Spataro and Buetti-Dinh
Translational Medicine Communications (2024) 9:4
https://doi.org/10.1186/s41231-024-00163-1

Translational Medicine Communications

RESEARCH

Open Access

Co-expression analysis of transcriptomic data from cancer and healthy specimens reveals rewiring of proteasome genes and an interaction with the XPO1 gene across several tumour types

Vito Spataro[1,2*] and Antoine Buetti-Dinh[3,4]

Abstract

Background The 26S proteasome is a large intracellular multiprotein complex, that plays a homeostatic role by degrading proteins that have been tagged by ubiquitin. It is composed of 64 subunits assembled according to a well-defined structure and stoichiometry. Several proteasome subunits have been found to be overexpressed in tumours. However, comprehensive data are lacking on the relative abundance of each subunit and the impact on proteasome composition or stoichiometry. In cancer treatment, proteasome inhibitors and inhibitors of XPO1 (Exportin-1) have unexpectedly a similar range of activity, but the interaction between the two pathways has not been studied.

Methods We performed gene co-expression analysis of 38 genes encoding proteasome subunits and 38 genes encoding proteins involved in nucleocytoplasmic transport in specimens from the Cancer Genome Atlas (33 tumour types) and from the Gene Tissue Expression database (32 healthy tissue types). We obtained 65 matrices, each containing Pearson correlation factors for 2964 gene pairs. We applied cluster analysis to the correlation matrices and compared the distribution of Pearson correlation coefficients of thirteen tumour types with their healthy tissue counterpart.

Results Strong positive correlation (R Pearson correlation > 0.8) was observed for pairs of proteasome genes in the majority of healthy tissues, whereas the correlation for co-expression was significantly lower ($R \leq 0.50$) for most gene pairs in the majority of cancer types. Cluster analysis based on gene co-expression allowed to distinguish cancers from healthy tissues in a clear-cut manner, and to identify the genes that contributed most to the separation. The crossed analysis between proteasome and nucleocytoplasmic transport genes showed that the expression of *XPO1* and a subset of proteasome genes, including in particular *PSMD14*, is correlated in several cancer types and not in their healthy counterpart.

Conclusions This analysis reveals that in cancer the co-expression of proteasome genes is significantly altered, highlighting the genes that are more often deregulated. In addition, it finds that *XPO1* expression is often correlated with the expression of proteasome genes. From a therapeutic perspective, these findings support the investigation

*Correspondence:
Vito Spataro
vito.spataro@hin.ch
Full list of author information is available at the end of the article

of novel targets within the proteasome and strategies of co-targeting of the proteasome and nucleocytoplasmic transport.

Keywords 26S proteasome, Nucleocytoplasmic transport, Gene co-expression, The cancer genome atlas, Gene tissue expression database, XPO1, PSMD14

Introduction

The proteasome is one of the largest multiprotein complexes in the eukaryotic cell; it is made of 64 subunits and has a well-defined structure and composition [1]. It has a crucial role in cellular homeostasis by degrading protein that have been tagged by ubiquitin in the ubiquitin-proteasome system (UPS) [2]. The assembly of large multiprotein complexes requires the coordinated transcription of several genes. The transcription factor NRF1 has been shown to induce de novo proteasome synthesis by promoting the concerted expression of genes encoding proteasome subunits (reviewed in [3]). Other transcription factors can regulate the expression of a set of subunits or individual subunits in specific contexts [4, 5].

In cancer cells the UPS is often deregulated and some select proteasome subunits have been found to be overexpressed in various tumour types [6–9]. However, comprehensive studies on all subunits and on their expression at the transcriptional level in a large set of tumour types are lacking.

The UPS is an attractive target for the development of new anticancer treatments [10] and a class of proteasome inhibitors (PI) is already registered and used in the clinic (bortezomib, carfilzomib, ixazomib). However, registered PI all target only one subunit (beta-5) of the complex and are active solely against myeloma and some B-cell malignancies and inactive in other tumour types. As a result, there is a need to explore whether other subunits might be candidate drug targets in solid tumours.

Gene co-expression analysis is an approach that helps to identify genes and proteins whose expression are significantly correlated to each other and has been used to discover new protein modules and networks in various biological contexts [11]. Differential gene co-expression analysis can be used to compare patterns of gene expression in different physiological or pathological conditions [12]. We used this methodology to investigate the abundance of every proteasome subunit relative to every other at the transcriptional level. This allowed us to compare alterations of the stoichiometry of the multiprotein complex in cancer versus healthy tissues, highlighting the subunits that are more often deregulated and thus of potential interest as therapeutic targets.

Selective inhibitors of nuclear export (SINE) are novel anticancer drugs that target Exportin-1 encoded by the *XPO1* gene. Like proteasome inhibitors, SINE drugs are more active in multiple myeloma than in other tumour types. Selinexor, the first-in-class SINE, is active in refractory multiple myeloma, similarly to what had been observed earlier with proteasome inhibitors [13]. It is registered by the FDA in combination with Dexamethasone for multi-refractory multiple myeloma and in combination with the proteasome inhibitor Bortezomib and Dexamethasone for the treatment of myeloma at first relapse [14]. Clinical trials combining the XPO1 inhibitor Selinexor and the proteasome inhibitors Bortezomib and Carfilzomib have resulted in high response rates in patients with multiple myeloma that were heavily pretreated [14, 15]. The biological basis for the similarity of the anticancer activity and the synergy of the two drug classes is unknown. Therefore, as a second aim of this study, we set out to measure by gene co-expression analysis the correlation between the expression of genes encoding proteins involved in nucleocytoplasmic transport and genes encoding proteasome subunits.

Methods

The genes encoding proteasome subunits were selected according to the HUGO classification and genes encoding XPO1-interacting proteins were selected according to the OpenCell database which provides a protein interactome describing molecular interactions of human intracellular proteins [16]. Among the 38 genes encoding proteasome subunits, 34 are known to encode subunits belonging to the 26S proteasome complex, made of the 20S proteasome core subcomplex (subunits encoded by *PSMA1, PSMA2, PSMA3, PSMA4, PSMA5, PSMA6, PSMA7, PSMB1, PSMB2, PSMB3, PSMB4, PSMB5, PSMB6, PSMB7*), and the 19S proteasome regulatory complex (subunits encoded by *PSMC1, PSMC2, PSMC3, PSMC4, PSMC5, PSMC6, PSMD1, PSMD2, PSMD3, PSMD4, PSMD5, PSMD6, PSMD7, PSMD8, PSMD9, PSMD10, PSMD11, PSMD12, PSMD13, PSMD14*). Three genes encode subunits found in immunoproteasomes (*PSMB8, PSMB9* and *PSMB10*) and one in spermatoproteasomes (*PSMA8*). Genes encoding XPO1-interacting proteins were *RANBP3, RGPD5, RGPD6, RGPD8, RANGAP, RANBP2, NUP107, SUB1, NUP98, SUMO2, RGPD3, NUP214, RCC1, RAN, SUMO1, LAMB1, UBE2I, NUP93, NUP155, NXF1, WBP11, NUP88, RSRC1, TTF2, SUPT16H, KPNB1, RAE1, SSRP1, NUP210, NUP133, SPATA5, NUP54, MAPRE1, NOP10, NUP62, SUMO3,*

SUMO4. Two genes were selected as controls (*ACTB* and *GAPDH*).

