RESEARCH ARTICLE

Open Access

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

Xena Giada Pappalardo[1], Pierpaolo Risiglione[1], Federica Zinghirino[1], Angela Ostuni[2], Daniela Luciano[1], Faustino Bisaccia[2], 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,

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 2 of 22

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

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 3 of 22

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 *PCDH11X* gene)	
VDAC1P4	7418	chr1:180403931–180405937	2007	+	weak and strong enhancers, weak transcriptional elongation, weak promoter (overlapping intronic region of *ACBD6* gene)	
VDAC1P5	10187	chr12:55195669–55197407	1739	−	Heterochromatin	
VDAC1P6	359800	chrY:5074406–5076119	1714	−	Heterochromatin (overlapping intronic region of *PCDH11Y* gene)	
VDAC1P7	100310839	chr3:77365047–77366767	1721	−	Insulator	Heterochromatin (overlapping intronic region of *ROBO2* gene)
VDAC1P8[a]	100310840	chr6:143811645–143813335	1691	+	Weak enhancer, weak transcriptional elongation (proximity to *PEX3* gene)	Weak and strong enhancer, weak transcriptional elongation (proximity to *FUCA2* 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 *ZNF169* gene)	
VDAC1P12	100874289	chr13:34656914–34657447	534	−	Heterochromatin	
VDAC1P13	100420568	chr2:42690279–42691137	859	−	Weak enhancers, weak transcriptional elongation (overlapping intronic region of *KCNG3* gene)	
(b)						
VDAC2P1	54015	chr21:17466509–17467792	1284	−	Heterochromatin (overlapping *LINC00478* 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, Polyocomb repressed	Weak enhancer, heterochromatin, Polyocomb 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*. [a]*Genomic 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 = isolator; dark green = transcriptional transition/elongation (Txn), light green = weak transcript

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 4 of 22

Fig. 1 (See legend on previous page.)

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 5 of 22

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

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 6 of 22

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 overexpressed 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, ENSG000278206,* 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.

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 7 of 22

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%5Bgroup%5D)show 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 Ensemble'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

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 8 of 22

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	*MYC*	Myc/Max factors	✓	X	[35, 38]
	Tal-related	TAL-HEN	*TAL1*	Tal-related factors	X	✓	[38, 43]
Basic leucine zipper factors bZIP	Jun-related	Jun	*JUND*	JunD proto-oncogene			[54]
	B-ATF-related		*BATF*	B-ATF-related			[48]
	Fos-related	Fos	*FOS*	Fos proto-oncogene	✓	X	[45]
	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			[37, 45]
			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	[41, 42]
		ZNF362-like	*ZNF384*	Zinc Finger Protein 384	X	✓	[39]
Tryptophan cluster factors	Myb-SANT	NCoR-like	*RCOR1*	REST corepressor 1	✓	X	[47]
	ETS-related	SPI-like	*SPI1*	Spi-1 proto-oncogene	✓		[45, 46]
		ETS-like	*ETS1*	Ets-protoncogene 1	X	✓	[44]
			ERG	Ets-related factors			[36, 38]
			FLI1	Fli-1 protoncogene			[40]
			IRF4	Interferon-regulatory factors			[34]
Runt domain factors	Runt-related		*RUNX1*	Runt domain factors	✓		[37, 44, 45]

	Type		Name		Tissue		Role in leukemogenesis Refs
					Bone marrow	Blood	
Other TFs	Epigenetic reader		*BRD4*	Bromodomain containing 4	✓		[38, 51]
	Heterochomatin protein 1 HP1 family		*CBX3*	Chromobox 3	✓	X	[50]
	Synthesizing messenger RNA		*POLR2A*	RNA polymerase II subunit A	✓		[52]
	CTCF-transcriptional coactivators		*RAD21*	RAD21 cohesin complex component	✓		[38]
			STAG1	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 housekeeping gene.

