Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard <u>Terms & Conditions</u> and the <u>Ethical guidelines</u> still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



www.rsc.org/molecularbiosystems

# Engineering a synthetic cell panel to identify signalling components reprogrammed by the cell growth regulator Anterior Gradient-2

Terry A. Gray<sup>1</sup>, Khaldoon Alsamman<sup>2</sup>, Euan Murray<sup>1</sup>, Andrew H. Sims<sup>3</sup>, & Ted Hupp<sup>1</sup>

From the University of Edinburgh, Institute of Genetics and Molecular Medicine, <sup>1</sup>Cell Signalling Unit, p53 Signal Transduction Laboratories; <sup>2</sup> Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, University of Dammam, Dammam, Saudi Arabia; and <sup>3</sup>Applied Bioinformatics of Cancer, Breakthrough Research Unit, Edinburgh Cancer Research Centre, Western General Hospital, Edinburgh, United Kingdom, EH4 2XR

*Correspondence to: Prof. T. Hupp, Cell Signalling Unit, p53 Signal Transduction Laboratories, Edinburgh Cancer Research Centre, Western General Hospital, Edinburgh, United Kingdom, EH4 2XR. E-mail: ted.hupp@ed.ac.uk* Tel: (+44) 0131 777 3500.

Running Title: Reprogramming of signalling pathways by AGR2

#### ABSTRACT

AGR2 forms an ER-resident signalling axis in cell development, limb regeneration, and in human diseases like asthma and cancer, yet molecular mechanisms underlying its effects remain largely undefined. A single integrated Flippase recombination target (FRT) site was engineered within the AGR2-non expressing A375 cell line to allow integration of a constitutively expressed AGR2 alleles. This allows an analysis of how AGR2 protein expression reprogrammes intracellular signalling. The engineered expression of AGR2 had marginal impact on global transcription signalling, compared to its paralogue AGR3. However, expression of AGR2 had a significant impact on remodelling the cellular proteome using a triple-labelled SILAC protocol. 29045 peptides were detected for the identification and relative quantitation of 3003 proteins across the experimental conditions. Ingenuity Pathway annotation highlighted the dominant pathway suppressed by wt-AGR2 was the p53-signalling axis. DNA damage induced p53 stabilization and p21 induction by cisplatin treatment confirmed that wt-AGR2 expression suppressed the p53 pathway. The furthest outlying SILAC protein expression change induced by AGR2 was the anti-viral and cell cycle regulator tumour susceptibility gene 101 (TSG101), confirmed by immunoblotting. Transfection of TSG101 into MCF7 (AGR2+, oestrogen dependent), A549 (AGR2+, oestrogen independent) or A375 (AGR2-) cells confirmed that TSG101 attenuates p53 signalling. These systems wide screens suggest that the most dominant landscape reprogrammed by low levels of AGR2 protein is the cellular proteome, rather than the transcriptome, and provide focus for evaluating its role in proteostasis.

Key words: AGR2, p53, TSG101, SILAC, gene reprogramming

#### INTRODUCTION

Anterior Gradient-2 (AGR2) is a key developmental protein expressed in the cement gland of *Xenopus* frogs central in the formation of the embryonic ectoderm<sup>1</sup>, and whose overexpression in the amputated limb of amphibians drives dedifferentiation of blastema cells in the process of limb regeneration through an interaction with PROD1<sup>2</sup>. In mammals, the human homologue AGR2 is primarily expressed in ectoderm-derived organs such as the lung, stomach, small intestine and prostate<sup>3</sup>, and has recently been implicated in a diverse range of pathological states and diseases including asthma, cancer, and inflammatory bowel disease<sup>4-8</sup>. The AGR2 gene, and its paralogue AGR3, is only known in the vertebrate lineage, thus precluding the use of classic genetics using eukaryotic models such as yeast, worms, or flies; this pair of genes have not expanded since their appearance suggesting important functions in vertebrate biology<sup>9</sup>.

The function and regulation of AGR2 is just emerging<sup>10</sup>. AGR2 protein is a endoplasmic reticulumresident member of the oxidoreductase family of protein disulphide isomerases (PDIs)<sup>9</sup>, with signature motifs including a hydrophobic, cleavable N-terminal sequence, a diverged thioredoxin like CxxS motif, and a non-optimal endoplasmic reticulum KTEL retention motif at the C-terminus. Expression supplementary to the endoplasmic reticulum has been described including cytosolic and extracellularly trafficked<sup>11, 12</sup>. An example of confirmed PDI function of AGR2 is in the catalysing and shuffling of disulphide bonds in the synthesis of the secreted cysteine-rich glycoprotein MUC2<sup>4, 6</sup>. Loss of AGR2 protein has been implicated in the loss of Paneth and goblet cell homeostasis, and subsequent onset of inflammatory bowel disease<sup>7</sup>. Reduced expression in the lung perturbs mucin maturation and thus may play a role in the aetiology of asthma<sup>4</sup>. Consequently, one key function for AGR2 appears to be in its ability to mediate folding of secretory target proteins<sup>13</sup>.

In addition to a physiological role in normal tissue homeostasis, AGR2 can be over-expressed in breast cancers where its paralogue, AGR3, was over-produced in breast cancer cell membrane fractions<sup>14</sup>. This might relate to the secretory requirements of pro-metastatic cancers<sup>15</sup>. Since then AGR2 has been shown to be over-produced in a range of cancers including breast, prostate, pancreatic, liver, oesophagus, GI and lung cancers<sup>3</sup>. The co-expression of AGR2 and AGR3 are uncoupled in ovarian cancers<sup>16</sup>, indicating that the two proteins are not always co-ordinately expressed and not necessarily linked to hormone signalling as they are in breast and prostate cancers. AGR2 was first linked mechanistically to p53 as an over-expressed protein present in Barrett's intestinal oesophageal metaplasia, an acid reflux disease linked syndrome, where AGR2 was identified to exert a negative pressure on p53 by suppressing p53 phosphorylation following DNA damage<sup>17</sup>. Subsequently, clonogenic assays have demonstrated that AGR2 enhances cancer

cell survival<sup>17</sup>, and suppression of AGR2 expression inhibits cell proliferation, invasion and survival in cultured pancreatic<sup>18</sup> and breast cancer cell lines<sup>19, 20</sup>. Intriguingly, conditioned media from cells with silenced AGR2 have reduced ability to stimulate proliferation of pancreatic cancer cells<sup>18</sup>. Protein interaction yeast-two hybrid studies have identified C4.4A and  $\alpha$ -dystroglycan (DAG-1) as interaction partners of AGR2, two proteins that are implicated in metastasis formation, influencing cell-cell or cell-matrix interactions of cancer and non-cancer cells<sup>21</sup>. Xenografts over producing AGR2<sup>22</sup>, or its orthologue AGR3<sup>16</sup>, can promote resistance to cisplatin, which might have relevance to ovarian cancers and other diseases treated with such DNA damaging agents. Pro-metastatic activity of AGR2 has gathered pace with these other studies; ectopic overexpression of AGR2 in rat mammary cells caused the induction of a metastatic phenotype<sup>23</sup>, and silenced AGR2 decreases anchorage-dependent growth and tumour xenograft size in a lung carcinoma cell line<sup>24</sup>.

AGR2 is now implicated in regulation of secretory stress protein response pathways<sup>13</sup>. AGR2 protein can be induced by chemical agents promoting endoplasmic reticulum stress inducing a CHOP dependent regulation of protein folding in the endoplasmic reticulum<sup>25</sup>. Additionally, the aqr2 gene can be induced by oestrogen <sup>19</sup> and the anti-oestrogen drug tamoxifen<sup>22</sup>. AGR2 (and AGR3) protein expression correlates strongly with oestrogen receptor status in human breast cancers<sup>21</sup>, yet are uncoupled in human ovarian cancers<sup>16</sup>. Additionally, the *agr2* gene can be suppressed by a SMAD4-TGF $\beta$  signalling axis, highlighting and oestrogen-independence to gene expression<sup>26</sup>. Subsequently, yeast-2-hybrid studies have been undertaken to begin to understand the interactome of AGR2 with the ultimate aim of deciphering signal transduction pathways and identifying whether AGR2 does indeed provide a potential drug target for p53-reactivation as an anti-cancer therapeutic<sup>11</sup>. A number of potential physiological interactors have been uncovered including HECTD1, RIP140, KCT3 and NRXN3<sup>15</sup>. However, the only well validated interacting protein from this screen is the AAA+ chaperone protein RuvBl2/Reptin<sup>27</sup>. This report characterised a small peptide motif, amino acid residues 104-111 on AGR2, which forms a peptide docking site for Reptin. The significance of this interaction in cancer growth is still under study, however a number of human breast cancers coexpress AGR2 and Reptin suggesting that they might have a level of co-operation in cancer progression<sup>27</sup>. A recent study has implicated AGR2 in the expression of amphiregulin, a growth promoting EGFR mitogen, via the Hippo pathway, resulting in promotion of cell growth<sup>28</sup>. However, despite AGR2 emerging as a compelling oncogenic driver<sup>15</sup>, there is very little known about the dominant mechanisms underlying its influence.

One of the difficulties in systems biology is the integration on relatively large amounts of data on signalling events and dynamic protein-protein interactions. An example of this apparent complexity

is the mass-spectrometric based screen that identified over 700 targets of the ATM/ATR protein kinases<sup>29</sup>. Developing approaches to define highly penetrant compared to less imbedded signalling events will facilitate shedding light on more compelling areas for further focus. We apply this challenge to the AGR2 gene in order to define dominant pathway interactions for future focus; we describe the generation of an engineered cell line that expresses relatively low amounts of the AGR2 protein which allow us to define the effects of the protein on cell signalling that are independent of its normal inputs. The output assays were transcriptomic and proteomic screens to define the dominant mode of action of AGR2 as it begins to reprogramme the cell into an enhanced prooncogenic state. The unbiased data demonstrates that AGR2 protein has a more dominant action on proteome remodelling rather than transcriptome effects and highlight a dominant effect on p53 pathway attenuation.

# **MATERIALS & METHODS**

#### Generating a stable Flp-In A375 cell line expressing the gene of interest

A375 cells were transfected with the pFRT/*lac*Zeo plasmid and incubated for 24 h prior to media exchange. The pFRT/*lac*Zeo construct incorporates a *lacZ-Zeocin* fusion gene, allowing positive integrant selection using the Zeocin antibiotic (Life Technologies), along with a flipase recombination target (FRT) sequence which mediates the recombination of the gene of interest into the host cell line. 48h post-transfection, cells were split into fresh medium at <25% confluency and incubated at 37°C. 400 µg/mL Zeocin was supplemented into tissue culture media and incubated with cells for 14 days with selective media replenished every 3-4 days. Subsequently, surviving clones were isolated and grown in 24-well plates and repopulated. The DNA of multiple clones were extracted and digested with HindIII (NEB) followed by Southern blot analysis using a <sup>32</sup>P radioactive labelled probe specified for the FRT site. Successful single integration is confirmed by the appearance of a single band on Southern blot corresponding to the FRT site, Southern blot band varies depending on its location in the genome and the different sizes of HindIII fragments. The pFRT/*lac*Zeo insertion occurs randomly, therefore will express different characteristics in different clones, some clones may incorporate multiple copies of the DNA sequence and the identification of single integrants is essential.