The quantitative level of mRNA expression of the above 78 genes was obtained by extracting publicly available data from the database of the biobank The Cancer Genome Atlas (TCGA) [17]. The TCGA database provides genomic data from the following 33 tumour types: adrenocortical carcinoma (ACC), bladder carcinoma (BLCA), breast cancer (BRCA), cervical squamous carcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), diffuse large-B cell lymphoma (DLBL), oesophageal squamous cancer (ESCA), glioblastoma multiforme (GBM), head and neck squamous carcinoma (HNSC), chromophobe renal cell carcinoma (KICH), clear cell renal carcinoma (KIRC), kidney papillary carcinoma (KIRP), acute myeloid leukaemia (LAM) L, low grade glioma (LGL), liver hepatocarcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous carcinoma (LUSC), mesothelioma (MESO), ovarian cancer (OVCA), pancreas adenocarcinoma (PAAD), paraganglioma and pheochromocytoma (PPCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), sarcoma (SARC), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), testicular germ cell cancer (TCGT), thyroid carcinoma (THCA), thymoma (THYM), uterus corpus endometrial carcinoma (UCEC), uterine carcinosarcoma (UCS), uveal melanoma (UVM). With regard to non-neoplastic tissue, we extracted transcriptomic data from the large database of the biobank of all body tissues that has been created by the Genotype-Tissue Expression Consortium GTEx [18, 19]. Specimens in the GTEx derive from the following tissues and organs: adipose tissue, adrenal gland, bladder, blood, blood vessel, bone marrow, brain, breast, cervix uteri, colon, oesophagus, fallopian tube, heart, kidney, liver, lung, muscle, nerve, ovary, pancreas, pituitary, prostate, salivary gland, skin, small intestine, spleen, stomach, testis, thyroid, uterus, vagina. Data extraction and gene co-expression analysis was carried out with a combination of the BASH/AWK/SED Unix utilities and the R programming language. The analysis was based on the mRNA expression levels of the set of 78 genes described above from all available 11'348 cancer specimens of the TCGA and 19'214 specimens from the GTEx (Table 1).

The correlation between the expression of each gene pair has been calculated and expressed with a Pearson correlation coefficient between gene's TPM (Transcripts per Million) value for N study samples (where N is reported in Table 1) and a matrix containing Pearson correlation coefficients of all gene pairs has been produced for each tumour type of the TCGA and each healthy tissue of the GTEX, obtaining 65 matrices, each containing Pearson correlation factors for 2964 gene pairs. The differences between the correlation matrices of gene co-expression analysis have been compared across tumors (33 matrices) and normal tissues (32 matrices). A workflow and the code used for making the correlation matrices are available at https://figshare.com/projects/ProteasomeGeneExpression/171357.

This has been optimized to enable an efficient classification of tumours as compared to healthy tissues. We have used supervised learning with the the xgboost library of Python3 on the TCGA and GTEx datasets to determine the factors with highest classification weights and to identify the genes which contribute most to the

Table 1 TCGA (left) and GTEx (right) study names and corresponding number of samples

TCGA Studies	Number of samples	GTEx Studies	Number of samples
ACC	79	ADIPOSE_TISSUE	1293
BLCA	433	ADRENAL_GLAND	274
BRCA	1256	BLADDER	21
CESC	309	BLOOD	1048
CHOL	45	BLOOD_VESSEL	1398
COAD	546	BONE_MARROW	204
DLBC	48	BRAIN	2931
ESCA	198	BREAST	482
GBM	175	CERVIX_UTERI	19
HNSC	548	COLON	822
KICH	91	ESOPHAGUS	1577
KIRC	618	FALLOPIAN_TUBE	9
KIRP	323	HEART	942
LAML	178	KIDNEY	98
LGG	532	LIVER	251
LIHC	424	LUNG	655
LUAD	601	MUSCLE	881
LUSC	555	NERVE	659
MESO	87	OVARY	195
OV	430	PANCREAS	360
PAAD	183	PITUITARY	301
PCPG	187	PROSTATE	263
PRAD	558	SALIVARY_GLAND	178
READ	177	SKIN	1940
SARC	265	SMALL_INTESTINE	193
SKCM	473	SPLEEN	255
STAD	453	STOMACH	384
TGCT	156	STUDY_NA	133
THCA	572	TESTIS	410
THYM	122	THYROID	706
UCEC	589	UTERUS	159
UCS	57	VAGINA	173
UVM	80		

separation. The corresponding genes have been used to compute heatmaps (with R heatmap.2 Pearson distance, using complete linkage).

For the purpose of a cancer *vs* healthy tissue comparison at the scale of the single tumour type, thirteen tumour types out of 33 in the TCGA were selected based on the availability of a corresponding unambiguous normal tissue counterpart in the GTEx database. These thirteen couples of matrices were the following: breast cancer (BRCA) vs. normal breast, prostate adenocarcinoma (PRAD) vs. normal prostate, lung adenocarcinoma (LUAD) and lung squamous carcinoma (LUSC) vs. normal lung, pancreas cancer (PAAD) vs. normal pancreas, liver hepatocarcinoma (LIHC) vs normal liver, stomach adenocarcinoma (STAD) vs. normal stomach, oesophageal carcinoma (ESCA) vs normal oesophagus, colon adenocarcinoma (COAD) and rectal adenocarcinoma (READ) vs normal colon, bladder carcinoma (BLCA) vs normal bladder, uterus endometrial cancer (UCEC) vs normal uterus, ovarian cancer (OVCA) vs normal ovary. For each of these 13 tumour types, the distribution of R Pearson correlation coefficients for gene co-expression was compared between the matrices of a given tumour type and its normal tissue counterpart. This was done separately for the set of 14 genes of the 20S proteasome, the set of 18 genes of the 19S proteasome and for the set of 38 genes encoding XPO1-interacting proteins. The statistical significance of the differences in the distribution of R values was evaluated with the T Student's test.