Putative TFBS suitable for driving *VDAC1P8* expression in leukemogenesis

The results of the expression analyses of *VDAC* pseudogenes showed that the pseudogene*s 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
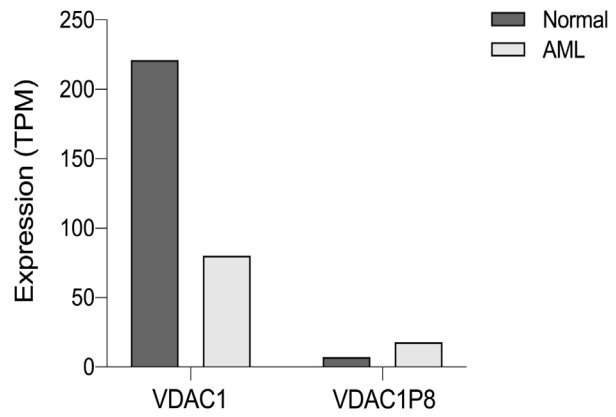
a

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
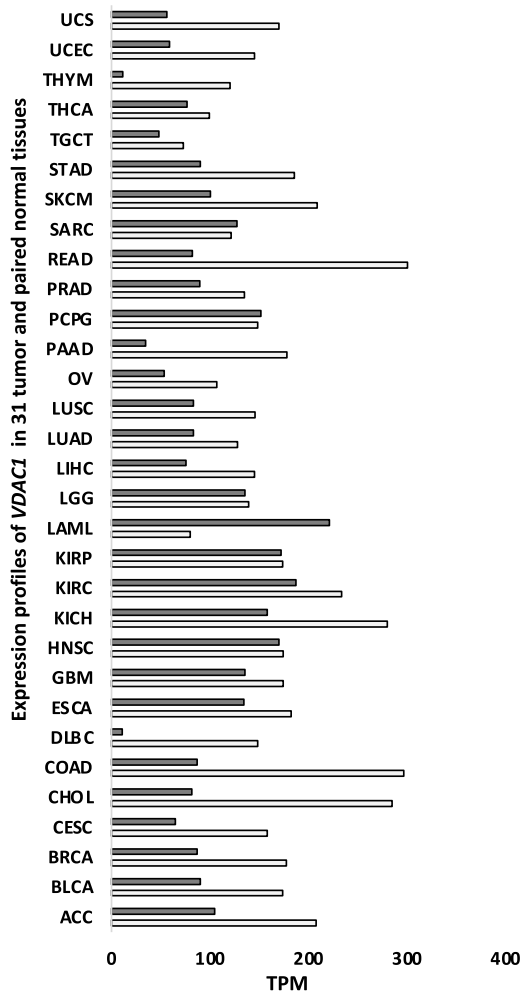
d
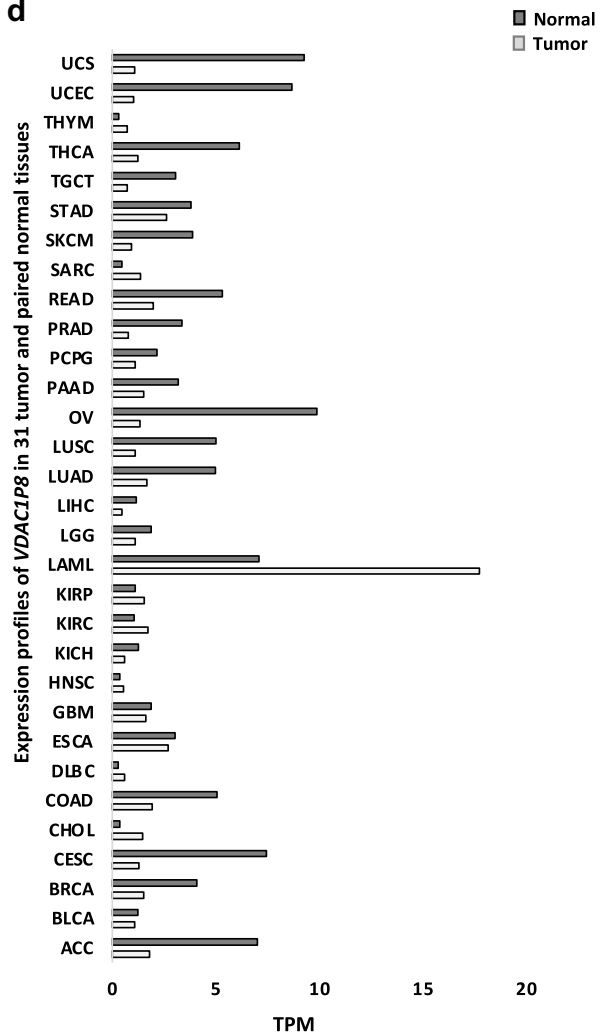
Fig. 2 (See legend on previous page.)
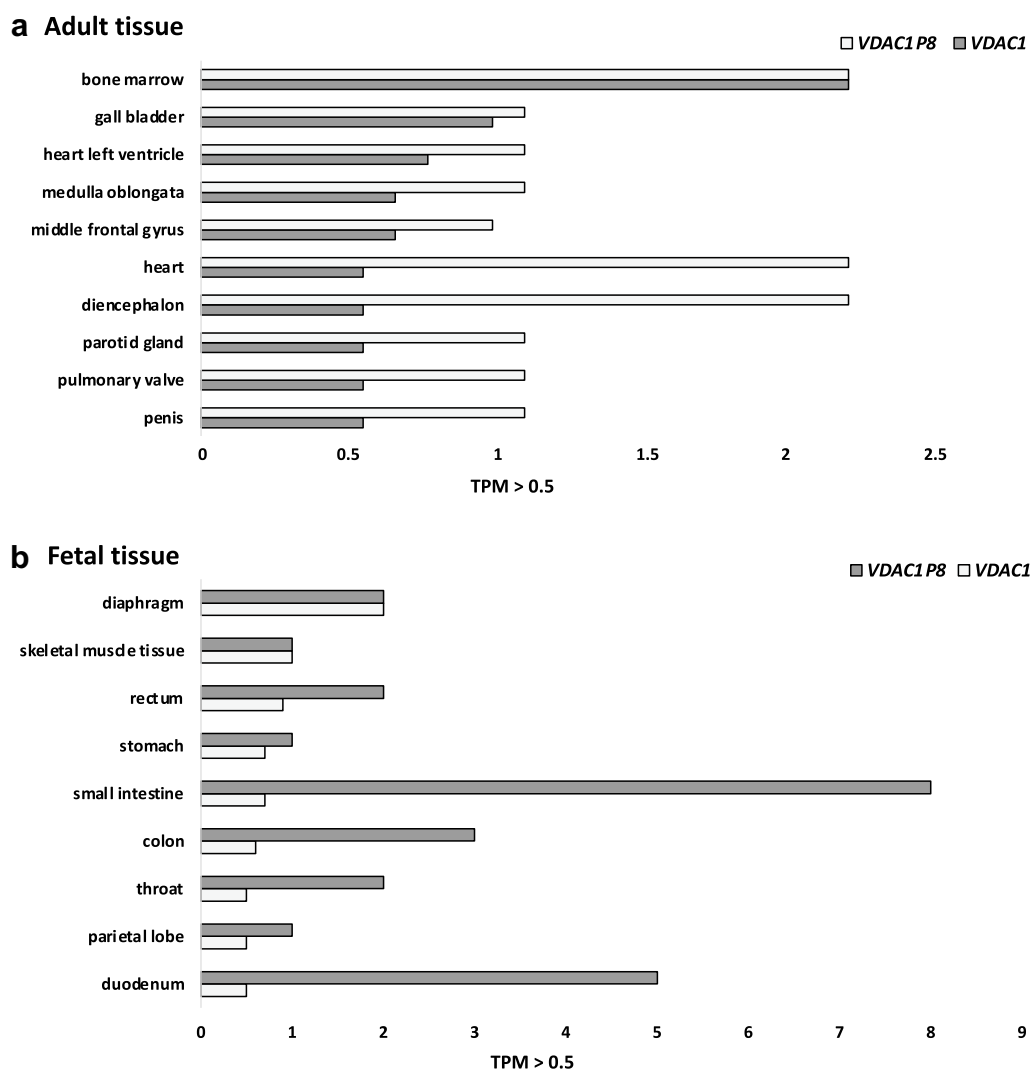
Fig. 3 a-b. Adult (**a**) and fetal (**b**) transcripts tissue expression levels of *VDAC1* and *VDAC1P8* from RNA-seq CAGE in RIKEN FANTOM 5 