Utilising *Gateway* technology (Life Technologies), primers (Sigma) were designed against the AGR2 gene, incorporating AttB sites, and amplified from cDNA by PCR. Following separation on a 1% agarose gel and gel extraction, a BP reaction was performed allowing recombination of the amplified

gene into the pDONr 221 plasmid. The BP reaction was heat-shock transformed into *E. coli* (DH5α) bacteria and selected on kanamycin supplemented agar. Resistant clones were isolated and grown in 5ml of LB for 16h and extracted by miniprep (Qiagen). The pDONr 221 construct was progressed into the LR reaction, allowing insertion of the AGR2 gene into the pcDNA5/FRT backbone. The pcDNA5/FRT plasmid incorporates a complementary FRT site to that cloned into the host cell line, facilitating the recombination event that will allow the AGR2 gene to stably integrate into the cell genome. A hygromycin resistance gene is incorporated into the pcDNA5/FRT plasmid to allow selection; however this lacks an ATG initiation codon or promoter to force the expression of the gene. Therefore, the transfection of the empty pcDNA5/FRT gene will not result in hygromycin resistance, rather once the recombination event is complete, the gene is brought into proximity and frame with the ATG start codon and SV40 promoter of the Flp-In host cell line.

In order to assist the recombination event, pcDNA5/FRT is cotransfected into the A375-FRT cells along with the pOG44 plasmid (Life Technologies). Host Flp-In cells were transfected with a 9:1 ratio of pOG44:pcDNA5/FRT construct to a final DNA mass of 10 µg using Attractene (Qiagen), following transfection, cells were incubated at 32°C, 10% CO<sub>2</sub>, the optimum conditions for the flipase enzyme, for 24h. Subsequently, cells were incubated with DMEM media supplemented with 200µg/ml hygromycin B (Life Technologies) until selection is complete. The pOG44 plasmid expresses the Flp recombinase enzyme which catalyses the FRT-site recombination event. Following co-transfection, the hygromycin resistance gene is brought in frame with the SV40 promoter and ATG start codon, the gene of interest (containing the CMV promoter) is inserted into the host cell genome and *lacZ-Zeocin* operon is disrupted by the loss of the SV40 promoter and ATG, and the gene of interest cassette. Therefore, positive clones are hygromycin resistant, Zeocin sensitive and express gene of interest.

#### Tissue Culture, Drug treatments and Immunoblotting

All tissue culture for engineered A375 cell lines was carried out at 37°C in 10% CO<sub>2</sub>. All other cell lines had optimal conditions at 37°C in 5% CO<sub>2</sub>. All cells were grown in DMEM (Gibco) supplemented with 10% foetal bovine serum and 1% penicillin/streptomycin (Invitrogen). DNA-damage caused by platinum containing compounds was carried out using 20  $\mu$ M cisplatin (Sigma) diluted into the complete tissue culture medium for a 16h time course. PolyI:C (LMW, Invivogen) activates the interferon response and was used at 25  $\mu$ g/mL in tissue culture media. All transfections and siRNA treatments were carried out using Attractene and DharmaFECT1 (Thermo Scientific) respectively, using optimised conditions. All siRNA was purchased from Dharmacon, and TSG101-pCMV-XL5 plasmid for mammalian cell expression was sourced from *Origene*. For cell lysis, a 1% NP40 lysis

buffer (1% NP40, 150 mM NaCl, 50 mM Tris pH 8.0, 50 mM NaF, 5 mM EDTA, 1X phosphatase inhibitor, 1 X protease inhibitor) or RIPA (1% Triton, 1% sodium deoxycholate, 0.1% SDS, 0.15M NaCl. 0.01 M Na<sub>3</sub>PO<sub>4</sub> pH7.2, 1 x protease inhibitor) were used. Cell pellets were harvested by scraping with a rubber policeman before chemical lysis with an appropriate volume of lysis buffer for 20 minutes on ice. Cell debris was pelleted and discarded. Protein quantitation was carried out by Bradford assay. Cell lysates were separated on either a 12% or a 13.5% SDS PAGE gel, and transferred to nitrocellulose for western blotting. Antibodies used for immunoblotting were anti-AGR2 (Abcam), anti-TSG101 (Sigma), anti-TSG101 (Abcam) anti-p53 (DO-1, Moravian Biotechnology), anti-p21<sup>WAF-1/CIP-1</sup> (Calbiochem), anti-CHOP (Santa Cruz), anti-IRF-1 (BD) and anti-β-Actin (Sigma). Densitometry calculations were calculated using ImageJ software (NIH).

Site Directed Mutagenesis of AGR2 to produce KDEL and  $\Delta$ KTEL mutants for the development of cell line panel

Mutation	Forward Primer	Reverse Primer
KDEL	ctcaagttgctgaag <mark>gat</mark> gaattgtaggaccc	gggtcctacaattcatccttcagcaacttgag
ΔΚΤΕL	ctcaagttgctg <mark>tag</mark> actgaattgtaggaccc	gggtcctacaattcagtctacagcaacttgag

A 25  $\mu$ L reaction mix of 1× pfu mastermix (Stratagene), 50 ng template DNA, 0.5  $\mu$ M forward primer, 0.5  $\mu$ M reverse primer, 2.5% DMSO was prepared for each mutation and subject to the thermocycler program of 95°C × 1 min, (95°C × 50s, 55°C × 1 min, 68°C × 18 min) × 15 cycles, 68°C × 30 min. The PCR reaction was digested with DpnI for 2h at 37°C to remove methylated DNA, before heat-inactivating the enzyme and transforming into *E. coli* (DH5 $\alpha$ ). Positive colonies were purified by miniprep (Qiagen) and sequenced using T7 forward primer.

# Quantitative RT-PCR confirming incorporation and transcription of the gene of interest

All cell lines were grown to confluency prior to mRNA extraction by RNeasy kit coupled with QIAshredder protocol (Qiagen). mRNA was then reverse transcribed to cDNA with Omniscript RT (Qiagen) and incubated for 1 h at 37°C. Utilising Solaris technology (Thermo Scientific), two mastermixes were prepared of Solaris qPCR mastermix, Solaris primer-probe set and water. One primer-probe set identified hybridised to an internal sequence of AGR2 while the other controlled for GAPDH. 1  $\mu$ L of prepared cDNA was added to each mastermix in triplicate and PCR run for 39 cycles.

Confocal Imaging and concentration of conditioned media to determine the subcellular localisation of wt-AGR2 and C-terminal mutants

Cells were seeded onto glass cover slips at 40% confluency and incubated at 37 °C overnight. Cover slips were incubated with 4% formaldehyde for 10 min to fix the cells and permeabilised in 2% BSA in 1 × TBS containing 1% Triton X-100. Fixed cells were probed with primary antibodies at the appropriate dilution. Primary antibodies used were mouse monoclonal anti-AGR2 (Abnova and AGR3.4 (13)) and the rabbit polyclonal endoplasmic reticulum marker anti-PDI antibody (Enzo Life Sciences). The cells were incubated with Alexa Fluor conjugated secondary antibodies goat antimouse (488 nm) and donkey anti-rabbit (594 nm) at 1/200 dilutions. TO-PRO-3 nuclear stain (Invitrogen) diluted in 1 × TBS was applied to cover slips, and incubated at 37 °C for 20 min providing a nuclear counterstain. Preparations were mounted in Fluorescent Mounting Medium (DakoCytomation), and viewed using an Olympus FV1000 confocal microscope with a 60 × oilimmersion objective and FV10-ASW software (Olympus). For concentrating conditioned media, cells were plated at constant number and allowed to grow to confluency over 3 days. Media was collected and concentrated 50× using 10kDa MWCO spin column filters (Millipore). Protein concentration of media was quantified by Nanodrop (Thermo Scientific) and standardised to 0.5 mg/mL, prior to boiling with SDS sample buffer and loading onto a polyacrylamide gel and western blotting.

#### Illumina HT-12 gene microarray and transcriptome analysis

RNA was extracted using RNeasy kit as described earlier, quality and quantity was assessed with Nanodrop and Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA). Purified RNA (0.5 μg) was biotin-labelled using the Illumina TotalPrep RNA Amplification Kit (Ambion) according to the manufacturer's instructions. Labelled RNA was hybridized to Illumina HT-12 v3 BeadChips and scanned at the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. Gene expression analysis was performed using the open source statistical programming language, R, and associated Bioconductor packages<sup>30</sup>. Probe bead summary data (Supplementary Figure with all raw data available upon request) was quantile-normalised using the Lumi<sup>31</sup> package. Differential expression was determined using Rank Products analysis<sup>32</sup> with 5% false discovery rate. Clustering and heatmaps were performed and generated using the Cluster and TreeView program<sup>33</sup>. The online Database for Annotation, Visualization and Integrated Discovery (DAVID) program<sup>34</sup> was used to identify over-represented KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways from lists of differentially expressed genes (data not shown).

