Integrative analysis of genomic and transcriptomic alterations of AGR2 and AGR3 in cancer

Delphine Fessart, Ines Villamor, Eric Chevet, Frederic Delom, Jacques Robert

OPEN BIOLOGY

royalsocietypublishing.org/journal/rsob

Research

Cite this article: Fessart D, Villamor I, Chevet E, Delom F, Robert J. 2022 Integrative analysis of genomic and transcriptomic alterations of *AGR2* and *AGR3* in cancer. *Open Biol.* **12**: 220068.
https://doi.org/10.1098/rsob.220068

Received: 8 March 2022
Accepted: 7 June 2022

Subject Area:
genomics/bioinformatics/cellular biology

Keywords:
AGR2, AGR3, transcriptomic, genomic, cancer

Authors for correspondence:
Delphine Fessart
e-mail: delphine.fessart@yahoo.fr
Frederic Delom
e-mail: frederic.delom@yahoo.fr
Jacques Robert
e-mail: J.Robert@bordeaux.unicancer.fr

Integrative analysis of genomic and transcriptomic alterations of *AGR2* and *AGR3* in cancer

Delphine Fessart[1,2], Ines Villamor[2], Eric Chevet[3,4], Frederic Delom[1] and Jacques Robert[1]

[1]ARTiSt, University Bordeaux, INSERM U1312, Bordeaux F-33000, France
[2]POETIC, University Bordeaux, INSERM U1312, Bordeaux F-33000, France
[3]INSERM U1242, 'Chemistry, Oncogenesis Stress Signaling', Université Rennes 1, Rennes, France
[4]Centre de Lutte Contre le Cancer Eugène Marquis, Rennes, France

DF, 0000-0001-7566-5670; EC, 0000-0001-5855-4522; FD, 0000-0002-4600-7633; JR, 0000-0002-3874-7965

The *AGR2* and *AGR3* genes have been shown by numerous groups to be functionally associated with adenocarcinoma progression and metastasis. In this paper, we explore the data available in databases concerning genomic and transcriptomic features of these two genes: the NCBI dbSNP database was used to explore the presence and roles of constitutional SNPs, and the NCI, Cancer Cell Line Encyclopedia (CCLE) and TCGA databases were used to explore somatic mutations and copy number variations (CNVs), as well as mRNA expression of these genes in human cancer cell lines and tumours. Relationships of *AGR2/3* expression with whole-genome mRNA expression and cancer features (i.e. mutations and CNVs of oncogenes and tumour suppressor genes (TSG)) were established using the CCLE and TCGA databases. In addition, the CCLE data concerning CRISPR gene extinction screens (Achilles project) of these two genes and a panel of oncogenes and TSG were explored. We observed that no functional polymorphism or recurrent mutation could be detected in *AGR2* or *AGR3*. The expression of these genes was positively correlated with the expression of epithelial genes and inversely correlated with that of mesenchymal genes. It was also significantly associated with several cancer features, such as *TP53* or *SMAD4* mutations, depending on the gene and the cancer type. In addition, the CRISPR screens revealed the absence of cell fitness modification upon gene extinction, in contrast with oncogenes (cell fitness decrease) and TSG (cell fitness increase). Overall, these explorations revealed that AGR2 and AGR3 proteins appear as common non-genetic evolutionary factors in the process of human tumorigenesis.

1. Introduction

Members of the protein disulfide isomerase (PDI) family, which are endoplasmic reticulum (ER)-resident enzymes interfering in the formation of disulfide bonds, cysteine-based redox reactions and quality control of proteins in the ER, play an essential role in ER homeostasis (proteostasis); in addition to their principal ER location, some of these enzymes are found in other localizations such as the extracellular milieu, in extracellular vesicles or the cytosol [1]. For instance, we have shown that PDIA2 is secreted into the lumen of the thyroid follicles by thyrocytes to control extracellular thyroglobulin folding and multimerization [2,3]. There is ample evidence supporting that PDI proteins are strongly associated with cancer either through their altered expression or through enhanced functions. Although they are among the most abundant cellular proteins, PDI expression is frequently upregulated in cancers and associated with metastasis and invasiveness [1].

THE ROYAL SOCIETY PUBLISHING

However, the functions of PDI proteins in the process of human oncogenesis remain to be understood. Among the most studied PDI in this respect are those belonging to the Anterior GRadient (AGR) family of proteins. The AGR family is composed of three proteins, namely AGR1 [gene *TXNDC12*], AGR2 and AGR3. Interestingly, AGR2, the prototypic member of the AGR family, is shown to play intracellular roles in the ER, contributing to proteostasis [4], but it remains unclear how this is related to oncogenesis. *AGR2* and *AGR3* genes are localized on chromosome 7, side by side (7p21.1), and their protein products are both overexpressed and their localizations deregulated in many types of adenocarcinomas [5–7]. We have shown that two non-canonical localizations: extracellular (eAGR2/3) [8–10] and cytosolic (cAGR2) [11] and exert pro-oncogenic gain-of-function to confer tumours specific and evolutive features (development, progression and aggressiveness). Moreover, the overexpression of AGR2 and AGR3 may be a prognosis factor for survival, which could be favourable or not favourable depending on the cancer type [7].

These observations raise the question of whether *AGR2* and *AGR3* could behave as '*cancer genes*', i.e. as oncogenes and/or tumour suppressor genes (TSG). To bring some answers to this question, we have explored publicly available databases to search for relationships between genomic variations of *AGR2* and *AGR3* and cancer. In a first attempt, polymorphisms were sought in germline DNA using the dbSNP database; then, somatic tumour variations were sought in the the cancer genome atlas (TCGA) tumour collection and in the Cancer Cell Line Encyclopedia (CCLE) cell line database, so as to elaborate a directory of potentially oncogenic mutations. In addition to the exploration of the sequence of these genes in constitutional and tumour DNA, we explored the expression pattern of both genes in tumour and cell lines of various tissue origins, and searched for relationships between *AGR2* and *AGR3* gene expression and several oncogenic determinants in various cancer types, tumours or cell lines, especially copy number variations (CNVs) and point mutations (single-nucleotide variations, SNV). We performed a comprehensive analysis of available data in order to better understand the role of AGR2 and AGR3 in cancer. All the analyses were conducted on the data available online as of April 2021.

2. Methods

2.1. Databases

The dbSNP database was accessed from the NCBI database using the followings links:

For AGR2: https://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=10551

For AGR3: https://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=155465

We restricted our analysis to exomic variations. Synonymous variations were not studied. TCGA (The Cancer Genome Atlas) was accessed through the cBioPortal for Cancer Genomics: https://www.cbioportal.org. Only data from the PanCancer Atlas were retrieved; they concern 32 different cancer types for a total of 10 945 tumours. Data concerning SNV, CNVs and mRNA expression (RSEM, batch normalized from Illumina HiSeq_RNASeqV2) were downloaded and converted into Excel sheets for analysis. We used the cancer type nomenclature of the TCGA (electronic supplementary material, table S1). The CCLE was accessed through a friendly user platform, https://discover.nci.nih.gov/cellminercdb/, established at NCI and gathering all publicly available data concerning cancer cell line molecular and pharmacological properties [12,13]. Rapid surveys of collections other than CCLE (namely GDSC, Genomics of Drug Sensitivity in Cancer, and CTRP, Cancer Therapeutics Response Portal) were performed in order to assess the accuracy of CCLE data. Most of the other analyses were conducted on the CCLE collection, which contained the highest number of cell lines, but all three collections are redundant and contain the same core cell lines, so that this restriction does not generate any bias.

2.2. Statistics

We used common statistical tests for data comparisons, mainly chi-squared and Student's *t*-test; all tests were two-sided and we considered that significance was obtained only at the 1% level. Large numbers of statistical tests were performed in several instances, and we took multiple testing into account by applying Bonferroni correction. For instance, as many as $12 \times 20\,000$ *p*-values were computed for gene association detection: in such cases, we considered only $p < 5 \times 10^{-8}$ as significant at the 1% level.

3. Results

3.1. *AGR2* and *AGR3* polymorphisms

In order to distinguish germline polymorphisms from potential mutations in tumour tissues, we first listed the *AGR2* and *AGR3* gene polymorphisms identified in the NCBI dbSNP database. In this database, 165 SNV or small insertion/deletion variations (indel) in the *AGR2* gene coding sequence are listed, affecting 115 of the 175 amino acids of the protein. When indicated in the database, none of them has a minor allele frequency (MAF) higher than 0.0002, with the exception of rs6842 (N147N), a synonymous variation with a MAF of 0.3355. These variations were synonymous (41 cases), missense (112), nonsense (6), frameshift (7) or in frame (1).

Similarly, in the NCBI dbSNP database, 214 SNV or indels have been described in the *AGR3* gene coding sequence, affecting 131 of the 166 amino acids of the protein. When indicated, none of them had a MAF higher than 0.0006, with the exception of rs55900499 (D40D), a synonymous variation with a MAF of 0.0505. These variations were synonymous (48 cases), missense (151), nonsense (11) or frameshift (4).

