenced cells compared to control cells (Figure 14b).

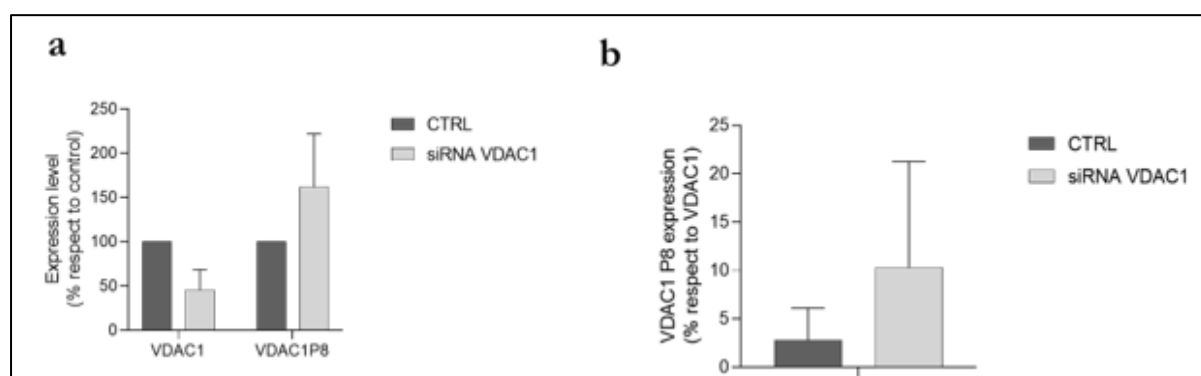


Figure 14a-b. *VDAC1* silenced in HCT116. (a) Silencing of endogenous *VDAC1* mRNA using siRNA results in a 2-fold reduction in *VDAC1* transcript levels compared to control cells. (b) In the same *VDAC1*-silenced cells, *VDACIP8* expression shows a 1.5-fold increase relative to control. These findings suggest that downregulation of *VDAC1* may relieve a repressive effect on *VDACIP8*, indicating a potential inverse regulatory relationship.

The level of *VDAC1P8* expression was also evaluated respect *VDAC1*. We observed that in cells where *VDAC1* is silenced, *VDAC1P8* is overexpressed compared to control cells reaching an increase of 3,3 folds (Figure 15a). On the contrary, overexpressing *VDAC1* mRNA, the level of this gene is increased of 2,5 folds while *VDAC1P8* is only slightly reduced (Figure 15a). Normalizing the *VDAC1P8* expression level respect its parental gene, in cells overexpressing *VDAC1*, we noticed a reduced level of 1,6 folds compared to control cells (Figure 15b). These results suggest a possible control mechanism of transcript regulation correlating the expression of *VDAC1P8* pseudogene and *VDAC1*.

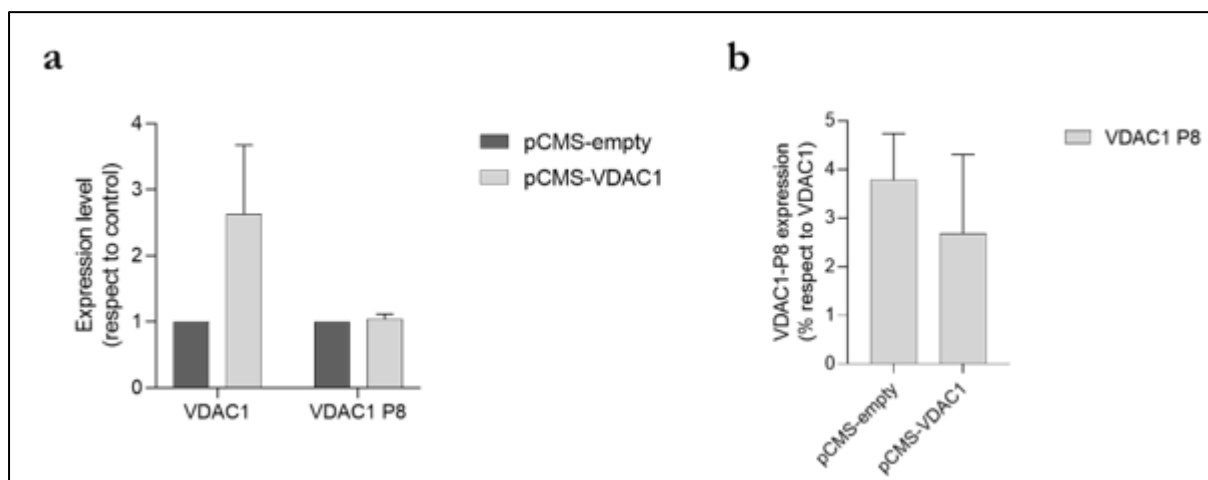


Figure 15a-b. *VDAC1* overexpression in HCT116. (a) qRT-PCR analysis showing that silencing *VDAC1* leads to a 3.3-fold increase in *VDAC1P8* expression compared to control cells. Conversely, *VDAC1* overexpression results in a 2.5-fold increase in *VDAC1* mRNA, while *VDAC1P8* expression shows a slight decrease. (b) Normalization of *VDAC1P8* expression to *VDAC1* levels reveals a 1.6-fold reduction in the *VDAC1P8/VDAC1* ratio in cells overexpressing *VDAC1* compared to controls. These results suggest an inverse regulatory relationship between *VDAC1* and its pseudogene *VDAC1P8*.

4.12 Epigenetic regulation of *VDAC1* gene and *VDAC1P8* expression in HCT116

In HCT116 cell line we tested the expression level of *VDAC1P8* pseudogene relatively to its parental gene *VDAC1* when cells were exposed to an epigenetic control mechanism induced by the demethylating agent 5'-AzaD (Figure 16a-c). Both *VDAC1* and *VDAC1P8* transcripts increased following the treatment, up to 2-fold and 0,5-fold respectively, each at its expression level in the control (Figure 16a-b). Moreover, normalizing the level of *VDAC1P8* expression to *VDAC1* in control and treated cells, we observed that the level of *VDAC1P8* in 5'-AzaD cells compared to control cells was reduced of 0,5-fold (Figure 16b). Although the expression level of *VDAC1P8* pseudogene increased following induced genomic demethylation, this result might suggest that *VDAC1* increase stimulates a reduction of *VDAC1P8* (Figure 16c). Indeed, we hypothesized the regulation of *VDAC1P8* expression is subjected to the level of its parental gene *VDAC1* and/or viceversa (Pappalardo et al. 2023).

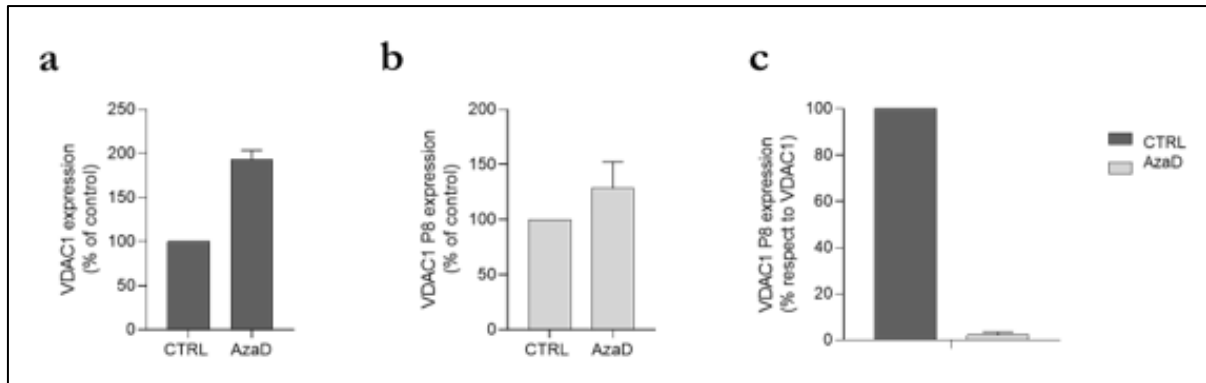


Figure 16a-c. Expression of *VDAC1* gene and *VDAC1P8* pseudogene in HCT116 cells. Expression analysis of the *VDAC1P8* pseudogene and its parental gene *VDAC1* in HCT116 cells after treatment with the DNA-demethylating agent 5'-AzaD. (a) qRT-PCR showing a 2-fold increase in *VDAC1* and a 0.5-fold increase in *VDAC1P8* transcript levels compared to untreated control cells. (b) Relative expression of *VDAC1P8* normalized to *VDAC1*, revealing a 0.5-fold reduction in *VDAC1P8*/*VDAC1* ratio in 5'-AzaD-treated cells versus controls. (c) Schematic interpretation suggesting that despite increased *VDAC1P8* expression upon demethylation, the proportionally greater upregulation of *VDAC1* may exert a regulatory influence, leading to a relative reduction in *VDAC1P8* levels. These results indicate a possible interplay between the pseudogene and its parental gene under epigenetic modulation.

5. Conclusions

In recent years, transcribed pseudogenes have garnered increasing attention for their functional roles in gene regulation, particularly through their integration into the complex non-coding RNA network (Ma et al. 2021, Polisenio et al. 2010, Proudfoot et al. 1980, Xiao-Jie et al. 2015). Despite growing interest in this field, our understanding of *VDAC* pseudogenes remains limited, especially considering the pivotal role of *VDAC* isoforms in cellular metabolism, apoptosis, and tumor progression. Investigating the regulation of these pseudogenes could shed light on the broader regulatory architecture governing *VDAC* gene activity, which has already been implicated in various disease contexts and is under exploration as both a prognostic biomarker and a therapeutic target (De Pinto et al. 2010, Magri et al. 18, Messina et al. 2012, Reina et al. 2020, Zinghirino et al. 2020, 2021).

To contribute to this emerging field, our study conducted a comprehensive in-silico analysis of *VDAC* pseudogenes, complemented by multi-omic integration of genomic and expression data, and further validated in AML cell lines. Among the pseudogenes examined, *VDACIP8* emerged as particularly noteworthy due to its genomic positioning within a transcriptionally active chromatin region flanked by expressed genes. Its localization may enable the exploitation of nearby regulatory elements, facilitating its transcription under pathological conditions such as AML. It is postulated that *VDACIP8*-201 arose from a retrotransposition event involving *VDACI*, embedding an exonic sequence at locus 6q24.2. The remaining *VDACIP8* transcript variants, however, do not appear to originate from the parental gene and likely represent alternative splicing products or arise from distinct transcription start sites.

Our expression profiling across healthy tissues confirmed that some pseudogenes—*VDACIP1*, *VDACIP2*, *VDACIP6*, *VDACIP8*, as well as *VDAC2P1* and *VDAC3P1*—exhibit widespread expression. Notably, *VDAC3P1* is prominently expressed in testicular germ cells, mirroring *VDAC3*'s essential role in spermatogenesis and male fertility. In tumor contexts, however, several *VDACI* pseudogenes—including *VDACIP1*, *VDACIP2*, *VDACIP4*, *VDACIP11*, and especially *VDACIP8*—exhibit altered expression patterns.

While *VDACIP1* and *VDACIP2* tend to parallel the expression dynamics of the parental gene, others such as *VDACIP4*, *VDACIP11*, and *VDACIP8* show opposing trends. Interestingly, *VDACIP8* is typically downregulated in most tumors but highly overexpressed in AML, indicating a potential functional divergence.

These findings were corroborated in five AML cell lines representing different hematopoietic precursors, where expression trends mirrored those seen in bulk tissue data. In the near-haploid leukemic HAP1 cell line, *VDAC1* and *VDACIP8* exhibited the expected inverse correlation. Importantly, in a *VDAC1*-knockout HAP1 line, *VDACIP8* expression rose significantly, suggesting a possible compensatory or regulatory interplay between the two loci.

To further explore this potential regulation, we mapped transcription factor binding sites (TFBS) in the regulatory regions of *VDAC1* and its pseudogenes. We identified 24 TFBS within *VDACIP8*, several of which—including *MYC*, *TAL1*, *JUND*, *BATF*, *FOS*, *CEBPB/D*, *GATA1/2/3*, *CTCF*, *ZNF384*, *RCOR1*, *SPI1*, *ETS1*, *ERG*, *FLI1*, *IRF4*, and *RUNX1*—are directly implicated in hematopoietic regulation and leukemogenesis. Remarkably, many of these elements are shared with *VDAC1*, reinforcing the hypothesis that both genes may be regulated by a common transcriptional program in AML.

Given the emerging evidence that pseudogenes can exert regulatory control via epigenetic and post-transcriptional mechanisms, we also examined DNA methylation patterns in *VDAC1* and *VDACIP8*. Our analysis revealed that *VDACIP8* generally displays higher methylation levels in both promoter and gene body regions compared to *VDAC1*, particularly in healthy tissues. In AML, however, *VDACIP8* exhibits a distinct methylation signature indicative of epigenetic “de-repression,” potentially facilitating its transcriptional activation independently of the parental gene. This de-repression may allow *VDACIP8* to function as a ceRNA, sequestering miRNAs that would otherwise target *VDAC1*, thereby reducing parental gene expression and promoting leukemic cell survival. This regulatory axis may be a key contributor to the anti-apoptotic phenotype characteristic of AML cells.

Consistent with previous findings in other tumor types—such as the epigenetic activation of the *VDACI* promoter by EPB41L4A-AS1 lncRNA in placental trophoblasts—our results suggest that *VDACI* regulation is tightly intertwined with chromatin status and non-coding RNA networks. This raises the possibility that *VDACIP8* not only contributes to transcriptional regulation but may also serve as a context-dependent biomarker or therapeutic target (Pappalardo et al. 2021).

Unlike in leukemia, where *VDACI* expression is relatively low and the *VDACIP8* pseudogene is more abundantly expressed, most other tumors exhibit overexpression of *VDACI* and poor representation of its pseudogene. To further investigate this inverse relationship, we analyzed the expression profiles of *VDACI* and its pseudogene *VDACIP8* in colorectal cancer HCT116 cells, a model in which *VDACIP8* expression is notably low compared to leukemia. This allowed us to explore the potential regulatory interplay between the two transcripts and to test the hypothesis that *VDACI* may exert a suppressive effect on its pseudogene. Our results reveal a reciprocal expression pattern: overexpression of *VDACI* leads to a decrease in *VDACIP8* levels, while silencing *VDACI* results in increased expression of the pseudogene. These findings suggest the existence of a finely tuned regulatory mechanism linking *VDACI* and *VDACIP8*, which may play a distinctive role in leukemia and warrants further investigation.

To support the validity of our expression data, it is important to clarify that the values reported are presented in linear expression units (e.g., TPM), not in log2 scale. Only fold changes, where applicable, are reported on a logarithmic scale. As such, these data are not directly comparable to log2TPM+1 values provided by tools like GEPIA. Furthermore, the TCGA database offers limited filtering capabilities: while we were able to stratify samples by ethnicity, we could not filter based on disease subtypes (e.g., de novo vs. secondary AML, Myelodysplastic syndromes-related cases), which may significantly influence gene expression profiles.

Despite these limitations, our results show that *VDACI* expression remains relatively stable across conditions, with average transcript levels of 12.40 TPM in normal samples and 12.72 TPM in AML.

Likewise, *VDACIP8* expression was largely unchanged between normal (8.76 TPM) and AML (8.74 TPM) tissues, suggesting no major overall transcriptional deregulation. In CRC samples, *VDACI* showed a slight increase (13.98 TPM vs. 13.93 TPM in normal), while *VDACIP8* decreased modestly (6.46 TPM vs. 6.55 TPM), reflecting minimal changes. However, a direct comparison between AML and CRC revealed that *VDACI* levels are higher in CRC, whereas *VDACIP8* is notably reduced in CRC compared to AML (6.46 vs. 8.74 TPM), pointing toward tissue-specific regulatory mechanisms.

When evaluated in representative cancer cell lines, *VDACI* remained consistently expressed (13.16 TPM in normal, 13.17 TPM in HL60, and 13.45 TPM in HTC116). In contrast, *VDACIP8* was significantly downregulated in both hematopoietic (HL60: 2.53 TPM) and epithelial (HTC116: 3.14 TPM) tumor contexts relative to normal tissues (7.65 TPM). This supports a model in which *VDACIP8* undergoes differential regulation in a cancer-type-specific manner and may be selectively repressed or activated depending on the cellular environment.

In conclusion, our findings establish *VDACIP8* as a potential epigenetic regulator of *VDACI* with functional implications in leukemia and possibly other cancers. Its unique expression profile, regulatory element overlap, and methylation landscape suggest a dynamic role in tumor biology. Future studies exploring the molecular interactions between pseudogenes, transcription factors, and chromatin modifiers could further unravel the complexities of pseudogene-mediated gene regulation and open novel avenues for cancer diagnostics and therapeutic interventions.

References

- Abu-Hamad S, Zaid H, Israelson A, Nahon E, Shoshan-Barmatz V. Hexokinase-I protection against apoptotic cell death is mediated via interaction with the voltage-dependent anion channel-1: mapping the site of binding. *J Biol Chem* 2008;283:13482–90. <https://doi.org/10.1074/jbc.M708216200>.
- Amodeo GF, Scorciapino MA, Messina A, De Pinto V, Ceccarelli M. Charged residues distribution modulates selectivity of the open state of human isoforms of the voltage dependent anion-selective channel. *PLoS One* 2014;9:e103879. <https://doi.org/10.1371/journal.pone.0103879>.
- Anflous K, Armstrong DD, Craigen WJ. Altered Mitochondrial Sensitivity for ADP and Maintenance of Creatine-stimulated Respiration in Oxidative Striated Muscles from VDAC1-deficient Mice*. *Journal of Biological Chemistry* 2001;276:1954–60. <https://doi.org/10.1074/jbc.M006587200>.
- Bayrhuber M, Meins T, Habeck M, Becker S, Giller K, Villinger S, et al. Structure of the human voltage dependent channel. *Proc Natl Acad Sci U S A* 2008;105:15370–5. <https://doi.org/10.1073/pnas.0808115105>.
- Ben-Hail D, Begas-Shvartz R, Shalev M, Shteinfer-Kuzmine A, Gruzman A, Reina S, et al. Novel Compounds Targeting the Mitochondrial Protein VDAC1 Inhibit Apoptosis and Protect against Mitochondrial Dysfunction. *J Biol Chem* 2016;291:24986–5003. <https://doi.org/10.1074/jbc.M116.744284>.
- Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., & Wittwer, C. T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry*, 55(4), 611–622. <https://doi.org/10.1373/clinchem.2008.112797>
- Cavalli, G., Heard, E. Advances in epigenetics link genetics to the environment and disease. *Nature* 571, 489–499 (2019). <https://doi.org/10.1038/s41586-019-1411-0>

- Cassar  MC, Menzel VA, Hinsch K-D, Wrenzycki C, Hinsch E. Voltage-dependent anion channels 1 and expressed in <https://doi.org/10.1042/BSR20090088>. porcine oocytes. *Biosci Rep* 2009;30:193–200. 68
- Charrier C, Joshi K, Coutinho-Budd J, et al. Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell* 2012;149:923–35.
- Checchetto V, Reina S, Magri A, Szabo I, De Pinto V. Recombinant human voltage dependent anion selective channel isoform 3 (hVDAC3) forms pores with a very small conductance. *Cell Physiol Biochem* 2014;34:842–53. <https://doi.org/10.1159/000363047>.
- Cheetham SW, Faulkner GJ, Dinger ME. Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nat Rev Genet* 2020;21:191–201.
- Chen S, Krinsky BH, Long M. New genes as drivers of phenotypic evolution. *Nat Rev Genet* 2013;14:645–60
- Chen X, Wan L, Wang W, Xi WJ, Yang AG, Wang T. Re-recognition of pseudogenes: from molecular to clinical applications. *Theranostics* 2020;10:1479–99.
- Cheng EHY, Sheiko TV, Fisher JK, Craigen WJ, Korsmeyer SJ. VDAC2 inhibits BAK activation and mitochondrial apoptosis. *Science* 2003;301:513–7. <https://doi.org/10.1126/science.1083995>.
- Chiang JJ, Sparrer KMJ, van Gent M, et al. Viral unmasking of cellular 5S rRNA pseudogene transcripts induces RIG-I-mediated immunity. *Nat Immunol* 2018;19:53–62.
- Chiefari E, Iiritano S, Paonessa F, et al. Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. *Nat Commun* 2010;1:40.
- Chin HS, Li MX, Tan IKL, Ninnis RL, Reljic B, Scicluna K, et al. VDAC2 enables BAX to mediate apoptosis and limit tumor development. *Nat Commun* 2018;9:4976. <https://doi.org/10.1038/s41467-018-07309-4>.
- Conti Nibali S, Di Rosa MC, Rauh O, Thiel G, Reina S, De Pinto V. Cell-free electrophysiology of human VDACs incorporated into nanodiscs: An improved method. *Biophys Rep (N Y)* 2021;1:None. <https://doi.org/10.1016/j.bpr.2021.100002>.