#### MS/MS Analysis of SILAC Samples and Quantitation

(i)Gel electrophoresis and in-gel digestion; SILAC analysis was processed by Dundee Cell Proteomics. Samples were combined in a 1:1:1 ratio before being reduced in SDS-PAGE loading buffer containing 10mM DTT, alkylated in 50 mM iodoacetamide prior to being boiled and separated by onedimensional SDS-PAGE (4–12% Bis-Tris Novex mini-gel, Invitrogen). Visualisation was carried out by colloidal Coomassie staining (Novex, Invitrogen). The entire protein gel lane was excised and cut into 10 gel slices. Each gel slice was subjected to in-gel digestion with trypsin. The resulting tryptic peptides were extracted by 1% formic acid, acetonitrile, lyophilized in a Speedvac, (Helena Biosciences) and resuspended in 1% formic acid. (ii)LC-MS/MS; Trypsin digested peptides were separated using an Ultimate U3000 (Dionex Corporation) nanoflow LC-system consisting of a solvent degasser, micro and nanoflow pumps, flow control module, UV detector and a thermostated autosampler. 10 ml of sample (a total of 2 mg) was loaded with a constant flow of 20 ml/min onto a PepMap C18 trap column (0.3 mm id x 5 mm, Dionex Corporation). After trap enrichment peptides were eluted off onto a PepMap C18 nano column (75 mm x 15 cm, Dionex Corporation) with a linear gradient of 5-35% solvent B (90% acetonitrile with 0.1% formic acid) over 65 minutes with a constant flow of 300 nL/min. The HPLC system was coupled to a LTQ Orbitrap velos (Thermo Fisher Scientific Inc) via a nano ES ion source (Proxeon Biosystems). The spray voltage was set to 1.2 kV and the temperature of the heated capillary was set to 200°C. Full scan MS survey spectra (m/z 335-1800) in profile mode were acquired in the Orbitrap with a resolution of 60,000 after accumulation of 500,000 ions. The five most intense peptide ions from the preview scan in the Orbitrap were fragmented by collision-induced dissociation (normalised collision energy 35%, activation Q 0.250 and activation time 30 ms) in the LTQ after the accumulation of 10,000 ions. Maximal filling times were 1,000 ms for the full scans and 150 ms for the MS/MS scans. Precursor ion charge state screening was enabled and all unassigned charge states as well as singly charged species were rejected. The dynamic exclusion list was restricted to a maximum of 500 entries with a maximum retention period of 90 seconds and a relative mass window of 10 ppm. The lock mass option was enabled for survey scans to improve mass accuracy. Data (Supplementary Figure 2) were acquired using the Xcalibur software. (iii) Protein Expression Quantitation and Bioinformatics Analysis; Quantitation was performed with MaxQuant version 1.0.7.4, and was based on two-dimensional centroid of the isotope clusters within each SILAC pair. To minimize the effect of outliers, protein ratios were calculated as the median of all SILAC pair ratios that belonged to peptides contained in the protein. The percentage variability of the quantitation was defined as the standard deviation of the natural logarithm of all ratios used for obtaining the protein ratio multiplied by a constant factor 100.

The generation of peak list, SILAC- and extracted ion current-based quantitation, calculated posterior error probability, and false discovery rate based on search engine results, peptide to protein group assembly, and data filtration and presentation was carried out using MaxQuant. The derived peak list was searched with the Mascot search engine (version 2.1.04; Matrix Science, London, UK) against a concatenated database combining 80,412 proteins from International Protein Index (IPI) human protein database version 3.6 (forward database), and the reversed sequences of all proteins (reverse database). Parameters allowed included up to three missed cleavages and three labelled amino acid residues (arginine and lysine). Initial mass deviation of precursor ion and fragment ions were up to 7 ppm and 0.5 Da, respectively. The minimum required peptide length was set to 6 amino acids. To pass statistical evaluation, posterior error probability (PEP) for peptide identification (MS/MS spectra) should be below or equal to 0.1. The required false positive rate (FPR) was set to 5% at the peptide level. False positive rates or PEP for peptides were calculated by recording the Mascot score and peptide sequence length-dependent histograms of forward and reverse hits separately and using Bayes' theorem in deriving the probability of a false identification for a given top scoring peptide. At the protein level, the false discovery rate (FDR) was calculated as the product of the PEP of a protein's peptides where only peptides with distinct sequences were taken into account. If a group of identified peptide sequences belong to multiple proteins and these proteins cannot be distinguished, with no unique peptide reported, these proteins are reported as a protein group in MaxQuant. Proteins were quantified if at least one MaxQuant-quantifiable SILAC pair was present. Identification was set to a false discovery rate of 1% with a minimum of two quantifiable peptides. The set value for FPR/PEP at the peptide level ensures that the worst identified peptide has a probability of 0.05 of being false; and proteins are sorted by the product of the false positive rates of their peptides where only peptides with distinct sequences are recognized. During the search, proteins are successively included starting with the best-identified ones until a false discovery rate of 1% is reached; an estimation based on the fraction of reverse protein hits. Enzyme specificity was set to trypsin allowing for cleavage of N-terminal to proline and between aspartic acid and proline. Carbamidomethylation of cysteine was searched as a fixed modification, whereas N-acetyl protein, and oxidation of methionine were searched as variable modifications.

#### Proteomic data analysis

(i) Pathway analysis using Ingenuity Pathway Analysis; All proteomic expression data showing a net change of <0.8- or >1.2-fold divergence from the control cell line, whose identification and quantitation are based on at least 3 peptides were mined and submitted to Ingenuity Pathway Analysis (IPA). Ingenuity scours literature databases to compile pathways that may be over-

represented in the dataset. The top transcriptional regulation pathways, their state of activation and the p-value are reported.

(ii) Data-driven analysis pro forma to interrogate proteomic data; All proteomic data with expression levels <0.8- or >1.2-fold divergence from the control cell line were deemed significant, these were ranked from greatest change of expression to lowest before outliers were reviewed and subject to literature-based analysis. Subsequently, intriguing data-driven targets were validated by independent biochemical means.

#### Fluorescence-activated cell sorting (FACS) to confirm overexpression of KI-67.

Parental and wt-AGR2 expressing cells in culture were trypsinised and diluted to  $5 \times 10^{6}$  cells/mL in PBS/10%FCS/1%NaN<sub>3</sub>. 100 µL of each the cell dilution was transferred to microfuge tube, before fixation with 0.01% formaldehyde and permeabilisation with 0.1% Triton X100. Cells were subsequently stained with KI-67 antibody conjugated to FITC fluorophore (e-bioscience) for 30 minutes in the dark, following 3 washes with PBS 0.05% Tween-20 cells were analysed by FACS.

#### X-CELLigence measuring of cell growth

Cells to be investigated were trypsinised when 80% confluent (still in logarithmic growth), before counting using a haemocytometer. 5000 cells of each cell type were seeded into wells of an E-plate-view (Roche) in quadruplicate and incubated at 37°C, 10% CO<sub>2</sub> for 72 hours. The X-CELLigence assay measures cell growth in real time as a measurement of electrical impedance across interdigitated electrodes on the base of the tissue culture plate. The software program was devised such that cell impedance readings were taken every 15 min over the complete incubation. Using RTCA software package (Roche), the data was analysed, averaged and plotted as changes in growth over time.

#### RESULTS

# Validating an engineered cell model to define the role of AGR2 in re-programming the cancer proteome.

The AGR2-negative A375 human melanoma cell line was used as a model system to define the ability of the artificially-recombined AGR2 allele to alter proteostasis and reprogramme a cell into a prooncogenic mode. This cell model has a well established p53 pathway responsiveness<sup>35</sup>, TLR3 interferon system signalling<sup>36</sup>, cell cycle checkpoint pathway, and DNA damage responsive network that provides a template from which to evaluate how the AGR2 gene product can alter growth control and link to the p53 pathway. We sought to use a cell line which stably integrates a single *agr2* allele using a promoter system that expressed relatively low levels of a target gene to determine how low level expression of an oncoprotein can alter cellular homeostasis. In addition, the stable integration of one *agr2* cassette per cell following selection of a population that outgrows over 10 doublings allows us to begin to ask how the cell can begin to be reprogrammed after the introduction of a single gene.

In stage I of the engineering process, A375 cells were first transfected with the pFRT/*lac*Zeo and subcultured into single clones for analysis for single site integration to allow an acceptor target in stage II (Figure 1A). The pFRT/*lac*Zeo plasmid contains a *lacZ-Zeocin* fusion gene allowing selection of stable integrants by Zeocin antibiotic resistance, resistant foci were isolated, genomic DNA extracted and screened by southern blot (Figure 1B). Due to the random nature of pFRT/*lac*Zeo plasmid integration in the genome, and our aim to insert a single allele of *agr2* into the cell, it is essential to identify and discard erroneous clones which contain more than one FRT site. Multiple integrated FRT sites may result in increased chromosomal rearrangements or unexpected recombination events as a result of complementary genomic FRT sites being in near proximity.

Once single site integrants were confirmed, the clonal cell line was tested for normal functioning pathways (Figure 1C). AGR2 was shown to promote cisplatin resistance in a xenograft system, as the protein can be expressed in cancers where cisplatin can be used as a therapy, and as cisplatin can activate p53, we evaluated whether the FRT-parent cell retains cisplatin responsiveness to p53<sup>16</sup>. Indeed the FRT integrated cell retains the ability to activate the tumour suppressor protein p53, resulting in upregulation of p21. Moreover, another p53-dependent signal responder is the TLR3-responsive interferon system; the addition of Polyinosinic:polycytidylic acid (poly I:C) to cells also stabilises the IRF-1 protein (Figure 1C), indicating that the parental FRT cells retains p53 signalling integrity.

Following confirmation of single FRT site integration, the AGR2 gene of interest cloned into the pcDNA5/FRT vector is co-transfected into the Flp-In host cell line with the recombinase-coding pOG44 plasmid. Selection with hygromycin was carried out until colonies of positive cells are apparent and these are tested for expression of the gene of interest (Figure 2A) and Zeocin sensitivity (data not shown). Comparison with known endogenous AGR2-expressing cell lines, indicated that the experimental engineered cells demonstrate a level of expression >5-fold less than endogenously expressing cells (Figure 2B). This infers that our engineered cell line is not expressing above the physiological level, the protein of interest. Indeed, the relatively low level expression of AGR2 allows us to ascertain how or whether subtle expression of AGR2 reprogrammes signalling pathways.

Subsequently, the role of AGR2 in ER stress induced by tunicamycin was validated, as the ER stress response of authentic AGR2 positive cells is the only physiological stress published to impinge upon AGR2 function as an ER-resident protein disulfide isomerase<sup>25</sup>. Data suggests that AGR2 silencing can restrict the unfolded protein response (UPR) through reduced induction of CHOP<sup>25</sup>. In the synthetically engineered cell lines (Figure 2C and 2D) we see an enhanced CHOP induction as a result of AGR2 expression compared to AGR2-negative isogenic cells. This role in ER stress indicates that these engineered cells are mimicking an authentic AGR2 positive cell with respect to ER-stress responsiveness and suggests that the single gene introduction is able to faithfully recapitulate the in vivo ER-chaperone function of AGR2.

#### Defining the extent to which AGR2 can alter proteostasis using engineered isogenic cell models.

Having established that this FRT-Flp cell system can recapitulate an authentic AGR2 stress responsiveness, we set out to determine how a key domain of AGR2 can impact on cell signalling (Figure 3A). As the only known domain of AGR2 that regulates localization is the C-terminal ER-retention motif<sup>11, 12</sup>, we developed a panel of cell lines that integrated either the wt-AGR2 or the C-terminal mutant AGR2 (e.g. AGR2- $\Delta$ KTEL; a four amino acid residue C-terminal KTEL deletion). The use of qRT-PCR demonstrated relatively similar levels of AGR2 and AGR2- $\Delta$ KTEL mRNA levels in the engineered cells (Figure 4A and B). Additionally, the cell panels demonstrated that although wt-AGR2 partially co-localizes with the ER marker PDI, as expected (Figure 3B and C), the AGR2- $\Delta$ KTEL mutant, lacking the ER-retention motif did not localize detectably to the ER (Figure 3B and C) but evidently traversed the plasma membrane resulting in secretion into the tissue culture media (Figure 3C). These latter data are consistent with a previous report that AGR2- $\Delta$ KTEL is secreted<sup>12</sup> and further suggests that this engineered FRT-FLP A375 cell model reconstitutes an authentic AGR2 trafficking response.