3.2. *AGR2* and *AGR3* somatic tumour variations in the TCGA

The TCGA database provides a unique comprehensive resource for exploring gene variations occurring in human tumours. Out of a total of 9888 tumours originating from 32 tumour types (list and abbreviations in the electronic supplementary material, table S1), we identified 32 samples bearing an *AGR2* gene variation (mutation or polymorphism) (figure 1*a*) and 35 bearing an *AGR3* gene variation (figure 1*b*). A total of 30 different variations involving 26 codons in *AGR2*, and 31 mutations involving 26 codons in *AGR3*, were present in these samples. Three samples presented two variations in the *AGR2* sequence

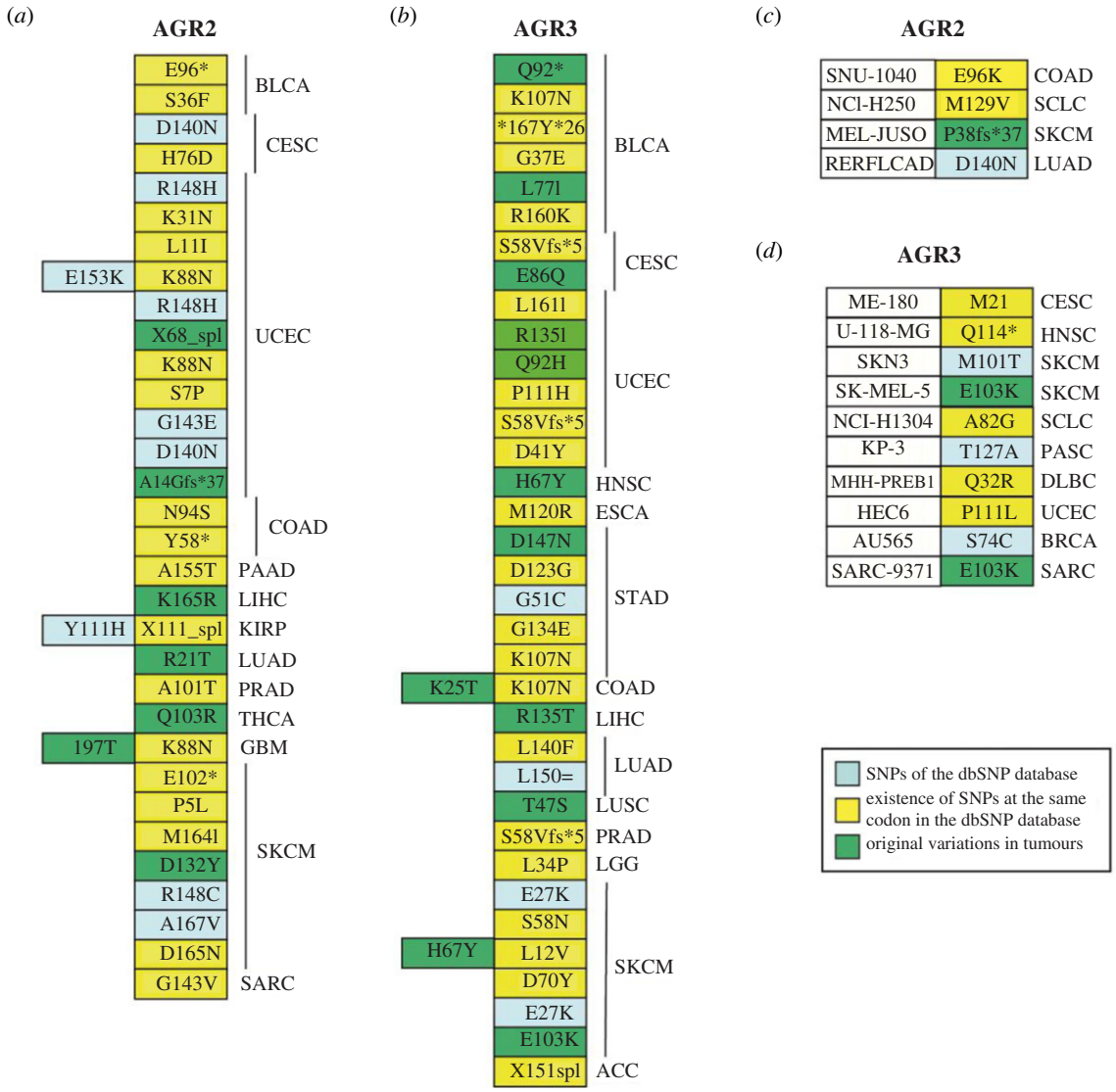

Figure 1. Point mutations of *AGR2* and *AGR3* are present in databases. (*a*,*b*) Point mutations of *AGR2* (*a*) and *AGR3* (*b*) genes in 10 376 tumour samples of the TCGA. The standard mutation nomenclature in molecular diagnostics can be found at https://www.hgvs.org/mutnomen/recs-prot.html. (*c*,*d*) Point mutations in *AGR2* (*c*) and *AGR3* (*d*) genes in 1036 cell lines of the CCLE and GDSC collections.

and two other samples in *AGR3* sequence (figure 2). Only three samples showed variations in both *AGR2* and *AGR3* genes (figure 2). Most variations were missense mutations; there were three nonsense mutations in *AGR2* and two in *AGR3*; two frameshift mutations in *AGR2* and one in *AGR3*; and two splice mutations in *AGR2* and one in *AGR3*. Some cancer types presented more mutations than others: skin cutaneous melanomas and endometrial carcinomas for *AGR2* (electronic supplementary material, table S2A), and the same plus stomach and bladder carcinomas for *AGR3* (electronic supplementary material, table S2B). Among the 30 *AGR2* variations found in the TCGA, seven were in the dbSNP list, 16 affected a codon where a SNP had been identified and seven concerned a codon not known as subject to polymorphic variation. Among the 31 *AGR3* variations found in TCGA, five were in the dbSNP list, 17 affected a codon where a SNP had been identified and nine concerned a codon not known as subject to a polymorphic variation.

Concerning CNVs, there were in TCGA 146 samples with *AGR2* gene amplifications and 14 with *AGR2* homozygous deletion (electronic supplementary material, table S2A); and 145 samples with *AGR3* gene amplification and 15 with *AGR3* homozygous deletion (electronic supplementary material, table S2B). Most of samples were amplified on both genes, nine samples presenting *AGR2* amplification only and seven *AGR3* amplification only. Similarly, only one sample had a homozygous deletion of only one of the two genes, *AGR3*.

3.3. *AGR2* and *AGR3* somatic tumour variations in cell line collections

In the collections of cell lines of GDSC (Genomics of Drug Sensitivity in Cancer) and CCLE, four tumour cell lines bear a variation in *AGR2* coding sequence, among which three are common to the two databases. One is listed in the NCBI dbSNP database, two occur at a codon where other SNPs are listed in the database and one is original (P38fs*37) (figure 1*c*).

In the collections of cell lines of GDSC and CCLE, 10 tumour cell lines bear a variation in *AGR3* coding sequence, among which five are common to the two databases. Three are listed in the NCBI dbSNP database, five occur at a codon where other SNPs are listed in the database and one is original (E103 K) and present in two cell lines SK-MEL-5 (human melanoma cell line) and SARC-9371 (human osteosarcoma cell line) (figure 1*d*).

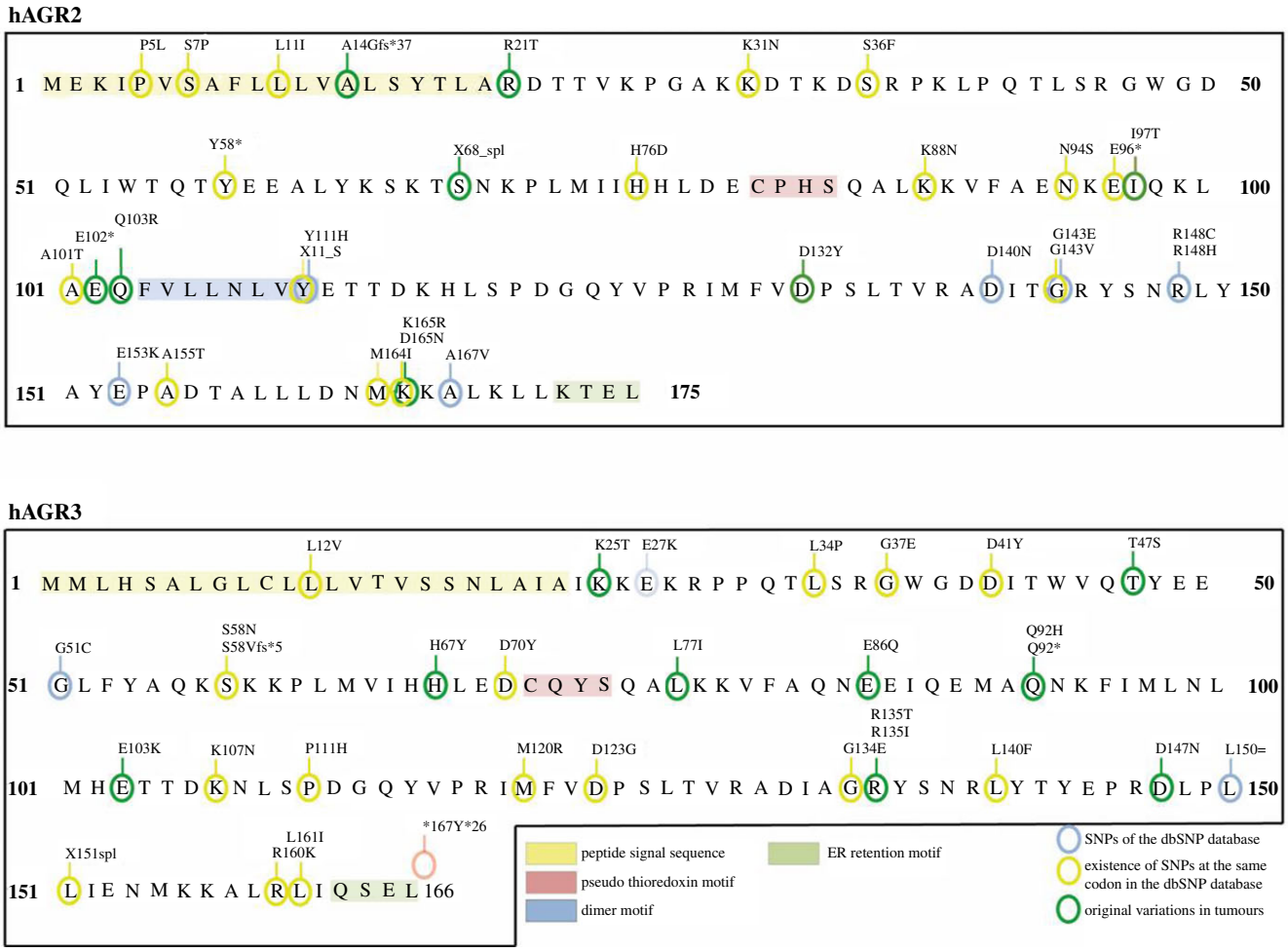

Figure 2. Localization of constitutional SNPs and tumour somatic SNVs in the sequence of AGR2 and AGR3 proteins. The functional domains of the proteins are indicated.