Results

Strong positive correlation (R Pearson correlation > 0.8) was observed for proteasome genes pairs in the majority of healthy tissues such as liver, pancreas, stomach, colon, kidney. In contrast, the correlation for co-expression is positive but significantly lower ($R \leq 0.50$) for most gene pairs in the majority of cancer specimens. Examples of matrices of gene co-expression analysis are shown in Fig. 1A and B and represent results for specimens originating from normal lung and normal pancreas (GTEx project) and from lung adenocarcinoma (LUAD) and pancreatic adenocarcinoma (PAAD) specimens within the TCGA. Such matrices have been obtained for all 33 tumour types represented in the TCGA and 32 normal tissues represented in the GTEx project (Supplementary Information available at https://figshare.com/projects/ProteasomeGeneExpression/171357). The position of the genes in the matrix from the top to the bottom is in the following order: proteasome genes of the core complex (*PSMA1-10* and *PSMB1-8*), proteasome genes of the 19 S complex (*PSMC1-6* and *PSMD1-14*), *XPO1*, genes encoding XPO-1 interacting proteins, control genes. In the right-hand side of the matrix each square represents the degree of correlation for the corresponding gene pair with the R factor indicated in the square and a colour scale that illustrates the level of correlation with darker brown for higher correlation and lighter brown for lower correlation. In the left part of the matrix each square indicates graphically the correlation between the number of transcripts of the gene on the x axis and the gene on the y axis.

In the two examples of matrices shown in Fig. 1A and B the degree of correlation is on average lower in the tumour than in the normal tissue, as suggested by the coloured squares. This is more evident for proteasome genes (upper part, left side) than for genes encoding XPO1-interacting proteins nuclear (lower part, right side).

It is of note that the expression of genes encoding subunits of the immunoproteasome (*PSMB8*, *PSMB9*, *PSMB10*) is not correlated with the expression of genes encoding the other 26S proteasome subunits, whereas the expression of the three genes of this subgroup is highly correlated to each other. Interestingly, in the spleen the three genes are co-expressed with all other proteasome genes, reflecting the abundance of immunoproteasomes in immune cells (data not shown, see Supplementary Information). Likewise, the expression of *PSMA8*, is not correlated with the other proteasome genes, except in testis, which is consistent with the function of the encoded subunit in spermatoproteasomes (data not shown, see Supplementary Information). As expected, the expression of control genes *ACTB* and *GAPDH* is also poorly correlated with the expression of proteasome genes.

In order to address whether the decrease in the degree of proteasome genes co-expression is a common theme distinguishing tumors and normal tissues, the matrices of gene co-expression analysis were compared across tumors (33 matrices) and normal tissues (32 matrices) and the Euclidean distance was used to cluster the correlation matrices into a dendrogram. When building the dendrogram either considering the set of 38 proteasome genes alone or the set of 38 genes encoding XPO1-interacting proteins alone, we obtained a clear separation with proteasome genes, and a much weaker separation between tumors and healthy tissues when considering the set of genes encoding XPO1-interacting proteins. Figure 2 shows how the 65 matrices clearly separate in two parts. The only tumor matrices that remain close to normal tissues with respect to gene co-expression analysis of this set of 78 genes are Kidney cancer with chromophobe cells (KICH), Uveal melanoma (UVM) and Cholangiocarcinoma (CHOL).

Interestingly, common cancers originating from epithelia and mucosae are close to each other in the dendrogram (for example breast cancer/BRCA, bladder cancer/

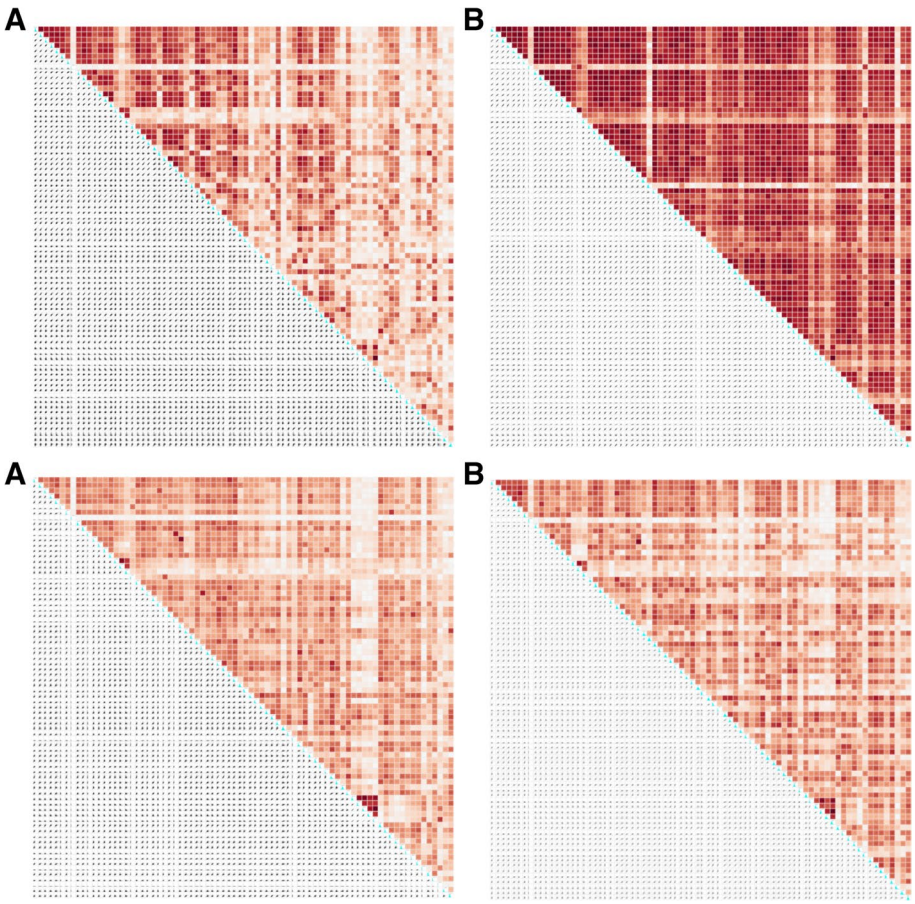

Fig. 1 **A** Matrices representing R Pearson correlation coefficients for gene co-expression within a set of 78 genes (38 proteasome genes, 38 genes encoding Exportin-1 interacting proteins and 2 control genes) for 605 specimens of healthy lung from the GTEx and 601 specimens of Lung adenocarcinoma from the TCGA. The cell's colour intensity is proportional to the R Pearson correlation between the gene pairs. Range of R values for Healthy Lung: 0.000083-0.967020; range for Lung Adenocarcinoma: 0.000152-0.962849. The gradient scale colour is available at https://figshare.com/projects/ProteasomeGeneExpression/171357. **B** Matrices representing R Pearson correlation coefficients for gene co-expression within a set of 78 genes (38 proteasome genes, 38 genes encoding Exportin-1 interacting proteins and 2 control genes) for 360 specimens of healthy pancreas from the GTEx and 183 specimens of Pancreas adenocarcinoma from the TCGA. The cell's colour intensity is proportional to the R Pearson correlation between the gene pairs. Range of R values for Healthy Pancreas: 0.003631-0.968954; range for Pancreas Adenocarcinoma: 0.000224-0.978171. The gradient scale colour is available at https://figshare.com/projects/ProteasomeGeneExpression/171357

BLCA, skin cutaneous melanoma/SKCM) and at large distance from matrices obtained with normal tissue; it is also of note that cancers with similar biology and cell of origins are close in the dendrogram (for example lung squamous cancer/LUSC and head and neck squamous carcinoma/HNSC or colon adenocarcinoma/COAD and rectal adenocarcinoma/READ).