- Craig WJ, Graham BH. Genetic strategies for dissecting mammalian and *Drosophila* voltage dependent anion channel functions. <https://doi.org/10.1007/s10863-008-9146-x>. *J Bioenerg Biomembr* 2008;40:207–12.
- De Pinto V, Guarino F, Guarnera A, Messina A, Reina S, Tomasello FM, et al. Characterization of human VDAC isoforms: a peculiar function for VDAC3? *Biochim Biophys Acta* 2010;1797:1268–75. <https://doi.org/10.1016/j.bbabi.2010.01.031>.
- De Pinto V, Reina S, Guarino F, Messina A. Structure of the voltage dependent anion channel: state of the art. *J Bioenerg Biomembr* 2008;40:139–47. <https://doi.org/10.1007/s10863-008-9140-3>. anion
- De Pinto V. Renaissance of VDAC: New Insights on a Protein Family at the Interface between Mitochondria and Cytosol. *Biomolecules* 2021;11:107. <https://doi.org/10.3390/biom11010107>.
- Dong, R.; Zhang, X.-O.; Zhang, Y.; Ma, X.-K.; Chen, L.-L.; Yang, Z. CircRNA-derived pseudogenes. *Cell Res.* 2016, 26, 747–750.
- Dong, X., et al. (2018). Pseudogene-derived lncRNAs and their roles in cancer. *Scientific Reports*, 8, 9057. <https://doi.org/10.1038/s41598-018-27180-6>
- Eddy MT, Yu T-Y, Wagner G, Griffin RG. Structural characterization of the human membrane protein VDAC2 in lipid bilayers by MAS NMR. *J Biomol NMR* 2019;73:451–60. <https://doi.org/10.1007/s10858-019-00242-8>.
- Fiddes IT, Lodewijk GA, Mooring M, et al. Human-specific NOTCH2NL genes affect notch signaling and cortical neurogenesis. *Cell* 2018;173:1356–1369.e22.
- Fiek C, Benz R, Roos N, Brdiczka D. Evidence for identity between the hexokinase-binding protein and the mitochondrial porin in the outer membrane of rat liver mitochondria. *Biochim Biophys Acta* 1982;688:429–40. [https://doi.org/10.1016/0005-2736\(82\)90354-6](https://doi.org/10.1016/0005-2736(82)90354-6).
- c
- Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol* 18, 83 (2017). <https://doi.org/10.1186/s13059-017-1215-1>

- Hardison, R.C.; Butler, E.T.; Lacy, E.; Maniatis, T.; Rosenthal, N.; Efstratiadis, A. The structure and transcription of four linked rabbit γ -like globin genes. *Cell* 1979, 18, 1285–1297.
- Hiller S, Garces RG, Malia TJ, Orekhov VY, Colombini M, Wagner G. Solution Structure of the Integral Human Membrane Protein VDAC-1 in Detergent Micelles. *Science* 2008;321:1206–10. <https://doi.org/10.1126/science.1161302>.
- Hinsch K-D, De Pinto V, Aires VA, Schneider X, Messina A, Hinsch E. Voltage-dependent anion selective channels VDAC2 and VDAC3 are abundant proteins in bovine outer dense fibers, a cytoskeletal component of the sperm <https://doi.org/10.1074/jbc.M313433200>. *J Biol Chem* 2004;279:15281–8.
- Huang P, Keller CA, Giardine B, et al. Comparative analysis of threedimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes Dev* 2017;31:1704–13.
- Huang, W., et al. (2017). The Emerging Role of Pseudogene-Expressed RNAs in Cellular Functions. *Cell Molecular Life Science*. DOI: 10.1007/s00018-017-2507-5
- Huizing M, Ruitenbeek W, Thinnies FP, DePinto V, Wendel U, Trijbels FJ, et al. Deficiency of the voltage-dependent anion channel: a novel cause of mitochondriopathy. *Pediatr Res* 1996;39:760–5. <https://doi.org/10.1203/00006450-199605000-00003>.
- Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* 2010;11:97–108.
- Israelson A, Arbel N, Da Cruz S, Ilieva H, Yamanaka K, Shoshan-Barmatz V, et al. Misfolded mutant SOD1 directly inhibits VDAC1 conductance in a mouse model of inherited ALS. *Neuron* 2010;67:575–87. <https://doi.org/10.1016/j.neuron.2010.07.019>.
- Jacq C, Miller JR, Brownlee GG. A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* 1977;12:109–20.
- Kaessmann H, Vinckenbosch N, Long M. RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet* 2009;10:19–31.

- Karreth FA, Reschke M, Ruocco A, et al. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell* 2015;161:319–32.
- Korneev SA, Park JH, O'Shea M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. *J Neurosci* 1999;19:7711–20.
- Kovalenko TF, Patrushev LI. Pseudogenes as functionally significant elements of the genome. *Biochemistry (Mosc)* 2018;83:1332–49.
- Leggio L, Guarino F, Magrì A, Accardi-Gheit R, Reina S, Specchia V, et al. Mechanism of translation control of the alternative *Drosophila melanogaster* Voltage Dependent Anion-selective Channel 1 mRNAs. *Sci Rep* 2018;8:5347. <https://doi.org/10.1038/s41598-018-23730-7>.
- Li, S., et al. (2020). Pseudogene expression profile analysis identifies a signature associated with prognosis and immune infiltration in colorectal cancer. *Frontiers in Cell and Developmental Biology*, 8, 603901. <https://doi.org/10.3389/fcell.2020.603901>
- Liu, W., et al. (2022). CTHRC1P1 promotes colorectal cancer cell EMT via the Wnt/ β -catenin pathway. *Frontiers in Oncology*, 12, 869123. <https://doi.org/10.3389/fonc.2022.869123>
- Luo, Z., Dai, Y., & Zheng, C. (2020). Regulation of Hematopoiesis by Non-Coding RNAs. *Frontiers in Oncology*. DOI: 10.3389/fonc.2020.00086
- Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science* 2000;290:1151–5.
- Ma Y, Liu S, Gao J, et al. Genome-wide analysis of pseudogenes reveals HBBP10s human-specific essentiality in erythropoiesis and implication in β -thalassemia. *Dev Cell* 2021;56:478–493.e11.
- Ma, Y., Chen, Z., & Yu, J. (2021). Pseudogenes and their potential functions in hematopoiesis. *Experimental Hematology*. DOI: 10.1016/j.exphem.2021.09.001
- Ma, Y., Chen, Z., & Yu, J. (2021). Pseudogenes and their potential functions in hematopoiesis. *Experimental Hematology*. DOI: 10.1016/j.exphem.2021.09.001

- Magri A, Di Rosa MC, Tomasello MF, Guarino F, Reina S, Messina A, et al. Overexpression of human SOD1 in VDAC1-less yeast restores mitochondrial functionality modulating beta-barrel outer membrane protein genes. *Biochim Biophys Acta* 2016;1857:789–98. <https://doi.org/10.1016/j.bbabbio.2016.03.003>.
- Magri A, Messina A. Interactions of VDAC with Proteins Involved in Neurodegenerative Aggregation: An Opportunity for Advancement on Therapeutic Molecules. *Curr Med Chem* 2017;24:4470–87. <https://doi.org/10.2174/0929867324666170601073920>.
- Magri A, Reina S, De Pinto V. VDAC1 as Pharmacological Target in Cancer and Neurodegeneration: Focus on Its Role in Apoptosis. *Front Chem* 2018;6:108. <https://doi.org/10.3389/fchem.2018.00108>.
- Magri A, Reina S, De Pinto V. VDAC1 as Pharmacological Target in Cancer and Neurodegeneration: Focus on Its Role in Apoptosis. *Front Chem* 2018;6:108. <https://doi.org/10.3389/fchem.2018.00108>.
- Majumder S, Fisk HA. VDAC3 and Mps1 negatively regulate ciliogenesis. *Cell Cycle* 2013;12:849–58. <https://doi.org/10.4161/cc.23824>.
- Maurya SR, Mahalakshmi R. Mitochondrial VDAC2 and cell homeostasis: highlighting hidden structural features and unique functionalities. *Biol Rev Camb Philos Soc* 2017;92:1843–58. <https://doi.org/10.1111/brv.12311>.
- Mazure NM. VDAC in cancer. *Biochim Biophys Acta Bioenerg* 2017;1858:665–73. <https://doi.org/10.1016/j.bbabbio.2017.03.002>.
- Messina A, Reina S, Guarino F, De Pinto V. VDAC isoforms in mammals. *Biochim Biophys Acta* 2012;1818:1466–76. <https://doi.org/10.1016/j.bbammem.2011.10.005>.
- Min CK, Yeom DR, Lee K-E, Kwon H-K, Kang M, Kim Y-S, et al. Coupling of ryanodine receptor 2 and voltage-dependent anion channel 2 is essential for Ca²⁺ transfer from the sarcoplasmic reticulum to the mitochondria in the heart. *Biochem J* 2012;447:371–9. <https://doi.org/10.1042/BJ20120705>. 69

- Naghdi S, Hajnóczky G. VDAC2-specific cellular functions and the underlying structure. *Biochim Biophys Acta* 2016;1863:2503–14. <https://doi.org/10.1016/j.bbamcr.2016.04.020>.
- Neumann, D., Bückers, J., Kastrup, L. et al. Two-color STED microscopy reveals different degrees of colocalization between hexokinase-I and the three human VDAC isoforms. *PMC Biophys* 3, 4 (2010). <https://doi.org/10.1186/1757-5036-3-4>.
- Ohno S. Evolution by gene duplication. London: George Alien & Unwin. Berlin/Heidelberg/New York: Springer-Verlag; 1970
- Okada SF, O’Neal WK, Huang P, Nicholas RA, Ostrowski LE, Craigen WJ, et al. Voltage-dependent anion channel-1 (VDAC-1) contributes to ATP release and cell volume regulation in murine cells. *J Gen Physiol* 2004;124:513–26. <https://doi.org/10.1085/jgp.200409154>.
- Okazaki M, Kurabayashi K, Asanuma M, Saito Y, Dodo K, Sodeoka M. VDAC3 gating is activated by suppression of disulfide-bond formation between the N-terminal region and the bottom of the pore. *Biochimica et Biophysica Acta* <https://doi.org/10.1016/j.bbamem.2015.09.017>. (BBA) - Biomembranes 2015;1848:3188–96. 70
- Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., et al. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature*. DOI: 10.1038/nature19348
- Pappalardo, X. G., Risiglione, P., Zinghirino, F., Ostuni, A., Luciano, D., Bisaccia, F., De Pinto, V., Guarino, F., & Messina, A. (2023). Human VDAC pseudogenes: an emerging role for VDAC1P8 pseudogene in acute myeloid leukemia. *Biological research*, 56(1), 33. <https://doi.org/10.1186/s40659-023-00446-1>
- Pastorino JG, Shulga N, Hoek JB. Mitochondrial binding of hexokinase II inhibits Bax-induced cytochrome release and <https://doi.org/10.1074/jbc.M109950200>. apoptosis. *J Biol Chem* 2002;277:7610–8.
- Pei B, Sisu C, Frankish A, et al. The GENCODE pseudogene resource. *Genome Biol* 2012;13:R51.

- Pei B, Sisu C, Frankish A, et al. The GENCODE pseudogene resource. *Genome Biol* 2012;13:R51.
- Pittalà MGG, Saletti R, Reina S, Cunsolo V, De Pinto V, Foti S. A High Resolution Mass Spectrometry Study Reveals the Potential of Disulfide Formation in Human Mitochondrial Voltage-Dependent Anion Selective Channel Isoforms <https://doi.org/10.3390/ijms21041468>. (hVDACs). *Int J Mol Sci* 2020;21:E1468.
- Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010;465:1033–8.
- Prasad M, Kaur J, Pawlak KJ, Bose M, Whittal RM, Bose HS. Mitochondria-associated endoplasmic reticulum membrane (MAM) regulates steroidogenic activity via steroidogenic acute regulatory protein (StAR)-voltage-dependent anion channel 2 (VDAC2) interaction. *J Biol Chem* 2015;290:2604–16. <https://doi.org/10.1074/jbc.M114.605808>.
- Proudfoot N. Pseudogenes. *Nature* 1980;286:840–1.
- Proudfoot, N.J.; Maniatis, T. The structure of a human β -globin pseudogene and its relationship to β -globin gene duplication. *Cell* 1980, 21, 537–544.
- Queralt-Martín M, Bergdoll L, Tejjido O, Munshi N, Jacobs D, Kuszak AJ, et al. A lower affinity to cytosolic proteins reveals VDAC3 isoform-specific role in mitochondrial biology. *J Gen Physiol* 2020;152:e201912501. <https://doi.org/10.1085/jgp.201912501>.
- Raghavan A, Sheiko T, Graham BH, Craigen WJ. Voltage-dependant anion channels: Novel insights into isoform function through genetic models. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 2012;1818:1477–85. <https://doi.org/10.1016/j.bbamem.2011.10.019>. 67
- Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge RM, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *Elife* 2013;2:e00762.

- Reddy PH. Is the mitochondrial outermembrane protein VDAC1 therapeutic target for Alzheimer's disease? *Biochim Biophys Acta* 2013;1832:67–75. <https://doi.org/10.1016/j.bbadis.2012.09.003>.
- Reina S, Checchetto V, Saletti R, Gupta A, Chaturvedi D, Guardiani C, et al. VDAC3 as a sensor of oxidative state of the intermembrane space of mitochondria: the putative role of cysteine residue modifications. *Oncotarget* 2016;7:2249–68. <https://doi.org/10.18632/oncotarget.6850>.
- Reina S, De Pinto V. Anti-Cancer Compounds Targeted to VDAC: Potential and Perspectives. *Curr Med Chem* 2017;24:4447–69. <https://doi.org/10.2174/0929867324666170530074039>.
- Reina S, Palermo V, Guarnera A, Guarino F, Messina A, Mazzoni C, et al. Swapping of the N-terminus of VDAC1 with VDAC3 restores full activity of the channel and confers anti-aging features to the cell. *FEBS Letters* 2010;584:2837–44. <https://doi.org/10.1016/j.febslet.2010.04.066>.
- Reina S, Pittalà MGG, Guarino F, Messina A, De Pinto V, Foti S, et al. Cysteine Oxidations in Mitochondrial Membrane Proteins: The Case of VDAC Isoforms in Mammals. *Front Cell Dev Biol* 2020;8. <https://doi.org/10.3389/fcell.2020.00397>.
- Rostovtseva TK, Gurnev PA, Protchenko O, Hoogerheide DP, Yap TL, Philpott CC, et al. α -Synuclein Shows High Affinity Interaction with Voltage-dependent Anion Channel, Suggesting Mechanisms of Mitochondrial Regulation and Toxicity in Parkinson Disease. *J Biol Chem* 2015;290:18467–77. <https://doi.org/10.1074/jbc.M115.641746>.
- Saletti R, Reina S, Pittalà MGG, Magrì A, Cunsolo V, Foti S, et al. Post-translational modifications of VDAC1 and VDAC2 cysteines from rat liver mitochondria. *Biochim Biophys Acta Bioenerg* 2018;1859:806–16. <https://doi.org/10.1016/j.bbabbio.2018.06.007>.
- Sampson MJ, Decker WK, Beaudet AL, Ruitenbeek W, Armstrong D, Hicks MJ, et al. Immobile Sperm and Infertility in Mice Lacking Mitochondrial Voltage-dependent Anion Channel Type 3*. *Journal of Biological Chemistry* 2001;276:39206–12. <https://doi.org/10.1074/jbc.M104724200>. 2 are

- Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. *PLoS Comput Biol*. 2013;9(10):e1003285. doi:10.1371/journal.pcbi.1003285
- Schredelseker J, Paz A, López CJ, Altenbach C, Leung CS, Drexler MK, et al. High resolution structure and double electron-electron resonance of the zebrafish voltage-dependent anion channel 2 reveal an oligomeric population. *J Biol Chem* 2014;289:12566–77. <https://doi.org/10.1074/jbc.M113.497438>.
- Shimizu H, Schredelseker J, Huang J, Lu K, Naghdi S, Lu F, et al. Mitochondrial Ca(2+) uptake by the voltage-dependent anion channel <https://doi.org/10.7554/eLife.04801>. 2 regulates cardiac rhythmicity. *Elife* 2015;4.
- Shoshan-Barmatz V, De Pinto V, Zweckstetter M, Raviv Z, Keinan N, Arbel N. VDAC, a multi functional mitochondrial protein regulating cell life and death. *Mol Aspects Med* 2010;31:227–85. <https://doi.org/10.1016/j.mam.2010.03.002>.
- Stasiak, M., Kolenda, T., Kozłowska-Masłoń, J., et al. (2021). The World of Pseudogenes: New Diagnostic and Therapeutic Targets in Cancers or Still Mystery Molecules? *Life*. DOI: 10.3390/life11121354
- Subedi KP, Kim J-C, Kang M, Son M-J, Kim Y-S, Woo S-H. Voltage-dependent anion channel 2 modulates resting Ca²⁺ sparks, but not action potential-induced Ca²⁺ signaling in cardiac myocytes. *Cell Calcium* 2011;49:136–43. <https://doi.org/10.1016/j.ceca.2010.12.004>.
- Suzuki IK, Gacquer D, Van Heurck R, et al. Human-specific NOTCH2NL genes expand cortical neurogenesis through Delta/Notch regulation. *Cell* 2018;173:1370–1384.e16.
- Tan S, Cardoso-Moreira M, Shi W, et al. LTR-mediated retroposition as a mechanism of RNA-based duplication in metazoans. *Genome Res* 2016;26:1663–75.
- To KKW, Zhang H, Cho WC. Competing endogenous RNAs (ceRNAs) and drug resistance to cancer therapy. *Cancer Drug Resist*. 2024;7:37. Published 2024 Sep 25. doi:10.20517/cdr.2024.66

- Tutar, L., Özgür, A., & Tutar, Y. (2018). Involvement of miRNAs and Pseudogenes in Cancer. *Methods in molecular biology* (Clifton, N.J.), 1699, 45–66. https://doi.org/10.1007/978-1-4939-7435-1_3
- Ujwal R, Cascio D, Colletier J-P, Faham S, Zhang J, Toro L, et al. The crystal structure of mouse VDAC1 at 2.3 Å resolution reveals mechanistic insights into metabolite gating. *Proc Natl Acad Sci U S A* 2008;105:17742–7. <https://doi.org/10.1073/pnas.0809634105>.
- Vander Heiden MG, Chandel NS, Schumacker PT, Thompson CB. Bcl-xL prevents cell death following growth factor withdrawal by facilitating mitochondrial ATP/ADP exchange. *Mol Cell* 1999;3:159–67. [https://doi.org/10.1016/s1097-2765\(00\)80307-x](https://doi.org/10.1016/s1097-2765(00)80307-x).
- Veresov VG, Davidovskii AI. Structural insights into proapoptotic signaling mediated by MTCH2, VDAC2, TOM40 and TOM22. <https://doi.org/10.1016/j.cellsig.2013.11.016>. *Cell Signal* 2014;26:370–82.
- Verges L, Vidal F, Geun E, Alemany-Schmidt A, Oliver-Bonet M, Blanco J. An exploratory study of predisposing genetic factors for DiGeorge/velocardiofacial syndrome. *Sci Rep* 2017;7:40031.
- Watanabe T, Totoki Y, Toyoda A, et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 2008;453:539–43
- Xiao-Jie, L., Ai-Mei, G., Li-Juan, J., & Jiang, X. (2015). Pseudogene in cancer: real functions and promising signature. *Journal of medical genetics*, 52(1), 17–24. <https://doi.org/10.1136/jmedgenet-2014-102785>
- Xu X, Decker W, Sampson MJ, Craigen WJ, Colombini M. Mouse VDAC isoforms expressed in yeast: channel properties and their roles in mitochondrial outer membrane permeability. *J Membr Biol* 1999;170:89–102. <https://doi.org/10.1007/s002329900540>.
- Zhang E, Mohammed Al-Amily I, Mohammed S, Luan C, Asplund O, Ahmed M, et al. Preserving Insulin Secretion in Diabetes by Inhibiting VDAC1 Overexpression and Surface Translocation in β Cells. *Cell Metab* 2019;29:64-77.e6. <https://doi.org/10.1016/j.cmet.2018.09.008>. 66

- Zhang Z, Harrison PM, Liu Y, Gerstein M. Millions of years of evolution preserved: A comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* 2003;13:2541–58.
- Zhang, Y., et al. (2020). DUSP5P1 inhibits colorectal cancer progression by negatively regulating ERK signaling pathway. *Journal of Cellular Physiology*, 235(6), 4751–4763. <https://doi.org/10.1002/jcp.29329>
- Zhang, Z.D.; Frankish, A.; Hunt, T.; Harrow, J.; Gerstein, M. Identification and analysis of unitary pseudogenes: Historic and contemporary gene losses in humans and other primates. *Genome Biol.* 2010, 11, R26.
- Zhao, J., et al. (2021). The emerging roles of pseudogenes in immune regulation and immune-related diseases. *Frontiers in Immunology*, 12, 679948. <https://doi.org/10.3389/fimmu.2021.679948>
- Zinghirino F, Pappalardo XG, Messina A, Guarino F, De Pinto V. Is the Secret of VDAC Isoforms in Their Gene Regulation? Characterization of Human VDAC Genes Expression Profile, Promoter Activity, and Transcriptional Regulators. *International Journal of Molecular Sciences* 2020;21:7388. <https://doi.org/10.3390/ijms21197388>.
- Zinghirino F, Pappalardo XG, Messina A, Nicosia G, De Pinto V, Guarino F. VDAC Genes Expression and Regulation in Mammals. *Front Physiol* 2021;12:708695. <https://doi.org/10.3389/fphys.2021.708695>.