In order to define the dominant pro-oncogenic pathways that are altered by AGR2 expression, we first evaluated whether wt-AGR2 altered the basal transcription programme of the cell. There is evidence, indeed, that forced expression of AGR2 by ectopic expression or siRNA depletion of the authentic protein from AGR2-positive cells can regulate EGFR pathway activation<sup>28</sup>. As controls for the transcriptomic screen, we used AGR2-ΔKTEL and AGR3 as control for specificity. Following the co-ordinated transfection of pcDNA5/FRT-AGR2, pcDNA5/FRT-AGR2-ΔKTEL, or pcDNA5/FRT-AGR3, along with plasmid pOG44, colony populations were isolated, propagated, and the basal transcriptome was evaluated in triplicate using Illumina HT-12 BeadChips (Figure 4C and D). The gene expression data demonstrate that the synthesis of wt-AGR2 has a negligible effect on global transcription homeostasis, giving a very similar signature to the FRT-parent or the AGR2-ΔKTEL cell

line (Figure 4C and D). By comparison, the AGR3 engineered cell induced significant changes in transcription. The small cluster of GAGE genes that exhibited significantly higher expression in the wt-AGR2 cells compared to FRT-parental and AGR2-ΔKTEL cells using Illumina BeadChips could not be reproduced using qRT-PCR (data not shown). The inability of the engineered AGR2 cells to induce EGFR pathway induction might be due to the relatively low level expression of the protein in the engineered cells, relative to the endogenous AGR2 protein in authentic cancer cell lines (Figure 2A). Thus, if threshold of AGR2 protein levels are important, then a hierarchy of AGR2 function would first impact on CHOP dependent ER-stress processes followed by transcription-dependent effects on EGFR pathway.

As AGR2 is fundamentally an ER-resident PDI, perhaps its chaperone function has a primary effect on cell signalling, especially if relatively low levels of AGR2 are synthesized, as in this engineered cell line. As such, we evaluated whether the addition of the single AGR2 expressing allele can reprogramme the cellular steady-state proteome under conditions in which the transcriptome changes are negligible (Figure 4C). In order to unravel the mechanism whereby AGR2 gene expression affects cell function, a quantitative proteomics approach was utilised whereby the effect of AGR2 expression on cellular protein levels could be analysed. As controls for the proteomic screen, we used AGR2-ΔKTEL expressing cells as control for specificity. As this requires the use of three cell lines simultaneously, we used triple-SILAC methodology to allow multiplexing of the experimental design, thus the protein expression of three isogenic cell lines could be investigated for peptide level quantitation (Supplementary Figure 1). Due to the cells being isogenic, any modification from the parental cell line indicates a potential role of the introduced gene in the reprogramming of the cell. 29045 peptides were detected and used for the identification and relative quantitation of 3003 proteins across the three experimental conditions (Figure 5A and supplementary figure 2). For unbiased downstream pathway analysis, an empirical level of change deemed significant was set at 20% up- or down-regulation, and a minimum of 3 peptides required for quantitation. Between conditions, the majority of protein expression levels did not change (Figure 5B). Wild-type AGR2 gene expression changed the expression of 735 proteins (333 up- and 402 down-regulated >20%), whereas AGR2-ΔKTEL modified 912 proteins (418 up- and 494 downregulated >20%), and of these only 226 were shared between the two distinct subcellularly localised gene products (Figure 5B).

Pathway annotation was carried out using Ingenuity Pathway Analysis (Ingenuity Systems; <sup>37</sup>), only proteins which met the previously stated proforma were submitted to the database for analysis. Ingenuity Pathway Analysis identified that wt-AGR2 reprogrammed the transcriptional regulation of

the cell, through the inhibition of the TP53 and activation of the E2F1 transcription regulator hubs (Figure 5C). Comparatively, Ingenuity Pathway Analysis did not identify any transcriptional pathway misregulated when AGR2- $\Delta$ KTEL was present. In addition to the Ingenuity Pathway Analysis, a data driven approach was sought, where protein expression changes from the parental cell were ranked and the most altered expression levels identified. The key identified protein expression changes that may give an insight into how wt-AGR2 functions as an oncoprotein, are the upregulation of tumour susceptibility gene 101 (TSG101) (3.5-fold) and Ki-67 (3.1-fold) (Figure 5D; Table 1A). AGR2- $\Delta$ KTEL expressing cells induced and suppressed a different set of proteins (Table 1B) but the possible role of these in mediating the signal of AGR2 that has naturally escaped the ER was not the subject of this report.

#### Validation of Ki-67 as a proliferation marker induced by AGR2.

The identification of Ki-67 as a dominant induced target links AGR2 expression to enhanced proliferation; Ki-67 is a nuclear protein, thought to be necessary for cellular proliferation and is one of the most commonly used immunohistochemical markers for proliferation in clinical tissue<sup>38</sup>. Ki-67 is evidently present in all cell types and is detectable during active phases of the cell cycle, but absent in resting (G<sub>0</sub>) cells. An increase in the expression levels of Ki-67 would suggest that more cell proliferation is taking place in this condition, and this may be coupled to the previously identified p53-activity attenuation. Subsequently, FACS analysis confirmed that compared to AGR2-null A375 cells, the wt-AGR2 expressing cells demonstrated an increased expression of Ki-67 (Figure 6A, summary). In addition, xCELLIgence real-time proliferation assays demonstrated that wt-AGR2 expressing cells generated using pcDNA5/FRT-AGR2 exhibited an increase in the rate of cell growth, relative to the parental FRT cells and the engineered control pcDNA5/FRT-STOP-AGR2 (Figure 6B-D). These results validate Ki-67 as an upregulated effector of AGR2 signalling and are consistent with the enhanced migration activity linked to AGR2<sup>39, 40</sup>.

In addition to Ki-67 as a biomarker that can be logically validated due to its links in proliferation, we also validated the p53 pathway, as this was the most significant pathway target identified using Ingenuity Pathway Analysis. Subsequent western blot analysis indicated that AGR2 expressing A375 cells suppressed the basal levels of p21, a read out of p53-dependent transcriptional activity, in the absence of exogenous DNA damage, relative to FRT parent (Figure 7A, lanes 2 vs 1). As controls, engineered cells expressing AGR2-ΔKTEL or AGR2-KDEL did not show significant suppression of p21 (Figure 7A, lanes 3 and 4 vs 2). When these cell panels were exposed to cisplatin (towards which AGR2 can mediate cisplatin resistance in xenografts<sup>22</sup>) and recovered for 24 hours, the classic induction of p53 and linked monoubiquitination can be observed using FRT-parent, AGR2-ΔKTEL, and

AGR2-KDEL cell lines, whereas wt-AGR2 expressing cells exhibited decreased p53 ubiquitination and p21 levels (Figure 7A, lanes 6 vs 5, 7, and 8). These data suggest that only wt-AGR2 can suppress p53, which is consistent with the Ingenuity Pathway Analysis from the triple SILAC screen (Figure 5). When a time course of the cellular response to cisplatin was performed, a similar trend was observed; the wt-AGR2 expressing cells exhibited the most suppressed p53 protein response, relative to FRT-parent, AGR2-ΔKTEL, or KDEL expressing cells (Figure 7B and C). Consistent with this data, siRNA-mediated depletion of authentically expressed AGR2 protein in MCF7 or A549 cancer cells (both with a wt-p53 pathway) resulted in steady-state p53 protein elevation (Figure 8A and B). Together, these data indicate that wt-AGR2 engineered A375 cells expressing relatively low levels of the protein (Figure 2) can attenuate the p53 pathway, induce Ki-67, elevate proliferation, and maintain the CHOP-dependent response to tunicamycin. However, no detectable changes in EGFR signalling could be observed, perhaps suggesting that more AGR2 protein needs to be titrated into the cell before there is an impact on this signalling pathway.

#### Validation of TSG101 as an AGR2 effector in the suppression of p53-dependent activity

In order to define a mechanism to explain how AGR2 can suppress p53 protein, we took advantage of the fact that Ingenuity Pathway Analysis identified components of the "TP53 pathway" (Figure 5) as significantly affected. Indeed, the most highly upregulated protein in AGR2-expressing cells was the previously identified p53 inhibitor, TSG101. Defects in the expression of TSG101 have previously been linked to multiple cancers<sup>41, 42</sup>, and steady-state expression of this gene product has been described as necessary for genome stability and cell cycle regulation<sup>43, 44</sup>. TSG101 forms part of the endosomal sorting complex required for transport-1 (ESCRT-1), and functions as a regulator of vesicular trafficking, whose linkage to AGR2 might provide an insight for the role of AGR2 as a PDI in protein maturation and protein trafficking. Intriguingly, TSG101 has previously been identified as participating in the well studied p53-MDM2 autoregulatory loop, by affecting protein decay. The Ubc domain of TSG101 disrupts the ubiquitination of MDM2, thus inhibiting MDM2 decay and elevating its steady-state level, and that these events are associated with down-regulation of p53 protein.

Subsequent validation of the role TSG101 in p53 protein levels was performed (Figure 8). Basal TSG101 levels are significantly greater in wt-AGR2 expressing cells, compared to the isogenic AGR2negative cells (Figure 8C), which is consistent with the SILAC data. Further, transient transfection of TSG101 into FRT-AGR2-A375 melanoma cells suppressed further endogenous p53 protein levels,

particularly when proteosomal degradation is blocked (Figure 8D), suggesting that TSG101 is predominantly stimulating p53 protein turnover under these conditions. We further probed this effect of TSG101 using cisplatin damaged in cells naturally expressing AGR2 and wt-p53.

A549 cells have a wt-p53 pathway and an active TGF-β dependent pathway that regulates AGR2 expression in an oestrogen-independent manner (data not shown). When this cell was transfected with TSG101, whether cells were damaged or not, p53 protein levels are suppressed (Figure 8E, lanes 2, 4, 6, 8 vs 1, 3, 5, 7). When A549 cells were damaged with cisplatin, there was still remaining a dominant effect of TSG101 over p53 protein levels (Figure 8E, lanes 6 vs 5). When MG132 was added to cells, then the effects of TSG101 on p53 protein turnover can still be seen in the presence of cisplatin (Figure 8E, lanes 8 vs 7). These data show that in this cell line, forced expression of TSG101 can suppress p53 protein induction and/or increase p53 protein turnover under a set of experimental conditions including basal state or cisplatin damage. We also evaluated the TSG101 effects in MCF7 cells, which also have a wt-p53 pathway, but the expression of AGR2 protein in this cell is oestrogen-dependent. Similar to A549 cells, under each condition, ectopic expression of TSG101 suppressed p53 protein levels (Figure 8F, lanes 2, 4, 6, 8 vs lanes 1, 3, 5, 7).