Figure 2 presents the localization of constitutional SNPs and tumour somatic SNVs extracted from the CCLE and TCGA databases, in the sequence of AGR2 and AGR3 proteins. With the exception of some known SNPs, none of them is present in the functional domains of the proteins.

3.4. *AGR2* and *AGR3* expression in TCGA and Cancer Cell Line Encyclopedia databases

Thanks to the cBioPortal facilities for TCGA and the CellMinerCDB portal for CCLE and GDSC, it was possible: (i) to compare the levels of *AGR2* and *AGR3* expressions in various tumour types; (ii) to identify *AGR2/3* expression variations in samples with SNV or CNV of these genes; (iii) to look for associations between *AGR2/3* expression and that of other genes in selected tumour types and (iv) to identify associations between *AGR2/3* and potentially oncogenic molecular features involving the whole exome, namely SNV and CNV. Since *AGR2* and *AGR3* expressions were highly correlated (figure 3) in all the TCGA tumour types studied as well as in the CCLE collection, we focused our interest on *AGR2* and simply indicated original features concerning *AGR3*.

(i) Expression levels

Among the 32 cancer types that are available in the PanCancer Atlas project of TCGA, only part of them displays a consistent expression of *AGR2* and *AGR3*. Non-epithelial cancers do not express this gene, and carcinomas from liver and kidney express these genes in a small part of the samples only, not always distinguishable from background noise; as a consequence, we concentrated our analysis on BLAD, CESC, UCEC, HNSC, STAD, ESCA, LUAD, LUSC, COAD-READ, PAAD, PRAD, BRCA and OVCA (figure 4*a*). In all cancer types, *AGR3* was expressed at a lower level than *AGR2*, and often not evaluable in samples from three cancer types: BLCA, HNSC and OVCA. The expression levels of the two genes were highly correlated in each cancer type. As a general feature, squamous cell carcinomas expressed *AGR2* and *AGR3* at a much lower level than adenocarcinomas (compare, for instance, LUAD with LUSC, ESCA with HNSC, CESC with UCEC).

In the CCLE collection, the levels of expression of *AGR2* and *AGR3* also vary considerably across cancer types. As a general feature, cancer cells derived from mesenchymal tissues express these genes at low levels, barely higher than background noise, whereas cancer cells derived from epithelial tissues have consistent expression levels. As a consequence, cancer cells from autonomic ganglia (neuroblastoma), bone (osteosarcoma and Ewing's sarcoma), central nervous system (glioma), haematopoietic and lymphoid tissue, pleura, skin (malignant melanoma) and soft tissue sarcomas were excluded from further analyses. Figure 4*b* presents the levels of expression of *AGR2* and *AGR3* in all other cancer cell line types. Cell lines derived

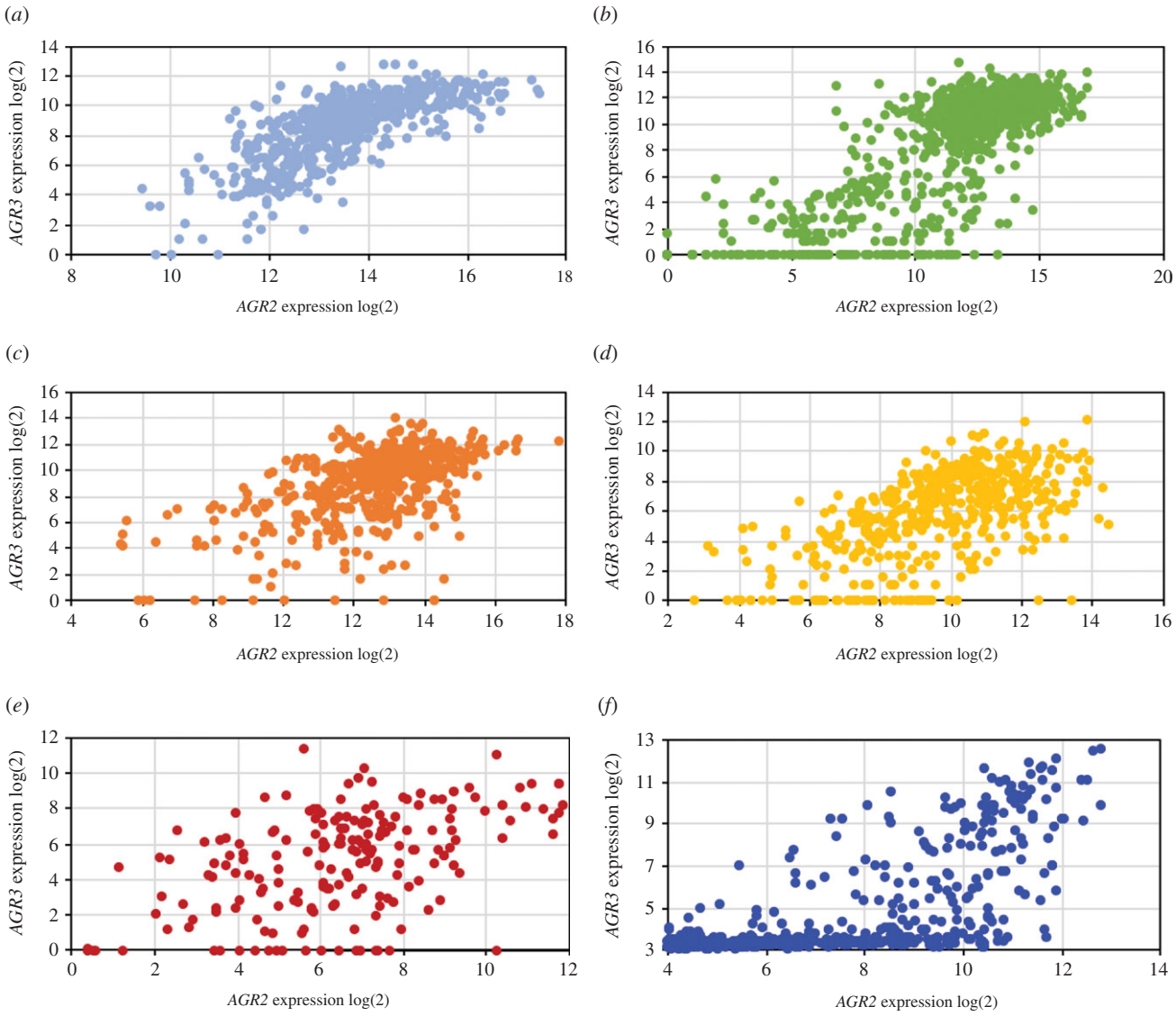

Figure 3. Correlation between *AGR2* and *AGR3* mRNA expressions in five TCGA tumour types ((*a*) COADREAD, (*b*) BRCA, (*c*) LUAD, (*d*) LUSC and (*e*) OVCA) and in the 634 carcinoma cell lines from the (*f*) CCLE.

from digestive tract cancers (with the exception of liver) had the highest expression levels, while cell lines derived from cancers of kidney, endometrium, ovary and thyroid carcinomas had the lowest expression levels.

(ii) *AGR2/3* expression variation

In the TCGA, the expression of *AGR2* and *AGR3* in genomic variants of these genes was not markedly different from that mentioned for the unaltered samples. Concerning CNV, looking for associations between *AGR2* or *AGR3* expression and copy number in five major tumour types (COADREAD, BRCA, LUAD, LUSC and OVCA), revealed no significant correlations between these two parameters (data not shown). In addition, when considering SNV, nonsense or frameshift mutations in gene sequence were not associated with loss of gene expression. Another way of analysing relationships between CNV and expression was to consider chromosome 7p losses in these cancer types; there were only three shallow 7p deletions in COADREAD out of 492 samples, not allowing comparisons, but in BRCA (66 samples with 7p loss out of 850 samples), there was significantly lower *AGR2* and *AGR3* expressions when chromosome 7p was lost ($p = 4.72 \times 10^{-13}$ and 2.7×10^{-9}, respectively); in LUAD and OVCA, barely significant lower expression values were noticed, and no significant results were obtained in LUSC (figure 5).

In the CCLE collection, there was no clear association between the presence of *AGR2/3* sequence variations in cell lines and the expression of these genes. In the MEL-JUSO melanoma cell line, the frameshift P38fs*37 AGR2 variation is accompanied by the lowest *AGR2* mRNA expression in melanoma cell lines, but only in the GDSC database. No other peculiarities could be discerned. By contrast, there was a significant correlation between *AGR2* expression and gene copy number ($p = 1.83 \times 10^{-9}$) when the whole set of cell lines was taken into consideration; however, this significance was lost when individual cancer types was studied in this respect, due to the relatively low number of cell lines in each cancer type.