Paper 1 enclosed to the present Ph.D. thesis

*Pappalardo, X. G., Risiglione, P., Zinghirino, F., Ostuni, A., **Luciano, D.**, Bisaccia, F., De Pinto, V., Guarino, F., & Messina, A. (2023). Human VDAC pseudogenes: an emerging role for VDACP8 pseudogene in acute myeloid leukemia. Biological research, 56(1), 33. <https://doi.org/10.1186/s40659-023-00446-1>*

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my coordinator, Prof. Vito De Pinto, and my supervisor, Prof. Francesca Guarino, for their continuous support, guidance, and encouragement throughout the course of this research. I also extend my sincere thanks to Prof. Angela Messina for her valuable insights over the years. Their expertise, constructive feedback, and patience have been invaluable in shaping this thesis.

I am grateful to my lab colleagues for providing the necessary resources, facilities, and a stimulating academic environment in which to conduct my research. Special thanks also go to the laboratory team, research assistants, and administrative staff, whose support made daily work both productive and enjoyable. In particular, I would like to thank Dr. Xena Pappalardo for her scientific support and advice, which were of great value throughout my work.

I would like to acknowledge the support of Dr. Orazia Diquattro, Head of the U.O.C. of Microbiology and Virology at the Riuniti Hospital Villa Sofia – Cervello, without which this research would not have been possible.

Finally, I would like to express my deepest love and appreciation to my family, my husband, and my three wonderful children for their unwavering support, patience, and encouragement. Their belief in me has been the foundation of all my achievements

RESEARCH ARTICLE

Open Access



Human *VDAC* pseudogenes: an emerging role for *VDAC1P8* pseudogene in acute myeloid leukemia

Xena Giada Pappalardo¹, Pierpaolo Risiglione¹, Federica Zinghirino¹, Angela Ostuni², Daniela Luciano¹, Faustino Bisaccia², Vito De Pinto^{1,3,4,5}, Francesca Guarino^{1,3,4,5†} and Angela Messina^{3,4,5,6*†} 

Abstract

Background Voltage-dependent anion selective channels (VDACs) are the most abundant mitochondrial outer membrane proteins, encoded in mammals by three genes, *VDAC1*, 2 and 3, mostly ubiquitously expressed. As 'mitochondrial gatekeepers', *VDACs* control organelle and cell metabolism and are involved in many diseases. Despite the presence of numerous *VDAC* pseudogenes in the human genome, their significance and possible role in *VDAC* protein expression has not yet been considered.

Results We investigated the relevance of processed pseudogenes of human *VDAC* genes, both in physiological and in pathological contexts. Using high-throughput tools and querying many genomic and transcriptomic databases, we show that some *VDAC* pseudogenes are transcribed in specific tissues and pathological contexts. The obtained experimental data confirm an association of the *VDAC1P8* pseudogene with acute myeloid leukemia (AML).

Conclusions Our *in-silico* comparative analysis between the *VDAC1* gene and its *VDAC1P8* pseudogene, together with experimental data produced in AML cellular models, indicate a specific over-expression of the *VDAC1P8* pseudogene in AML, correlated with a downregulation of the parental *VDAC1* gene.

Keywords Pseudogene, Voltage-dependent anion selective channels (VDACs) gene, Acute myeloid leukemia (AML), Disease animal model, Competing endogenous RNAs (ceRNAs)

[†]Francesca Guarino and Angela Messina co-last author and jointly supervised this work.

*Correspondence:
Angela Messina
mess@unict.it

¹ Department of Biomedical and Biotechnological Sciences, University of Catania, Via Santa Sofia 97, 95123 Catania, Italy

² Department of Sciences, University of Basilicata, 85100 Potenza, Italy

³ we.MitoBiotech S.R.L., C.so Italia 172, 95125 Catania, Italy

⁴ I.N.B.B., National Institute for Biostructures and Biosystems, Interuniversity Consortium, Catania, Italy

⁵ Research Centre on Nutraceuticals and Health Products (CERNUT), University of Catania, 95125 Catania, Italy

⁶ Department of Biological, Geological and Environmental Sciences, University of Catania, Via Santa Sofia 97, 95123 Catania, Italy

Introduction

Voltage dependent anion selective channels (*VDACs* or mitochondrial porins) are a phylogenetically conserved family of genes encoding channel proteins embedded into the outer mitochondrial membrane (OMM). *VDAC* proteins are considered the master regulators of the mitochondrial function because the important functions they perform [1]. The principal role of *VDACs* is to allow the exchange of important ions and metabolites to and from the mitochondria [2–4]. In mammals, three *VDAC* isoforms have been identified and named *VDAC1*, *VDAC2* and *VDAC3* [5]. *VDACs* genes are located in different autosomes but sharing the same number and size of exons. The only exception is *VDAC2* gene which contains an additional coding exon at the N-terminal [5, 6]. Also,



the size of introns is variable [7]. We recently reported that *VDAC* genes show a different expression pattern in human tissues, as well as a distinct set of transcription factor binding sites (TFBSs) on each promoter [8, 9]. Furthermore, it is known that altered *VDAC1* expression is associated with many diseases, like cancer, neurodegenerative and cardiovascular diseases [10–14]. In particular, the observation that *VDAC1* is over-expressed in most tumors suggests that this protein may play a pivotal role in this pathological condition [15]. And concerning the tumor phenotype, the interaction of *VDAC1* with hexokinase (HK), the first glycolysis enzyme, grants the enzyme immediate access to newly synthesized ATP and confers protection from apoptosis [16]. It is therefore not surprising that *VDAC1* has become a primary target in the fight against cancer.

In recent years, the involvement of transcribed pseudogenes in the development and progression of many types of cancer has become well established [17]. Due to the vast amount of data available in the literature and in public databases, specific pseudogenes, whose expression varies considerably in certain cancers, can be classified as predictive, heritable or prognostic biomarkers.

Pseudogenes have long been considered function-less elements arisen during genome evolution as failed copies of genes. However, a large number of recent evidence indicate that most of them are not functionally inert but rather, once transcribed, have a consistent role in protein-coding genes regulation [18]. Pseudogene expression pattern can vary under different pathological conditions including diabetes and cancer, even though, due to the high sequence similarity with the parental genes. However, measurements of their transcription level remain difficult to perform [19, 20]. Not surprisingly, considering the homology between pseudogene and its parental gene, pseudogene transcripts can play a part as competitive endogenous RNAs (ceRNAs). Generally, any transcript that competitively binds a common microRNA (miRNA) behaves as a ceRNA that, acting as a “sponge”, reduces the free miRNA available. As a result, this can lead to the release or attenuation of miRNA target gene(s) repression. Alternatively, RNA resulting from pseudogene transcription may be processed into a short interfering RNA (siRNA) or act as antisense RNA. Regardless of the mechanism, the pseudogene would in any case be an important factor in the regulation of protein-coding genes. Recently, it has been proposed that pseudogenes represent primate- or human-specific regulatory elements, especially in hemopoiesis, given their higher expression in bone marrow than in other tissues [21].

In the human genome there are 13 processed pseudogenes (or retrocopies) of *VDAC1*, distributed in 10 different chromosomes and annotated in GENECODE and

NCBI Entrez Gene [22]. Processed pseudogenes originate when a mRNA transcript is reverse-transcribed and integrated into the genome at a new position [23, 24]. Therefore, lacking its own promoter sequence and introns as well, their transcriptional activity depends on the presence of promoter(s) and regulatory sequences nearby. Ido and co-workers, after a first characterization of 16 putative rat *VDAC1* pseudogenes [25], reported a comparison of *VDAC1* pseudogenes among rat, mouse and human showing that no synteny exists between humans and rodents [26]. Apart from these literature reports, nothing more is known about human *VDAC* pseudogenes. Considering the many regulatory roles described for pseudogenes in cancer [17, 20, 27] and the known dysregulation of *VDAC* in cancer pathologies [13, 15, 16], we wanted to investigate whether there is an association of any pathology, and in particular cancer, with *VDAC* pseudogenes expression. The assumption is that, in pathological situations, the expression of *VDAC* proteins could be post-transcriptionally regulated and/or influenced by the transcription of related pseudogenes. This information could be of great importance because it could certainly be useful in identifying new diagnostic, prognostic and therapeutic approaches for the pathologies under consideration.

In this work, we present a whole, comprehensive analysis of in-silico data concerning the human *VDAC* pseudogenes together with a careful *in-cellulo* validation of the expression data. Overall, our results allow us to hypothesize for some *VDAC1* pseudogenes a role in the physiology or pathology of certain tissues. In particular, the specific and marked expression in AML of the *VDAC1P8* pseudogene suggests its potential role in leukaemogenesis and makes it a possible biomarker of acute myeloid leukaemia.

Results

Human *VDACs* pseudogenes and their relative chromatin state

All the genomic databases (NCBI Entrez Gene, UCSC and Ensembl) queried in this work report the presence in the human genome of 13 *VDAC1* pseudogenes, 4 *VDAC2* pseudogenes and only one *VDAC3* pseudogene. Evidence in human genome of putative regulatory elements located in the proximity of *VDAC* pseudogenes was obtained from UCSC Genome Browser (GRCh38/hg38).

The data on regulatory elements found for each *VDAC* pseudogene are summarized in Table 1. This analysis revealed that *VDAC1P4*, *VDAC1P8*, *VDAC1P11* and *VDAC1P13* pseudogenes are located at transcriptionally active genomic loci (Table 1). Interestingly, the *VDAC1P8* sequence is located between the 3' end of the *PEX3* gene and the 3' end of the *FUCA2* gene, on the opposite strand

Table 1 Features and chromatin status of genomic sequences containing *VDAC* pseudogenes, from UCSC Genome Browser GRCh38/hg38

Name	Entrez ID	Genomic coordinates			Regulatory build	
		Location	Size (bp)	Strand	5'-Flanking region	3'-Flanking region
(a)						
<i>VDAC1P1</i>	642585	chrX:80184923–80186706	1784	+	Heterochromatin	Insulator and weak enhancer
<i>VDAC1P2</i>	10064	chrX:49396238–49397962	1725	–	Heterochromatin	
<i>VDAC1P3</i>	100310838	chrX:91236801–91238513	1713	–	Weak enhancers, weak and active promoter (overlapping intronic region of <i>PCDH11X</i> gene)	
<i>VDAC1P4</i>	7418	chr1:180403931–180405937	2007	+	weak and strong enhancers, weak transcriptional elongation, weak promoter (overlapping intronic region of <i>ACBD6</i> gene)	
<i>VDAC1P5</i>	10187	chr12:55195669–55197407	1739	–	Heterochromatin	
<i>VDAC1P6</i>	359800	chrY:5074406–5076119	1714	–	Heterochromatin (overlapping intronic region of <i>PCDH11Y</i> gene)	
<i>VDAC1P7</i>	100310839	chr3:77365047–77366767	1721	–	Insulator	Heterochromatin (overlapping intronic region of <i>ROBO2</i> gene)
<i>VDAC1P8^a</i>	100310840	chr6:143811645–143813335	1691	+	Weak enhancer, weak transcriptional elongation (proximity to <i>PEX3</i> gene)	Weak and strong enhancer, weak transcriptional elongation (proximity to <i>FUCA2</i> gene)
<i>VDAC1P9</i>	391106	chr1:157693990–157694816	827	–	Heterochromatin	Weak enhancer, weak transcriptional elongation
<i>VDAC1P10</i>	643536	chr1:215549841–215551581	1741	+	Flanking heterochromatin and weak enhancers inside gene body	
<i>VDAC1P11</i>	100310841	chr9:97048814–97050837	2024	–	Weak transcriptional elongation, weak and strong enhancer (overlapping intronic region of <i>ZNF169</i> gene)	
<i>VDAC1P12</i>	100874289	chr13:34656914–34657447	534	–	Heterochromatin	
<i>VDAC1P13</i>	100420568	chr2:42690279–42691137	859	–	Weak enhancers, weak transcriptional elongation (overlapping intronic region of <i>KCNQ3</i> gene)	
(b)						
<i>VDAC2P1</i>	54015	chr21:17466509–17467792	1284	–	Heterochromatin (overlapping <i>LINC00478</i> gene)	
<i>VDAC2P2</i>	643996	chr12:9188920–9190119	1200	–	Heterochromatin	
<i>VDAC2P3</i>	401959	chr1:118182962–118184349	1388	–	Insulator, weak enhancer, heterochromatin	Weak transcriptional elongation, polycomb repressed, heterochromatin
<i>VDAC2P4</i>	100420574	chr2:135554739–135555449	711	+	Heterochromatin, insulator, Polycomb repressor	
(c)						
<i>VDAC3P1</i>	341965	chr14:100430801–100431912	1112	+	Weak transcriptional elongation, heterochromatin, Polycomb repressed	Weak enhancer, heterochromatin, Polycomb repressed

The table shows in (a) information on *VDAC1* pseudogenes *VDAC1P1–13*, in (b) on *VDAC2* pseudogenes *VDAC2P1–4* and in (c) about the only one *VDAC3* pseudogene *VDAC3P1*. ^aGenomic position is referred to splicing variant with ENST00000406025.2 ID at Ensembl database Additional file 1: Fig. S1a

(Table 1, Fig. 1). Therefore, *VDAC1P8* gene sequence is found within a likely active region characterized by lax chromatin states, the presence of strong and weak enhancers and a weak transcriptional elongation capacity

(ENCODE genome segmentation). Although the state of the chromatin is not comparable to that in which housekeeping genes, as *GAPDH*, are located, the features of chromatin region hosting *VDAC1P8* pseudogene

(See figure on next page.)

Fig. 1 Chromatin state and genomic features of *VDAC1P8* pseudogene from UCSC Genome Browser GRCh37/hg19. The genomic context of *VDAC1P8* pseudogene set around 1000 Kb upstream and downstream of the annotated Refseq is shown. The selected regulatory hub tracks are Pseudogene Annotation Set from GENCODE v.38lift37 Ensemble 104, Eukaryotic Promoter Database EPD v.4–6, CpG island track, Genotype-Tissue Expression GTEx RNA-seq v.8 2019, ChIP-Seq data for RNA polymerase II, H3K4me3 and H3K4me1, used as markers of transcriptional activation, while H3K9me3 and H3K27me3 are markers of transcriptional repression, and chromatin state segmentation by Hidden Markov Model from the ENCODE/Broad project of nine different cell lines (GM12878, H1-hESC, HepG2, HMEC, HUVEC, K562, NHEK, NHLF) identified using the following different colors: yellow = weak/poised enhancer; blue = insulator; dark green = transcriptional transition/elongation (Txn), light green = weak transcript

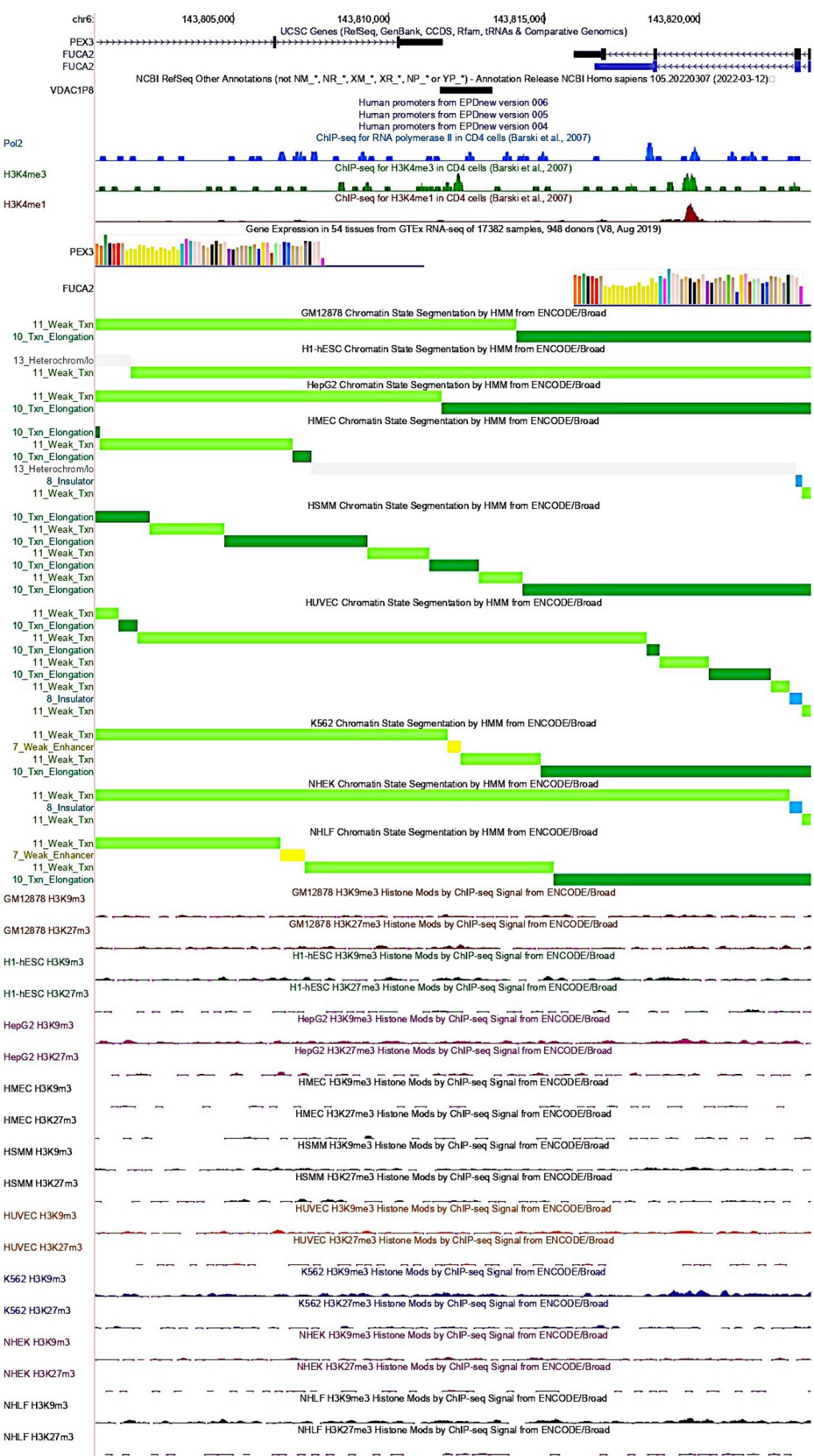


Fig. 1 (See legend on previous page.)

suggest the possibility of transcriptional activation of this genomic stretch, unlike the chromatin region where *VDAC1P5* is found (Fig. 1, Additional file 2: Fig. S2, Additional file 3: Fig. S3). Although the genomic regions of the pseudogenes *VDAC1P3*, *VDAC1P4*, *VDAC1P11* and *VDAC1P13* are associated with moderate transcriptional activity, their sequences overlap with introns of other genes (*PCDH11-X*, *ACBD6*, *ZNF169* and *KCNG-3*, respectively) apparently unrelated to *VDAC*. Their localization suggests that, in the absence of a specific maintenance mechanism, the fate of these sequences is to be degraded once spliced. The *VDAC1P1*, *VDAC1P6*, *VDAC1P7*, *VDAC1P9* and *VDAC1P10* sequences are instead associated with heterochromatin, reduced transcriptional activity and weak enhancers (Table 1). Likewise, the remaining *VDAC1* pseudogenes (*VDAC1P2*, *VDAC1P5*, *VDAC1P12*) along with all four *VDAC2* pseudogenes and the only one from *VDAC3* are located near transcriptionally inactive domains and associated with polycomb repressors (Table 1).