To complement the experiments above that evaluate how ectopic expression of TSG101 can impact on cells authentically expressing both wt-p53 and AGR2, we used targeted siRNA to determine whether reduction in endogenous TSG101 impacted on p53 protein turnover. In A549 cells, depletion of TSG101 using siRNA only had a significant impact on stabilizing p53 protein levels in the presence of MG132 (Figure 8G, lanes 5 and 6 vs 1-4), which is consistent with the effect of transfected TSG101 on p53 levels in FRT-AGR2-A375 cells (Figure 8D). Similarly, in MCF7 cells, TSG101 depletion most impacted on p53 protein in the presence of MG132 (Figure 8H, lanes 5 and 6 vs 1-4), although an effect on p21 protein was observed under all conditions (Figure 8H, lanes 2, 4, and 6 vs 1, 3, and 5). Together, these data suggest that the over production of TSG101 induced by the reprogrammed FRT-AGR2-A375 cell can in part relate to the ability of TSG101 to stimulate the turnover of p53 in a range of cell types.

### DISCUSSION

Anterior Gradient-2 is emerging as a protein whose aberrant expression influences the development of several human diseases<sup>15</sup> in part through protein secretion systems<sup>13</sup>. Mechanisms whereby the protein exhibits these functions is thought to include its propensity to chaperone nascent proteins in the ER<sup>4, 6, 25</sup>, and its ability to mediate induction of oncogenic signals<sup>28, 45</sup>. However, co-factors that mediate these functions are currently poorly understood. The majority of data on the molecular

basis of AGR2 function comes primarily from protein interaction yeast-two hybrid studies which have implicated C4.4A,  $\alpha$ -DAG<sup>21</sup> and Reptin<sup>27</sup> along with other yeast-2-hybrid interactions which remain to be fully validated<sup>15</sup>. In oncology, AGR2 has been implicated in tamoxifen and cisplatin drug resistance<sup>19, 22</sup>, as a pro-metastatic factor<sup>18, 24</sup>, as well as disrupting homeostatic signalling pathways<sup>17, 28</sup> and remains under study in cancer drug discovery programmes<sup>15</sup>. The core function of AGR2 in biology might relate to its ability to reprogramme cell migration, as highlighted in studies that identified its ability to mediate limb regeneration in amphibia<sup>2, 46-48</sup>.

In order to drive the understanding of the effects of AGR2 protein expression that might be useful in fields such as regenerative medicine or disease, we engineered an isogenic FLP-FRT cell panel constitutively expressing the recombined gene(s) of interest. These cell lines expressed our AGR2 alleles at relatively low levels (Figure 2), compared to alternative techniques such as transient transfections, allowing a subtle introduction of the interrogated gene. In evaluation of the effects of AGR2 expression on cell phenotypes, transcriptomic screens did not reveal significant informant (Figure 4). As such we applied quantitative proteomics; SILAC was used as the method which has several obvious advantages over the classical proteomic quantitation methods (using radioactivity or dyes) since it removes the requirement for high-resolution protein separation, predominantly by 2D-SDS-PAGE, which limits the quantification to only the most highly abundant, soluble proteins<sup>49, 50</sup>. SILAC does suffer from the drawback that some cell lines are sensitive to changes in media composition; however this was not the case in this study. The coupling of SILAC and isogenic cells has been described previously<sup>51</sup>, and indicates how a mutation in a single gene can alter the diversity of protein expression in the cell. Due to the well-characterised nature of the cell models, including CHOP induction induced by ER stress and the sensitivity of AGR2- $\Delta$ KTEL to secretion, we can consider divergence of the proteomic profile from the control cell line to be fundamental to the function of AGR2.

Non-biased pathway analysis using Ingenuity Pathway Analysis identified the main biological pathways affected by overexpression of AGR2 protein. The TP53 signalling pathway was the most significantly attenuated in the experimental wt-AGR2 expressing cell line compared to the AGR2-null cell line and the AGR2-ΔKTEL control cell proteomic profiles (Figure 5; Table 1). This complements previous studies of the p53-suppressing nature of AGR2 and its potential as a novel drug target for p53 reactivation<sup>11, 17</sup>. Further, these data show that the synthetically engineered cells can be utilised in unravelling, a least one mechanism of action of AGR2 which regards it as a proto-oncoprotein. Other transcriptional pathways disrupted include the E2F1 (activation) and TFEB (suppressed). E2F1 plays a significant role in the regulation of cell proliferation, the cell cycle, and is often misregulated

**Molecular BioSystems Accepted Manuscript** 

in developing tumours<sup>52</sup>. Intriguingly, E2F1 is reported to co-operate with p53, via activation of ARF and subsequent inhibition of MDM2, stabilising p53 levels<sup>53</sup>. Our data suggest that AGR2 might uncouple this interplay, resulting in p53 levels becoming attenuated.

Among the protein expression levels modified by wt-AGR2 (Figure 5), proteins apart from TSG101 were identified a provide intriguing links to AGR2 signalling. For example, Mucin18 is a cysteine rich glycoprotein whose expression is increased by 36% following wt-AGR2 expression. As previously described<sup>4, 5</sup>, AGR2 participates in disulphide reshuffling of mucin proteins (like mucin-5) resulting in an increase in the maturity of mucin proteins and a reduction in the degradation of nascent chains<sup>4</sup>. Another disease relevant gene product was Huntington protein (Table 1A); for example GRP78 chaperone can regulate the extent of misfolding of the relatively large Huntington gene product<sup>54</sup> and it is possible that AGR2 similarly assists its solubility through its' chaperone activities. By contrast, using our cell model with AGR2 protein with a 4 amino acid deletion in its ER-retention site resulted in the over-production of proteins such as TMEM9 and WNK1 kinase (Table 1B). WNK1 can stimulate pro-oncogenic signalling and it is possible that, physiologically, the isoform of authentic AGR2 that escapes the ER might indeed remodel the expression of kinases such as WNK1. Indeed, Huntington protein is actually downregulated by AGR2 protein the ER-retention site (Table 1B) suggesting that the balance between AGR2 ER residence and ER escape can regulate steady-state levels of the Huntington protein. Previously validated AGR2 interacting proteins, Ruvbl-2 (Reptin), dystroglycan and CD59 (PROD-1) expression levels do not modulate significantly in this dataset. HECTD1, a putative yeast-2-hybrid interaction candidate of AGR2<sup>15</sup> demonstrates a reduced expression profile. KI-67 demonstrates expression >3× higher compared to the parental cell line; the KI-67 protein is associated with cell proliferation, present in all active stages of the cell cycle but absent in  $G_0$  cells<sup>38</sup>. Previous studies of the oncogenic properties of AGR2 make this a relevant overexpressed protein, as it, and subsequent FACS analysis validation of expression levels, indicates wt-AGR2 expressing cells have enhanced a key immunohistochemical marker of cell proliferation.

Of all the target proteins identified, we focused on validation of TSG101 as this target provided the most direct link to the p53 pathway suppression in the Ingenuity pathway profiling. TSG101 is involved in proteostasis and protein trafficking, due to its involvement in the ESCRT-1 (endosomal complex required for transport) complex of proteins regulating the trafficking of components of late endosomal compartments<sup>55</sup>. TSG101 contains a ubiquitin conjugating (Ubc)-like domain, which recognises ubiquitylated substrates and incorporates them into the multi-vesicular body<sup>56</sup>. TSG101 overexpression has recently, somewhat controversially been implicated in cancer growth, despite initial categorization as a tumour suppressor gene<sup>57, 58</sup>. The majority of literature now describes

TSG101 as a tumour enhancer in various cancers including breast, ovarian and prostate cancers (reviewed in<sup>42</sup>). Steady-state levels of TSG101 are maintained within narrow limits<sup>59</sup>, and overexpression of TSG101 can lead to neoplastic transformation<sup>60</sup>. In complementary studies, *tsg101* silencing in mice reported defective cell proliferation and p53 protein accumulation (and subsequent p21<sup>WAF-1/CIP-1</sup> build-up)<sup>61</sup>, while a reported regulatory loop involving TSG101/MDM2 resulted in MDM2 degradation modulation with a downstream effect on MDM2/p53<sup>62</sup>. Our data also support the interplay between levels of TSG101 and the subsequent effect on p53 levels and activity.

Our study defines a mechanism for how the emerging chaperone, AGR2<sup>15</sup>, can remodel the cell using an isogenic system to evaluate the effects of engineered AGR2 expression on the transcriptome and the proteome (Figure 9). Our cellular system titrates a relatively small amount of AGR2 into the cell (Figure 2), where the protein has its most dominant effect on the proteome rather than gene expression. In relevance to cell growth control, AGR2 expression upregulates Ki-67 and TSG101, and Ingenuity Pathway Analysis indicates that the most dominantly suppressed pathway is p53. Ongoing investigations will evaluate the co-expression of AGR2 and TSG101 in an array of cell types to determine whether the coupling of the expression of these cofactors is a contributor to disease phenotypes in vivo.

Conflicts of Interest: The authors declare no conflicts of interest.

#### ACKNOWLEDGMENTS

We wish to thank Elisabeth Freyer, Flow Cytometry Facility, MRC IGMM, University of Edinburgh, for FACS analysis; Lenka Hernychova, RECAMO, Masaryk Memorial Cancer Institute, Czech Republic for discussion regarding proteomics, and Judith Nicholson, Australian Proteomics Facility, Macquarie University for proteomic data analysis. T.A.G. was supported by a PhD studentship from the BBSRC.

#### **Reference List**

- 1. F. Aberger, G. Weidinger, H. Grunz and K. Richter, *Mech Dev*, 1998, **72**, 115-130.
- A. Kumar, J. W. Godwin, P. B. Gates, A. A. Garza-Garcia and J. P. Brockes, Science, 2007, 318, 772-777.
- 3. V. Brychtova, B. Vojtesek and R. Hrstka, *Cancer Lett*, 2011, **304**, 1-7.
- 4. B. W. Schroeder, C. Verhaeghe, S. W. Park, L. T. Nguyenvu, X. Huang, G. Zhen and D. J. Erle, *Am J Respir Cell Mol Biol*, 2012, **47**, 178-185.
- W. Zheng, P. Rosenstiel, K. Huse, C. Sina, R. Valentonyte, N. Mah, L. Zeitlmann, J. Grosse, N. Ruf, P. Nurnberg, C. M. Costello, C. Onnie, C. Mathew, M. Platzer, S. Schreiber and J. Hampe, *Genes Immun*, 2006, 7, 11-18.
- 6. S. W. Park, G. Zhen, C. Verhaeghe, Y. Nakagami, L. T. Nguyenvu, A. J. Barczak, N. Killeen and D. J. Erle, *Proc Natl Acad Sci U S A*, 2009, **106**, 6950-6955.