In order to identify the genes that contributed most to the distinction between tumours and healthy tissues, we used a machine learning algorithm (XGBoost) trained on all pairs of tumour-healthy tissues. The model was subsequently applied to the same dataset for predicting the dataset category (tumour or healthy). Unsurprisingly, the model was able to classify the training set with 100% accuracy, thereby revealing the dataset features (i.e., the gene pair correlations) that contributed most to the successful classification. The genes with highest classification weights are presented in Table 2.

Further, these genes (PSMA1, PSMA5, PSMA7, PSMB1, PSMB3, PSMB4, PSMB5, PSMB6, PSMB7, PSMC2, PSMC3, PSMC4, PSMD1, PSMD7, PSMD8, PSMD11, GAPDH, XPO1, TTF2, UBE2I) involved in the correlation pairs for which XGBoost feature contribution was higher than a cutoff value of 0.01 were used as a gene subset for hierarchical clustering. Sixteen out of the top 20 genes in this list encode proteasome subunits, either of the 20S (core) complex (9 genes) or of the regulatory 19S complex (7 genes) and this set includes

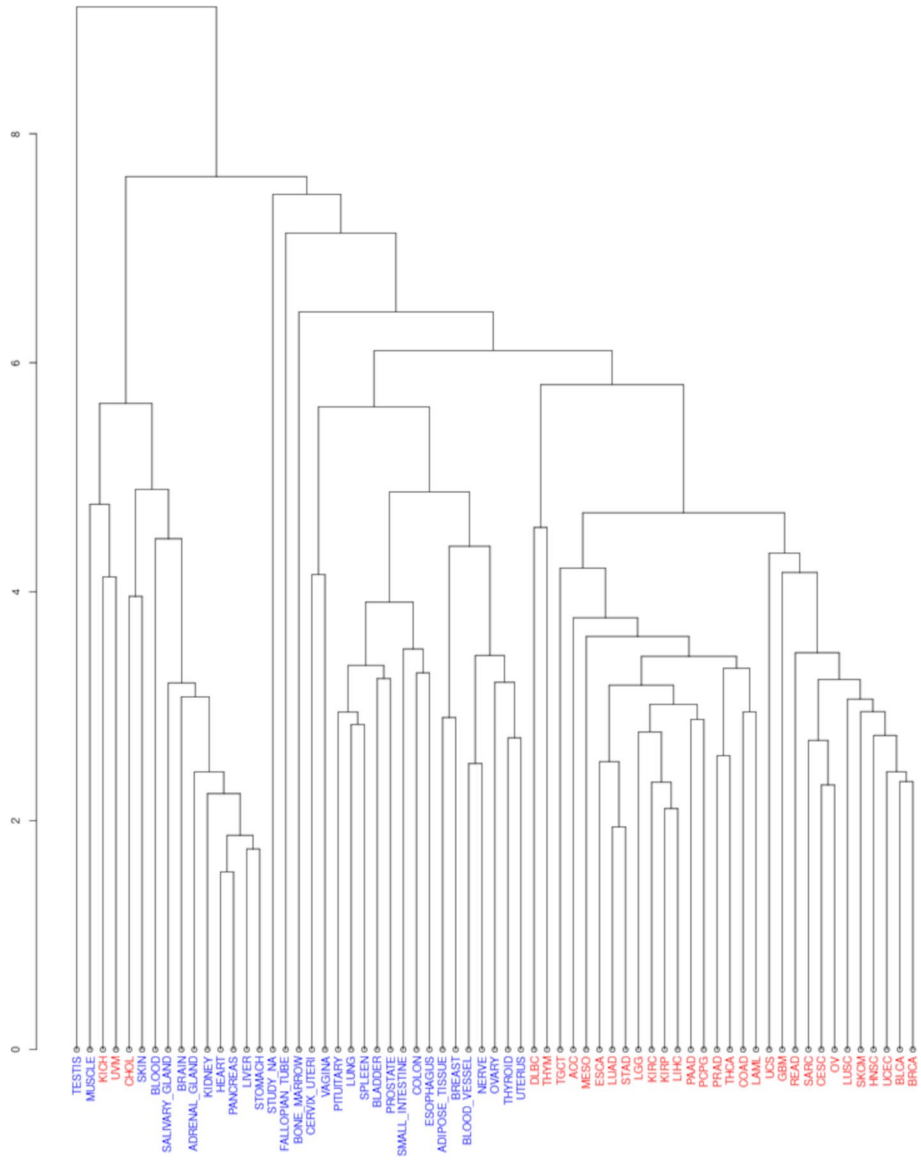

Fig. 2 Dendrogram representing results of cluster analysis of 65 matrices (33 matrices from tumour types from TCGA in red colour and 32 matrices of gene co-expression from healthy tissues from the GTEx project in blue colour)

the 3 genes encoding the subunits with catalytical activity for substrate degradation (*PSMB5* encoding beta-5 with chymotrypsin-like activity, *PSMB6* encoding beta-1 with caspase-like activity and *PSMB7* encoding beta-2 with trypsin-like activity). Other genes in this set have important functions such as *PSMD8* encoding the subunit which allows the incorporation of the lid subcomplex into the base of the 19S, *PSMD1* which encodes RPN2 protein which acts as a docking site for the ubiquitin receptor RPN13, and *PSMD7*, which encodes a protein which dimerizes with PSMD14 and assists in deubiquitylation of proteasome substrates. Among the set of the top 20 genes that separate cancers and healthy tissues we found 3 genes of the group of XPO1-interacting proteins: *XPO1* itself, *UBE2I* which encodes UBC9, a protein with E2-conjugating activity and *TTF2*, a gene encoding the Transcription Termination Factor 2, a protein involved in DNA-protein interactions which is found to physically interact with XPO1 in the OpenCell database [16].