Expression of *VDAC* genes and their pseudogenes in tumour and normal tissues

To find support for the chromatin status analysis data and those obtained by querying UCSC Genome Browser, we analysed the transcriptomic data of *VDAC* genes (Additional file 8: Table S1) and their pseudogenes (Additional file 9: Table S2) released from the GEPIA repository. A comparative analysis of expression profiles in tumour and normal samples was performed for most of the *VDAC1* pseudogenes. As suggested by the active chromatin status where *VDAC1P8* is located, we found that this pseudogene is widely expressed in both normal and tumour tissues, with an expression value in the range of 1 to 9 transcript per million (TPM) in normal tissues, with the exception of a few cases (Additional file 9: Table S2). In particular, the highest expression level is reported in uterus, ovary, lung, rectum, colon, thyroid. Also, *VDAC1P8* is downregulated in almost all tumour tissues considered, compared with its normal counterpart. An exception is acute myeloid leukemia (AML) where, surprisingly, *VDAC1P8* is overexpressed with a TPM value reaching the highest value of 17.75 TPM (2.5 times higher than normal tissue). *VDAC1P1* is also a pseudogene expressed in all tissues and their tumour counterpart, in most case at comparable levels and in the range of 0.01–0.9 TPM (Additional file 9: Table S2). The *VDAC1P2* pseudogene is expressed, with the exception of AML, in all tumours examined, at levels between 0.11 and 0.44 TPM in the stomach and kidney, respectively. In addition, *VDAC1P2* is expressed in some normal tissues, at levels between 0.01 and 0.26 TPM. (Additional file 9: Table S2). Similarly, the *VDAC1P6* pseudogene,

located in a region of repressed chromatin is expressed in the range of 0.01–0.07 TPM in normal tissues (head, neck, kidney, bone marrow, bone and "soft" tissues—fat, muscle, nerves) and in their tumour counterparts with the highest value in colon and rectum adenocarcinoma (Additional file 9: Table S2).

Also, by examining the data retrieved from the GEPIA server, we determined that the pseudogenes *VDAC1P3*, *VDAC1P4*, *VDAC1P9*, *VDAC1P13*, and *VDAC1P11* are expressed in only very few tissues and at very low levels (0.01–0.1 TPM). No expression data were found for the pseudogenes *VDAC1P5*, *VDAC1P7*, *VDAC1P10*, *VDAC1P12* and *VDAC2P2-4*. Indeed, these pseudogenes are located in regions of repressed chromatin or where only weak active chromatin elements are present.

Furthermore, we observed that *VDAC2P1* and *VDAC3P1* are weakly expressed (range of 0.03–0.07 TPM) only in some normal connective tissues (i.e. tissues subject to sarcomas), although *VDAC3P1* is also expressed in testicular germ cells (Additional file 9: Table S2).

Our search in the GEPIA server for the expression profiles of *VDAC1-3* genes provided us with useful information as to whether the transcription of active *VDAC* pseudogenes could be linked to the expression of their parental genes. Exactly, this analysis showed that all three *VDAC* isoforms are overexpressed in many tumors, with the exception of AML, where their transcriptional levels were lower than in healthy tissue (Additional file 8: Table S1). Focusing only on expression alterations in AML, we have collected in Fig. 2a, b the results of the comparative analysis of the transcriptional profiles of *VDAC1* gene and its pseudogenes specifically reported for AML and the normal counterpart. In particular, we observed a correlation in AML between the significant downregulation of the *VDAC1* gene (from 221.12 TPM in normal tissue to 80.01 TPM in tumour) in AML and the expression levels of some pseudogenes, especially *VDAC1P8* pseudogene. Indeed, *VDAC1P1* and *VDAC1P2* show a reduction in AML, while *VDAC1P4* and *VDAC1P11* an increase over their normal counterpart but with expression levels below 1 TPM. An exception is *VDAC1P11*, which shows a change of 1.69 TPM units in AML compared with normal tissue. Interestingly, the expression level of *VDAC1P8* in AML increases of 2.5 folds, reaching the value of 17.50 TPM compared to its normal counterpart of 7.10 TPM. These intriguing data prompted us to pay attention to the putative role of *VDAC1* pseudogenes in AML and, in particular, to further investigate the involvement of the *VDAC1* gene/pseudogene *VDAC1P8* pair.

Indeed, the expression pattern of *VDAC1* gene in tumours revealed that AML is the only type in which

this gene is exclusively downregulated. On the contrary, it is even more interesting to note that in AML there is a significant increase in *VDAC1P8* pseudogene expression (Fig. 2c, d).

The putative involvement of 6q24.2 locus in AML

AML is a common haematological disorder with abnormal proliferation and differentiation of immature myeloid cells, characterised by genetic and clinical heterogeneity and high mortality [28]. Despite the development of new therapies, most patients with AML continue to relapse, highlighting the existence of an unresolved problem. More recently, the role of non-coding RNAs (ncRNAs) in the biology of AML and in the mechanisms of resistance to therapy has begun to be explored [29]. The full potential of circular RNA (circRNA), miRNA and long non-coding RNA (lncRNA) as diagnostic, therapeutic and prognostic factors in AML is thus emerging [30–32]. Also, the involvement of specific pseudogenes is beginning to appear. Our data therefore point in this direction. In particular, they show that, interestingly, AML is the only tumor type in which the *VDAC1* gene is downregulated and its *VDAC1P8* pseudogene is clearly up-regulated (Fig. 2a, b).

Furthermore, since *VDAC1P8* pseudogene is over-expressed exclusively in AML (Fig. 2a, b) we then wished to investigate whether other genes located in the same genomic locus and putatively under the effect of the same regulatory elements are also associated with the same disease. By querying the GeneCards database (<https://www.genecards.org/>) we found that the best enhancer associated with *VDAC1P8* also controls the expression of 13 other target genes (*ADAT2*, *PEX3*, *HSALNG0054032*, *RF00001-290*, *RNA5SP221*, *LTV1*, *ENSG00000270890*, *PLAGL1*, *TUBB8P2*, *ENSG00000278206*, *CM03496-315*, *FUCA2*, *MN298114-190*), located in the same 6q24.2 cytogenetic region. Interestingly, 5 genes (*ADAT2*, *ENSG00000278206*, *CM03496-315*, *FUCA2*, *MN298114-190*) as *VDAC1P8* are also correlated with AML (Additional file 10: Table S3).

These results were confirmed by querying the GEPIA database from which we also extrapolated altered expression data in AML for the *ENSG00000270890*, *TUBB8P2* and *LTV1* genes. Thus, the 6q24.2 locus would appear to have a strong association with AML. It is also noteworthy that as many as 7/13 genes (*RNA5SP221*, *HSALNG0054032*, *ENSG00000270890*, *TUBB8P2*, *ENSG00000278206*, *CM03496-315*, *MN298114-216*) with the same cytogenetic localization of *VDAC1P8* pseudogene produce ncRNA (3 pseudogenes and 4 lncRNAs). Unfortunately, there is little or no existing expression data for these genes.

Sequence analysis of *VDAC1* pseudogenes transcripts

From an initial analysis on human *VDAC* genes and their respective pseudogenes, we noticed an interesting downregulation in AML of the *VDAC1* gene, compared to its expression in normal tissue, associated with increased expression of the *VDAC1P8*, *VDAC1P1* and *VDAC1P11* pseudogenes. This prompted us to study in more detail the relationships between the *VDAC1* gene and its pseudogenes. We thus proceeded to analyze the sequences of the *VDAC1* pseudogenes by comparing them with the transcript of the parental gene. The multi-alignment shows the high homology between the pseudogenes and *VDAC1* sequences (Additional file 4: Fig. S4a), in particular the *VDAC1P1* transcript matches exactly the *VDAC1* ORF with which it has a very high sequence homology (97%). Multi-alignment also reveals the existence of additional sequence segments to the *VDAC1* ORF in only a few pseudogenes (*VDAC1P10*, *VDAC1P12*, *VDAC1P13*). The *VDAC1P8* pseudogene was not included in this multi-alignment analysis because, unlike other *VDAC1* pseudogenes that were assigned a single processed transcript, *VDAC1P8* has multiple transcripts (Additional file 4: Fig. S4b). Indeed, fourteen non-coding splicing variants, classified according to their biotypes, are annotated for this *VDAC1* pseudogene in GENCODE v.39 (Ensembl 105) (Additional file 5: Fig. S5a), as in all other databases interrogated. Among these transcript variants, 8/14 variants are reported as “intron retained”, 5/14 as “processed transcripts” and only one as a “transcribed processed pseudogene”. By analyzing the sequences of the *VDAC1P8* splice variants, emerges that the *VDAC1P8-201* transcript consists of a single exon aligning almost perfectly (89% homology) with *VDAC1* ORF sequence (Additional file 4: Fig. S4b, Additional file 5: Fig. S5). In contrast, *VDAC1P8-210* and *VDAC1P8-205* transcripts contain 5 and 4 exons, respectively, of which only the first exon overlaps with the *VDAC1P8-201* exon (87.53% and 89.28% sequence homology between the *VDAC1* ORF and the 1st exon of *VDAC1P8-205* and *VDAC1P8-210*, respectively), whereas the other exons have unrelated sequences to the parental *VDAC1* gene (Additional file 4: Fig. S4b, Additional file 5: Fig. S5). Similarly, the remaining alternative transcripts from *VDAC1P8* contain exon sequences that are not shared with the *VDAC1* ORF, so they should not be considered as correlated to this gene (Additional file 4: Fig. S4b, Additional file 5: Fig. S5). Later in the text, the term *VDAC1P8* is used to refer to the—201 variant of this pseudogene, unless otherwise specified.

Search for orthologs of *VDAC1* pseudogenes in the highest primates

Literature [5, 6, 25, 25] combined with data in the NCBI Entrez Gene database (https://www.ncbi.nlm.nih.gov/gene/?Term=related_functional_gene_22333%5Bgrop%5D%5Dshow) that several *VDAC* pseudogenes are predicted in the *Mus musculus* genome, exactly: thirteen for *mVDAC1*, three for *mVDAC2* and three more for *mVDAC3*. In their work, Ido et al. [25, 26] suggest the existence of syntenic relationships in the mouse and rat genomes between two putative rat *VDAC1* pseudogenes and one mouse *VDAC1* pseudogene, respectively. These authors also point out the lack of synteny between rodent *VDAC1* pseudogenes and humans. However, apart from the results from this outdated study focused mainly on rodent *VDAC* pseudogenes, nothing to date is known about the presence of homologs of *VDAC1* pseudogenes in higher primates.

To compensate for the lack of this and other important information, by using the Ensembl BLASTN tool we searched primate genomes for the presence of orthologous genes to the 13 human *VDAC1* pseudogenes. Sequences homologous to *VDAC1* pseudogenes were identified in all genomes examined (9 primate species, belonging to different evolutionary groups: Marmoset, Macaque, Drill, Baboon, Gibbon, Orangutan, Gorilla, Bonobo, Chimpanzee). Unfortunately, the sequences identified mostly fall into genomic tracts that are either unannotated or overlapping with coding gene sequences or, in some cases, correspond to genes for ncRNAs that have not yet been characterized. Therefore, no firm data were found on the possible expression of orthologs of the *VDAC1* pseudogenes. This lack of information combined with the high homology of the 13 query sequences used in the BLAST analysis led to the identification in several primate species of the same sequence as homologous to various human pseudogenes of *VDAC1*. For these reasons but also considering the specific overexpression of *VDAC1P8* in AML, we restricted the homology analysis to this pseudogene only. This BLAST analysis showed that in all primate species examined there is a sequence highly homologous to the human *VDAC1P8* pseudogene, which is very well maintained starting from new world monkeys to hominids (Additional file 11: Table S4).

Then, starting from the genomic location of each *VDAC1P8* ortholog and using Ensembl's "Synteny" tool, we assessed whether each genomic locus fell within the known synteny blocks between human and other primate species' chromosomes. We thus found that in all primate species analyzed, the orthologs of *VDAC1P8* are maintained within the same genomic synteny blocks with humans (Additional file 11: Table S4). These genomic regions correspond to stretches of active chromatin in

which the other genes associated with the human 6q24.2 locus and shown in Sect. 3.3 are also retained. The presence in all higher primates of sequences highly homologous to *VDAC1P8* in potentially expressed genomic regions allows us to hypothesize an important role for this backcopying, already in the physiological context.

Expression patterns and methylation profile of *VDAC1* gene and *VDAC1P8* pseudogene in healthy adult and fetal human tissues

The expression profiles data of the *VDAC1* gene and its pseudogenes, in particular *VDAC1P8*, obtained by querying the GEPIA server prompted us to further investigate their physiological expression in human adult and fetal tissues. For this aim, we analyzed the RNA-Seq CAGE (Cap Analysis of Gene Expression) RIKEN FANTOM5 project from Expression Atlas (www.ebi.ac.uk/gxa/home). Thus, we confirmed that *VDAC1* gene and *VDAC1P8* pseudogene are both normally expressed in any healthy adult and fetal tissues reported. The expression level of the *VDAC1* gene is comparable between adult and fetal tissues while the *VDAC1P8* pseudogene is more expressed in fetus (Fig. 3b). Exactly, in adult tissues the highest expression levels of *VDAC1P8* are found in bone marrow, diencephalon and heart (Fig. 3a); whereas in the fetus, there is an abundant expression in the colon, duodenum, small intestine and throat, which is up to ten times higher than *VDAC1* gene transcript (Fig. 3b). Therefore, *VDAC1P8* is mainly expressed in fetal tissues, although it is during embryogenesis that pseudogenes silencing generally occurs [33]. In particular, by performing a correlation analysis of the expression levels of *VDAC1* and its pseudogene *VDAC1P8* in the reported tissues, it appears that adult and fetal tissues with subthreshold expression of *VDAC1P8* have high levels of *VDAC1*. Likewise, tissues with high levels of *VDAC1P8* expression have reduced levels of *VDAC1* (with the exception of bone marrow in adults). No data were available in the database for the other pseudogenes of *VDAC1*.

Furthermore, the methylation profile of putative promoter sequences and gene body of *VDAC1* gene and *VDAC1P8* pseudogene was investigated across the results retrieved from MethBank. As summarized in the Fig. 4a, b, the β -value expressing the methylation level is clearly higher in the promoter region and in the gene body of *VDAC1P8* pseudogene in comparison to *VDAC1* gene for all the tissue samples derived from the central nervous system, circulation system, digestive system and lymphatic system. In particular, the promoter region of *VDAC1P8* presents a degree of methylation which is about tenfold higher than *VDAC1* gene. The methylation profile of the other *VDAC1* pseudogenes regulated

Table 2 List of transcription factors TFs for *VDAC1P8* detected in blood and bone marrow

Class	Family	Subfamily	Name		Tissue		Role in leukemogenesis (Ref.)
					Bone marrow	Blood	
Basic helix-loop-helix factors bHLH)	bHLH-ZIP	AP-4	<i>MYC</i>	Myc/Max factors	✓	X	[35, 38]
	Tal-related	TAL-HEN	<i>TAL1</i>	Tal-related factors	X	✓	[38, 43]
Basic leucine zipper factors bZIP	Jun-related	Jun	<i>JUND</i>	JunD proto-oncogene			[54]
	B-ATF-related		<i>BATF</i>	B-ATF-related			[48]
	Fos-related	Fos	<i>FOS</i>	Fos proto-oncogene	✓	X	[45]
	CEBP-related	CEBP	<i>CEBPB</i>	CCAAT/enhancer binding protein beta			
			<i>CEBPD</i>	CCAAT/enhancer binding protein delta			
Other C4 zinc finger-type factors	GATA-related	GATA double	<i>GATA1</i>	GATA binding protein 1			[37, 45]
			<i>GATA2</i>	GATA binding protein 2	X	✓	
			<i>GATA3</i>	GATA binding protein 3			
C2H2 zinc finger factors	More than 3 adjacent zinc fingers	CTCF-like	<i>CTCF</i>	CCCTC-binding factor	✓	X	[41, 42]
		ZNF362-like	<i>ZNF384</i>	Zinc Finger Protein 384	X	✓	[39]
Tryptophan cluster factors	Myb-SANT	NCoR-like	<i>RCOR1</i>	REST corepressor 1	✓	X	[47]
	ETS-related	SPI-like	<i>SPI1</i>	Spi-1 proto-oncogene	✓		[45, 46]
		ETS-like	<i>ETS1</i>	Ets-protoncogene 1	X	✓	[44]
			<i>ERG</i>	Ets-related factors			[36, 38]
			<i>FLI1</i>	Fli-1 protoncogene			[40]
			<i>IRF4</i>	Interferon-regulatory factors			[34]
Runt domain factors	Runt-related		<i>RUNX1</i>	Runt domain factors	✓		[37, 44, 45]
Type		Name			Tissue		Role in leukemogenesis Refs
					Bone marrow	Blood	
Other TFs	Epigenetic reader		<i>BRD4</i>	Bromodomain containing 4	✓		[38, 51]
	Heterochromatin protein 1 HP1 family		<i>CBX3</i>	Chromobox 3	✓	X	[50]
	Synthesizing messenger RNA		<i>POLR2A</i>	RNA polymerase II subunit A	✓		[52]
	CTCF-transcriptional coactivators		<i>RAD21</i>	RAD21 cohesin complex component	✓		[38]
			<i>STAG1</i>	Stromal antigen 1	X	✓	

In bold are the names of the transcription factors identified

The analysis was performed by querying the hTFtarget database <http://bioinfo.life.hust.edu.cn/hTFtarget#!/> and selecting only candidate TFs found in blood and bone marrow

in AML is similar to that of *VDAC1P8* (Additional file 6: Fig. S6a, b). Altogether, the information available on methylation profile of *VDAC* genes are very poor. However, the level of CG methylation of the human *VDAC* genes revealed that, in human and in mouse, the *VDAC1* promoter is less methylated than the other two isoforms confirming its ubiquitous and stable expression as house-keeping gene.

Putative TFBS suitable for driving *VDAC1P8* expression in leukemogenesis

The results of the expression analyses of *VDAC* pseudogenes showed that the pseudogenes *VDAC1P8*, *VDAC1P11* and *VDAC1P4* are over-expressed in AML, whereas *VDAC1P1* and *VDAC1P2* are under-expressed, relative to their level in the corresponding normal tissues. We therefore searched for further evidence to

(See figure on next page.)