(iii) Associations with cancer genes

In the TCGA, we also identified the genes that were co-expressed with *AGR2* or *AGR3* in five major tumour types (COADREAD, BRCA, LUAD, LUSC and OVCA). Each of them had a specific set of genes positively and negatively associated with that of *AGR2/3*. In table 1, we present the

Figure 4. mRNA expression levels of *AGR2* and *AGR3* extracted from databases. (*a*) Expression in 13 major cancer types from the TCGA database. (*b*) Expression in 17 cancer cell types from the CCLE database.

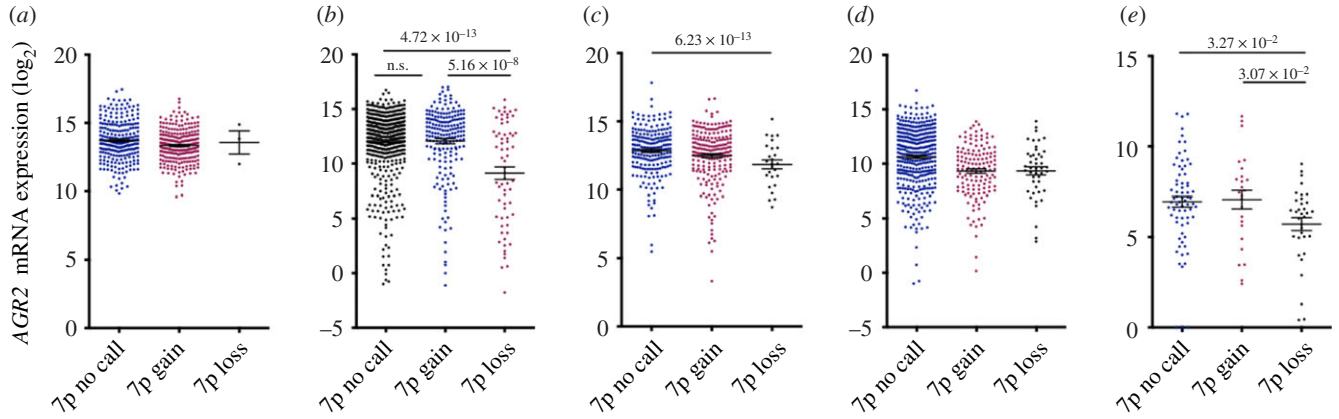

Figure 5. Association between chromosome 7p gains and losses and *AGR2* mRNA levels in five cancer types: (*a*) COADREAD, (*b*) BRCA, (*c*) LUAD, (*d*) LUSC and (*e*) OVCA.

significance level of the correlations between *AGR2* expression and that of selected representative genes. As a general feature, the expression of epithelial genes (e.g. *TJP3*, *TSPAN13*, *CLDN7* and *EPCAM*) was positively correlated with the expression of *AGR2/3* and the expression of mesenchymal genes (e.g. *VIM* and *MSN*) was negatively correlated, with specific correlations according to cancer type. The expression of the genes encoding the transcription factors involved in EMT (*SNAI*, *ZEB* and *TWIST* families) were often negatively correlated with *AGR2* expression, but

Table 1. List of selected genes whose expression is correlated with that of *AGR2* in five major cancer types of TCGA. Gene selection was arbitrary; we have selected genes representative of epithelial features in yellow (*TJP3*, *TSPAN13* and *CLDN7*), of mesenchymal features in green (*MSN* and *VIM*), of EMT in pink (*SNAI1*, *ZEB1* and *TWIST1*) as well as *TC2N* and *ESR1*, which are already known to be associated with *AGR2* in colon and breast carcinomas, respectively. In addition, genes encoding proteins known to interact with AGR2 [14,15] were studied (spotted in blue). Threshold for significance was set at 10^{-8} because of multiple testing, but we indicated *p*-values down to 10^{-4} to indicate trends at the limit of significance. *r*: Pearson coefficient of correlation; *p*: degree of significance of the correlation.

		COADREAD		BRCA		LUAD		LUSC		OVCA	
gene	cytoband	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
TSPAN13	7p21.1	0.746	2.89×10^{-94}	0.752	1.78×10^{-181}	0.486	3.09×10^{-31}	0.611	4.31×10^{-49}	0.599	5.87×10^{-21}
TJP3	19p13.3		n.s.	0.397	7.00×10^{-39}	0.405	2.69×10^{-21}	0.476	1.03×10^{-27}		n.s.
CLDN7	17p13.1	0.337	2.32×10^{-15}		n.s.	0.372	5.61×10^{-18}	0.279	9.36×10^{-10}		n.s.
MSN	Xq11.1		n.s.	−0.436	2.42×10^{-47}	−0.331	2.32×10^{-14}		n.s.		n.s.
VIM	10p13		n.s.	−0.227	4.60×10^{-13}	−0.262	2.38×10^{-9}		n.s.		n.s
SNAI1	20q13.2	−0.252	4.62×10^{-9}	−0.305	6.63×10^{-23}	−0.172	1.03×10^{-4}		n.s.		n.s.
ZEB1	10p11.2	−0.161	2.20×10^{-4}	−0.169	8.60×10^{-8}		n.s.	−0.195	2.15×10^{-5}		n.s.
TWIST1	7p21.2	−0.181	3.08×10^{-5}		n.s.		n.s.		n.s.		n.s.
TC2N	14q32.12	0.643	1.60×10^{-62}	0.234	7.69×10^{-14}	0.443	1.40×10^{-25}	0.386	5.68×10^{-18}		n.s.
ESR1	6q25.1-2		n.s.	0.557	5.21×10^{-82}		n.s.		n.s.		n.s.
TFF1	21q22.3	0.511	3.04×10^{-36}	0.632	7.41×10^{-112}	0.479	3.09×10^{-30}	0.559	1.05×10^{-39}	0.314	5.69×10^{-6}
TFF3	21q22.3	0.352	9.87×10^{-17}	0.610	3.32×10^{-102}	0.395	3.15×10^{-20}	0.542	6.16×10^{-37}	0.469	2.09×10^{-12}
FOXA1	14q12-q13	0.365	5.36×10^{-18}	0.564	1.99×10^{-84}		n.s.	0.564	1.61×10^{-40}		n.s.
DAG1	3p21	0.200	3.88×10^{-6}		n.s.		n.s.		n.s.		n.s.
LYPD3	19q13.31		n.s.	0.186	3.31×10^{-9}		n.s.		n.s.		n.s.
EGFR	7p12		n.s.	−0.418	2.17×10^{-43}	−0.243	3.43×10^{-8}		n.s.		n.s.
SPDEF	6p21.3	0.589	3.13×10^{-50}	0.483	3.49×10^{-59}	0.478	4.57×10^{-30}	0.583	8.40×10^{-44}		n.s.
FABP2	4q28-q31	0.423	3.99×10^{-24}				n.s.		n.s.		n.s.
AREGB	4q13.3	−0.275	1.52×10^{-10}	0.224	8.84×10^{-13}		n.s.		n.s.		n.s.
P4HB	17q25	0.306	8.20×10^{-13}	−0.183	6.26×10^{-9}	0,247	2.10×10^{-8}		n.s.		n.s.
HSP90B1	12q24	0.259	1.71×10^{-9}	−0.148	2.82×10^{-6}		n.s.		n.s.		n.s.
PDIA6	2p25.1		n.s.	−0.352	1.94×10^{-30}	0.203	4.27×10^{-6}		n.s.		n.s.
HSPG2	1p36.1-p34	−0.239	3.19×10^{-8}		n.s.		n.s.		n.s.		n.s.
AGRN	1p36.33		n.s.		n.s.		n.s.		n.s.		n.s.
DMD	Xp21.2	−0.392	1.19×10^{-20}	−0.322	2.33×10^{-25}		n.s.		n.s.		n.s.

(*Continued*.)

Table 1. (*Continued.*)

		COADREAD		BRCA		LUAD		LUSC		OVCA	
gene	cytoband	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
UTRN	6q24		n.s.	0.144	4.84×10^{-6}	−0.250	1.24×10^{-8}		n.s.		n.s.
LAMA2	6q22-q23		n.s.	0.151	1.75×10^{-6}		n.s.		n.s.		n.s.
CALR	19p13.11		n.s.	−0.268	8.07×10^{-18}		n.s.		n.s.		n.s.
KDELR1	19q13.3	0.383	8.41×10^{-20}		n.s.	0.290	3.29×10^{-11}		n.s.		n.s.
KDELR2	7p22.1	0.377	3.99×10^{-19}		n.s.	0.374	3.61×10^{-18}	0.221	1.38×10^{-6}	0.312	6.37×10^{-6}
TMED2	12q24.31	0.409	1.39×10^{-22}		n.s.	0.204	3.96×10^{-6}		n.s.		n.s.
MUC1	1q21	0.361	1.38×10^{-17}	0.315	2.44×10^{-24}	0.321	1.57×10^{-13}	0.485	6.31×10^{-29}		n.s.
MUC2	11p15.5	0.551	6.27×10^{-43}	0.172	5.23×10^{-8}	0.265	1.58×10^{-9}	0.276	1.40×10^{-9}		n.s.
MUC5B	11p15.5	0.338	1.69×10^{-15}		n.s.	0.474	1.40×10^{-29}	0.379	2.08×10^{-17}		n.s.
CD59	11p13		n.s.	−0.256	2.77×10^{-16}		n.s.		n.s.		n.s.
EPCAM	2p21		n.s.	−0.129	4.55×10^{-5}		n.s.		n.s.		n.s.
UNG	12q23-q24	0.228	1.28×10^{-7}		n.s.		n.s.		n.s.		n.s.
TAB2	6q25.1		n.s.		n.s.		n.s.		n.s.		n.s.
RUVBL2	19q13.3		n.s.		n.s.		n.s.		n.s.		n.s.
FGF2											
VEGFA	6p12	−0.271	2.70×10^{-10}	−0.294	2.98×10^{-21}		n.s.		n.s.		n.s.
HIF1A	14q23.2	0.369	2.25×10^{-18}	0.262	5.23×10^{-17}	−0.225	3.22×10^{-7}		n.s.		n.s.