When plotting the distribution of the R Pearson correlation coefficients obtained for the co-expression of genes encoding the 20S subunit (14 genes), the 19S subunit (18 genes), and the genes of XPO-1 interacting proteins (38 genes) and comparing the distribution of R

Table 2 Genes contributing most to the clustering of cancer versus healthy tissue matrices

GENE	Protein product with subcomplex location (20S/19S) and function
PSMA1	20S, encodes alpha-6 subunit (alpha ring), involved in 19S docking
PSMA5	20S, encodes alpha-5 subunit (alpha-ring), involved in 19S docking
PSMA7	20S, encodes alpha-4 subunit (alpha-ring), involved in 19S docking and 20S gating
PSMB1	20S, encodes beta-6 subunit, part of beta-ring, likely structural role
PSMB3	20S, encodes beta-3 subunit, part of beta-ring, likely structural role
PSMB4	20S, encodes beta-4 subunit, part of beta-ring, likely structural role
PSMB5	20S, chymotrypsin activity, the target of bortezomib, carfilzomib, ixazomib
PSMB6	20S, encodes beta-1 subunit, caspase-like activity (degradation)
PSMB7	20S, encodes beta-2 subunit, trypsin-like activity (degradation)
PSMC2	19S, resides in the base, «unfoldase» activity
PSMC3	19S, resides in the base, «unfoldase» activity
PSMC4	19S, resides in the base, «unfoldase» activity
PSMD1	19S, encodes RPN2, docking of the ubiquitin receptor RPN13
PSMD7	19S, dimerises with PSMD14, assists in deubiquitylation
PSMD8	19S, allows incorporation of the lid to the base of proteasome cap
PSMD11	19S, encodes PCI protein regulating ATPase activity (translocation)
TTF2	Encodes transcription terminator factor 2, various roles in protein-DNA interactions
UBE2I	Encodes UBC9, which has E2 conjugating activity
XPO1	Encodes Exportin-1, the target of Selinexor and other SINE-drugs
GAPDH	Encodes Gliceraldehyde-3P –dehydrogenase (included as a control)

The top 20 genes contributing most to the clustering of cancer *vs* healthy tissue matrices. Dark green colour highlights proteins that are established targets in the treatment of cancer, light green colour highlights proteins with functional and/or enzymatic activity that is potentially targetable. Non-highlighted proteins are structural proteins, one control protein and one transcription factor

values obtained for 13 tumor types and their healthy tissue counterparts, we found different results for the co-expression of genes encoding subunits of the proteasome and for genes encoding XPO1-interacting proteins. There was a statistically significant decrease of R values for the co-expression of 20S genes and 19S genes in all tumors as compared to their corresponding healthy tissue, with the only exception of the genes encoding 19S subunits in endometrial carcinoma of the uterine corpus. For pairs of genes encoding 20S subunits the median correlation factor is quite high in healthy tissues (at least 0.70) whereas it decreases to medium-low values in corresponding tumour specimens (around 0.50 or less). The decrease of the correlation in gene co-expression is observed both for the set of gene pairs encoding 20S subunits and for gene pairs encoding 19S subunits (as shown by violin plots in Fig. 3A, B, D, E referring to normal lung and lung adenocarcinoma and to normal pancreas and pancreas

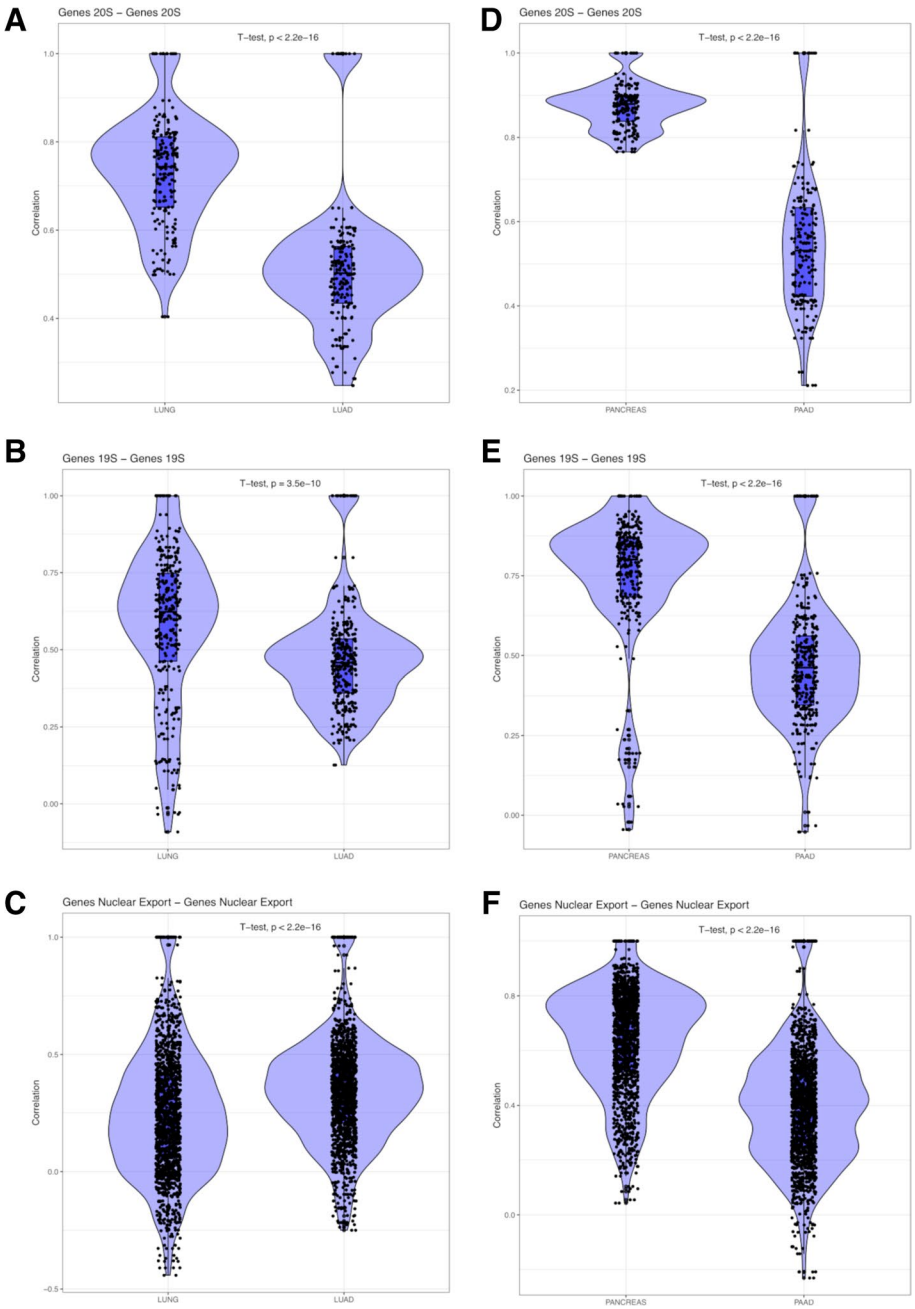

Fig. 3 **A** Distribution of R Pearson correlation factors for co-expression of genes encoding 20S proteasome subunits in 655 specimens of normal lung and 601 specimens of lung adenocarcinoma. **B** Distribution of R Pearson correlation factors for co-expression of genes encoding 19S proteasome subunits in 655 specimens of normal lung and 601 specimens of lung adenocarcinoma. **C** Distribution of R Pearson correlation factors for co-expression of genes encoding XPO1-interacting proteins in 655 specimens of normal lung and 601 specimens of Lung adenocarcinoma. **D** Distribution of R Pearson correlation factors for co-expression of genes encoding 20S proteasome subunits in 360 specimens of normal pancreas and 183 specimens of pancreas adenocarcinoma (PAAD). **E** Distribution of R Pearson correlation factors for co-expression of genes encoding 19S proteasome subunits in 360 specimens of normal pancreas and 183 specimens of pancreas adenocarcinoma (PAAD). **F** Distribution of R Pearson correlation factors for co-expression of genes encoding XPO1-related proteins in 360 specimens of normal pancreas and 183 specimens of pancreas adenocarcinoma (PAAD)