project. Data cut-off is > 0.05 TPM. Data shown consist of RNA-Seq CAGE (Cap Analysis of Gene Expression) analysis from RIKEN FANTOM5 project (www.ebi.ac.uk/gxa/home). 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
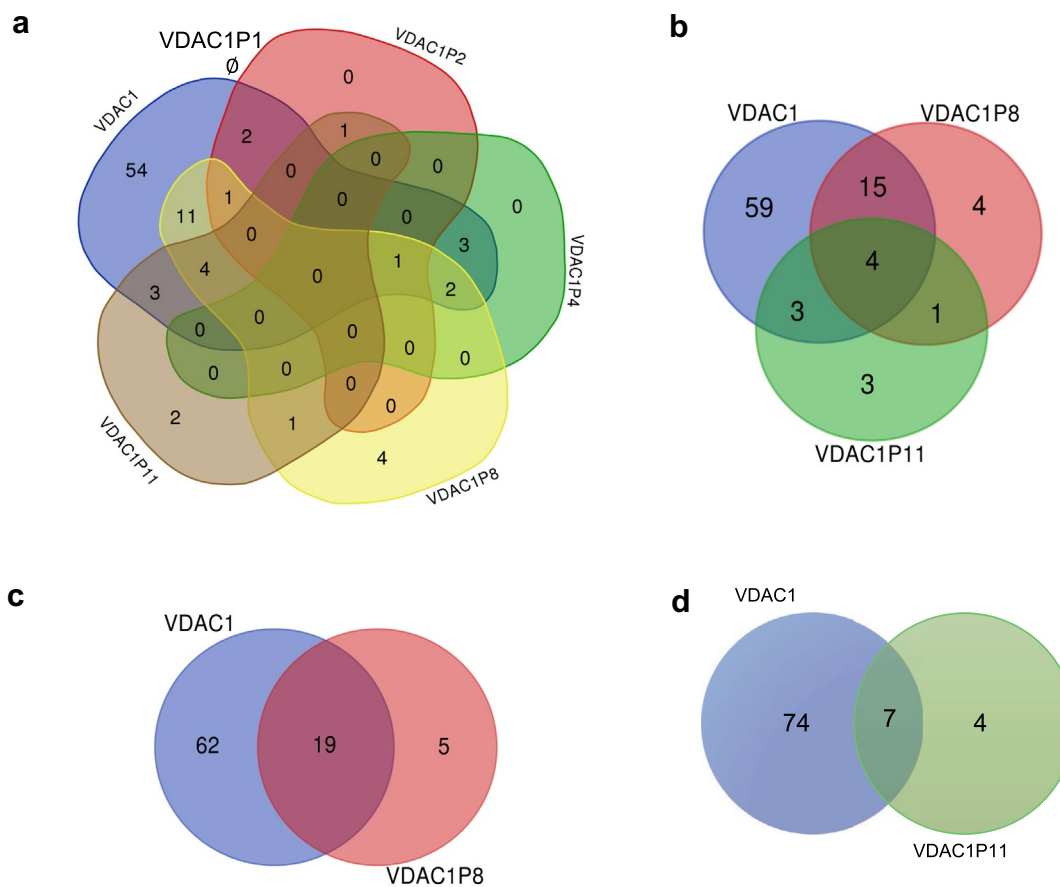
Pappalardo *et al. Biological Research*   (2023) 56:33