Table 2. List of genes whose expression is highly positively or negatively correlated with that of *AGR2* in the whole set of cell lines of the CCLE collection. Gene selection was arbitrary; we have selected genes representative of epithelial features in yellow (*TJP3*, *TSPAN13* and *CLDN7*), of mesenchymal features in green (*MSN* and *VIM*), of EMT in pink (*SNAI1*, *ZEB1* and *TWIST1*) as well as *TC2N* and *ESR1*, which are already known to be associated with *AGR2* in colon and breast carcinomas, respectively. In addition, genes encoding proteins known to interact with AGR2 [14,15] were studied (spotted in blue). Threshold for significance was set at 10^{-8} because of multiple testing, but *p*-values down to 10^{-4} were assumed to indicate trends at the limit of significance. *r*: Pearson coefficient of correlation; *p*: degree of significance of the correlation.

		AGR2		AGR3	
gene		*r*	*p*	*r*	*p*
TSPAN13	7p21.1	0.507	1.03×10^{-68}	0.337	6.95×10^{-29}
TJP3	19p13.3	0.736	4.94×10^{-177}	0.565	1.71×10^{-88}
CLDN7	17p13.1	0.719	2.05×10^{-165}	0.480	6.61×10^{-61}
MSN	Xq11.1	−0.521	3.39×10^{-73}	−0.494	7.62×10^{-65}
VIM	10p13	−0.581	1.19×10^{-94}	−0.448	2.67×10^{-52}
SNAI1	20q13.2	−0.181	4.89×10^{-9}	−0.143	4.05×10^{-6}
ZEB1	10p11.2	−0.592	4.20×10^{-99}	−0.428	2.05×10^{-47}
TWIST1	7p21.2	−0.285	8.20×10^{-21}	−0.187	1.32×10^{-9}
TC2N	14q32.12	0.727	4.95×10^{-171}	0.556	3.53×10^{-85}
ESR1	6q25.1-2	0.196	2.08×10^{-10}	0.170	3.85×10^{-8}
TFF1	21q22.3	0.701	9.50×10^{-154}	0.592	5.15×10^{-99}
TFF3	21q22.3	0.460	2.89×10^{-55}	0.542	4.37×10^{-80}
FOXA1	14q12-q13	0.669	2.79×10^{-135}	0.432	2.67×10^{-48}
DAG1	3p21	0.249	4.02×10^{-16}	0.123	6.79×10^{-5}
LYPD3	19q13.31	0.428	1.99×10^{-47}	0.166	7.04×10^{-8}
EGFR	7p12	0.270	9.01×10^{-19}	n.s.	
SPDEF	6p21.3	0.471	2.69×10^{-58}	0.340	1.77×10^{-29}
FABP2	4q28-q31	n.s.		n.s.	
AREGB	4q13.3	0.430	5.47×10^{-48}	0.269	1.26×10^{-18}
P4HB	17q25	n.s.		n.s.	
HSP90B1	12q24	n.s.		n.s.	
PDIA6	2p25.1	n.s.		n.s.	
HSPG2	1p36.1-p34	n.s.		n.s.	
AGRN	1p36.33	0.328	1.92×10^{-27}	n.s.	
DMD	Xp21.2	−0.174	1.66×10^{-8}	n.s.	
UTRN	6q24	n.s.	−0.147	1.9×10^{-6}	
LAMA2	6q22-q23	n.s.		n.s.	
CALR	19p13.11	n.s.		−0.134	1.42×10^{-5}
KDELR1	19q13.3	0.193	4.07×10^{-10}	n.s.	
KDELR2	7p22.1	0.191	6.32×10^{-10}	n.s.	
TMED2	12q24.31	0.141	5.28×10^{-6}	n.s.	
MUC1	1q21	0.505	3.87×10^{-68}	0.382	2.01×10^{-37}
MUC2	11p15.5	0.393	1.16×10^{-39}	0.452	3.21×10^{-53}
MUC5B	11p15.5	0.387	2.73×10^{-38}	0.296	2.25×10^{-22}
MUC5AC	11p15.5	0.401	2.69×10^{-41}	0.266	2.68×10^{-18}
CD59	11p13	n.s.		−0.124	6.28×10^{-5}
EPCAM	2p21	0.641	3.46×10^{-121}	0.427	2.93×10^{-47}
UNG	12q23-q24	n.s.		0.144	3.42×10^{-6}
TAB2	6q25.1	n.s.		n.s.	
RUVBL2	19q13.3	n.s.		n.s.	
FGF2	4q26	−0.348	6.07×10^{-31}	−0.305	8.17×10^{-24}
VEGFA	6p12	n.s.		n.s.	
HIF1A	14q23.2	n.s.		−0.132	2.07×10^{-5}

this generally remains slightly below the level of significance we have chosen for 1% risk. It was remarkable that *ESR1* (oestrogen receptor) was highly significantly associated with *AGR2* in BRCA, but not in other malignancies. Similarly, *FOXA1* and *AGR2* expressions were correlated in BRCA, COADREAD and LUSC, but not in LUAD or OVCA. The expression of genes encoding mucins (*MUC1*, *MUC2* and *MUC5A*) or involved in mucosa protection (*TFF1* and *TFF3*) were positively correlated with *AGR2* expression in most tumour types. In addition, genes encoding proteins known to interact with AGR2 [14,15] were studied. There was a clear specificity in their co-expression pattern with *AGR2*: some genes were co-expressed in colon adenocarcinoma, others in breast adenocarcinoma, etc. It should be mentioned that *EGFR*, *CD59* and *VEGFA* gene expressions were, in contrast, negatively correlated with *AGR2* expression in breast adenocarcinomas.

In the CCLE collection as in TCGA, the genes significantly positively co-expressed with *AGR2* and *AGR3* in the whole set of 1036 cell lines of the CCLE were mostly epithelial genes, according to the list established by Kohn *et al*. [16]. Conversely, the expression of mesenchymal genes was inversely correlated with *AGR2* and *AGR3* gene expressions (table 2). This is not surprising, in view of the fact that these genes were expressed to a much higher level in epithelial tissue-derived cell lines than in mesenchymal tissue-derived ones. However, when cancer types were studied independently (namely breast, colorectal, lung and ovarian adenocarcinomas), the same positive correlation between *AGR2* and *AGR3* expressions and those of epithelial genes was maintained, as well as the negative correlation between *AGR2* and *AGR3* expressions and those of mesenchymal genes (data not shown). In addition to epithelial/mesenchymal genes, some interesting associations could be identified: *AGR2* and *AGR3* mRNA levels are positively associated with high significance with *FOXA1* expression, *TFF1/2/3* and *ESR1*. It is interesting to note that the expressions of genes encoding the transcription factors of EMT are negatively correlated with those encoding *AGR2* and *AGR3*: *ZEB1/2* with a very high significance, *TWIST1/2* and *SNAI1/2* with lower *p*-values, but still highly significant. The genes encoding AGR2 protein interactants were positively co-expressed with *AGR2* for some of them such as *KDELR*, *TMED2*, *DAG1*, *LYPD3* and *MUC1/2/5AC/5B*) but negatively correlated for others such as *DMD* or *FGF2*. For AGR3 interactants, a distinct pattern was observed, with positive co-expressions with *DAG1*, *LYPD3*, *MUC1/2/5AC/5B* or *UNG*, and negative correlations with *UTRN*, *CALR*, *CD59*, *FGF2* or *HIF1A*.

(iv) Association with oncogenic features

We wanted to know whether some oncogenic alterations in various pathways were associated with *AGR2* and *AGR3* expressions. Indeed, the oncogenic status of these genes is not clear and the possible association with established oncogenic features could shed some light upon this status. In this respect, we have selected in the TCGA the five tumour types (COADREAD, BRCA, LUAD, LUSC and OVCA) and the set of genes that are the most commonly mutated in these malignancies (*KRAS*, *APC*, *TP53*, *SMAD4*, *BRAF* and *PIK3CA* for COADREAD; *TP53*, *PIK3CA*, *BRCA1/2* and *PTEN* for BRCA; *KRAS* and *TP53* for LUAD and LUSC; *TP53*, *BRCA1/2* and *RB1* for OVCA).

Concerning COADREAD (figure 6*a*), it appeared that the presence of a *KRAS* mutation in a tumour was not associated with *AGR2* expression, whereas the presence of *APC* or *TP53* mutation was negatively associated with *AGR2* expression, and the presence of *SMAD4*, *BRAF* or *PIK3CA* mutation was positively associated with *AGR2* expression. It was the same for *MTOR*, *MLH1* and *MSH2* mutations (data not shown). Very similar associations were found between *AGR3* expression and oncogenic mutations in COADREAD, the only difference being the exact level of significance (data not shown).