adenocarcinoma). In contrast, the degree of co-expression of genes encoding XPO1-interacting proteins was for many tumor types either not significantly different (for example for prostate cancer or pancreas adenocarcinoma, Fig. 3F) or even in some cases higher in tumors than in corresponding healthy tissue (for example for lung adenocarcinomas, lung squamous carcinomas, bladder carcinomas, ovarian carcinomas, uterine endometrial carcinomas) (Fig. 3C and Supplementary Information).

In order to explore the hypothesis of an interplay between the proteasome and nucleocytoplasmic transport, we screened for gene pairs made of two genes belonging each to one of the two families of genes (proteasome genes and genes encoding proteins interacting with XPO1) that had a high degree of co-expression. We found several cases of such gene pairs that were distributed among various tumour types. Interestingly, the majority of these pairs of genes involved, as the proteasome partner of the pair, a gene encoding a subunit of the 19S regulatory complex, such as PSMD7 (correlated with *NUP93* with a R of 0.70 in breast cancer), and *PSMD14* (correlated with *XPO1* with $R > 0.70$ in rectum adenocarcinoma, stomach adenocarcinoma, uveal melanoma and with *SUMO1* with $R = 0.78$ in skin melanoma). The expression of the *XPO1* gene is not correlated with the expression of proteasome genes in healthy tissues, whereas it is so in tumour tissues, as shown in Table 3 for select gene pairs in breast cancer, prostate cancer and endometrial cancer of the uterus. The proteasome gene with which the expression of *XPO1* has the highest degree of correlation in cancer samples is *PSMD14*. When focusing on the gene pair *PSMD14/XPO1*, we find that the correlation is always higher in cancer specimens than in the tissue of origin (Table 4). In some instances (breast cancer, prostate cancer, endometrial cancer), there is no correlation at all between *PSMD14* and *XPO1* expression in the tissue of origin and a highly correlated expression in the cancer specimens. In contrast, in most normal tissues considered for the comparison, the *PSMD14* gene is highly correlated ($R > 0.8$) with *RAN*, encoding the small GTPase RAN, a key XPO1 partner in mediating nucleocytoplasmic transport (Table 4).

Discussion

These findings show that proteasome genes are co-expressed in a highly correlated manner in healthy tissues and confirm that gene co-expression analysis can be used to investigate the composition of multiprotein

Table 3 Correlation between select gene pairs in three cancer types and healthy counterparts

Gene pair	Breast N	Breast CA	Prostate N	Prostate CA	Uterus N	Uterus CA
PSMA2/PSMB5	0.74	0.30	0.68	0.65	0.76	0.60
PSMA2/PSMD7	0.58	0.47	0.80	0.61	0.62	0.57
PSMA2/PSMD14	0.66	0.65	0.82	0.54	0.78	0.72
PSMB5/PSMD7	0.52	0.40	0.81	0.59	0.67	0.50
PSMB5/PSMD14	0.67	0.42	0.82	0.69	0.68	0.59
PSMD7/PSMD14	0.71	0.57	0.81	0.52	0.55	0.58
PSMA2/XPO1	<0.1	0.44	<0.1	0.30	0.15	0.58
PSMB5/XPO1	<0.1	0.17	<0.1	<0.1	0.23	0.31
PSMD7/XPO1	<0.1	0.41	<0.1	0.29	0.34	0.37
PSMD14/XPO1	<0.1	0.60	<0.1	0.56	0.11	0.62

R Pearson correlation coefficients for select gene pairs within the set of proteasome genes (*PSMA2, PSMB5, PSMD7, PSMD14*) and between this subset of proteasome genes and the *XPO1* gene in breast, prostate and endometrial cancer relative to normal breast, normal prostate and normal uterus. The higher R correlation factor in the comparison Normal vs Cancer is coloured in darker blue and the lower correlation factor in lighter blue. Differences of R values ≥ 0.50 are highlighted in yellow and differences ≥ 0.3 are highlighted in orange

Table 4 Correlation between gene pairs *PSMD14/RAN* and *PSMD14/XPO1* in thirteen cancer types and healthy counterparts

	PSMD14/RAN		PSMD14/XPO1	
NORMAL TISSUE(N)/TUMOUR TYPE (CA)	N	CA	N	CA
BLADDER/BLCA	0.99	0.54	0.12	0.53
BREAST/BRCA	0.81	0.60	<0.1	0.60
COLON/COAD	0.86	0.71	0.28	0.58
ESOPHAGUS/ESCA	0.90	0.61	0.25	0.60
LIVER/LIHC	0.88	0.64	0.35	0.62
LUNG/LUAD	0.89	0.72	0.21	0.48
LUNG/LUSC	0.89	0.63	0.21	0.41
OVARY/OVCA	0.77	0.46	0.01	0.39
PANCREAS/PAAD	0.85	0.62	0.34	0.58
PROSTATE/PRAD	0.88	0.65	<0.1	0.56
COLON/READ	0.86	0.71	0.28	0.70
STOMACH/STAD	0.91	0.75	0.49	0.71
UTERUS/UCEC	0.81	0.65	0.11	0.62

R Pearson correlation coefficients for co-expression of gene pairs *PSMD14/RAN* and *PSMD14/XPO1* in 13 tumour types (CA) and in their respective normal tissue counterparts (N). Higher values in the comparison are in darker blue and lower values in lighter blue. Differences of R values ≥ 0.5 between CA and N are highlighted in yellow and differences of R values ≥ 0.3 are highlighted in orange

complexes. Specialized tissues such as spleen show a high correlation between regular proteasome genes and genes encoding special subunits of immunoproteasomes found in immune cells with the function of producing peptides to be presented by the MHC complex. Conversely the expression of these genes is unrelated to the expression of other proteasome genes in all other tissues.

In contrast with normal tissues, the pattern of co-expression of proteasome genes is extensively altered in almost all tumour types included in the TCGA and the degree of correlation between genes' expression is lower both for genes encoding subunits of the 20S core of the complex and for genes encoding subunits of the 19S regulatory subcomplex. This observation holds true not only at a "pancancer" level but also when comparing single tumour types with their normal tissue counterpart, representing their tissue of origin. As a result, this data supports the hypothesis that not only the ubiquitination and deubiquitylation processes can be altered in cancer [20] but also the proteasome complex itself.