Page 11 of 22

a   **Average methylation of *VDAC1* and *VDAC1P8* promoter**

	Adipose	Adrenal gland	Bladder	Breast	Central nervous system	Circulatory system	Digestive system	Lung	Lymphatic system	Placenta	Prostate	Psoas
VDAC1	0.19	0.23	0.15	0.16	0.16	0.16	0.17	0.16	0.19	0.22	0.19	0.15
VDAC1P8	0.91	0.97	0.96	0.99	0.95	0.94	0.91	0.91	0.95	1	0.97	0.94

b   **Average methylation of *VDAC1* and *VDAC1P8* gene body**

	Adipose	Adrenal gland	Bladder	Breast	Central nervous system	Circulatory system	Digestive system	Lung	Lymphatic system	Placenta	Prostate	Psoas
VDAC1	0.65	0.64	0.62	0.60	0.58	0.61	0.61	0.68	0.64	0.63	0.63	0.54
VDAC1P8	0.93	0.93	0.9	0.9	0.94	0.92	0.92	0.93	0.92	0.9	0.89	0.93

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* pseudogenes. 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.htpl

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 12 of 22

Intersection among *VDAC1, VDAC1P8* and *VDAC1P11 (b)*		
Genes	**Total elements**	**TFs**
VDAC1 VDAC1P11 VDAC1P8	4	***JUND CBX3 RCOR1 GATA1***
VDAC1 VDAC1P8	15	***IRF4 ETS1 GATA2 STAG1 FLI1 CTCF BRD4 RUNX1 SPI1 RAD21 ERG CEBPB MYC TAL1 POLR2A***
VDAC1 VDAC1P11	3	*JUN TEAD4 MAX*
VDAC1P11 VDAC1P8	1	***FOS***
VDAC1	59	*SIN3A YY1 TCF12 RUNX3 ZNF263 SPIB PAX5 NOTCH1 TCF3 CDK7 GLIS1 KDM1A KLF1 UBTF RELA KDM5B E2F6 TP53 MED1 SRF MAZ STAT1 BHLHE40 MYH11 EP300 GABPA CTCFL PBX3 POU2F1 BCL6 TRIM28 HDAC1 TBL1XR1 EGR1 HCFC1 CBFB KMT2A ELF1 BRD3 SP1 NFIC IRF1 PBX1 USF1 ESRRA TFAP4 MXI1 TAF1 CREBBP EBF1 RBBP5 E2F1 CDK9 PHF8 E2F4 ZBTB7A HDAC2 BCL11A MYB*
VDAC1P8	4	***CEBPD BATF ZNF384 GATA3***
VDAC1P11	3	*JUNB FOSL1 NR2F2*
In bold are highlighted the main AML-associated TFs reported in Table 2.		

Fig. 5 (See legend on previous page.)

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 13 of 22

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 Methyl-Cap-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
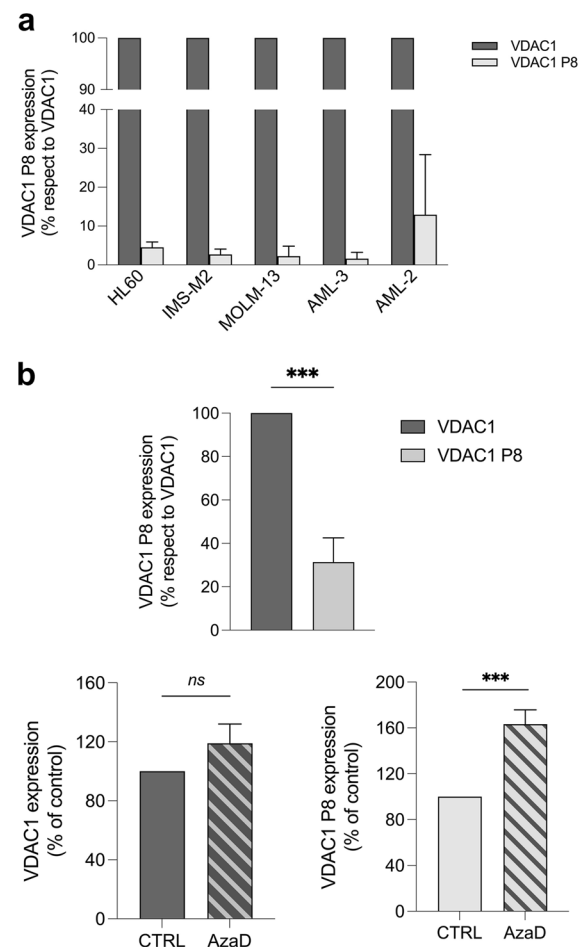
Fig. 6 a-b. Expression of *VDAC1* gene and *VDAC1P8* pseudogene in HL60 AML cell line. In (**a**), the expression of *VDAC1* gene and *VDAC1P8* pseudogene was evaluated in different models of leukemia cell line (HL60, IMS-M2, MOLM-13, AML-3, AML-2) by real-time PCR relative quantification. In (**b**), 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 ΔΔCt 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