Concerning BRCA (figure 6*a*), the same was observed for *TP53* and *PIK3CA*: negative association between *AGR2* expression and *TP53* mutations, positive association for *PIK3CA*; no significant association was found between *PTEN* or *BRCA1/2* mutations and *AGR2* expression. Concerning lung tumours (figure 6*a*), there was no significant association between *KRAS* mutations and *AGR2* expression, while there was, as in COADREAD and BRCA, a negative association between *TP53* mutation and *AGR2* expression in LUAD samples (but this was not the case in LUSC samples). No association between *AGR2* expression and oncogenic mutations were noticed in OVCA (figure 6*a*). There again, similar associations were found between *AGR3* expression and oncogenic mutations in these cancer types (data not shown).

In the CCLE taken as a whole, an increase in *AGR2* and *AGR3* expressions was systematically associated with several oncogenic mutations (electronic supplementary material, table S3). As an illustration, we present in figure 6*b* the significative associations existing between the expressions of *AGR2* and the presence of representative oncogene and TSG mutations, namely those occurring in *KRAS*, *SMAD4*, *APC* and *PIK3CA*. However, this significance was lost when individual cancer types were studied in this respect, due to the relatively low number of cell lines in each cancer type.

Looking further into the associations that could be found between *AGR2* or *AGR3* gene expression and oncogenic features, we also analysed the relationships between *AGR2* and *AGR3* expressions and the CNV of a set of oncogenes and TSG that are activated in cancers by copy gains (including amplifications) and losses (including deletions), respectively.

In the COADREAD samples of TCGA, a significant correlation is obvious between *AGR2* expression and *FOXA1* expression, in relation to the correlation observed between *AGR2* gene expression and *FOXA1* copy number. Also, a significant change in *AGR2* gene expression accompanied several CNV features known to drive colorectal cancers, especially those involved in cell cycle control (*TP53*, *FBXW7*, *RB1*, *CDC27* and *AURKA*), in WNT signalling (*APC*, *WNT4*, *FZD3* and *AJUBA*) and others (*SMAD4* and *SMAD2*). Figure 7*a* presents a selection of representative associations and electronic supplementary material, table S4A a list of significant associations (down to $p < 10^{-4}$) between oncogene or TSG gene copy numbers and *AGR2* expression in COADREAD. Some oncogenes and TSG of this list are not known to be frequently altered in colorectal cancers; it should be noticed that they belong to 14q or 18q chromosome arms, which, respectively, harbour *FOXA1* and *SMAD2/4*, suggesting that this correlation might in fact be related to the same event of gain or loss of a whole chromosome arm and has no functional meaning. Very similar results were obtained with *AGR3* expression (data not shown), slight differences

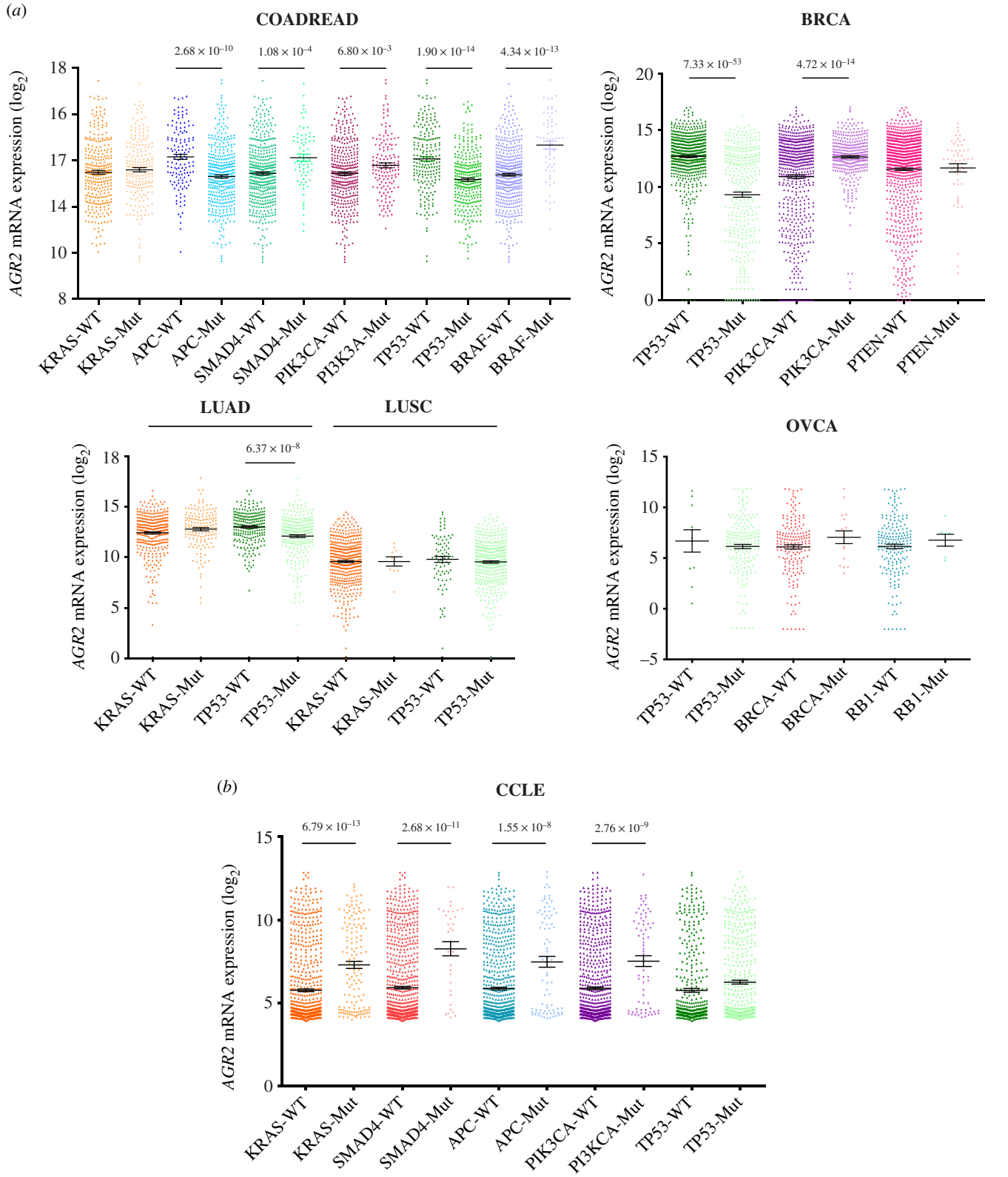

Figure 6. *AGR2* gene expression levels are associated with oncogene and TSG mutation status in different cancer types from the TCGA database and the CCLE database. (*a*) Box plots displaying *AGR2* expression levels in COADREAD, BRCA, LUAD, LUSC and OVCA tumours with a mutation in genes *KRAS*, *APC*, *SMAD4*, *PIK3CA*, *BRAF*, *TP53*, *PTEN*, *BRCA* and *RB1*. *p*-Values were assessed using Student's *t*-test. (*b*) Box plots displaying *AGR2* expression levels in cancer cell lines with a mutation in genes *KRAS*, *SMAD4*, *APC*, *PIK3CA* and *TP53*. *p*-Values were assessed using Student's *t*-tests.

occurring for the genes that were just below or just above the limit of significance chosen (10^{-4}).

In the BRCA samples of TCGA, we also noticed a significant relationship between *AGR2* expression and *FOXA1* gene copy number, as well as several cancer gene copy numbers localized at 14q such as *NFKBIA*, *SAV1*, *CHD8* or *AJUBA*, which are not known as driver oncogenes or TSG in breast cancer (figure 7*b*; electronic supplementary material, table S4A). A highly significant association of *AGR2* expression was seen with *APC*, *JUN*, *CCNE1*, *ERBB2*, *MDM2* or *RAD21* copy numbers, which may have more functional implications. By contrast, copy numbers of *RB1* or *TP53* were not associated with *AGR2* expression, showing that the relationship between *AGR2* expression and oncogenic features in breast cancer is

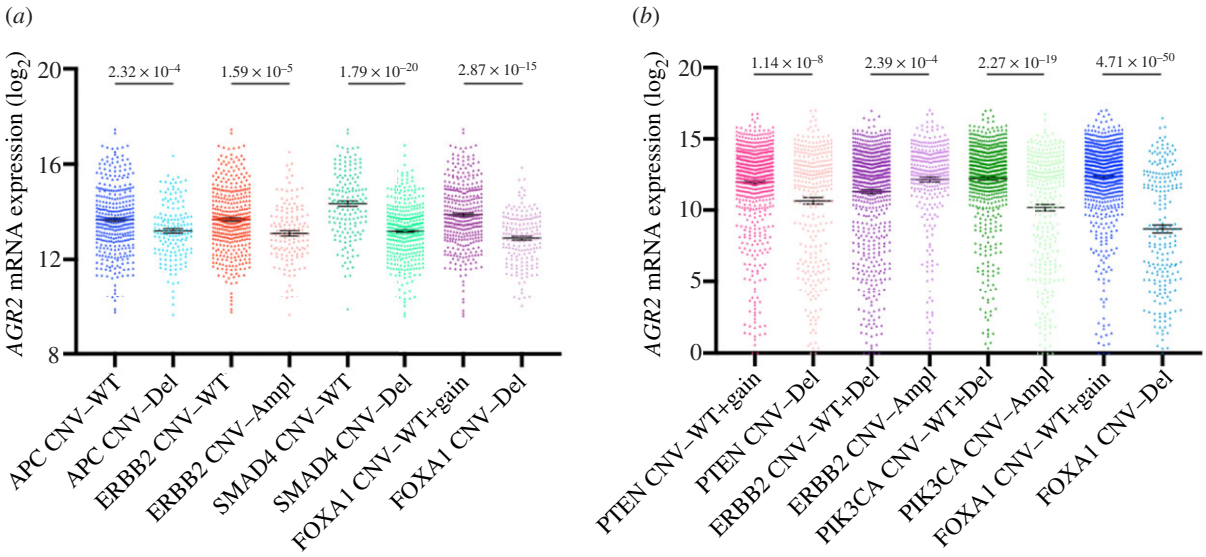

Figure 7. *AGR2* gene expression levels are associated with oncogene CNV in (*a*) COADREAD and (*b*) BRCA samples from the TCGA database. Only some examples are given, concerning principally genes known as oncogenic drivers in these cancer types; see electronic supplementary material, table S4A for more details.

certainly complex and requires more in-depth analysis. Similar results were obtained with *AGR3* gene expression (data not shown), the differences between the two genes appearing to be marginal. No significant relationship between *AGR2* or *AGR3* expression and oncogene or TSG CNV was observed in LUAD, LUSC and OVCA (data not shown).