- 7. F. Zhao, R. Edwards, D. Dizon, K. Afrasiabi, J. R. Mastroianni, M. Geyfman, A. J. Ouellette, B. Andersen and S. M. Lipkin, *Dev Biol*, 2010, **338**, 270-279.
- 8. E. Di Valentin, C. Crahay, N. Garbacki, B. Hennuy, M. Gueders, A. Noel, J. M. Foidart, J. Grooten, A. Colige, J. Piette and D. Cataldo, *Am J Physiol Lung Cell Mol Physiol*, 2009, **296**, L185-197.
- 9. S. Persson, M. Rosenquist, B. Knoblach, R. Khosravi-Far, M. Sommarin and M. Michalak, *Mol Phylogenet Evol*, 2005, **36**, 734-740.
- 10. E. Chevet, D. Fessart, F. Delom, A. Mulot, B. Vojtesek, R. Hrstka, E. Murray, T. Gray and T. Hupp, *Oncogene*, 2013, **32**, 2499-2509.
- 11. A. Fourtouna, E. Murray, J. Nicholson, M. M. Maslon, L. Pang, D. T. F. Dryden and T. R. Hupp, *Curr Chem Biol*, 2009, **3**, 124-137.
- 12. A. Gupta, A. Dong and A. W. Lowe, *J Biol Chem*, 2012, **287**, 4773-4782.
- 13. N. Dejeans, S. Manie, C. Hetz, F. Bard, T. Hupp, P. Agostinis, A. Samali and E. Chevet, *Trends Mol Med*, 2014.
- P. J. Adam, R. Boyd, K. L. Tyson, G. C. Fletcher, A. Stamps, L. Hudson, H. R. Poyser, N. Redpath, M. Griffiths, G. Steers, A. L. Harris, S. Patel, J. Berry, J. A. Loader, R. R. Townsend, L. Daviet, P. Legrain, R. Parekh and J. A. Terrett, *J Biol Chem*, 2003, **278**, 6482-6489.
- 15. E. Chevet, D. Fessart, F. Delom, A. Mulot, B. Vojtesek, R. Hrstka, E. Murray, T. Gray and T. Hupp, *Oncogene*, 2012.
- T. A. Gray, N. J. MacLaine, C. O. Michie, P. Bouchalova, E. Murray, J. Howie, R. Hrstka, M. M. Maslon, R. Nenutil, B. Vojtesek, S. Langdon, L. Hayward, C. Gourley and T. R. Hupp, *J Immunol Methods*, 2012, **378**, 20-32.
- 17. E. Pohler, A. L. Craig, J. Cotton, L. Lawrie, J. F. Dillon, P. Ross, N. Kernohan and T. R. Hupp, *Mol Cell Proteomics*, 2004, **3**, 534-547.
- 18. V. Ramachandran, T. Arumugam, H. Wang and C. D. Logsdon, *Cancer Res*, 2008, **68**, 7811-7818.
- R. Hrstka, R. Nenutil, A. Fourtouna, M. M. Maslon, C. Naughton, S. Langdon, E. Murray, A. Larionov, K. Petrakova, P. Muller, M. J. Dixon, T. R. Hupp and B. Vojtesek, *Oncogene*, 2010, 29, 4838-4847.
- 20. K. E. Vanderlaag, S. Hudak, L. Bald, L. Fayadat-Dilman, M. Sathe, J. Grein and M. J. Janatpour, *Breast Cancer Res*, 2010, **12**, R32.
- G. C. Fletcher, S. Patel, K. Tyson, P. J. Adam, M. Schenker, J. A. Loader, L. Daviet, P. Legrain, R. Parekh, A. L. Harris and J. A. Terrett, *Br J Cancer*, 2003, 88, 579-585.
- 22. S. M. Hengel, E. Murray, S. Langdon, L. Hayward, J. O'Donoghue, A. Panchaud, T. Hupp and D. R. Goodlett, *J Proteome Res*, 2011, **10**, 4567-4578.
- 23. D. Liu, P. S. Rudland, D. R. Sibson, A. Platt-Higgins and R. Barraclough, *Cancer Res*, 2005, **65**, 3796-3805.
- 24. Z. Wang, Y. Hao and A. W. Lowe, *Cancer Res*, 2008, **68**, 492-497.
- 25. A. Higa, A. Mulot, F. Delom, M. Bouchecareilh, D. T. Nguyen, D. Boismenu, M. J. Wise and E. Chevet, *J Biol Chem*, 2011, **286**, 44855-44868.
- 26. A. M. Norris, A. Gore, A. Balboni, A. Young, D. S. Longnecker and M. Korc, *Oncogene*, 2012.
- 27. M. M. Maslon, R. Hrstka, B. Vojtesek and T. R. Hupp, *J Mol Biol*, 2010, **404**, 418-438.
- 28. A. Dong, A. Gupta, R. K. Pai, M. Tun and A. W. Lowe, *J Biol Chem*, 2011, **286**, 18301-18310.
- 29. S. Matsuoka, B. A. Ballif, A. Smogorzewska, E. R. McDonald, 3rd, K. E. Hurov, J. Luo, C. E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S. P. Gygi and S. J. Elledge, *Science*, 2007, **316**, 1160-1166.
- R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang and J. Zhang, *Genome Biol*, 2004, **5**, R80.
- 31. P. Du, W. A. Kibbe and S. M. Lin, *Bioinformatics*, 2008, **24**, 1547-1548.

- 32. R. Breitling, P. Armengaud, A. Amtmann and P. Herzyk, *FEBS Lett*, 2004, **573**, 83-92.
- 33. M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, *Proc Natl Acad Sci U S A*, 1998, **95**, 14863-14868.
- 34. W. Huang da, B. T. Sherman and R. A. Lempicki, *Nat Protoc*, 2009, 4, 44-57.
- 35. J. P. Blaydes, A. L. Craig, M. Wallace, H. M. Ball, N. J. Traynor, N. K. Gibbs and T. R. Hupp, *Oncogene*, 2000, **19**, 3829-3839.
- 36. M. Taura, A. Eguma, M. A. Suico, T. Shuto, T. Koga, K. Komatsu, T. Komune, T. Sato, H. Saya, J. D. Li and H. Kai, *Mol Cell Biol*, 2008, **28**, 6557-6567.
- 37. S. Draghici, P. Khatri, A. L. Tarca, K. Amin, A. Done, C. Voichita, C. Georgescu and R. Romero, *Genome Res*, 2007, **17**, 1537-1545.
- 38. T. Scholzen and J. Gerdes, *J Cell Physiol*, 2000, **182**, 311-322.
- 39. L. Sweeny, Z. Liu, B. D. Bush, Y. Hartman, T. Zhou and E. L. Rosenthal, *Exp Cell Res*, 2012, **318**, 1788-1798.
- 40. K. Park, Y. J. Chung, H. So, K. Kim, J. Park, M. Oh, M. Jo, K. Choi, E. J. Lee, Y. L. Choi, S. Y. Song, D. S. Bae, B. G. Kim and J. H. Lee, *Exp Mol Med*, 2011, **43**, 91-100.
- 41. R. T. Liu, C. C. Huang, H. L. You, F. F. Chou, C. C. Hu, F. P. Chao, C. M. Chen and J. T. Cheng, *Oncogene*, 2002, **21**, 4830-4837.
- 42. Y. Jiang, Y. Ou and X. Cheng, *Front Biosci*, 2013, **18**, 279-288.
- 43. G. Zhu, R. Gilchrist, N. Borley, H. W. Chng, M. Morgan, J. F. Marshall, R. S. Camplejohn, G. H. Muir and I. R. Hart, *Int J Cancer*, 2004, **109**, 541-547.
- 44. Q. Zhong, Y. Chen, D. Jones and W. H. Lee, *Cancer Res*, 1998, **58**, 2699-2702.
- 45. E. Pohler, A. L. Craig, J. Cotton, L. Lawrie, J. F. Dillon, P. Ross, N. Kernohan and T. R. Hupp, *Mol Cell Proteomics*, 2004, **3**, 534-547.
- 46. A. Kumar and J. P. Brockes, *Trends Neurosci*, 2012, **35**, 691-699.
- 47. A. Kumar, J. P. Delgado, P. B. Gates, G. Neville, A. Forge and J. P. Brockes, *Proc Natl Acad Sci U S A*, 2011, **108**, 13588-13593.
- 48. A. Kumar, G. Nevill, J. P. Brockes and A. Forge, *J Anat*, 2010, **217**, 16-25.
- 49. S. E. Ong, L. J. Foster and M. Mann, *Methods*, 2003, **29**, 124-130.
- 50. S. E. Ong and M. Mann, *Methods Mol Biol*, 2007, **359**, 37-52.
- 51. J. R. Hart, L. Liao, L. Ueno, J. R. Yates, 3rd and P. K. Vogt, *Cell Cycle*, 2011, **10**, 971-976.
- 52. D. Ginsberg, *FEBS Lett*, 2002, **529**, 122-125.
- J. D. Weber, L. J. Taylor, M. F. Roussel, C. J. Sherr and D. Bar-Sagi, *Nat Cell Biol*, 1999, 1, 20-26.
- 54. Y. Jiang, H. Lv, M. Liao, X. Xu, S. Huang, H. Tan, T. Peng, Y. Zhang and H. Li, *Neurosci Lett*, 2012, **516**, 182-187.
- 55. M. Babst, G. Odorizzi, E. J. Estepa and S. D. Emr, *Traffic*, 2000, **1**, 248-258.
- 56. S. Dupre, C. Volland and R. Haguenauer-Tsapis, *Curr Biol*, 2001, **11**, R932-934.
- 57. L. Li, U. Francke and S. N. Cohen, *Cell*, 1998, **93**, following 660.
- 58. L. Li, X. Li, U. Francke and S. N. Cohen, *Cell*, 1997, **88**, 143-154.
- 59. G. H. Feng, C. J. Lih and S. N. Cohen, *Cancer Res*, 2000, **60**, 1736-1741.
- 60. L. Li and S. N. Cohen, *Cell*, 1996, **85**, 319-329.
- 61. J. Ruland, C. Sirard, A. Elia, D. MacPherson, A. Wakeham, L. Li, J. L. de la Pompa, S. N. Cohen and T. W. Mak, *Proc Natl Acad Sci U S A*, 2001, **98**, 1859-1864.
- 62. L. Li, J. Liao, J. Ruland, T. W. Mak and S. N. Cohen, *Proc Natl Acad Sci U S A*, 2001, **98**, 1619-1624.
- 63. S. O'Gorman, D. T. Fox and G. M. Wahl, *Science*, 1991, **251**, 1351-1355.
- 64. M. M. Maslon, R. Hrstka, B. Vojtesek and T. R. Hupp, J Mol Biol, 2010, 404, 418-438.