Fig. 2 a-d. Analysis of gene expression profiles of *VDAC1* and its pseudogenes from GEPIA repository. Specific-AML expression changes of *VDAC1* compared to its pseudogenes was shown (a). Only values for *VDAC1* and *VDAC1P8* are plotted (b). Gene expression trend changes in tumor lines of *VDAC1* and *VDAC1P8* and in their normal counterparts are graphically represented (c-d). Data shown are included in Additional file 8, 9: Tables S1–S2. The height of bars is the median expression of Log2TPM + 1 of normal and AML tissue based on GTEX and TCGA respectively, available from GEPIA repository. *TPM* transcripts per kilobase million

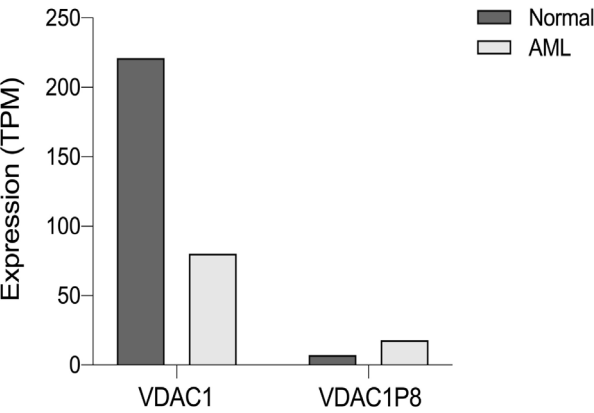
a

Gene	Normal	AML
VDAC1	221.12	80.01
VDAC1P1	0.78	0.23
VDAC1P2	0.11	0
VDAC1P3	0	0
VDAC1P4	0	0.36
VDAC1P6	0.01	0
VDAC1P7	0.01	0
VDAC1P8	7.1	17.75
VDAC1P9	0	0
VDAC1P11	0.13	1.69
VDAC1P13	0	0

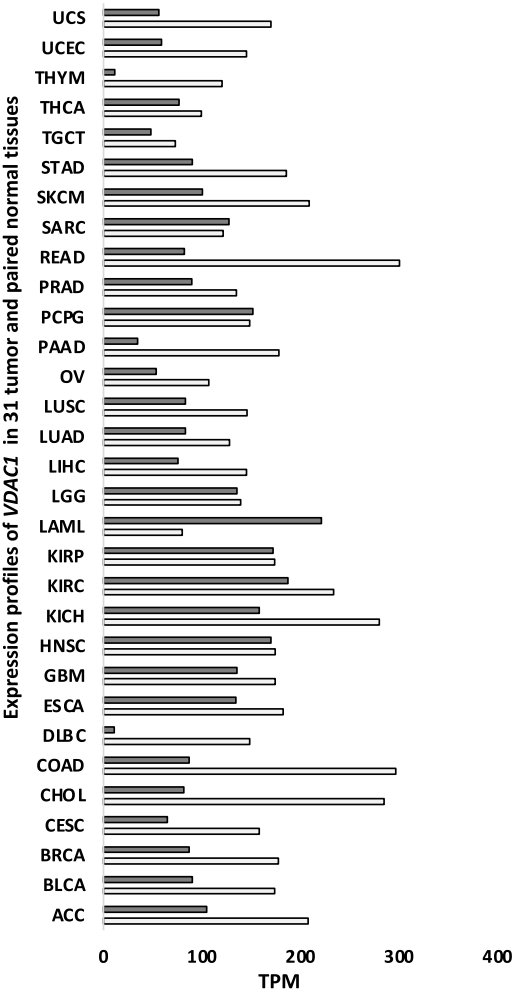
Significant gene expression changes of VDAC1 and VDAC1P8 are boxed.

upregulation downregulation

b



c



d

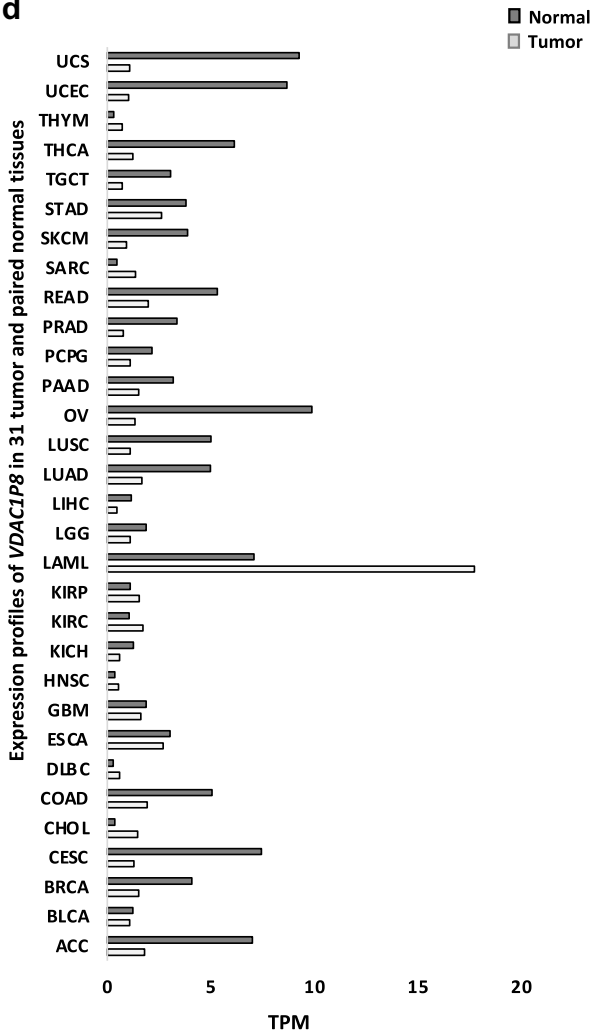


Fig. 2 (See legend on previous page.)

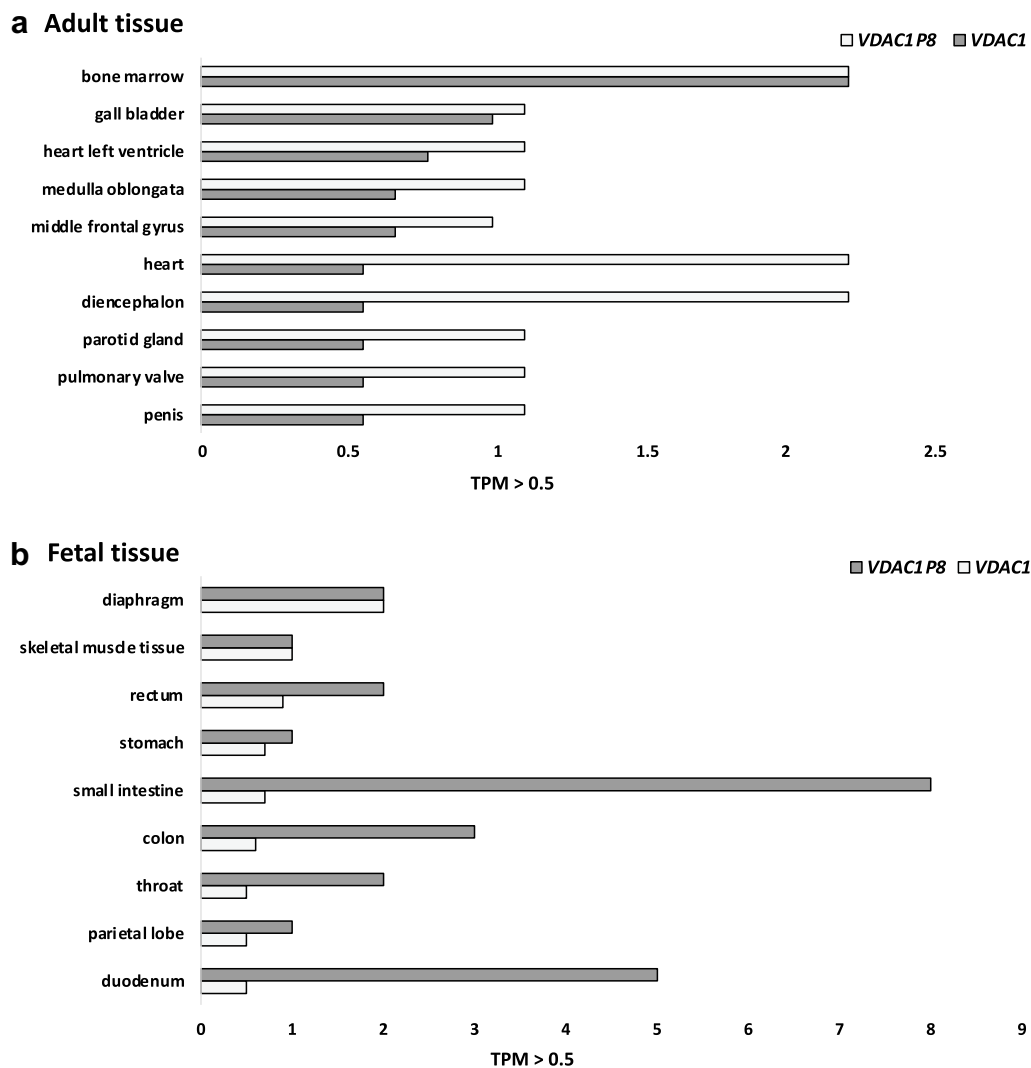


Fig. 3 a-b. Adult (a) and fetal (b) transcripts tissue expression levels of *VDAC1* and *VDAC1P8* from RNA-seq CAGE in RIKEN FANTOM 5 project. Data cut-off is > 0.05 TPM. Data shown consist of RNA-Seq CAGE (Cap Analysis of Gene Expression) analysis from RIKEN FANTOM5 project (www.ebi.ac.uk/gxa/home). TPM=transcripts per kilobase million

support these gene expression data. Specifically, we analyzed the putative TFBS pattern of the *VDAC1* pseudogenes and compared it with that of the *VDAC1* gene. This analysis was performed by querying the hTFtarget database (<http://bioinfo.life.hust.edu.cn/hTFtarget#!/>) and selecting each candidate TF in blood and bone marrow. Interestingly, only the pseudogenes *VDAC1P8* and *VDAC1P11*, upregulated in AML, have binding sites for transcription factors involved in development of the tumor counterpart. We identified 24 TFBSs in the target regulatory region of *VDAC1P8* (Table 2), some of which (*MYC*, *TAL1*, *JUND*, *BATF*, *FOS*, *CEBPB/D*, *GATA1/2/3*, *CTCF*, *ZNF384*, *RCOR1*, *SPI1*, *ETS1*, *ERG*, *FLI1*, *IRF4* and *RUNX1*) had already been associated with

leukemogenesis [34–49], while for others (*BRD4*, *CBX3*, *POLR2A*, *RAD21* and *STAG1*) the role of epigenetic modulators of the hematopoietic system, from normal to disease state, has only recently been demonstrated [38, 50–52]. In the regulatory region of *VDAC1P11*, six TFBS associated with leukemogenesis (*JUND*, *CBX3*, *RCOR1*, *GATA1*, *JUN*, *FOS*), have identified as possible regulators. To understand whether a common or differentiated regulatory pattern for the *VDAC1* gene and its pseudogenes *VDAC1P8* and *VDAC1P11* is associated with blood and bone marrow tumor development, the distribution of selected TFBS was compared (Fig. 5). The results clearly show that 19 of 24 TFBS found in the *VDAC1P8* regulatory region and involved in leukemogenesis are also

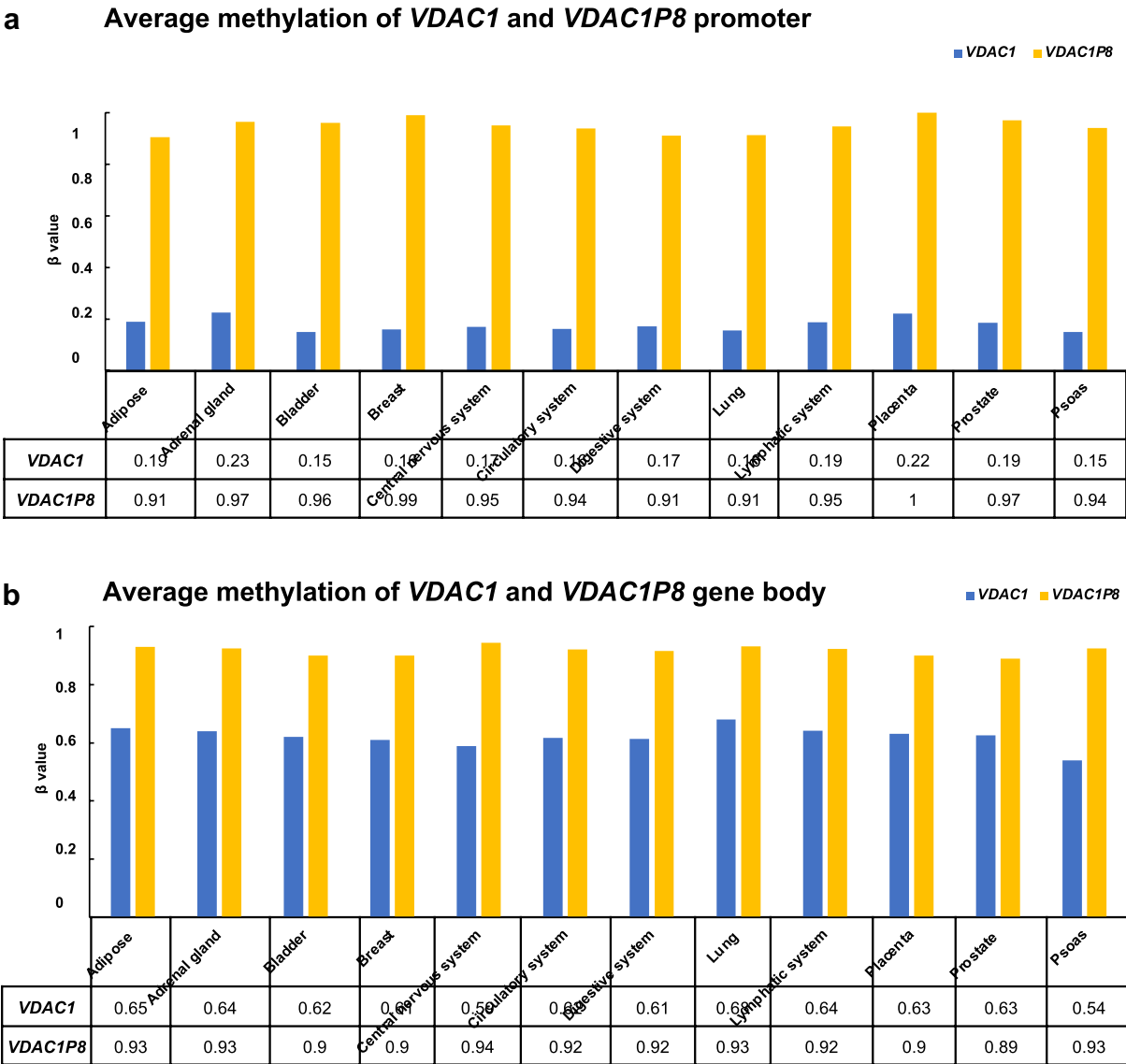


Fig. 4 a-b. Methylation levels of *VDAC1P* and *VDAC1P8* gene in putative promoter (**a**) and gene body (**b**). *The average methylation of single-based resolution methylomes (SRMs) data provided by MethBank v.4.1 (<https://ngdc.cncb.ac.cn/methbank>) are calculated as β -Value that reflects the methylation intensity at each CpG site. β -Values of 0–1 (represented from 0 to 1) indicate signifying percent methylation, from 0 to 100%, respectively, for each CpG site

putative regulators of *VDAC1* gene expression; while, only 3 of 8 were counted for *VDAC1P11*. Among all the transcription factors found, *JUND*, *CBX3*, *RCOR1* and *GATA1* are regulators commonly shared by the *VDAC1* gene and both *VDAC1P8* and *VDAC1P11* pseudo-genes. In contrast, *VDAC1P1* appears to be associated with lamin B1 (*LMNB1*) transcription factor, which is known to be overexpressed in several cancers, such

(See figure on next page.)
Fig. 5 a-d. Comparative transcriptional analysis between *VDAC1* and its pseudogenes. The analysis of TFs was performed by querying the hTFtarget database (<http://bioinfo.life.hust.edu.cn/hTFtarget#!/>) selecting only candidate TFs found in blood and bone marrow. In (**a**), the Venn diagram shows the intersection among *VDAC1*, *VDAC1P2-P4-P8* and *P11* with a null intersection of *VDAC1P1*. According to hTFtarget database, the TF found in *VDAC1P1* is lamin B1 (*LMNB1*), not shared by other pseudogenes. In (**b**), it is illustrated the intersection among *VDAC1*, *VDAC1P8* and *VDAC1P11*, which elements are represented in the table. In (**c**), the intersection between *VDAC1* and *VDAC1P8*. In (**d**), the intersection between *VDAC1* and *VDAC1P11*. The Venn diagrams made by https://bioinformatics.psb.ugent.be/cgi-bin/liste/Venn/calculate_venn.html

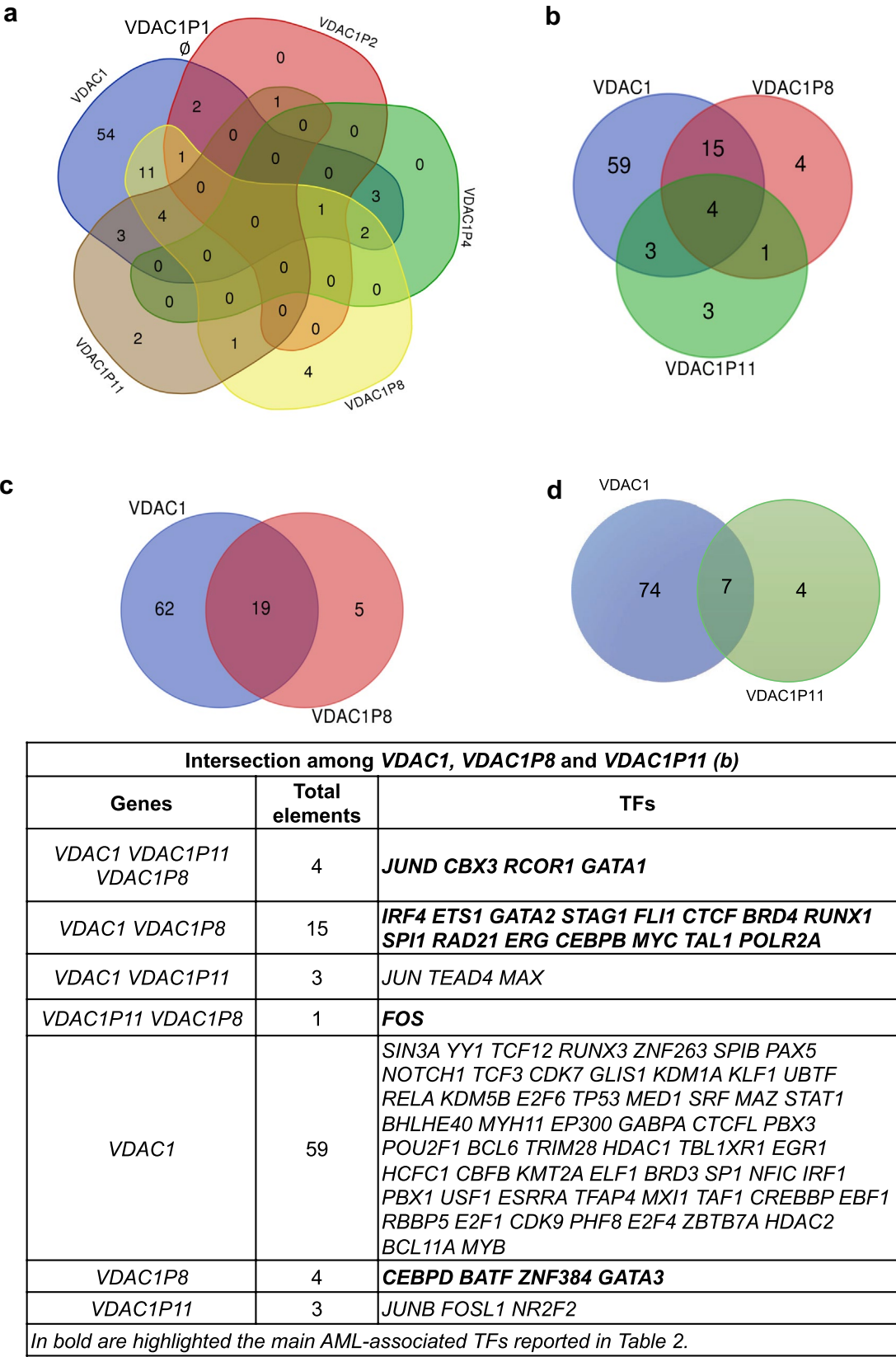


Fig. 5 (See legend on previous page.)

as lung adenocarcinoma, breast cancer and leukemia [53]. Overall, these results suggest that *VDAC1* and the pseudogenes *VDAC1P8* and *VDAC1P11* share a common transcriptional regulatory pattern in the onset of leukemia.

Expression of *VDAC1P8* pseudogene in leukemia cell lines

In light of the expression data of the *VDAC1* gene and its pseudogene *VDAC1P8* retrieved from public repository, we decided to experimentally evaluate their expression in 5 different acute myeloid leukemia cell lines. Exactly, we tested HL60, OCI-AML/2 and OCI-AML/3 (as myelocytic AML cell lines), MOLM13 (as monocytic AML cell line) and IMS-M2 (as megakaryocytic AML cell line) [55–57].