In the CCLE collection, CNV were not classified as gains or losses but copy numbers were given; we observed positive correlations between *AGR2* expression and gene copy numbers of several oncogenes such as *FOXA1*, *ERBB2*, *CCND1* and *MYC*, whereas a negative correlation was found between *AGR2* expression and copy numbers of several TSG such as *SMAD4* (electronic supplementary material, table S4B). However, this general trend was not constant over the whole set of oncogenes and GST. Similar results were obtained for *AGR3* with a lower number of cancer genes whose CNVs were associated with *AGR3* than with *AGR2* expression. In both cases, there was an overrepresentation of genes located on the 14q and 19p chromosome arms, which may indicate that the association concerns a whole chromosome arm and not specific cancer genes. There again, this significance was lost when individual cancer types of the CCLE were studied, due to the relatively low number of cell lines in each cancer type.

(v) Pattern of AGR2 extinction in the Cancer Cell Line Encyclopedia as studied by clustered regularly interspaced short palindromic repeats (CRISPR) screens

The Broad Institute has set-up CRISPR screens to study vulnerability targets through gene extinction screens in 769 cell lines of the CCLE collection [17]. It integrates data obtained by knocking-out each gene of the genome to analyse its consequences on cell viability and proliferation (regrouped as 'cell fitness'). A friendly user access has been made available by NCI on the CellMinerCDB site. The pattern of *AGR2* and *AGR3* gene extinction over cell lines can therefore be extracted and compared to the extinction pattern of other genes. The pattern of cell fitness alterations associated with *AGR2* and *AGR3* extinction are highly correlated ($r = 0.331$, $p = 4.58 \times 10^{-21}$) and did not reveal any preferential vulnerability towards a given cancer type represented in the cell line panels of the CCLE. No preferential effect was seen in epithelial versus mesenchymal cell lines or in adenocarcinoma versus squamous cell carcinomas, as was the case for expression data. The mean values of cell fitness alteration over 769 cell lines after *AGR2* and *AGR3* extinction are 1.003 ± 0.088 and 1.090 ± 0.075, whereas the same parameter is largely lower than 1 when oncogenes are knocked out (e.g. 0.305 for *MYC*, 0.685 for *CDK4*, 0.778 for *MDM2*, 0.701 for *KRAS*, 0.414 for *MTOR*) and higher than 1 when TSG are knocked out (1.411 for *TP53*, 1.792 for *PTEN*, 1.170 for *RB1*, 1.227 for *CDKN1A*, 1.136 for *BAX*), all values being highly significantly different from those of *AGR2* and *AGR3* (p-values ranging from 10^{-9} to 0). As a consequence, *AGR2* and *AGR3* appear in this respect as 'neutral' genes, whose knock-outs have very moderate influence on cell fitness. However, when building a heat map with normalized ranked values of cell fitness alterations induced by 10 major oncogenes and 10 major TSG (figure 8), a good segregation between oncogenes and TSG clearly appears, with *AGR2* and *AGR3* segregating together among oncogenes. We also evaluated the correlations that could exist between the extinction patterns of *AGR2* and *AGR3* to those of other genes (electronic supplementary material, table S5). It appeared that, among the 60 genes presenting a pattern of extinction significantly correlated (down to 10^{-6}) with that of *AGR2*, 44 are localized on chromosome arm 7p, indicating a topological rather than a functional relationship. Whereas there was no oncogene or TSG among the genes located on chromosome arm 7p, there were three oncogenes (*KLF5*, *TCF7L2* and *CTNNB1*) and one TSG (*SOX9*) located in other chromosome arms, all presenting an extinction pattern similar to that of *AGR2* among the CCLE collection (positive correlation) and playing a role in transcription. *AGR3* displayed a distinct pattern of gene extinction, with only nine genes not located on 7p chromosome arm out of 105 whose extinction pattern was correlated with that of this gene, which does not bring information on the functional relationship.

4. Discussion

The question underlying the development of this work is whether *AGR2* and *AGR3* can be considered as playing a

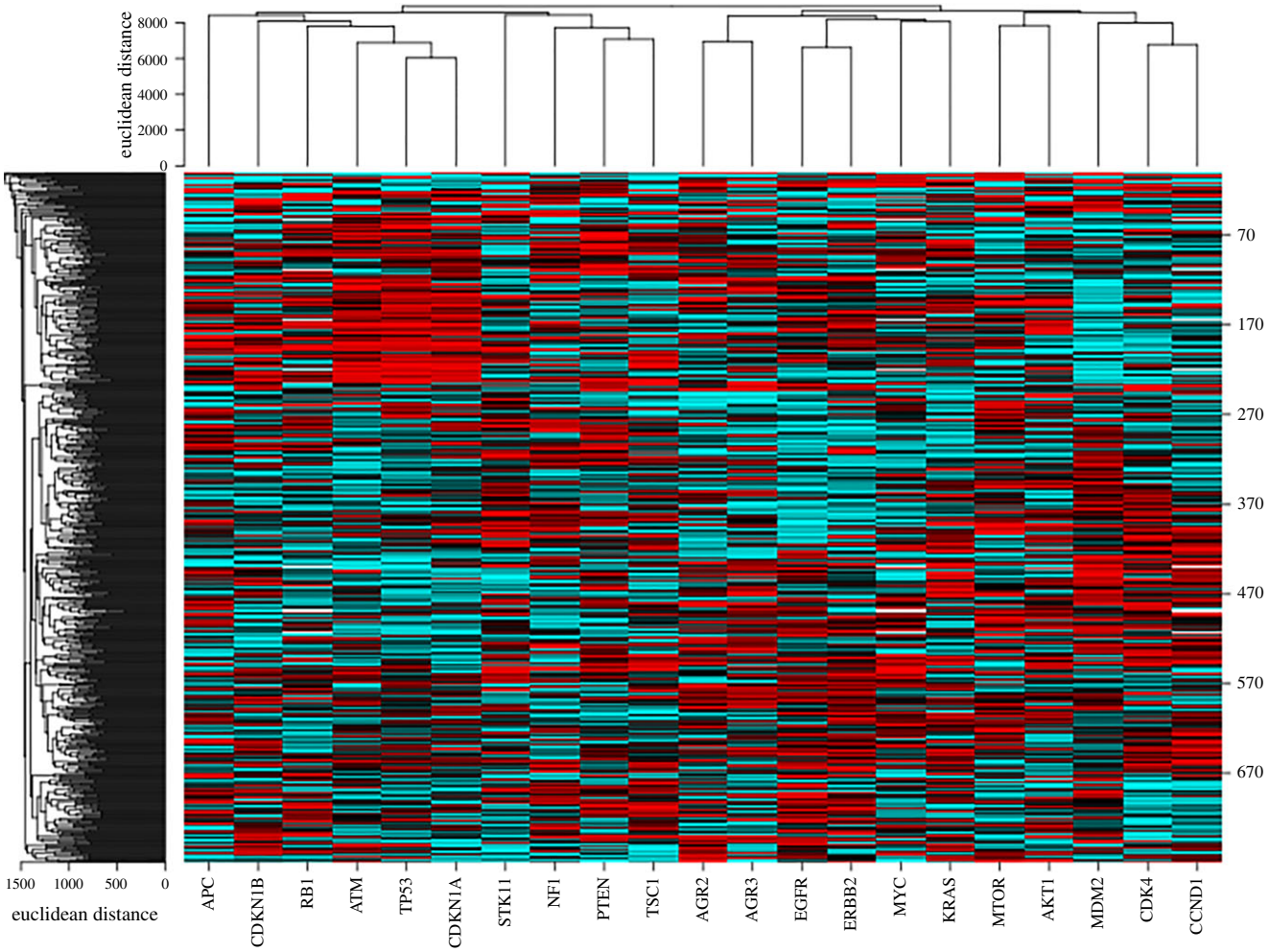

Figure 8. Clustering of cell fitness alterations in various oncogenes and TSG. Clustering was performed using CIMminer on the NCI Genomics and Pharmacology Facility (https://discover.nci.nih.gov/cimminer/oneMatrix.do). Fitness values were downloaded and normalized by ranking before building the heat map.

major role in oncogenesis and progression of cancers; in other terms, whether they can be considered as oncogenes and/or TSG. Known polymorphisms in *AGR2* and *AGR3* sequences as well as variations encountered at a known polymorphic site are not likely to confer oncogenic properties to AGR2 or AGR3 proteins. Only five SNV in *AGR2* and six in *AGR3* sequences deserve some attention: those that are supposed to result in a truncated or different protein (nonsense, frameshift variations). These variations are not recurrent and cannot be considered as oncogenic variations since the tumours and cell lines bearing these variations do not behave differently than the others in terms of *AGR2/3* gene expression.