#### **FIGURE LEGENDS**

Figure 1. Generation of an FRT-integrated cell line. (A). Generation of an FRT-integrated A375 cell line. (1) Diagrammatic representation of the FRT-recombination site containing plasmid pFRT/lacZeo whose insertion incorporates a 34 base pair motif that provides a binding and cleavage site for the Flp-recombinase enzyme<sup>63</sup>. The lacZ-Zeocin fusion gene incorporated into the vector allows selection for Zeocin resistance. Individual Zeocin-resistant cell colonies are screened by Southern blotting for the presence of a single integrated plasmid (as in B, clone 1). FRT-integrated cells are re-screened to ensure they still maintain the p53 and IRF-1 responsiveness to cisplatin and polyI:C, respectively (as in C, below). FRT-parental A375 cells with well-defined integrity are subjected to transfection using (2) the gene-of-interest (in this case AGR2) cloned into pcDNA5/FRT by the LR reaction into the Gateway expression system (Invitrogen) along with (3) pOG44 expressing the Flp recombinase that mediates (4) a homologous recombination event at the FRT sites, thus the gene of interest is integrated into the host cell genome at the FRT site. Cells are selected for Hygromycin resistance and screened for Zeocin sensitivity that represents cones with the integrated gene-of-interest. (B). Identification of the A375 cell lines with a single FRT site integrated in the genome using Southern blot. A375 cells transfected with pFRT/lacZeo plasmid and cultured in a medium containing 400µg/ml Zeocin. After 14 days, over 48 single surviving clones were inoculated into 24 wells plate and repopulated. The DNA of the clones was extracted, digested with HindIII followed by Southern blot analysis using a <sup>32</sup>P radioactive labelled probe specified for the FRT site. Successful single integration is confirmed by the appearance of a single band on Southern blot corresponding to the FRT site. Results show 3 representative clones with a single FRT site integration (black arrows), which demonstrate a distinct location within the genome of the host cell line. Some Zeocin resistant cells did not express identifiable FRT insertion plasmids (as in 2) and some incorporated more than one FRT site (as 3). (C) Integrity of the A375-FRT parental cells. (top panel) Evaluation of a functional wild-type p53 pathway was defined by the ability of p53 to be activated by incubating cells with 20  $\mu$ M cisplatin over a 24h timescale. The induction of p53 protein results in a subsequent induction of the p53 activity-dependent gene product p21. Cells were lysed in 1% NP40 lysis buffer and 20 µg of protein lysate was loaded onto a 13.5% polyacrylamide gel and probed for p53 (DO-1 monoclonal antibody), p53-dependent activity using p21 (WAF-1 antibody, Calbiochem). (Bottom panel) Cells also demonstrate an intact IRF-1 response to  $25\mu$ g/ml polyI:C treatment over a 6h time course.  $\beta$ actin (Sigma) is a loading control.

**Figure 2.** Measuring the integrity of the AGR2-FRT cells by examining the effects of Tunicamycin on CHOP levels. (A). Immunoblotting of an FRT-A375 cell with an integrated AGR2 gene. A375 cells

were co-transfected with pcDNA5/FRT-AGR2 and pOG44 as in the legend of Figure 1 and hygromycin-resistant cell populations were isolated and evaluated for AGR2 protein expression relative to the parental FRT-A375 cells. (B). The levels of AGR2 in the isogenic cell population was evaluated relative to authentic AGR2-expressing cells including MCF7, OE19, A549, and T47D to ensure that above "physiological" levels of AGR2 protein were not being synthesized. In fact, relatively low levels of AGR2 protein are produced as defined by the immunoblot signal for AGR2 protein upon loading equivalent protein lysates (10  $\mu$ g) on the immunoblot thus ensuring that effects of low level expression of AGR2 on the cancer cell landscape can be evaluated.  $\beta$ -actin is a loading control. (C and D). Measuring the integrity of the AGR2-FRT cell by examining the effects of Tunicamycin on CHOP levels. A prior report identified the first stress-induced function for endogenous AGR2 in the ER (22); that being attenuated CHOP levels induced by Tunicamycin stress in cells depleted of AGR2 using targeted siRNA. A time course of CHOP levels after Tunicamycin incubation at 1 µg or 4 µg/ml (C or D) demonstrates that AGR2 protein elevates the level of CHOP protein induced in response to stress (lanes 5 and 6). These data are consistent with the reduced levels of CHOP protein induced by Tunicamycin when AGR2 protein levels are depleted <sup>25</sup> and suggest that this artificially engineered AGR2-FRT cell line faithfully recapitulates the ER-stress response.

Figure 3. Measuring the integrity of the engineered FRT-AGR2 cell by evaluating the fate of AGR2-**ΔKTEL protein**. A prior report highlighted that AGR2 lacking its C-terminal ER retention site is secreted (11). A domain map of AGR2 is summarized in A and includes; (i) the N-terminal leader sequence; (ii) the CxxS thioredoxin module; (iii) the substrate (Reptin) binding loop (FVLLNLVY); and the (iv) KTEL endoplasmic retention motif. We determined whether the engineered A375 cells expressing AGR2-ΔKTEL has an active AGR2 secretion system in place as a metric to define integrity of the AGR2 pathway. Subcellular protein expression was visualised by immunofluorescence microscopy to identify the intracellular distribution. AGR2 mutants were constructed for transient mammalian transfection (pcDNA3-CW) and transfected into H1299 cells. (B) AGR2 mouse monoclonal antibody<sup>16</sup>; detected by goat anti-mouse 488nm secondary antibody) probed the localisation of the protein with ER-marker PDI (Enzo, detected by donkey anti-rabbit 594nm secondary antibody) and nuclear TO-PRO3 (Invitrogen). Wild-type AGR2 is frequently co-localised with the PDI endoplasmic reticulum marker in a perinuclear location, however, wt-AGR2 does not appear to always colocalise with PDI. (C). Immunoblotting of parental FRT, AGR2, and AGR2-ΔKTEL expressing cells demonstrate that, in lanes 1-3, wt-AGR2 is more easily detected that AGR2-ΔKTEL which is minimally detectable within the cell lysate. By contrast, in lanes 4-6, when conditioned media from cell culture was collected and concentrated 50 X using 10,000 kDa molecular weight cut

off filters, and 100  $\mu$ L of this concentrated media was loaded onto a 12% polyacrylamide gel and probed for AGR2 detection using the monoclonal AGR2 antibody, the majority of AGR2- $\Delta$ KTEL protein was found to be secreted into the exterior milieu.

**Figure 4. The effects of AGR2 expressing cells on the basal transcriptome**. (A and B). Evaluation of AGR2 and AGR2-ΔKTEL expression in FRT-cell panels. RNA was isolated from the indicated three cell panels and the extent of *agr2* mRNA (A) and *gapdh* mRNA (B) was evaluated using qRT-PCR as indicated in the methods. The data are potted as the hybridization signal as a function of PCR cycle number. (C) The circles of the Venn diagram indicate the numbers of significantly differentially expressed genes in the AGR2, AGR2-ΔKTEL, and AGR3 cell lines relative to the FRT-parent (Rank Products, percentage false present=0.05). (D) Heatmap to show transcriptional landscape of the triplicate AGR2, AGR2-ΔKTEL, and AGR3 cell lines relative to FRT-parent (colours are relative to mean of FRT-parental expression, red=increased, green=decreased, black=no change). The data demonstrate that a small cluster of predominantly GAGE related genes is differentially expressed in the wt-AGR2 cells, whilst only AGR3 mediates significant global changes in basal gene expression.

Figure 5. Summary of dominant protein expression changes as a result of AGR2 expression. (A).

Volcano plot of proteomic data presenting quantifiable proteins expression changes as a function of the statistical probability, the posterior error probability (PEP). (B). Using a proforma of expression changes of ± 20% compared to the parental cell line, and a minimum of 3 peptides used for protein identification and quantitation, a Venn diagram demonstrating relative changes in protein expression due to the expression of wild-type or C-terminal mutant AGR2 was composed. (C). Pathway annotation using *Ingenuity* highlighted that the dominant pathway with a predicted activation state (*'inhibited'*) by wt-AGR2 is the p53 pathway. E2F1 has a predicted activation state of *'activated'* and the activation state of HNF4a was undefined. (D). Summary of up and down-regulation of previously defined AGR2 interacting proteins<sup>15</sup> including DAG; CD59; HECTD1; MUC18 (Mucin family); and RUVBL2. Of these proteins, DAG and MUC18 protein levels are elevated selectively by wt-AGR2 protein.

**Figure 6. Validation of Ki-67 as a protein upregulated by AGR2**. (A) Cell cycle analysis of FRT-parent and FRT-AGR2 cells using FACS. The cells were harvested then fixed in ethanol followed by staining with propidium iodide and FITC-conjugated anti-Ki-67 antibody. Cell DNA contents were determined by FACS and the reactivity of Ki-67 was detected at 488nm and analyzed with FlowJo7 software. The data summarize the Ki-67 levels using FACS. (B). X-CELLigence growth curves of FRT-Parent cells, FRT-AGR2, and FRT-*STOP*-AGR2 plotted as impedance as a function of time. (C) Summary of the rate of cell growth of FRT-Parent cells, FRT-AGR2, and FRT-*STOP*-AGR2, from B, and doubling time (D).

**Figure 7. The effects of AGR2 on p53 protein levels**. As Ingenuity Pathway Analysis demonstrated that p53 was the most down-regulated pathway (Figure 6C), we evaluated the steady state levels of p53 protein in the panel of cells indicated. (A). Lysates from FRT-parent, FRT-AGR2, AGR2-ΔKTEL, and AGR2-KDEL were immunoblotted to define the levels of p53 protein in the basal state (lane 1-4) of 24 hours after cisplatin damage (lanes 5-8); as AGR2 and AGR3 are both known to independently mediate cisplatin resistance in xenografts<sup>16, 22</sup>. The levels of p53 protein were quantified and expressed as a ratio relative to β-actin. (B and C). A time course of p53 induction after cisplatin treatment in lysates from cells derived of (B) left set of lanes; FRT-parent, (B) right set of lanes FRT-AGR2, (C) left set of lanes; AGR2-ΔKTEL, and (C) right set of lanes AGR2-KDEL. The levels of p53 protein were quantified and expressed as a ratio relative to β-actin set of lanes are a ratio relative to β-actin.

**Figure 8. The effects of TSG101 expression on p53 protein levels**. (A and B). Targeted siRNA depletion of AGR2 protein in A549 cells (left panel) or MCF7 cells (middle panel) results in p53 protein stabilization in authentically expressing AGR2 cells. (C). TSG101 protein is upregulated in the engineered FRT-AGR2 cell line relative to FRT parent, consistent with the SILAC data in Figure 6. (D). Transfection of increasing amount of the TSG101 expression plasmids (as indicated) was followed by twenty four hour incubation and then a 3 hour incubation with MG132 to determine whether TSG101 effects p53 turnover in A375 cells. Lysates were blotted with antibodies to p53, TSG101, and a β-actin loading control. (E and F). TSG101 was transfected into authentically expressing AGR2 cells (as indicated, A549 or MCF7) and followed by the addition of carrier, cisplatin, or MG132, as indicated, to determine whether TSG101 suppresses p53 protein in response to cisplatin damage. Lysates were blotted with antibodies to p53, p21, TSG101, and a β-actin loading control. (G and H). Targeted siRNA was transfected into cells to determine whether depletion of TSG101 in authentically expressing AGR2 cells altered the turnover of p53 protein, as defined by p53 protein levels without or with MG132. Lysates were blotted with antibodies to p53, p21, TSG101, and a β-actin loading control.