Confirming the data obtained in silico, both *VDAC1* and *VDAC1P8* were found to be expressed in all cell lines tested, with a higher transcript level of *VDAC1* than pseudogene (Fig. 6a). By sequencing the amplification products obtained by the specific primer pairs (Additional file 5: Fig. S5), we confirmed the identity of *VDAC1P8* pseudogene and *VDAC1* gene.

Also, the expression level of *VDAC1P8/VDAC1* pair in HL60 was evaluated following treatment with 5-aza-2'-deoxycytidine (AzaD). By inducing demethylation, no change in the transcript level of *VDAC1* was produced while the expression of the *VDAC1P8* pseudogene increased by an average 1.8-fold (Fig. 6b).

To confirm these data, we searched genome-wide for detailed information on the location and extent of any methylation changes induced by the demethylating agent AzaD. In particular, the methylation profile of the promoter and gene body regions of the *VDAC1* gene and the *VDAC1P8* pseudogene in HL60 cells, treated and untreated with AzaD, was analyzed. Exactly, the MethylCap-seq approach combining “methylated DNA capture” with next-generation sequencing was used for the methylome analysis of the genomic DNA from the cell samples under study.

Results obtained for *VDAC1* show an equivalent level of chromatin methylation between AzaD-treated cells and control cells in both the 600 bp of the promoter considered and the entire gene body (Fig. 7a).

In contrast, methylome analysis of four different putative regulatory regions (indicated as promoter 1, 2, 3 or 4) of *VDAC1P8* pseudogene revealed a lower degree of enrichment of methylated fragments in the genomic DNA of AzaD-treated HL60 compared to control cells (Fig. 7b–e). In particular, the putative *VDAC1P8* promoter 3, of all those positioned closest to the pseudogene body, presents a region of approximately 600 bp with a distinctly different level of methylation between the chromatin samples compared (Fig. 7d). It is also noteworthy

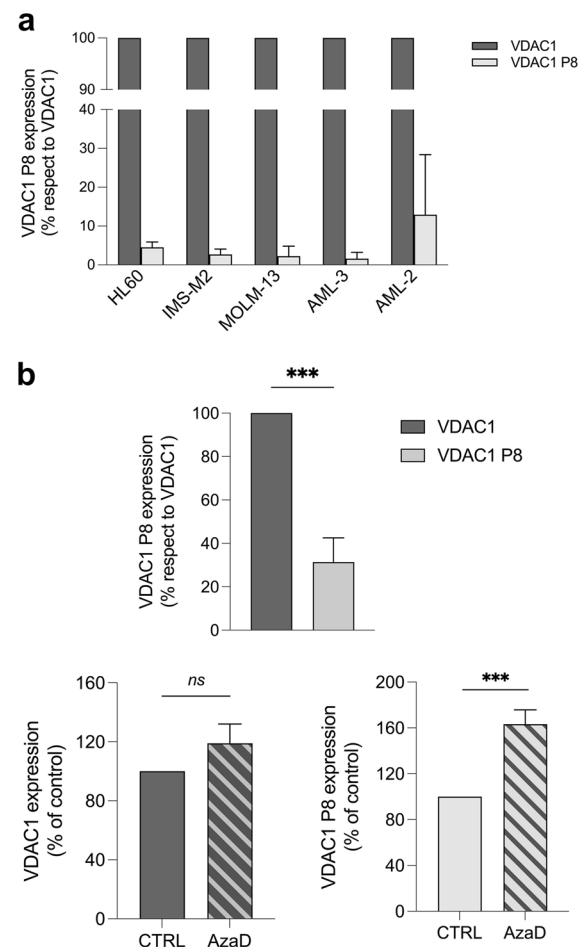


Fig. 6 a-b. Expression of *VDAC1* gene and *VDAC1P8* pseudogene in HL60 AML cell line. In (a), the expression of *VDAC1* gene and *VDAC1P8* pseudogene was evaluated in different models of leukemia cell line (HL60, IMS-M2, MOLM-13, AML-3, AML-2) by real-time PCR relative quantification. In (b), the expression of *VDAC1* and *VDAC1P8* both assessed in HL60 by real-time PCR relative quantification were also measured following 48 h of AzaD treatment. The variation of mRNA in treated samples were expressed relatively to control samples, after normalization with the housekeeping gene *GAPDH*, by the $\Delta\Delta C_t$ method. *VDAC1P8* expression is reported as % respect to *VDAC1* indicating with a value of 100 the level of *VDAC1*. Data were statistically analyzed by T-test and a value of $P < 0.05$ was considered significant.

that the whole sequence (“gene body”) of the *VDAC1P8* pseudogene was demethylated in the genome of the AzaD-treated HL60 cells in contrast to the control cells (Fig. 7f).

Therefore, methylome analysis results are well in line with the methylation profile found in MethBank and with the modulation data of *VDAC1* and *VDAC1P8* expression we obtained in HL60 cells. In particular, the *VDAC1P8* methylation profile suggests a control

mechanism for its transcriptional activity which could be triggered in specific pathological conditions as in AML.

Modulation of *VDAC1P8* pseudogene expression in *VDAC1* KO leukemic cell lines.

We also evaluated the expression of the *VDAC1-VDAC1P8* gene-pseudogene pair in the wild-type HAP1 and in the HAP1Δ*VDAC1* cell lines. HAP1 is a nearly haploid cell line of leukemic origin where by CRISPR/Cas9 editing the *VDAC1* gene has been silenced (www.horizondiscovery.com) [58, 59]. The HAP1 cells devoid of *VDAC1* were here exploited to analyze, in a leukemic context, the contribution of the *VDAC1* levels to the expression of *VDAC1P8* pseudogene, and vice versa. Interestingly, in HAP1Δ*VDAC1* cells we found that the expression level of *VDAC1P8* pseudogene, relatively to the parental *VDAC1* gene, is significantly increased in HAP1Δ*VDAC1* cells compared to HAP1 WT cells (Fig. 8a). Indeed, in HAP1 WT cells, where the *VDAC1* transcript is normally expressed, the expression level of the *VDAC1P8* pseudogene is almost tenfold lower as in above-mentioned cell lines. In contrast, in HAP1Δ*VDAC1* cells, where *VDAC1* expression is suppressed, the level of the *VDAC1P8* pseudogene is significantly doubled compared to the parental *VDAC1* gene (Fig. 8b).

Discussion

In recent years, transcribed pseudogenes have been increasingly proposed as key regulators of parental gene expression, becoming part of the non-coding RNA network. Since their discovery, almost nothing is known about *VDAC* pseudogenes [26, 60]. As the factors and circumstances leading to the regulation of *VDAC* genes are still ongoing [5, 8, 9], understanding whether there is an involvement or a role for their pseudogenes could provide an important key to understanding the whole regulatory network of *VDAC* isoforms expression. This is even more important considering that *VDAC* is already considered a main player in several diseases, and is becoming a prognostic factor and a possible therapeutic target [61, 62].

To compensate for the lack of information, in this work we investigated *VDAC* pseudogenes by means of an *in-silico* analysis integrating multi-omic data in terms of genomic features and expression profile, and also validated data in AML cell lines.

We found that among all *VDAC* pseudogenes, *VDAC1P8* appears to reside in an active chromatin region rich in regulatory elements. Exactly, *VDAC1P8* is located between two expressed genes that are controlled by regulatory elements that may direct the expression of the pseudogene transcripts. In this regard, we can speculate that the original *VDAC1* retrotranscription event led to the insertion at the 6q24.2 locus of the exonic sequence that then gave rise to the *VDAC1P8-201* pseudogene. In fact, considering the sequence and the structure of the other 13 annotated alternative *VDAC1P8* transcripts, they do not appear to originate from the *VDAC1* gene and their expression is probably due to the use of different splicing and transcription start sites.

Also, we have analyzed the transcript expression profile of each pseudogene and its parental *VDAC* gene. Expression analysis in healthy tissues, showed ubiquitous activation of some *VDAC1* pseudogenes (*VDAC1P1*, *VDAC1P2*, *VDAC1P6*, *VDAC1P8*), of *VDAC2P1* and of *VDAC3P1*. Intriguingly, the latter is also expressed in testicular germ cells, a tissue in which the cognate gene *VDAC3* plays important functions in spermiogenesis and male fertility [63, 64]. Some *VDAC1* pseudogenes have been also found to be expressed in cancer, among them some exclusively or predominantly associated with tumor phenotypes. In particular, pseudogenes *VDAC1P1*, *VDAC1P2*, *VDAC1P4*, *VDAC1P11* and *VDAC1P8* correlate with AML. Specifically, the expression profile of *VDAC1P1* and *VDAC1P2* in this analysis largely mirrors that of *VDAC1*, in both healthy and tumor tissues. In contrast, *VDAC1P4*, *VDAC1P11* and, more interestingly, *VDAC1P8* have an expression trend exactly opposite to that of *VDAC1*. In general, *VDAC1P8* is under-expressed compared to healthy tissue in all tumor samples tested, with the exception of AML where it is highly overexpressed.

These expression data of the *VDAC1P8/VDAC1* pair were confirmed in five AML cell lines (myeloblastic, monoblastic, megakaryoblastic) precursors of different blood cells.

The same expression ratio of *VDAC1* and *VDAC1P8* transcripts is also found in the nearly haploid HAP1 wild-type human leukemic line. In contrast, in the HAP1-*VDAC1* knock-out line, the level of *VDAC1P8* increases significantly, suggesting that the expression of the *VDAC1* gene and its pseudogene *VDAC1P8* may be related to each other.

(See figure on next page.)

Fig. 7 a-f. Methylation patterns (β values) of the regulatory regions of *VDAC1* and *VDAC1P8* in control or AzaD-treated HL60 cells. Data was obtained by MethylCap-seq protocol and processed using the bioinformatic pipeline analysis described in Methods section. The promoters and gene sequences were tiled into 300 bp non overlapping windows; the color scale (range 0–1) represents the methylation levels of 300 bp sequence. Red boxes represent fully methylated windows and blue boxes represent unmethylated windows. Windows with insufficient readings to estimate a β -value are in grey (NA)

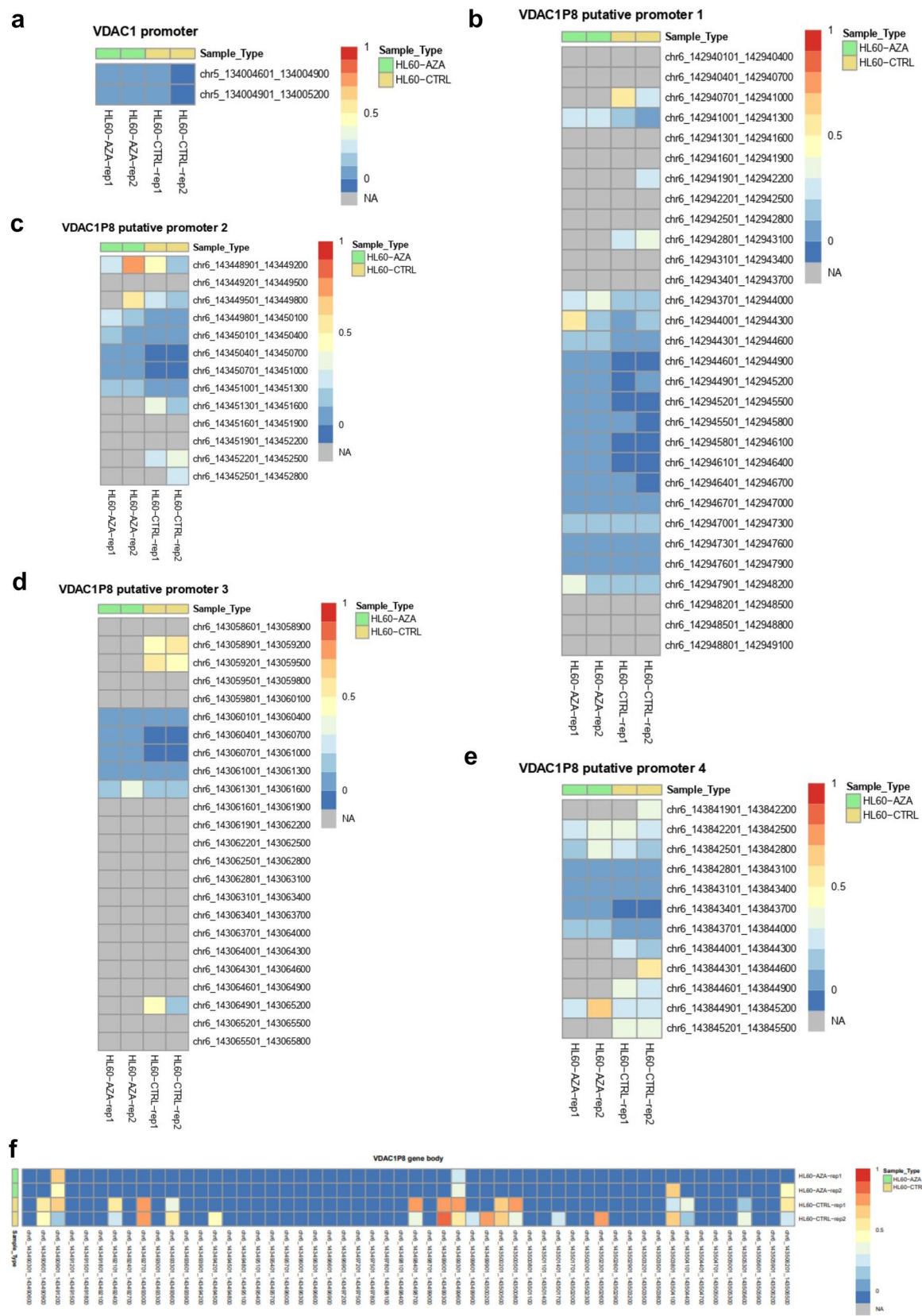


Fig. 7 (See legend on previous page.)

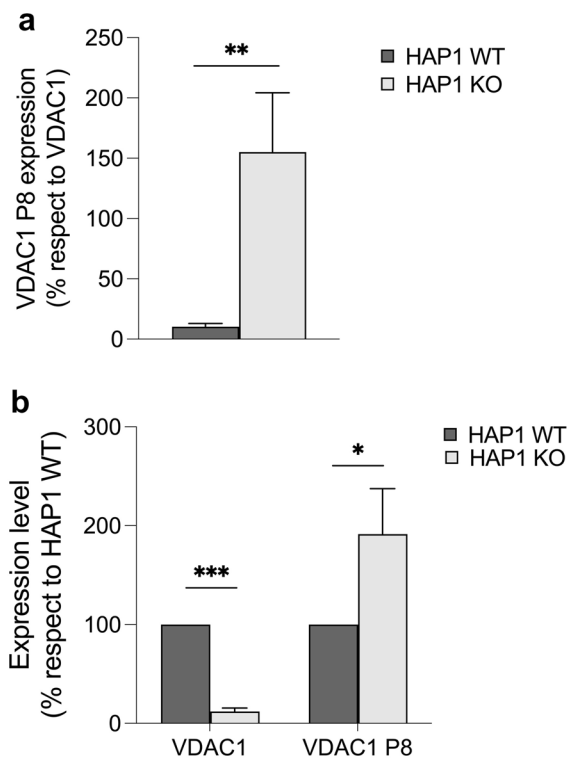


Fig. 8 a, b. Expression of *VDAC1* gene and *VDAC1P8* pseudogene in *VDAC1* KO leukemic cell lines. The level of *VDAC1* and *VDAC1P8* transcripts expression were evaluated in HAP1 WT and HAP1 *VDAC1* cells by real-time PCR relative quantification. In (a) *VDAC1P8* transcript level was measured relatively to that of *VDAC1*. In (b) the expression level of both *VDAC1* and *VDAC1P8* in HAP1 *VDAC1* cells was measured relatively to their relative amount found in HAP1 WT. After normalization with the housekeeping gene *GAPDH*, by the $\Delta\Delta C_t$ method. *VDAC1P8* expression is reported as % respect to *VDAC1* indicating with a value of 100 the level of *VDAC1*. Data were statistically analyzed by T-test and a value of $P < 0.05$ was considered significant

Also, in regulatory sequences that could control *VDAC1P8* gene we identified 24 putative TFBSs associated with bone marrow, blood and leukemogenesis. Interestingly, common transcriptional factors involved in leukemia are shared in the regulatory sequences of *VDAC1* gene and *VDAC1P8* pseudogene, suggesting that both could play a role in this pathology.

Taken together, these results highlight that *VDAC1P8* is closely associated with leukemia, a finding mentioned in the literature when cancer-specific pseudogenes were identified in chronic lymphocytic leukemia [27].

It is also important to note that previous studies have shown the existence of different DNA methylation rates for pseudogenes and their parent genes [33, 65, 66]. To better understand the significance of *VDAC1* pseudogenes activation patterns in relation to parent gene expression levels, we focused our attention on the DNA

methylation status of these pseudogenes and *VDAC1* gene. Exactly, we observed that the methylation levels of the *VDAC* pseudogenes presumably depended on their position within the genome. In particular, the expression of the *VDAC1P8* pseudogene undergoes no or little variation and is presumably under the transcriptional control of adjacent genes, as observed with other pseudogenes (i.e. *LIN28* and *PTEN*) [33, 65, 66]. Indeed, five other genes located at the 6q24.2 locus and controlled by the same regulatory elements acting on *VDAC1P8* correlate with AML.

In addition, we searched in the MethBank database for data on the promoter and gene body methylation profile of the *VDAC1* gene and its pseudogenes, particularly *VDAC1P8*. We found that all *VDAC1*-related pseudogenes exhibit a higher promoter and gene body methylation profile than the *VDAC1* gene, suggesting a normally repressed status but eventually subjected to regulation. Indeed, our experimental data in HL60 show that the regulatory regions of the *VDAC1* gene are essentially under-methylated unlike those of its pseudogene *VDAC1P8*. This result suggests that in AML the pseudogene undergoes a methylation pattern independent of that of the parental gene. In this regard, we could also speculate that chromatin methylation, and epigenetic modifications in general, may contribute to the transcriptional changes of *VDAC1P8* and *VDAC1* in AML pathology. Although, the unusual downregulation of the *VDAC1* gene in AML, could be the result of different mechanisms of gene expression regulation. The epigenetic control mechanisms hypothesized here are reflected in other examples found in the literature concerning the transcriptional control of the *VDAC1* gene under extremes conditions. For example, in tumor cells and in placental trophoblasts from patients with recurrent miscarriages, the increase of *EPB41L4A-AS1* lnc-RNA induced the enhancement of *VDAC1* promoter activity affecting histone modification. In this condition, *EPB41L4A-AS1* lnc-RNA is considered to regulate and to reprogramme tumor and trophoblast cells metabolism. As a consequence, the activation of oxidative metabolism was triggered by enhanced mitochondrial function [67, 68].

Overall, our data suggest a role for some *VDAC1* pseudogenes in both physiological and pathological processes, particularly in AML. In this context, comparing the expression level of the *VDAC1* gene with the two groups of pseudogenes, those overexpressed and those down-regulated, reveals a possible role from ceRNA for these pseudogenes in AML. Considering the inverse relationship between the expression of *VDAC1P4*, *VDAC1P11* and in particular *VDAC1P8* and *VDAC1*, these pseudogenes in AML could function as competing-endogenous RNA (ceRNA), or rather as a lncRNA able to generate

siRNA acting on the transcript of the parental gene, as described in [18]. Furthermore, there are numerous examples in the literature of lncRNAs related to leukemia progression [69]. With regard to *VDACIP1*, *VDACIP2* these pseudogenes have an expression profile in AML that follows that of the *VDAC1* gene. It is therefore likely that these pseudogenes can act as competitive lncRNA for binding miRNAs or regulatory RNA-binding proteins (RBP) directed to the parental gene, like other pseudogenes [19]. This putative mechanism of action correlates very well with the observation that several miRNAs modulating myeloid differentiation (e.g. *miR-20a*, *miR-106b* and *miR-125b*) [69] or AML [70, 71] are also predicted to control *VDAC1* expression [72] using mirSystem, a miRNA target prediction software (<http://mirsystem.cgm.ntu.edu.tw/index.php>) [73]. This effect could have important consequences especially considering the role of *VDAC1* protein in apoptosis. Since in leukemia the expression of *VDACIP8* is considerably higher than in healthy tissue, this pseudogene could be responsible for, or at least contribute to, the down-regulation of *VDAC1* in AML. Given the pro-apoptotic role of *VDAC1* protein, in this pathological context a reduced expression could be compatible with the increased cell resistance to apoptosis typical of cancer phenotype. It is also useful to remember that active retrocopies produce lncRNAs, molecules playing many regulatory roles in cancer [74], especially by influencing various cellular mechanisms that promote energy metabolism in cancer [75, 76]. At the OMM, *VDAC* serves as a docking site for cytosolic enzymes, such as hexokinase, and is a key protein in mitochondrial apoptosis. Considering the significance of hexokinase for tumor metabolism, its interaction with *VDAC* and the latter's role in apoptosis, it is conceivable that lncRNAs modifying the activity of glycolytic enzymes and energy metabolism in cancer have an impact on *VDAC* function and apoptosis, thus helping to define the tumor phenotype.