Similarly, the *AGR2/3* CNVs encountered in TCGA did not seem to affect *AGR2/3* gene expression. However, we observed a significant negative correlation between *AGR2* expression and chromosome 7p deletions in BRCA, which could be expected since this is the chromosome location of *AGR2/3*. In the CCLE, when the whole set of cell lines was taken in consideration, there was a significant correlation for both genes between *AGR2* gene copy number and expression. When ranking the copy number values from highest to lowest values, there was no preferential contribution of the cancer types to presenting high or low *AGR2/3* copy numbers. The overall conclusion of these explorations of *AGR2* and *AGR3* genomic variations in tumours and cancer cell lines is that it is quite unlikely that they could behave as *bona fide* oncogenes or TSG.

The associations we noticed between *AGR2* gene expression and that of a large series of genes reveal in contrast several important features in relation to oncogenesis and cancer progression. A common general feature is the fact that both genes appear as epithelial markers, in TCGA different cancer types as well as in the whole set of CCLE cell lines and in cell lines of different cancer types. In addition, there was a negative correlation between *AGR2* expression and that of the main transcription factors of epithelial-to-mesenchymal transition. Another point of interest is that some of the known partners of AGR2 and AGR3 proteins are co-expressed with them, but this is not a general feature, and concerns the different cancer types in a specific way, with the exception of mucins whose expression appears to be strongly positively correlated to that of *AGR2/3* in all cancer types, in agreement with their known functional association.

It appears from our explorations that AGR2 and AGR3 are connected to the cancer phenotype. In clinical samples as well as in CCLE cancer cell lines, the presence of oncogenic mutations and CNVs in various driver genes is associated with variations in *AGR2/3* expression, depending both on the cancer gene and the tumour type.

AGR2 gene extinction in CRISPR screens of the CCLE is followed by a mitigate, low-amplitude consequence on cell survival and proliferation, with a null average value, whereas oncogene or TSG extinction is followed by significant effects, either in favour (oncogenes) or to the detriment (TSG) of cell

fitness. The *AGR2* gene extinction pattern appears to be correlated with that of several cancer genes, reinforcing the participation of this protein in cancer phenotypes.

It is commonly assumed that somatic mutations drive the multi-step tumour development process. Although *AGR2* and *AGR3* genes present no recurrent mutations, both proteins are often overexpressed, have non-canonical localizations (extracellular, cytosol) and are associated with different tumour processes such as differentiation, proliferation, migration, invasion and metastasis, in almost all epithelial cancer types. Cancer follows an evolutionary trajectory, characterized by stepwise acquisition of mutations that allow the tumour cells to increase their fitness, from the pre-cancer lesion to tumour metastasis. However, the non-genetic gain-of-function alterations, acquired by overexpression and non-canonical localizations of AGR2 and AGR3 proteins, may be pivotal for tumour development and progression.

Thus, AGR2 and AGR3 proteins appear as common non-genetic evolutionary factors in the process of human tumorigenesis. Complex and dynamic adaptation mechanisms and evolutionary processes take place during the process of human epithelial tumorigenesis (tumour initiation, development and progression). Although cancer has been considered mainly, for decades, as a process governed by genetic mechanisms, it is becoming clearer that non-genetic mechanisms may also play an important role in cancer progression. Tumours are constantly evolving, displaying highly variable patterns resulting in extremely complex genetic and non-genetic phenotypic diversification. Therefore, when dealing with such a complex system that is barely understood, common hallmarks are rare. Thus, it is of crucial importance to identify and investigate the functional role of novel unexpected common hallmarks that will undoubtedly aid the development of therapeutic approaches. Overexpression and non-canonical localizations of AGR2 and AGR3 may reflect a non-genetic evolutive process, which is indeed a common feature in human epithelial tumorigenesis. We believe that further in-depth functional studies of cancer development from an AGR2/3 expression and localization perspective may enable us to progress in the understanding of the epithelial cancer evolutionary framework, which might result in the discovery of new original therapeutic perspectives.

Data accessibility. This article has no additional data.

Authors' contributions. D.F.: conceptualization, formal analysis, investigation, project administration, visualization and writing—original draft; I.V.: data curation and visualization; E.C.: validation, visualization and writing—original draft; F.D.: conceptualization, data curation, funding acquisition, investigation, project administration, resources, software and writing—original draft; J.R.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation and writing—original draft.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. F.D. was supported by a grant from the 'Fondation ARC pour la recherche sur le cancer', F.D. and I.V. from the 'Site de recherche intégrée sur le cancer de Bordeaux' (SIRIC Brio). This work has been supported by grants from the 'Région Nouvelle-Aquitaine' (D.F. and F.D.), by the 'Agence Nationale de la Recherche' (ANR) (D.F.) and by the 'Ligue contre le cancer Gironde' (F.D.). This work was also funded by grants from the 'Institut National du Cancer' (INCa, PLBIO), 'Fondation pour la Recherche Médicale' (FRM, DEQ20180339169) and 'Agence Nationale de la Recherche' (ANR; ERAAT) to E.C.

Acknowledgements. We gratefully acknowledge the members from ARTiSt group for their critical remarks.

References

1. Lee E, Lee DH. 2017 Emerging roles of protein disulfide isomerase in cancer. *BMB Rep.* **50**, 401–410. (doi:10.5483/BMBRep.2017.50.8.107)
2. Delom F, Lejeune PJ, Vinet L, Carayon P, Mallet B. 1999 Involvement of oxidative reactions and extracellular protein chaperones in the rescue of misassembled thyroglobulin in the follicular lumen. *Biochem. Biophys. Res. Commun.* **255**, 438–443. (doi:10.1006/bbrc.1999.0229)
3. Delom F, Mallet B, Carayon P, Lejeune PJ. 2001 Role of extracellular molecular chaperones in the folding of oxidized proteins. Refolding of colloidal thyroglobulin by protein disulfide isomerase and immunoglobulin heavy chain-binding protein. *J. Biol. Chem.* **276**, 21 337–21 342. (doi:10.1074/jbc.M101086200)
4. Higa A, Mulot A, Delom F, Bouchecareilh M, Nguyen DT, Boismenu D, Wise MJ, Chevet E. 2011 Role of pro-oncogenic protein disulfide isomerase (PDI) family member anterior gradient 2 (AGR2) in the control of endoplasmic reticulum homeostasis. *J. Biol. Chem.* **286**, 44 855–44 868. (doi:10.1074/jbc.M111.275529)
5. Chevet E, Fessart D, Delom F, Mulot A, Vojtesek B, Hrstka R, Murray E, Gray T, Hupp T. 2013 Emerging roles for the pro-oncogenic anterior gradient-2 in cancer development. *Oncogene* **32**, 2499–2509. (doi:10.1038/onc.2012.346)
6. Fessart D, Robert J, Hartog C, Chevet E, Delom F, Babin G. 2021 The anterior GRadient (AGR) family proteins in epithelial ovarian cancer. *J. Exp. Clin. Cancer Res.* **40**, 271. (doi:10.1186/s13046-021-02060-z)
7. Obacz J, Takacova M, Brychtova V, Dobes P, Pastorekova S, Vojtesek B, Hrstka R. 2015 The role of AGR2 and AGR3 in cancer: similar but not identical. *Eur. J. Cell Biol.* **94**, 139–147. (doi:10.1016/j.ejcb.2015.01.002)
8. Fessart D *et al.* 2016 Secretion of protein disulphide isomerase AGR2 confers tumorigenic properties. *eLife* **5**, e13887. (doi:10.7554/eLife.13887)
9. Fessart D *et al.* 2021 Extracellular AGR2 triggers lung tumour cell proliferation through repression of p21(CIP1). *Biochim. Biophys. Acta Mol. Cell Res.* **1868**, 118920. (doi:10.1016/j.bbamcr.2020.118920)
10. Obacz J *et al.* 2019 Extracellular AGR3 regulates breast cancer cells migration via Src signaling. *Oncol. Lett.* **18**, 4449–4456. (doi:10.3892/ol.2019.10849)
11. Sicari D *et al.* 2021 Reflux of endoplasmic reticulum proteins to the cytosol inactivates tumor suppressors. *EMBO Rep.* **22**, e51412. (doi:10.15252/embr.202051412)
12. Rajapakse VN *et al.* 2018 CellMinerCDB for integrative cross-database genomics and pharmacogenomics analyses of cancer cell lines. *iScience* **10**, 247–264. (doi:10.1016/j.isci.2018.11.029)
13. Luna A *et al.* 2021 CellMiner cross-database (CellMinerCDB) version 1.2: exploration of patient-derived cancer cell line pharmacogenomics. *Nucleic Acids Res.* **49**, D1083–D1093. (doi:10.1093/nar/gkaa968)
14. Delom F, Mohtar MA, Hupp T, Fessart D. 2020 The anterior gradient-2 interactome. *Am. J. Physiol. Cell Physiol.* **318**, C40–C47. (doi:10.1152/ajpcell.00532.2018)
15. Szklarczyk D *et al.* 2021 The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612. (doi:10.1093/nar/gkaa1074)
16. Kohn KW, Zeeberg BM, Reinhold WC, Pommier Y. 2014 Gene expression correlations in human cancer cell lines define molecular interaction networks for epithelial phenotype. *PLoS ONE* **9**, e99269. (doi:10.1371/journal.pone.0099269)
17. Ghandi M *et al.* 2019 Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569**, 503–508. (doi:10.1038/s41586-019-1186-3)