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 14 of 22

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 were 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 fertily [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.
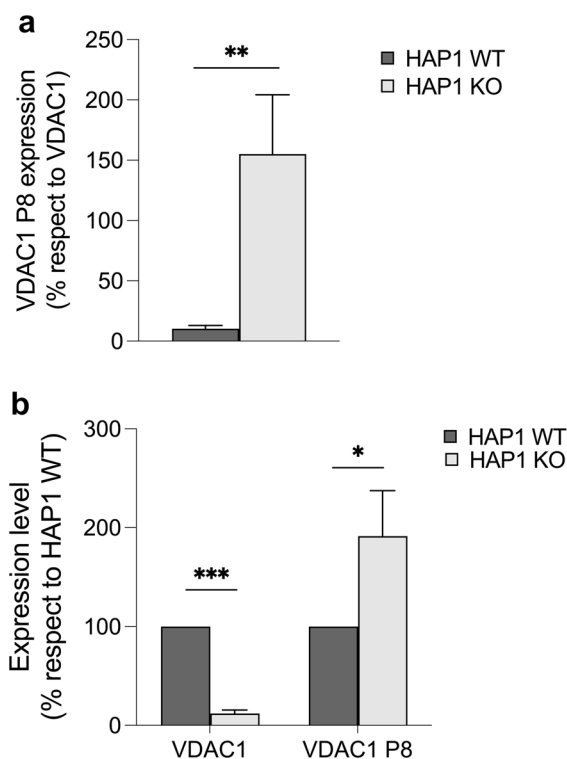
The same expression ratio of *VDAC1* and *VDAC1P8* transcripts is also found in the nearly haploid HAP1 wild-type human leukemic line. In contrast, in the HAP1-*VDAC1* knock-out line, the level of *VDAC1P8* increases significantly, suggesting that the expression of the *VDAC1* gene and its pseudogene *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 promotors 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)