Conclusion

The total number of pseudogenes in the human genome and their function are not yet fully known. It is also unclear why some human genes, such as *VDAC* genes, possess several pseudogenes, differently localized from their parental genes, while others possess only one or a few, and still others none. Our search for orthologous genes indicated strong conservation throughout primate evolution of all *VDAC1* pseudogenes. This leads us to consider *VDAC1* pseudogenes advantageous to primates and to hypothesize that there was selective pressure towards their maintenance already at birth in ancestral primates. This is not surprising considering that *VDAC1* is a very important housekeeping gene under

stress conditions, when mitochondrial function must be ensured in cells [77].

Overall, this is the first comprehensive survey of human *VDAC* pseudogenes reported in the literature, and the results presented here provide important insights into their relevance to human genome expression. We found evidence for their activation and importance as human-specific regulatory elements. In particular, *VDACIP8* appears to be particularly involved in hematopoiesis and emerges as a putative AML-specific diagnostic and prognostic factor. This leads to the evaluation of its regulatory mechanism as a possible target in the fight against the disease. The low survival rate of AML patients ignites the demand for new therapies for this widespread malignancy [84]. More in-depth studies are therefore desirable to elucidate the mechanism by which *VDACIP8* (and other *VDAC* pseudogenes) influence the expression of parental genes in AML, i.e. how their ncRNAs act on regulatory RNA, DNA or RBPs. So far, we have provided data supporting the association of the *VDACIP8* pseudogene with AML as well as a suitable detection method to distinguish it from its parental gene *VDAC1*. This lays the foundation for its possible use as a biomarker for AML.

Methods

Data collection process

Specific repositories used to retrieve data on *VDAC* pseudogenes were Ensemble (<https://www.ensembl.org/index.html>), NCBI Entrez Gene (<https://www.ncbi.nlm.nih.gov/>), UCSC Genomic Browser (<https://genome.ucsc.edu>), Hoppsigen database (<http://pbil.univ-lyon1.fr/databases/hoppsigen.html>) [78], pseudogene.org, by Yale Gerstein Group [79], DreamBase (<https://rna.sysu.edu.cn/dreamBase/>) [80] and PseudoFun (https://integrativeomics.shinyapps.io/pseudofun_app/) [81].

Genomic data analysis

UCSC Genomic Browser (assembly GRCh38/hg38) was used to analyze the genomic structure of *VDAC* pseudogenes and gene regulatory data from hub tracks' selection. This included Pseudogene Annotation Set of GENCODE v.38lift37 (Ensemble 104), Eukaryotic Promoter Database (EPD) v.4–6, CpG islands track, Genotype-Tissue Expression (GTEx) RNA-seq v.8 (2019), ChIP-Seq data for RNA polymerase II, H3K4me3 and H3K4me1, and chromatin state segmentation by Hidden Markov Model from ENCODE/Broad project. GRCh37/hg19 assembly was used because it provided a more suitable representation of the genomic trait for data visualization and interpretation. The range of analysis of the genomic context of each pseudogene was set about 1000 kb upstream and downstream from the Refseq annotation. In addition, comparative genomics data were

acquired from the Ensembl vertebrate genome browser (<https://www.ensembl.org/index.html>) to syntenic search between human *VDAC* genes or their pseudogenes and more evolved primates. Also, the GeneCards—The Human Gene Database (<https://www.genecards.org/>) was queried to refine the genomic information related to *VDAC* pseudogenes.

Transcriptomic data analysis

GEPIA (Gene Expression Profiling Interactive Analysis) web-based tool and GEPIA2 version were accessed in order to collect gene expression profiles of *VDAC* pseudogenes and their parental genes, in both tumor and healthy tissues [82, 83]. GEPIA provides data of TCGA [84] and GTEx (Genotype-Tissue Expression project) dataset release v.8 (<https://www.gtexportal.org/home/>). Expression Atlas (www.ebi.ac.uk/gxa/home) [85] was also exploited for the analysis of adult and fetal human tissues in RIKEN functional annotation of the mammalian genome 5 (FANTOM5) project (<https://fantom.gsc.riken.jp/5/>) [86]. In addition, GWAS Catalog at <https://www.ebi.ac.uk/gwas/home> was queried for gene-linked phenotypes/functions in genome-wide association studies.

Sequence alignment analysis

Sequence alignment analysis was running at free software online (<http://multalin.toulouse.inra.fr/multalin/>). The percentage of homology between the examined sequences was obtained by BLAST website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The search for orthologous genes was performed using the BLAST/BAT tool at Ensembl.

Prediction analysis of transcription factor binding sites (TFBSs) in the *VDAC* pseudogenes regulatory regions

The prediction of potential transcription factors (TFs) and their transcription factor binding sites (TFBSs) on sequences associated with human *VDAC1* pseudogenes was conducted by hTFtarget database (<http://bioinfo.life.hust.edu.cn/hTFtarget#!/>) [87]. Each record of candidate TFs for query gene provides the chromosomal location of chromatin immunoprecipitation sequencing (ChIP-seq) peaks in different experimental conditions or cell lines. The association between candidate TF and target sequences is computed through a machine learning method and accumulated data of ChIP-seq peaks, epigenetic modification status of TFBSs and targets, co-regulation of TFs of the same targets, and co-association of TFs.

DNA methylation bioinformatic analysis

DNA methylation profiles of both regulatory sequences and gene bodies of *VDAC1* and *VDAC1P8* were

investigated using data of single-base resolution methylomes (SRMs) retrieved from MethBank v.4.1 (Sep 2020) (<https://ngdc.cncb.ac.cn/methbank>) [88]. Since MethBank contains a large amount of SRM replicates both from healthy and pathological condition (prevalently lung and prostate carcinomas), our search strategy filtered SRM data for ‘human healthy tissues’ selecting epigenetic change of guanine–cytosine (‘GC’) base pair found in the ‘promoter’ and ‘gene body’. The average methylation data are calculated as β -value that reflects the methylation intensity at each CpG site. β -values of 0–1 (represented from 0 to 1) indicate signifying percent methylation, from 0 to 100%, respectively, for each CpG site. For illustration purpose, in case of some tissue replicates (adipose, adrenal gland, bladder, breast, lung, placenta, prostate, psoas) we averaged their β -values showing a single record for each sample. We also averaged samples belonging to the same apparatus clustering them into following systems: circulatory (aorta, carotid, right atrium, right and left ventricles, whole blood); digestive (pancreas, small and large intestine, stomach, colon, sigmoid colon, esophagus, esophagus squamous epithelium, adjacent esophageal tissue, gastric); lymphatic (spleen, thymus), nervous (anterior cingulate cortex, cerebral cortex (fetal), frontal lobe, hippocampus, nucleus accumbens, prefrontal cortex (Brodmann area 9), retina).

MethylCap-seq analysis

A full description of the MethylCap-seq procedure can be found in [89]. In summary, in our experiments 100 ng of genomic DNA from HL60 cells was subjected to enzymatic fragmentation, end repair/A-tailing and adapter ligation using Sure Select SureSelect XT HS2 DNA library preparation kit (Agilent). Adapter ligated DNA was subjected to methylation enrichment using MethylCap Kit (Diagenode) following the manufacturer’s protocol, with some little modifications. The amounts of MethylCap protein and magnetic beads were decreased proportionally according to the recommended DNA to protein and beads ratio (0.2 μ g protein and 3 μ l beads per 100 ng DNA input). Single fraction elution with High Elution Buffer was applied. The eluted fraction was purified by Ipure kit V2 (Diagenode). Then, the methylation-enriched DNA libraries were amplified using Sure Select SureSelect XT HS2 DNA library preparation kit (amplification step). During amplification, a unique index from the primer was added to the sequencing adapter for each sample. The amplified libraries were purified using 1 $\times$ AMPure XP beads (Agilent). All final libraries were first quantified using D1000 Screen Tape Assay (Agilent). Libraries were then sequenced by NextSeq 2000 platform (Illumina; San Diego, CA, USA), 2 $\times$ 150 bp paired-end read.

FASTQ files were processed using a bioinformatic pipeline for data analysis described in Chemi et al. [90]. Briefly, FASTQ files were trimmed using Agilent AGeNT tool (<https://www.agilent.com/cs/library/usermanuals/public/G9983-90000.pdf>), they were aligned to the GRCh38 reference genome using bwa mem [91]. BAM files were deduplicated using Picard markduplicates (<https://gatk.broadinstitute.org/hc/en-us/articles/360037052812-MarkDuplicates-Picard>) followed by running Samtools [92] fixmate to assign mate quality scores. The QSEA R package [93] was used to analyze BAM files, with the use of a custom R package to extend QSEA, as described in Chemi et al. [90]. The entire genome was tiled into 300-bp non-overlapping windows, with the removal of windows lying within the encode exclusion list regions. Reads were then uniquely assigned into these almost nine millions bins according to their midpoint location. Reads were filtered by keeping read pairs where either end of the pair mapped with a mapping quality (MAPQ) score of at least 10. For paired reads, a fragment length between 50 and 1000 bp was required and a distance along the reference genome of at least 30 bp was required. Copy number variations were calculated from the non-enriched input sequencing for each sample, using HMMcopy with base parameters over 1-Mbp windows. Each sample was normalized for library size using TMM (trimmed mean of M values, part of QSEA) with a pooled reference sample. β -values (a scaled measure of methylation between 0 and 1) for each window in each sample were estimated within QSEA using the 'blind calibration' method; windows with insufficient reads to estimate a β -value were returned as NAs.

Cell cultures and treatment

The acute myeloid leukemia model cell lines (OCI AML2, OCI AML3, MOLM13, IMS M2) from American Type Culture Collection (Manassas, VA, USA) and HL60 (from DSMZ cell cultures, Braunschweig, German) were cultured in RPMI-1640 medium (Sigma Chemical Co.—St. Louis, MO, USA) supplemented with 10% fetal bovine serum, 2 mM L-Glutamine, 100 U/ml Penicillin and 100 µg/ml Streptomycin (EuroClone, Pero, MI, Italy). The cells were grown at 37 °C in a humidified atmosphere containing 5% CO₂ and dissociated with 0,25% trypsin-0,02% EDTA solution. The HAP1 parental (C631) and knock-out for VDAC1 (HZGHC005706c002) cell lines were purchased from Horizon Discovery (Waterbeach, UK). Cells were maintained as in [58], as was the monitoring of their proliferation. For cell treatments, 5-Aza-2'-deoxycytidine (AzaD) was purchased from Sigma Chemical Co. (St. Louis, MO, USA) and resuspended in dimethyl sulfoxide (DMSO) at the concentration of 50 mg/mL, respectively. The cells were seeded at

0,1×10⁶/ml in 25 cm² culture flasks (NUNC, Thermo Scientific, Roskilde, Denmark) and treated with 5 µM AzaD for 48 h.

Real-time amplification

Real-time amplification was performed in a Mastercycler EP Realplex (Eppendorf) in 96-well plates. cDNAs from leukemia cells line HL60 were used as template for the experiments. The reaction mixture was composed by 1 µl cDNA, 0.2 µM gene specific primers pairs for *VDAC1*, *VDAC1P8-201* and *GAPDH* (Additional file 12: Table S5) and 12.5 µl of master mix (QuantiFast SYBR Green PCR kit, Qiagen). Analysis of relative expression level was performed using the housekeeping *GAPDH* gene as internal calibrator by the $\Delta\Delta C_t$ method. Three independent experiments were performed. Data were statistically analyzed by Student t-test. Values of * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ were taken as significant.

Abbreviations

AzaD	5-Aza-2'-deoxycytidine
AML	Acute myeloid leukemia
ceRNA	ChIP-seq, chromatin immunoprecipitation sequencing
circRNA	Circular RNA
ceRNA	competitive endogenous RNA
EPD	Eukaryotic Promoter Database
GEPIA	Gene Expression Profiling Interactive Analysis
GTEx	Genotype-Tissue Expression
HK	Hexokinase
lncRNA	Long non-coding RNA
miRNA	MicroRNA
OMM	Outer mitochondrial membrane
TF	Transcription factor
TFBS	Transcription factor binding site
siRNA	Short interfering RNAs
RBP	RNA-binding proteins
SRMs	Single-base resolution methylomes
TPM	Transcript per million
VDACs	Voltage dependent anion selective channels

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40659-023-00446-1>.

Additional file 1: Figure. S1a, b Summary of annotated transcripts for the *VDAC1P8* pseudogene from GENCODE v.39 Ensembl 105. In (a) the table reports gene annotations of 14 transcripts for *VDAC1P8* reported on Ensembl database release 105. In (b) the screenshot modified by GENCODE v.39 illustrates some main structural and regulatory features of the variants' alignment.

Additional file 2: Figure. S2. Chromatin state and genomic features of *GAPDH* gene from UCSC Genome Browser GRCh37/hg19. The genomic context of *GAPDH* set around 1000 Kb upstream and downstream of the annotated Refseq is shown. The selected regulatory hub tracks are Pseudogene Annotation Set from GENCODE v.38lift37 Ensemble 104, Eukaryotic Promoter Database EPD v.4-6, CpG island track, Genotype-Tissue Expression GTEx RNA-seq v.8 2019, ChIP-Seq data for RNA polymerase II, H3K4me3 and H3K4me1, used as markers of transcriptional activation, while H3K9me3 and H3K27me3 are markers of transcriptional repression, and chromatin state segmentation by Hidden Markov Model from the ENCODE/Broad project of nine different cell lines (GM12878, H1-hESC,

HepG2, HMEC, HUVEC, K562, NHEK, NHLF) identified using the following different colors: bright red=active promoter; light red= weak promoter; orange = strong enhancer; dark green = transcriptional transition/elongation (Txn).

Additional file 3: Figure. S3. Chromatin state and genomic features of *VDAC1P5* gene from UCSC Genome Browser GRCh37/hg19. The genomic context of *VDAC1P5* set around 1000 Kb upstream and downstream of the annotated Refseq is shown. The selected regulatory hub tracks are Pseudogene Annotation Set from GENCODE v.38lift37 Ensemble 104, Eukaryotic Promoter Database EPD v.4-6, CpG island track, Genotype-Tissue Expression GTEx RNA-seq v.8 2019, ChIP-Seq data for RNA polymerase II, H3K4me3 and H3K4me1, used as markers of transcriptional activation, H3K9me3 and H3K27me3 are markers of transcriptional repression, and chromatin state segmentation by Hidden Markov Model from the ENCODE/Broad project of nine different cell lines (GM12878, H1-hESC, HepG2, HMEC, HUVEC, K562, NHEK, NHLF) colored in grey to indicate the heterochromatin state.

Additional file 4: Figure. S4a, b Sequence multi-alignments among *VDAC1* and pseudogenes. In (a) the multi-alignment encompasses *VDAC1* against *VDAC1P1-7* and *VDAC1P9-13*. In this picture, the *VDAC1* mRNA sequence is shown starting at nucleotide 201 because its start ATG is located at 203-205 position and also because the previous sequence stretches nucleotides 1-200 does not align with the *VDAC1* pseudogenes sequences. In (b), the multi-alignment is among *VDAC1* and splicing variants of *VDAC1P8* *VDAC1P8-201*, *-205* and *-210*. The start and end codons of the *VDAC1* coding sequence are boxed in yellow.

Additional file 5: Figure. S5. Multi-alignment between the *VDAC1* mRNA and some *VDAC1P8* alternative transcripts sequences. Alignment achieved by <http://multalin.toulouse.inra.fr/multalin/> setting the default parameters. The regions with high consensus value (90% setting) are shown in red, while regions with low consensus value (50% setting) are shown in blue.

Additional file 6: Figure. S6a, b Methylation levels of *VDAC1P1*, *VDAC1P4*, *VDAC1P8* and *VDAC1P11* gene in putative promoter (a) sequence and gene body (b). *The average methylation of different healthy human samples from single-based resolution methylomes (SRMs) are provided by MethBank v.4.1 (<https://ngdc.cncb.ac.cn/methbank>). SRMs data are calculated as β -Value that reflects the methylation intensity at each CpG site. β -Values of 0–1 (represented from 0 to 1) indicate signifying percent methylation, from 0 to 100%, respectively, for each CpG site.

Additional file 7: Figure. S7. Mapping of the best putative promoters for the *VDAC1P8* pseudogene (Promo 1-4) inferred by GeneHancer and found in Genecards (hg38). Genomic regions indicated in figure as Promo 1, 2, 3 and 4 correspond to putative *VDAC1P8* promoters (GH06J142940, GH06J143449, GH06J143058, GH06J143842) selected from GeneCards (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=VDAC1P8>) with the best GH score (from 2.1 to 1.9). They map around the NCBI Refseq of *VDAC1P8* gene and the transcript *VDAC1P8-201* (ENST00000406025.2). All four putative promoters for *VDAC1P8* fall within transcriptionally active chromatin regions as indicated by the levels of activating methylation at histone H3 (H3K4m1 and H3K4m3, histone H3 methyl-lysine 4 and histone H3 trimethyl-lysine 4, respectively). Bottom right, an enlargement of the genomic region between *VDAC1P8* promoter 3 and the gene for the *VDAC1P8-201* transcript, a region potentially under the control of the regulatory sequences of the adjacent genes *ADAT2*, *PEX1* and *FUCA2*.

Additional file 8: Table S1. Expression profiles of *VDAC1-3* genes from GEPIA server across 31 tumor and paired normal tissues. The median expression of Log2TPM+1 of certain tumor type or normal tissue based on TCGA and GTEx respectively. TPM = transcript per million.

Additional file 9: Table S2. Expression profiles of *VDAC* genes from GEPIA server across 31 tumor and paired normal tissues. No data was found for *VDAC1P5*, *VDAC1P7*, *VDAC1P10*, *VDAC1P12* and *VDAC2P2* pseudogenes. (a) *VDAC1* pseudogenes; (b) *VDAC2* pseudogene; (c) *VDAC3* pseudogene. The median expression of Log2TPM+1 of certain tumor type or normal tissue based on TCGA and GTEx respectively. TPM = transcript per million.

Additional file 10: Table S3. Genes at the 6q24.2 locus controlled by the same elements in the *VDAC1P8* pseudogene. Data shown below

are from GeneCards database. *From GWAS catalog; **From GEPIA database; ***From ENA (European Nucleotide Archive at EBI); ND= no data available; lncRNA=long non-coding RNA; ncRNA= non-coding RNA; rRNA=ribosomal RNA.

Additional file 11: Table S4. BLAST analysis of human *VDAC1P8* pseudogene in all primate species.

Additional file 12: Table S5. List of primers used for Real-Time amplification.

Acknowledgements

The authors thank Cogentech S.R.L.—Catania for the support provided in the methyloma data analysis. Authors are grateful to Dr. Daniele Tibullo (University of Catania) for the gift of the AML cell lines used in the work. Authors also acknowledge Fondi di Ateneo 2020–2022 and Linea Open Access, University of Catania.

Author contributions

XGP developed the pipelines, did most of the bioinformatic analyses, and wrote together PR the initial drafts of the manuscript. PR performed also some bioinformatics analysis and analysed the data. FG performed the sequence alignments. FZ and FG performed the analysis to search for TFs. AO and DL performed all cell experiments. AM conceived the study, performed some bioinformatics analysis and wrote the final drafts of the manuscript. VDP and FB revised and edited the text. XGP, FG, AM contributed to the final version of the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the Italian Ministry of University and Research—Proof of Concept 2018, grant number PEPSLA POC 01_00054 to A.M., by University of Catania—PIACERI line, grants number “ARVEST” to A.M. and “VDAC” to VDP, Fondi di Ateneo 2020–2022, Linea Open Access and Linea CHANCE to VDP.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 21 January 2023 Accepted: 8 June 2023

Published online: 22 June 2023

References

- Lemasters JJ, Holmuhamedov E. Voltage-dependent anion channel (VDAC) as mitochondrial governor—thinking outside the box. *Biochim Biophys Acta*. 2006;1762(2):181–90.
- De Pinto V. Renaissance of VDAC: new insights on a protein family at the interface between mitochondria and cytosol. *Biomolecules*. 2021;11(1):107.
- Reina S, Magri A, Lolicato M, Guarino F, Impellizzeri A, Maier E, et al. Deletion of β -strands 9 and 10 converts VDAC1 voltage-dependence in an asymmetrical process. *Biochim Biophys Acta*. 2013;1827(6):793–805.
- Aiello R, Messina A, Schiffler B, Benz R, Tasco G, Casadio R, et al. Functional characterization of a second porin isoform in *Drosophila melanogaster*.