One possible interpretation of these findings could be that the stoichiometry of the proteasome complex, i.e. the composition and the assembly of the subunits, is altered in cancer cells. It is known that several species of proteasomes exist in different tissues (e.g., the immunoproteasome in immune cells) and that some parts of the complex (the 19S) are very dynamic. Assembly of non-canonical proteasome complexes has been described in yeast [21]. Therefore, it is plausible that cancer cells harbour non-canonical proteasome species in order to comply with the increased demand for protein turn-over in the transformed cell or to obtain a selective advantage of growth and survival. This model could be investigated by biochemical studies on purified proteasomes, although such studies are labour-intensive, technologically demanding and require large amounts of tissue [22]. To our knowledge, our study is the first attempt to evaluate the entire proteasome complex at a large scale in cancer, albeit by an approach based on computational analysis

of transcriptomic data, whereas no comparable studies have been reported so far at the proteomic level.

It is also possible that some proteasome subcomplexes or single subunits exist in cancer cells in addition to the regular full complex. It has been already hypothesized that such subcomplexes might have a role in cell signalling, given the similarity between the COP9/signalosome and parts of the proteasome [23]. With regard to single subunits, it has been shown that some proteasome subunits have non-canonical roles (i.e. roles independent from the full proteasome complex), for example in DNA-repair [24]. It is possible that these non-canonical functions are amplified in cancer cells and explain an uncoupling of the transcription pattern of the respective genes in comparison with other proteasome genes. These models could operate differently depending on tumour type, grade and stage and some subunits might have a more important oncogenic role in some contexts and not in others. The data presented here do not enable us to investigate these issues, as this would require further studies focused on single tumour types and considering their clinical features and prognosis.

When interpreting this data, we should also consider the possibility of methodological biases. We have used tools such as recount3 [25] that have significantly eased the effort to compare diverse datasets such as TCGA and GTEx in a uniform way, but biases due to experimental factors such as "batch effects" should be taken into account. However, the intrinsic feature of our analysis should limit such biases, since we addressed the correlation analysis of co-expression rather than the analysis of absolute expression levels linked to measurements. Second, a systematic bias seems unlikely, since the second set of genes included in this work (those encoding XPO1-interacting proteins) behaves differently from proteasome genes and does not show a general decrease of the degree of correlated co-expression, unlike proteasome genes (Fig. 3A-F). Moreover, the findings of this work are biologically plausible and consistent with several reports on the overexpression of single subunits in cancer relative to normal tissue [7–9].

From a therapeutical perspective, this work highlights 20 genes out of the 78 genes considered that are more often deregulated in comparison with normal tissues and among these genes we find 16 encoding subunits of the proteasome. These include PSMB5 which is already a therapeutic target for the treatment of myeloma and other subunits that are known to have important functional roles. This work therefore supports further investigation of these targets for anticancer treatment [10].

Our study shows that there is a high degree of correlation in terms of gene co-expression between some proteasome genes and genes encoding proteins that interact with XPO1 and are involved in nuclear export. This part of the study is exploratory and the findings are only hypothesis generating. Nevertheless, very interestingly one of the genes in this group that is more often correlated with proteasome genes is the *XPO1* gene itself, raising the possibility that in cancer cells proteasome biosynthesis is co-regulated with the XPO1 protein, which acts as the main nuclear export protein. This might be explained by the necessity to support the shuttling of proteasome complex between cellular compartments and by a cooperation between some proteasome subunits and the nuclear export proteins (RAN, XPO1). Several lines of evidence suggest that there is an interplay between nuclear transport and the ubiquitin-proteasome system in several biological contexts (reviewed in [26]). In cancer cells that have increased needs in terms of protein quality control and protein turn-over, this interaction might gain importance. Another interesting observation is the high correlation for co-expression between *PSMD14* and *RAN* in normal tissues and between *PSMD14, RAN* and *XPO1* in cancer tissues. PSMD14 is a key subunit involved in deubiquitylation of proteasome substrates and the high correlation of co-expression with RAN and XPO1 suggests a possible interplay between the processes of deubiquitylation and nuclear export of substrates.

The findings of this study are consistent with the observation of a similar range of anticancer activity and the synergy between the XPO1 inhibitor Selinexor and the PSMB5 inhibitors Bortezomib and Carfilzomib [14, 15]. Based on our findings the concomitant inhibition of PSMD14, which has a deubiquitinase activity that is druggable, and XPO1 could also be a combination of therapeutic interest.

Conclusions

This study shows that the expression of proteasome genes is extensively altered in solid tumours, regardless of tumour type. Cancer specimens have a significantly lower degree of co-expression of proteasome genes as compared with the tissue of origin, which suggests a substantial rewiring of proteasome genes and a change in the stoichiometry of proteasome subunits. Genes that are more often involved encode subunits of the complex that are known to have critical roles within the complex, such as subunits beta-1 (PSMB6), beta-2(PSMB7), beta-5 (PSMB5), PSMD8, PSMD1 and PSMD7.

These data could be a basis to envision further studies on proteasome complexes in cancer by proteomic approaches. They also support the development of compounds or treatment strategies for targeting the proteasome not only in multiple myeloma (where registered proteasome inhibitors target the beta-5 subunit) but also in other cancer types, in which other subunits,

with deregulated expression as shown in this work, could be attractive targets. Indeed, several pieces of evidence show that the inhibition or downregulation of critical subunits might have an antiproliferative effect both in solid tumours [27, 28] and in multiple myeloma resistant to inhibitors of the beta-5 subunit [29].

In addition, our results show that there are genes involved in nucleocytoplasmic transport such as *XPO1* that have a highly correlated co-expression with proteasome genes, and in particular with a subset of them such as *PSMD14*, in cancer specimens as compared to normal tissues. The expression of the *PSMD14* gene is also highly correlated to that of the *RAN* gene in normal tissues, suggesting that a coordination of the two pathways exists in normal cells and gains importance in cancer cells. These findings are consistent with the observation that drugs inhibiting XPO1 synergize with proteasome inhibitors against multiple myeloma and lend support to the development of combination treatments that co-target proteins involved in nucleocytoplasmic transport and the proteasome also for the treatment of other tumour types.

Acknowledgements
The computations were performed at University of Geneva on "Baobab" and/or "Yggdrasil" HPC clusters.

Authors' contributions
Concept and design of the work: VS and ABD. Data acquisition and computational analysis: ABD. Interpretation of the data: VS and ABD. Draft and revision of the manuscript: VS and ABD.

Funding
The authors did not receive funding for this work.