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 15 of 22
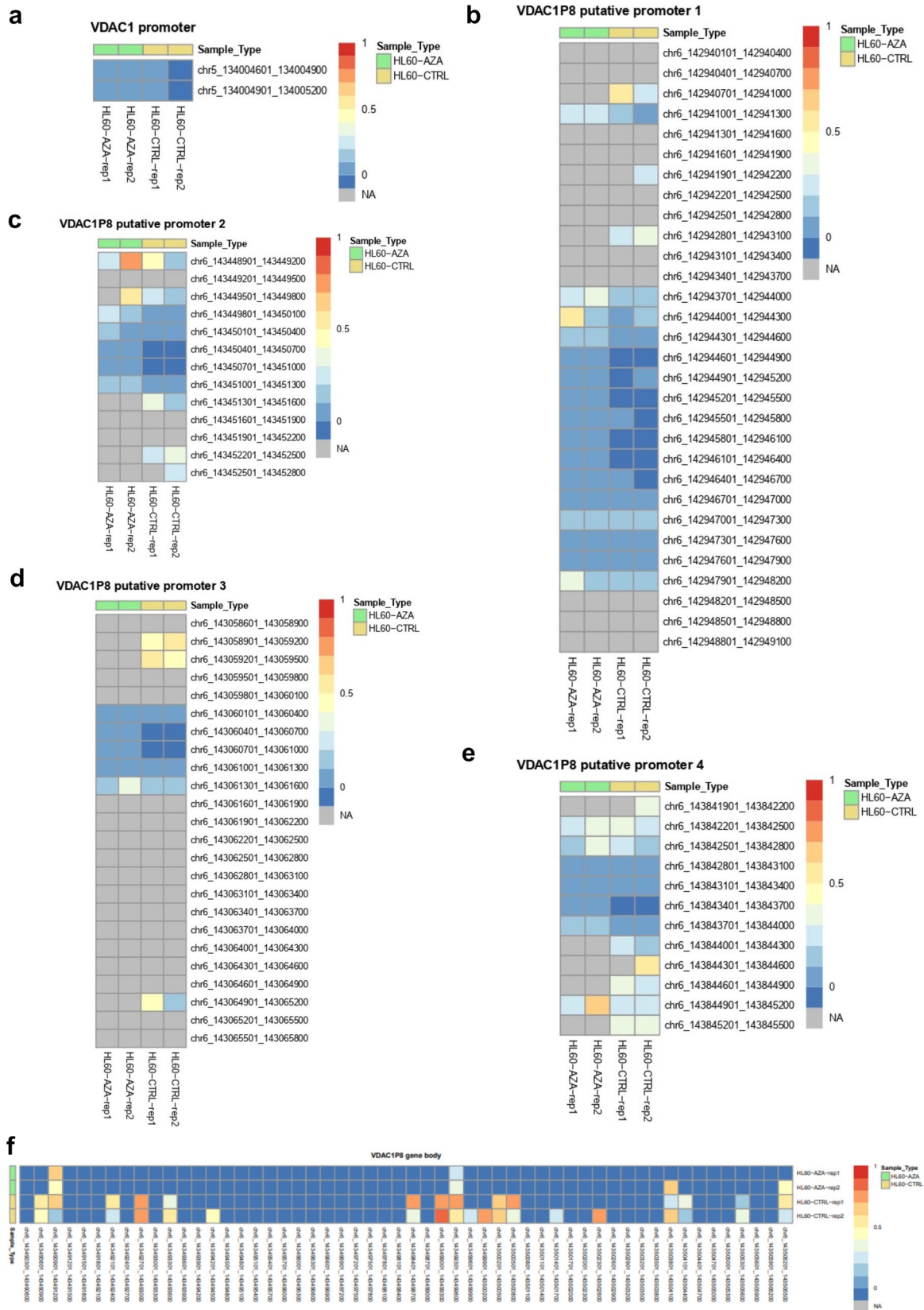
Fig. 7 (See legend on previous page.)

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 ΔΔCt 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 undermethylated 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, *EPB41L4AAS1* 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 downregulated, 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 *VDAC1P1, VDAC1P2* 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 miRNAtarget 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 *VDAC1P8* 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, *VDAC1P8* 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 *VDAC1P8* (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 *VDAC1P8* 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

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 18 of 22

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.

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 lobus, hippocampus, nucleus accumbens, prefrontal cortex (Broadman 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 Methyl-Cap 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 μg protein and 3 μl 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×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), 2×150 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 milions 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% CO_2 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 cm^2 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 ΔΔCt 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.

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.

References
1. Lemasters JJ, Holmuhamedov E. Voltage-dependent anion channel (VDAC) as mitochondrial governator–thinking outside the box. Biochim Biophys Acta. 2006;1762(2):181–90.
2. De Pinto V. Renaissance of VDAC: new insights on a protein family at the interface between mitochondria and cytosol. Biomolecules. 2021;11(1):107.
3. Reina S, Magrì 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.
4. 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, Magrì 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. Magrì 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

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 noncoding 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.

Pappalardo *et al. Biological Research*   (2023) 56:33

Page 22 of 22

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, Magrì A, Messina A, De Pinto V. Voltage dependent anion channel 3 (VDAC3) protects mitochondria from oxidative stress. Redox Biol. 2022;51: 102264.

59. Magrì 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, Ozolinya 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-1α 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/gki084.

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. Johnson 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