- DmPorin2 forms voltage-independent cation-selective pores. *J Biol Chem*. 2004;279(24):25364–73.
5. Messina A, Reina S, Guarino F, De Pinto V. VDAC isoforms in mammals. *Biochimica et Biophysica Acta (BBA) Biomembranes*. 2012;1818(6):1466–76.
 6. Messina A, Oliva M, Rosato C, Huizing M, Ruitenbeek W, van den Heuvel LP, et al. Mapping of the human voltage-dependent anion channel isoforms 1 and 2 reconsidered. *Biochem Biophys Res Commun*. 1999;255(3):707–10.
 7. Raghavan A, Sheiko T, Graham BH, Craigen WJ. Voltage-dependant anion channels: Novel insights into isoform function through genetic models. *Biochimica et Biophysica Acta (BBA) Biomembranes*. 2012;1818(6):1477–85.
 8. Zinghirino F, Pappalardo XG, Messina A, Guarino F, De Pinto V. Is the secret of VDAC Isoforms in their gene regulation? Characterization of human VDAC genes expression profile, promoter activity, and transcriptional regulators. *Int J Mol Sci*. 2020;21(19):E7388.
 9. Zinghirino F, Pappalardo XG, Messina A, Nicosia G, De Pinto V, Guarino F. VDAC genes expression and regulation in mammals. *Front Physiol*. 2021;12: 708695.
 10. Cuadrado-Tejedor M, Vilariño M, Cabodevilla F, Del Río J, Frechilla D, Pérez-Mediavilla A. Enhanced expression of the voltage-dependent anion channel 1 (VDAC1) in Alzheimer's disease transgenic mice: an insight into the pathogenic effects of amyloid- β . *J Alzheimers Dis*. 2011;23(2):195–206.
 11. Liao Z, Liu D, Tang L, Yin D, Yin S, Lai S, et al. Long-term oral resveratrol intake provides nutritional preconditioning against myocardial ischemia/reperfusion injury: involvement of VDAC1 downregulation. *Mol Nutr Food Res*. 2015;59(3):454–64.
 12. Manczak M, Reddy PH. Abnormal interaction of VDAC1 with amyloid beta and phosphorylated tau causes mitochondrial dysfunction in Alzheimer's disease. *Hum Mol Genet*. 2012;21(23):5131–46.
 13. Mazure NM. VDAC in cancer. *Biochim Biophys Acta Bioenerg*. 2017;1858(8):665–73.
 14. Risiglione P, Zinghirino F, Di Rosa MC, Magri A, Messina A. Alpha-synuclein and mitochondrial dysfunction in Parkinson's disease: the emerging role of VDAC. *Biomolecules*. 2021;11(5):718.
 15. Shoshan-Barmatz V, De Pinto V, Zweckstetter M, Raviv Z, Keinan N, Arbel N. VDAC, a multi-functional mitochondrial protein regulating cell life and death. *Mol Aspects Med*. 2010;31(3):227–85.
 16. Magri A, Reina S, De Pinto V. VDAC1 as pharmacological target in cancer and neurodegeneration: focus on its role in apoptosis. *Front Chem*. 2018;6(6):108.
 17. Stasiak M, Kolenda T, Kozłowska-Masłoń J, Sobocińska J, Poter P, Guglas K, et al. The world of pseudogenes: new diagnostic and therapeutic targets in cancers or still mystery molecules? *Life*. 2021;11(12):1354.
 18. Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Carter DRF. Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA*. 2011;17(5):792–8.
 19. Chiefari E, Iiritano S, Paonessa F, Le Pera I, Arcidiacono B, Filocamo M, et al. Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. *Nat Commun*. 2010;27(1):40.
 20. Suo G, Han J, Wang X, Zhang J, Zhao Y, Zhao Y, et al. Oct4 pseudogenes are transcribed in cancers. *Biochem Biophys Res Commun*. 2005;337(4):1047–51.
 21. Ma Y, Chen Z, Yu J. Pseudogenes and their potential functions in hematopoiesis. *Exp Hematol*. 2021;103:24–9.
 22. Tutar Y. Pseudogenes. *Comp Funct Genomics*. 2012;2012: 424526.
 23. D'Errico I, Gadaleta G, Saccone C. Pseudogenes in metazoa: origin and features. *Brief Funct Genomic Proteomic*. 2004;3(2):157–67.
 24. Maestre J, Tchénio T, Dhellin O, Heidmann T. mRNA retroposition in human cells: processed pseudogene formation. *EMBO J*. 1995;14(24):6333–8.
 25. Ido Y, Yamamoto T, Yoshitomi T, Yamamoto A, Obana E, Ohkura K, et al. Pseudogenes of rat VDAC1: 16 gene segments in the rat genome show structural similarities with the cDNA encoding rat VDAC1, with 8 slightly expressed in certain tissues. *Mamm Genome*. 2012;23(3–4):286–93.
 26. Ido Y, Yoshitomi T, Ohkura K, Yamamoto T, Shinohara Y. Utility of syntenic relationships of VDAC1 pseudogenes for not only an understanding of the phylogenetic divergence history of rodents, but also ascertaining possible pseudogene candidates as genuine pseudogenes. *Genomics*. 2014;104(2):128–33.
 27. Kalyana-Sundaram S, Kumar-Sinha C, Shankar S, Robinson DR, Wu Y-M, Cao X, et al. Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell*. 2012;149(7):1622–34.
 28. Chopra M, Bohlander SK. The cell of origin and the leukemia stem cell in acute myeloid leukemia. *Genes Chromosomes Cancer*. 2019;58(12):850–8.
 29. Izadirad M, Jafari L, James AR, Unfried JP, Wu Z-X, Chen Z-S. Long non-coding RNAs have pivotal roles in chemoresistance of acute myeloid leukemia. *Drug Discov Today*. 2021;26(7):1735–43.
 30. Liu Z, Spiegelman VS, Wang H-G. Distinct noncoding RNAs and RNA binding proteins associated with high-risk pediatric and adult acute myeloid leukemias detected by regulatory network analysis. *Cancer Rep*. 2021;4: e1592.
 31. Bhatnagar B, Garzon R. Clinical applications of MicroRNAs in acute myeloid leukemia: a mini-review. *Front Oncol*. 2021;11: 679022.
 32. Kirtonia A, Ashrafizadeh M, Zarrabi A, Hushmandi K, Zabolian A, Bejandi AK, et al. Long noncoding RNAs: a novel insight in the leukemogenesis and drug resistance in acute myeloid leukemia. *J Cell Physiol*. 2022;237(1):450–65.
 33. Davis AP, Benninghoff AD, Thomas AJ, Sessions BR, White KL. DNA methylation of the LIN28 pseudogene family. *BMC Genomics*. 2015;11(16):287.
 34. Adamaki M, Lambrou GI, Athanasiadou A, Tzanoudaki M, Vlahopoulos S, Moschovi M. Implication of IRF4 aberrant gene expression in the acute leukemias of childhood. *PLoS ONE*. 2013;8(8): e72326.
 35. Ahmadi SE, Rahimi S, Zarandi B, Chegeni R, Safa M. MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. *J Hematol Oncol*. 2021;14(1):121.
 36. Chen DW-C, Saha V, Liu J-Z, Schwartz J-M, Krstic-Demonacos M. Erg and AP-1 as determinants of glucocorticoid response in acute lymphoblastic leukemia. *Oncogene*. 2013;32(25):3039–48.
 37. Churpek JE, Bresnick EH. Transcription factor mutations as a cause of familial myeloid neoplasms. *J Clin Invest*. 2019;129(2):476–88.
 38. Fang C, Rao S, Crispino JD, Ntziachristos P. Determinants and role of chromatin organization in acute leukemia. *Leukemia*. 2020;34(10):2561–75.
 39. Hirabayashi S, Ohki K, Nakabayashi K, Ichikawa H, Momozawa Y, Okamura K, et al. ZNF384-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype. *Haematologica*. 2017;102(1):118–29.
 40. Kornblau SM, Qiu YH, Zhang N, Singh N, Faderl S, Ferrajoli A, et al. Abnormal expression of FLI1 protein is an adverse prognostic factor in acute myeloid leukemia. *Blood*. 2011;118(20):5604–12.
 41. Li Y, Liao Z, Luo H, Benyoucef A, Kang Y, Lai Q, et al. Alteration of CTCF-associated chromatin neighborhood inhibits TAL1-driven oncogenic transcription program and leukemogenesis. *Nucleic Acids Res*. 2020;48(6):3119–33.
 42. Qiu Y, Huang S. CTCF-mediated genome organization and leukemogenesis. *Leukemia*. 2020;34(9):2295–304.
 43. Sanda T, Leong WZ. TAL1 as a master oncogenic transcription factor in T-cell acute lymphoblastic leukemia. *Exp Hematol*. 2017;53:7–15.
 44. Shrivastava T, Mino K, Babayeva ND, Baranovskaya O, Rizzino A, Tahirov TH. Structural basis of Ets1 activation by Runx1. *Leukemia*. 2014;28(10):2040–8.
 45. Takao S, Forbes L, Uni M, Cheng S, Pineda JMB, Tarumoto Y, et al. Convergent organization of aberrant MYB complex controls oncogenic gene expression in acute myeloid leukemia. *Elife*. 2021;2(10): e65905.
 46. Takei H, Kobayashi SS. Targeting transcription factors in acute myeloid leukemia. *Int J Hematol*. 2019;109(1):28–34.
 47. Yao H, Goldman DC, Fan G, Mandel G, Fleming WH. The corepressor Rcor1 is essential for normal myeloerythroid lineage differentiation. *Stem Cells*. 2015;33(11):3304–14.
 48. Zhang Y, Xiao L. Identification and validation of a prognostic 8-gene signature for acute myeloid leukemia. *Leuk Lymphoma*. 2020;61(8):1981–8.
 49. Zhou Z, Li X, Liu Z, Huang L, Yao Y, Li L, et al. A Bromodomain-containing protein 4 (BRD4) inhibitor suppresses angiogenesis by regulating AP-1 expression. *Front Pharmacol*. 2020;10(11):1043.
 50. Nordlund J, Syvänen A-C. Epigenetics in pediatric acute lymphoblastic leukemia. *Semin Cancer Biol*. 2018;51:129–38.
 51. Roe J-S, Vakoc CR. The essential transcriptional function of BRD4 in acute myeloid leukemia cells. *Cold Spring Harb Symp Quant Biol*. 2016;81:61–6.

52. Yu Q, Xu Y, Zhuang H, Wu Z, Zhang L, Li J, et al. Aberrant activation of RPB1 is critical for cell overgrowth in acute myeloid leukemia. *Exp Cell Res*. 2019;384(2): 111653.
53. Klymenko T, Bloehdorn J, Bahlo J, Robrecht S, Akylzhanova G, Cox K, et al. Lamin B1 regulates somatic mutations and progression of B-cell malignancies. *Leukemia*. 2018;32(2):364–75.
54. Zhou C, Martinez E, Di Marcantonio D, Solanki-Patel N, Aghayev T, Peri S, et al. JUN is a key transcriptional regulator of the unfolded protein response in acute myeloid leukemia. *Leukemia*. 2017;31(5):1196–205.
55. Astolfi A, Milano F, Palazzotti D, Brea J, Pismataro MC, Morlando M, et al. From serendipity to rational identification of the 5,6,7,8-Tetrahydrobenzo[4,5]thieno[2,3-d]pyrimidin-4(3H)-one Core as a New chemotype of AKT1 inhibitors for acute myeloid leukemia. *Pharmaceutics*. 2022. <https://doi.org/10.3390/pharmaceutics14112295>.
56. Koya J, Kataoka K, Sato T, Bando M, Kato Y, Tsuruta-Kishino T, et al. DNMT3A R882 mutants interact with polycomb proteins to block haematopoietic stem and leukaemic cell differentiation. *Nat Commun*. 2016;24(7):10924.
57. Zou Q, Tan S, Yang Z, Wang J, Xian J, Zhang S, et al. The human nucleophosmin 1 mutation A inhibits myeloid differentiation of leukemia cells by modulating miR-10b. *Oncotarget*. 2016;7(44):71477–90.
58. Reina S, Nibali SC, Tomasello MF, Magri A, Messina A, De Pinto V. Voltage dependent anion channel 3 (VDAC3) protects mitochondria from oxidative stress. *Redox Biol*. 2022;51: 102264.
59. Magri A, Cubisino SAM, Battiato G, Lipari CLR, Conti Nibali S, Saab MW, et al. VDAC1 knockout affects mitochondrial oxygen consumption triggering a rearrangement of ETC by impacting on complex I activity. *Int J Mol Sci*. 2023;24(4):3687.
60. Rahmani Z, Maunoury C, Siddiqui A. Isolation of a novel human voltage-dependent anion channel gene. *Eur J Hum Genet*. 1998;6(4):337–40.
61. Magri A, Messina A. Interactions of VDAC with proteins involved in neurodegenerative aggregation: an opportunity for advancement on therapeutic molecules. *Curr Med Chem*. 2017;24(40):4470–87.
62. Shoshan-Barmatz V, Keinan N, Zaid H. Uncovering the role of VDAC in the regulation of cell life and death. *J Bioenerg Biomembr*. 2008;40(3):183–91.
63. Hinsch K-D, De Pinto V, Aires VA, Schneider X, Messina A, Hinsch E. Voltage-dependent anion-selective channels VDAC2 and VDAC3 are abundant proteins in bovine outer dense fibers, a cytoskeletal component of the sperm flagellum. *J Biol Chem*. 2004;279(15):15281–8.
64. Shimada K, Park S, Miyata H, Yu Z, Morohoshi A, Oura S, et al. ARMC12 regulates spatiotemporal mitochondrial dynamics during spermiogenesis and is required for male fertility. *Proc Natl Acad Sci USA*. 2021;118(6): e2018355118.
65. Kovalenko TF, Morozova KV, Ozolins LA, Lapina IA, Patrushev LI. The PTENP1 pseudogene, unlike the PTEN Gene, is methylated in normal endometrium, as well as in endometrial hyperplasias and carcinomas in middle-aged and elderly females. *Acta Naturae*. 2018;10(1):43–50.
66. Kovalenko TF, Morozova KV, Pavlyukov MS, Anufrieva KS, Bobrov MYu, Gamisoniya AM, et al. Methylation of the PTENP1 pseudogene as potential epigenetic marker of age-related changes in human endometrium. *PLoS ONE*. 2021;16(1):e0243093.
67. Liao M, Liao W, Xu N, Li B, Liu F, Zhang S, et al. LncRNA EPB41L4A-AS1 regulates glycolysis and glutaminolysis by mediating nucleolar translocation of HDAC2. *EBioMedicine*. 2019;41:200–13.
68. Zhu Y, Liu Q, Liao M, Diao L, Wu T, Liao W, et al. Overexpression of lncRNA EPB41L4A-AS1 induces metabolic reprogramming in trophoblast cells and placenta tissue of miscarriage. *Mol Ther Nucleic Acids*. 2019;6(18):518–32.
69. Liu Y, Sun P, Zhao Y, Liu B. The role of long non-coding RNAs and downstream signaling pathways in leukemia progression. *Hematol Oncol*. 2021;39(1):27–40.
70. Chen Y-L, Zhang Z-X, Shou L-H, Di J-Y. Regulation of DNA methylation and tumor suppression gene expression by miR-29b in leukemia patients and related mechanisms. *Eur Rev Med Pharmacol Sci*. 2018;22(1):158–65.
71. Zhang T-J, Zhang L-C, Xu Z-J, Zhou J-D. Expression and prognosis analysis of DNMT family in acute myeloid leukemia. *Aging*. 2020;12(14):14677–90.
72. Roshan R, Shridhar S, Sarangdhar MA, Banik A, Chawla M, Garg M, et al. Brain-specific knockdown of miR-29 results in neuronal cell death and ataxia in mice. *RNA*. 2014;20(8):1287–97.
73. Lu T-P, Lee C-Y, Tsai M-H, Chiu Y-C, Hsiao CK, Lai L-C, et al. miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets. *PLoS ONE*. 2012;7(8): e42390.
74. Tan Y-T, Lin J-F, Li T, Li J-J, Xu R-H, Ju H-Q. LncRNA-mediated posttranslational modifications and reprogramming of energy metabolism in cancer. *Cancer Commun*. 2021;41(2):109–20.
75. Li Y, Li L, Wang Z, Pan T, Sahni N, Jin X, et al. LncMAP: Pan-cancer atlas of long noncoding RNA-mediated transcriptional network perturbations. *Nucleic Acids Res*. 2018;46(3):1113–23.
76. Liu J, Liu Z-X, Wu Q-N, Lu Y-X, Wong C-W, Miao L, et al. Long noncoding RNA AGPG regulates PFKFB3-mediated tumor glycolytic reprogramming. *Nat Commun*. 2020;11(1):1507.
77. Guarino F, Zinghirino F, Mela L, Pappalardo XG, Ichas F, De Pinto V, et al. NRF-1 and HIF-1 α contribute to modulation of human VDAC1 gene promoter during starvation and hypoxia in HeLa cells. *Biochim Biophys Acta Bioenerg*. 2020;1861(12): 148289.
78. Khelifi A, Adel K, Duret L, Laurent D, Mouchiroud D, Dominique M. HOPPSIGEN: a database of human and mouse processed pseudogenes. *Nucleic Acids Res*. 2005. <https://doi.org/10.1093/nar/gkl084>.
79. Karro JE, Yan Y, Zheng D, Zhang Z, Carriero N, Cayting P, et al. Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res*. 2007. <https://doi.org/10.1093/nar/gkl851>.
80. Zheng L-L, Zhou K-R, Liu S, Zhang D-Y, Wang Z-L, Chen Z-R, et al. dreamBase: DNA modification, RNA regulation and protein binding of expressed pseudogenes in human health and disease. *Nucleic Acids Res*. 2018;46(D1):D85–91.
81. Johnston TS, Li S, Franz E, Huang Z, Dan Li S, Campbell MJ, et al. PseudoFuN: Deriving functional potentials of pseudogenes from integrative relationships with genes and microRNAs across 32 cancers. *Gigascience*. 2019. <https://doi.org/10.1093/gigascience/giz046>.
82. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98–102.
83. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res*. 2019;47(W1):W556–60.
84. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. 2013;45(10):1113–20.
85. Papatheodorou I, Moreno P, Manning J, Fuentes AM-P, George N, Fexova S, et al. Expression Atlas update: from tissues to single cells. *Nucleic Acids Res*. 2020;48(D1):D77–83.
86. Noguchi S, Arakawa T, Fukuda S, Furuno M, Hasegawa A, Hori F, et al. FANTOM5 CAGE profiles of human and mouse samples. *Sci Data*. 2017;29(4): 170112.
87. Zhang Q, Liu W, Zhang H-M, Xie G-Y, Miao Y-R, Xia M, et al. hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. *Genomics Proteomics Bioinform*. 2020;18(2):120–8.
88. Li R, Liang F, Li M, Zou D, Sun S, Zhao Y, et al. MethBank 3.0: a database of DNA methylomes across a variety of species. *Nucleic Acids Res*. 2018;46(D1):D288–95.
89. De Meyer T, Mampaey E, Vlemmix M, Denil S, Trooskens G, Renard JP, et al. Quality evaluation of Methyl Binding Domain based kits for enrichment DNA-methylation sequencing. *PLoS ONE*. 2013;8(3): e59068.
90. Chemi F, Pearce SP, Clipson A, et al. cfDNA methylome profiling for detection and subtyping of small cell lung cancers. *Nat Cancer*. 2022;3:1260–70.
91. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
92. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021. <https://doi.org/10.1093/gigascience/giab008>.
93. Lienhard M, Grasse S, Rolff J, Frese S, Schirmer U, Becker M, et al. QSEA—modelling of genome-wide DNA methylation from sequencing enrichment experiments. *Nucleic Acid Res*. 2017;45(6): e44.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.