**Figure 9. Dominant effects of engineered expression AGR2 on the cellular proteome.** (I) AGR2 is an ER resident PDI that can regulate the secretion of the Mucin family of proteins<sup>6</sup> as well as interacting with pro-oncogenic receptors as defined by yeast-two hybrid, including Dystroglycan and C4.4A, that might contribute to its pro-metastatic properties<sup>21</sup>. (II) AGR2 also has a role outwith the ER where it can signal to nuclear transcription pathways to upregulate EGFR receptor pathways<sup>28</sup> and interact with the chaperone protein Reptin<sup>64</sup>. (III) The relatively low level expression of AGR2 in this report has a comparatively minimal effect on the transcriptome of the cell, but SILAC analysis has shown the dominant reprogramming of the proteome that occurs by virtue of AGR2 production. This

includes the up-regulation of the proliferation associated protein Ki67 and TSG101, the latter of which provides a signalling mechanism that can explain in part how AGR2 can attenuate the specific activity of p53 in response to DNA damage.

**Table 1. Proteomics changes induced by AGR2 isoforms.** Lists of the top ten up regulated or down regulated proteins in the (A) AGR2 or (B) AGR2-ΔKTEL expressing cells. The protein IDs are summarized with respect to normalized expression, Log2 ratios, Uniprot ID, peptides identified, and % coverage of the target.



218x296mm (150 x 150 DPI)



133x281mm (150 x 150 DPI)



250x193mm (150 x 150 DPI)



210x281mm (150 x 150 DPI)



С	Transcription Regulator	p-value of Overlap	Predicted Activation State
	HNF4a	$1.70 \times 10^{-5}$	
	TP53	$1.42 \times 10^{-4}$	Inhibited
	E2F1	$2.20 \times 10^{-4}$	Activated
	E2F2	$6.73 \times 10^{-4}$	
	TFEB	$1.16 \times 10^{-3}$	Inhibited
	Transcription Regulator	p-value of Overlap	Predicted Activation State
F	Transcription Regulator HNF4A	<b>p-value of Overlap</b> $6.18 \times 10^{-8}$	Predicted Activation State
	Transcription Regulator HNF4A SREBF1	p-value of Overlap $6.18 \times 10^{-8}$ $3.81 \times 10^{-5}$	Predicted Activation State
	Transcription Regulator HNF4A SREBF1 MCYN	p-value of Overlap $6.18 \times 10^8$ $3.81 \times 10^{-5}$ $6.90 \times 10^{-5}$	Predicted Activation State
	Transcription Regulator HNF4A SREBF1 MCYN IRF2	p-value of Overlap $6.18 \times 10^{-8}$ $3.81 \times 10^{-5}$ $6.90 \times 10^{-5}$ $1.04 \times 10^{-5}$	Predicted Activation State





203x275mm (150 x 150 DPI)





228x164mm (150 x 150 DPI)



211x289mm (150 x 150 DPI)



185x111mm (150 x 150 DPI)

Protein	Protein	Normalized	Log2	Uniprot	Peptides	Sequence
ID	Name	Relative	Ratios			Coverage
		Expression				(%)
		(UP)				
IPI00018434	TSG101	3.5344	1.8215	Q99816-1	2	4.4
IPI00004233	KI-67	3.1386	1.6501	P46013-1	3	2.8
IPI00002335	Huntington	2.7066	1.4365	P42858	3	1.9
IPI00941193	HLA-B	2.5402	1.3449	A3F719	4	12.2
IPI00743894	ZFC3H1	2.2781	1.878	060293-1	3	3.6
IPI00747462	FLJ54004	2.2355	1.1606	B4DJ76	9	43.9
IPI00295772	P450-51	1.9671	0.9761	Q16850-1	4	14.7
IPI00550882	PYCR1	1.9460	0.9605	B4DMU0	3	10.4
IPI00844014	C90RF114	1.8037	0.8510	Q5T280	2	9.6
IPI00018196	NLE1	1.7595	0.8151	Q9NVX2	1	2.3
Protein	Protein	Normalized	Log2	Uniprot	Peptides	Sequence
ID	Name	Relative	Ratios			Coverage
		Expression				(%)
		(DOWN)				
IPI00021634	171.00					
	KLC2	0.0313	-4.9992	Q9H0B6	6	17.4
IPI00554525	KLC2 NHS protein	0.0313 0.0516	-4.9992 -4.2773	Q9H0B6 Q6T4R5-2	6 1	17.4 1
IPI00554525 IPI00027280	KLC2 NHS protein TOP2B	0.0313 0.0516 0.0725	-4.9992 -4.2773 -3.7851	Q9H0B6 Q6T4R5-2 Q02880-1	6 1 2	17.4 1 2.8
IPI00554525 IPI00027280 IPI00027232	KLC2 NHS protein TOP2B IGF1R	0.0313 0.0516 0.0725 0.2203	-4.9992 -4.2773 -3.7851 -2.1825	Q9H0B6 Q6T4R5-2 Q02880-1 P08069	6 1 2 3	17.4 1 2.8 6.7
IPI00554525 IPI00027280 IPI00027232 IPI00409601	KLC2 NHS protein TOP2B IGF1R RALGAPB	0.0313 0.0516 0.0725 0.2203 0.2573	-4.9992 -4.2773 -3.7851 -2.1825 -1.9585	Q9H0B6 Q6T4R5-2 Q02880-1 P08069 Q86X10-1	6 1 2 3 3	17.4 1 2.8 6.7 4.7
IPI00554525 IPI00027280 IPI00027232 IPI00409601 IPI00296388	KLC2 NHS protein TOP2B IGF1R RALGAPB BAZ2A	0.0313 0.0516 0.0725 0.2203 0.2573 0.2918	-4.9992 -4.2773 -3.7851 -2.1825 -1.9585 -1.7772	Q9H0B6 Q6T4R5-2 Q02880-1 P08069 Q86X10-1 Q9UIF9-1	6 1 2 3 3 1	17.4 1 2.8 6.7 4.7 1.5
IPI00554525 IPI00027280 IPI00027232 IPI00409601 IPI00296388 IPI00032038	KLC2 NHS protein TOP2B IGF1R RALGAPB BAZ2A CPT1A	0.0313         0.0516         0.0725         0.2203         0.2573         0.2918         0.3068	-4.9992 -4.2773 -3.7851 -2.1825 -1.9585 -1.7772 -1.7046	Q9H0B6 Q6T4R5-2 Q02880-1 P08069 Q86X10-1 Q9UIF9-1 P50416-1	6 1 2 3 3 1 4	17.4 1 2.8 6.7 4.7 1.5 8.7
IPI00554525 IPI00027280 IPI00027232 IPI00409601 IPI00296388 IPI00032038 IPI00386208	KLC2 NHS protein TOP2B IGF1R RALGAPB BAZ2A CPT1A YA61	0.0313         0.0516         0.0725         0.2203         0.2573         0.2918         0.3068         0.3234	-4.9992 -4.2773 -3.7851 -2.1825 -1.9585 -1.7772 -1.7046 -1.6285	Q9H0B6 Q6T4R5-2 Q02880-1 P08069 Q86X10-1 Q9UIF9-1 P50416-1 Q9NZ23	6 1 2 3 3 1 4 8	17.4 1 2.8 6.7 4.7 1.5 8.7 80.9
IPI00554525 IPI00027280 IPI00027232 IPI00409601 IPI00296388 IPI00032038 IPI00386208 IPI00301139	KLC2 NHS protein TOP2B IGF1R RALGAPB BAZ2A CPT1A YA61 MED17	0.0313         0.0516         0.0725         0.2203         0.2573         0.2918         0.3068         0.3234         0.3520	-4.9992 -4.2773 -3.7851 -2.1825 -1.9585 -1.7772 -1.7046 -1.6285 -1.5064	Q9H0B6 Q6T4R5-2 Q02880-1 P08069 Q86X10-1 Q9UIF9-1 P50416-1 Q9NZ23 Q9NVC6-1	6 1 2 3 3 1 4 8 3	17.4 1 2.8 6.7 4.7 1.5 8.7 80.9 8.6

Page	38	of	38
------	----	----	----

Protein	Protein	Normalized	Log2	Uniprot	Peptides	Sequence
ID	Name	Relative	Ratios			Coverage
		Expression				(%)
		(UP)				
IPI00941193	HLA-B	3.899543	1.963305	A3F719	4	12.2
IPI00890779	DYS-F	2.877175	1.524653	075923-13	31	20.8
IPI00935722	DUBA5	2.676563	1.420381	Q8N6M0	1	5.8
IPI00845373	NFKB2	2.303684	1.203943	Q00653-1	12	19.9
IPI00409607	CAMP2	2.210841	1.144595	Q08AD1-1	2	2.4
IPI00304232	WDR12	2.180031	1.124348	Q9GZL7	2	7.1
IPI00294578	TGM2	2.167843	1.11626	P21980-1	29	51.8
IPI00101374	TM9S1	2.140867	1.098195	015321	1	1.5
IPI00031768	НООКЗ	2.094993	1.066945	Q86VS8	2	3.2
IPI00004472	WNK1	2.044053	1.031432	Q9H4A3-1	6	3.1
Protein	Protein	Normalized	Log2	Uniprot	Peptides	Sequence
ID	Name	Relative	Ratios			Coverage
		Expression				(%)
		(DOWN)				
IPI00027280	ТОР2В	0.129716	-2.94657	Q02880	2	2.8
IPI00072377	SET	0.196326	-2.34868	Q01105-1	8	41
IPI00554525	NHS protein	0.22967	-2.12237	Q6T4R5-2	1	1
IPI00002335	Huntington	0.235531	-2.08601	P42858	3	1.9
IPI00410034	SLC38A2	0.238448	-2.06826	Q96QD8-1	5	16.6
IPI00217468	Histone H1.5	0.254236	-1.97576	P16401	3	16.4
IPI00012545	TGN51	0.277091	-1.85157	043493-1	5	13.3
IPI00644502	GARL3	0.2878	-1.79686	Q5VVW2-1	2	2
IPI00026154	FLJ59211	0.311942	-1.68065	B4DJQ5	29	63.4
IPI00333015	SPTBN1	0.32176	-1.63594	Q01082-3	140	74.3