Availability of data and materials
The datasets generated and analysed during the current study are available in the figshare repository at the following link: https://figshare.com/projects/ProteasomeGeneExpression/171357.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Service of Medical Oncology, Oncology Institute of Southern Switzerland (IOSI), Ospedale San Giovanni, Via Gallino, CH-6500 Bellinzona, Switzerland. [2]Practice of Medical Oncology, Viale Stazione 23, CH-6500, Bellinzona, Switzerland. [3]Institute of Microbiology, Department of Environmental Constructions and Design, University of Applied Sciences and Arts of Southern Switzerland (SUPSI), via Mirasole 22a, CH-6500 Bellinzona, Switzerland. [4]Swiss Institute of Bioinformatics, Quartier Sorge, Batiment Genopode, CH-1015 Lausanne, Switzerland.

Received: 6 September 2023 Accepted: 17 January 2024
Published online: 07 February 2024

References

1. Bard JAM, Goodall EA, Greene ER, Jonsson E, Dong KC, Martin A. Structure and Function of the 26S Proteasome. Annu Rev Biochem. 2018;87:697–724.
2. Mao Y. Structure, Dynamics and Function of the 26S Proteasome. Subcell Biochem. 2021;96:1–151.
3. Koizumi S, Hamazaki J, Murata S. Transcriptional regulation of the 26S proteasome by Nrf1. Proc Jpn Acad Ser B Phys Biol Sci. 2018;94(8):325–36.
4. Vilchez D, Boyer L, Morantte I, Lutz M, Merkwirth C, Joyce D, et al. Increased proteasome activity in human embryonic stem cells is regulated by PSMD11. Nature. 2012;489(7415):304–8.
5. Vangala JR, Dudem S, Jain N, Kalivendi SV. Regulation of PSMB5 protein and beta subunits of mammalian proteasome by constitutively activated signal transducer and activator of transcription 3 (STAT3): potential role in bortezomib-mediated anticancer therapy. J Biol Chem. 2014;289(18):12612–22.
6. Pilarsky C, Wenzig M, Specht T, Saeger HD, Grutzmann R. Identification and validation of commonly overexpressed genes in solid tumors by comparison of microarray data. Neoplasia. 2004;6(6):744–50.
7. Salah Fararjeh A, Al-Khader A, Al-Saleem M, Abu Qauod R. The Prognostic Significance of Proteasome 26S Subunit, Non-ATPase (PSMD) Genes for Bladder Urothelial Carcinoma Patients. Cancer Inform. 2021;20:11769351211067692.
8. Zhou C, Li H, Han X, Pang H, Wu M, Tang Y, et al. Prognostic Value and Molecular Mechanisms of Proteasome 26S Subunit, Non-ATPase Family Genes for Pancreatic Ductal Adenocarcinoma Patients after Pancreaticoduodenectomy. J Invest Surg. 2022;35(2):330–46.
9. Spataro V, Buetti-Dinh A. POH1/Rpn11/PSMD14: a journey from basic research in fission yeast to a prognostic marker and a druggable target in cancer cells. Br J Cancer. 2022;127(5):788–99.
10. Wertz IE, Wang X. From discovery to bedside: targeting the ubiquitin system. Cell Chem Biol. 2019;26(2):156–77.
11. Li H, Sun Y, Zhan M. Exploring pathways from gene co-expression to network dynamics. Methods Mol Biol. 2009;541:249–67.
12. Jiang Z, Dong X, Li ZG, He F, Zhang Z. Differential coexpression analysis reveals extensive rewiring of arabidopsis gene coexpression in response to pseudomonas syringae infection. Sci Rep. 2016;6:35064.
13. Chari A, Vogl DT, Gavriatopoulou M, Nooka AK, Yee AJ, Huff CA, et al. Oral selinexor-dexamethasone for triple-class refractory multiple myeloma. N Engl J Med. 2019;381(8):727–38.
14. Grosicki S, Simonova M, Spicka I, Pour L, Kriachok I, Gavriatopoulou M, et al. Once-per-week selinexor, bortezomib, and dexamethasone versus twice-per-week bortezomib and dexamethasone in patients with multiple myeloma (BOSTON): a randomised, open-label, phase 3 trial. Lancet. 2020;396(10262):1563–73.
15. Gasparetto C, Schiller GJ, Tuchman SA, Callander NS, Baljevic M, Lentzsch S, et al. Once weekly selinexor, carfilzomib and dexamethasone in carfilzomib non-refractory multiple myeloma patients. Br J Cancer. 2022;126(5):718–25.
16. Cho NH, Cheveralls KC, Brunner AD, Kim K, Michaelis AC, Raghavan P, et al. OpenCell: Endogenous tagging for the cartography of human cellular organization. Science. 2022;375(6585):eabi6983.
17. Hutter C, Zenklusen JC. The cancer genome atlas: creating lasting value beyond its data. Cell. 2018;173(2):283–5.
18. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580–5.
19. Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318–30.
20. Ge Z, Leighton JS, Wang Y, Peng X, Chen Z, Chen H, et al. Integrated Genomic Analysis of the Ubiquitin Pathway across Cancer Types. Cell Rep. 2018;23(1):213-26 e3.
21. Hammack LJ, Kusmierczyk AR. Assembly of proteasome subunits into non-canonical complexes in vivo. Biochem Biophys Res Commun. 2017;482(1):164–9.

22. Bousquet-Dubouch MP, Uttenweiler-Joseph S, Ducoux-Petit M, Matondo M, Monsarrat B, Burlet-Schiltz O. Purification and proteomic analysis of 20S proteasomes from human cells. Methods Mol Biol. 2008;432:301–20.
23. Dubiel W, Chaithongyot S, Dubiel D, Naumann M. The COP9 Signalosome: A Multi-DUB Complex. Biomolecules. 2020;10(7):1082.
24. Butler LR, Densham RM, Jia J, Garvin AJ, Stone HR, Shah V, et al. The proteasomal de-ubiquitinating enzyme POH1 promotes the double-strand DNA break response. EMBO J. 2012;31(19):3918–34.
25. Wilks C, Zheng SC, Chen FY, Charles R, Solomon B, Ling JP, et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. Genome Biol. 2021;22(1):323.
26. Enenkel C, Kang RW, Wilfling F, Ernst OP. Intracellular localization of the proteasome in response to stress conditions. J Biol Chem. 2022;298(7):102083.
27. Li D, Lu Y, Sun P, Feng LX, Liu M, Hu LH, et al. Inhibition on Proteasome beta1 Subunit Might Contribute to the Anti-Cancer Effects of Fangchinoline in Human Prostate Cancer Cells. PLoS One. 2015;10(10):e0141681.
28. Shi K, Zhang JZ, Zhao RL, Yang L, Guo D. PSMD7 downregulation induces apoptosis and suppresses tumorigenesis of esophageal squamous cell carcinoma via the mTOR/p70S6K pathway. FEBS Open Bio. 2018;8(4):533–43.
29. Shi CX, Zhu YX, Bruins LA, Bonolo de Campos C, Stewart W, Braggio E, et al. Proteasome Subunits Differentially Control Myeloma Cell Viability and Proteasome Inhibitor Sensitivity. Mol Cancer Res. 2020;18(10):1453-64.

Publisher's Note