

METHOD

Open Access

A genome scale overexpression screen to reveal drug activity in human cells

Anthony Arnoldo^{1,2,3}, Saranya Kittanakom^{1,2,3}, Lawrence E Heisler^{1,2,3,4}, Anthony B Mak^{1,3}, Andrey I Shukalyuk⁵, Dax Torti⁴, Jason Moffat^{1,3}, Guri Giaever^{1,3,6,7} and Corey Nislow^{1,2,3,4,7*}

Abstract

Target identification is a critical step in the lengthy and expensive process of drug development. Here, we describe a genome-wide screening platform that uses systematic overexpression of pooled human ORFs to understand drug mode-of-action and resistance mechanisms. We first calibrated our screen with the well-characterized drug methotrexate. We then identified new genes involved in the bioactivity of diverse drugs including antineoplastic agents and biologically active molecules. Finally, we focused on the transcription factor RHOXF2 whose overexpression conferred resistance to DNA damaging agents. This approach represents an orthogonal method for functional screening and, to our knowledge, has never been reported before.

Background

Biological systems tend to remain phenotypically stable in the face of environmental challenges and genetic changes [1]. As such, genetic perturbation has become an efficient technique to dissect cellular functions. In this study, we used the concept of modulating gene dosage in human cells to gain insight into drug mode of action. Understanding the primary mechanism of action as well as the potential polypharmacological effects of a drug can provide insight into how to reduce detrimental side effects, uncover new applications for novel indication and explain resistance mechanisms [2-4].

Drug resistance in malignant tissues can be categorized into three main mechanisms: (i) drug distribution/metabolism (pharmacokinetics), (ii) heterogeneity of cancer cells and (iii) tumor micro-environment [5,6]. Among the other cellular mechanisms, gene overexpression (for example, by amplification of the drug target) can titrate a drug's effect. This is exemplified by the classic case of methotrexate resistance through the amplification of the gene *DHFR* in neoplastic tissue from an individual with disseminated small-cell lung cancer that relapsed during methotrexate chemotherapy [7]. Other forms of overexpression resistance

include the up-regulation of 1) efflux pumps (for example, ATP-binding cassette transporters), 2) survival mechanisms (for example, anti-apoptotic proteins), 3) DNA damage repair, 4) pathways for drug inactivation and 5) the overexpression of target isotypes. Additionally, proteins downstream of the inhibited target can be modulated in such a manner as to bypass the toxic effect of a drug.

To dissect some of these mechanisms by which drugs act within the cell, we postulated that, when overexpressed, genes conferring resistance to a lethal chemical treatment can illuminate the drug mode of action. Overexpressing genes to confer resistance to an otherwise toxic compound is a well-established concept. Early studies showed that a streptomycin resistance gene cassette could function in a bacterial tetracycline resistant plasmid [8] and it was shown in other studies that cloning of the mouse dihydrofolate reductase into a bacterial plasmid provided resistance to trimethoprim in *Escherichia coli* [9]. A large-scale approach using the overexpression of a yeast genomic DNA library to identify genes conferring resistance to specific drugs has validated the concept of gene dosage for the discovery of drug targets in eukaryotes [10].

A logical extension of this concept is to employ newly available biological tools that enable the systematic perturbation of all protein function in human cultured cells using RNA interference (RNAi), CRISPR (clustered regularly interspaced short palindromic repeats), cDNA libraries, transposons or small molecule inhibitors in combination

* Correspondence: corey.nislow@ubc.ca

¹Department of Molecular Genetics, University of Toronto, Toronto, M5S 3E1, Canada

²Banting and Best Department of Medical Research, University of Toronto, Toronto, M5S 3E1, Canada

Full list of author information is available at the end of the article

with gene overexpression. Although testing all important therapeutic drugs against all proteins from the approximately 20,000 genes of the human genome in all differentiated cell types is not currently feasible, we report the initial development of a gain-of-function approach to identify drug mode of action by the overexpression of 12,200 human ORFs (hORFs) [11,12] in the human cell line HEK293.

In summary, we have developed a new experimental pipeline to identify genes whose up-regulation suppresses the toxic effect of chemicals in human cells and have successfully applied this strategy to seven pharmacological compounds.

Methods

Cell culture and plasmids

Human embryonic kidney HEK293 (ATCC), HEK293T (ATCC), non-small cell lung carcinoma NCI-H1299 (ATCC), chondrosarcoma SW1353 (gift from Johanne Martel-Pelletier, University of Montreal, Canada) and pancreatic adenocarcinoma HPAC (ATCC) cells were maintained in DMEM (Wisent Inc. Montreal, Quebec, Canada) supplemented with 10% fetal bovine serum (Gibco, Carlsbad, California, US) at 37°C and 5% CO₂. The rtTA expressing HEK293_M2 (ATCC), human breast adenocarcinoma MCF7_M2 (ATCC) and adenocarcinomic human alveolar basal epithelial A549_M2 (ATCC) cells were maintained in DMEM (Wisent Inc.) supplemented with 10% fetal bovine serum (Gibco) at 37°C and 5% CO₂. Blast phase chronic myelogenous leukemia K562 cells (gift from Reinhart Reithmeier, University of Toronto, Canada) were maintained in RPMI1640 (Sigma, St. Louis, Missouri, US) supplemented with 10% fetal bovine serum (Gibco) at 37°C and 5% CO₂. Multiple myeloma U266B1 cells (gift from Aaron Schimmer, Princess Margaret Cancer Centre, Canada) were maintained in Iscove's modified Dulbecco's medium (IMDM; Gibco) supplemented with 10% fetal bovine serum (Gibco) at 37°C and 5% CO₂. Thyroid gland medullary carcinoma TT cells (gift from Gilbert Cote, MD Anderson Cancer Center, USA) were maintained in Ham's F-12 K (Gibco) supplemented with 10% fetal bovine serum (Gibco) at 37°C and 5% CO₂.

The lentiviral destination vector pLD-T-IRES-Venus-WPRE-STOP was cloned using the following procedure: the Gateway cassette from pLS-Dest-EcF (gift of Dr Tony Pawson, University of Toronto, Canada) was subcloned into pLJM17 (J Moffat, unpublished) using the restriction enzymes MluI and XbaI, resulting in the pLD-puro-TRE (T) vector. IRES-Venus-WPRE were PCR amplified from the pSLIK-Venus [13] using the following primers: XcmI_IRES_F (5'-GCGCCTTTTCCAAGGCAGCCCTGGAATTCGCCCCCTCTCCCTCC) and NsiI_WPRE_R (5'-AAACAATGCATGTCGACGCGGGGAGGCGGCCCAAAGGGAATCC). The IRES-Venus-WPRE amplicon was cloned

into the pLD-puro-T vector using the restriction enzymes XcmI and NsiI to replace the human phosphoglycerate kinase promoter and puromycin resistance gene. A stop codon was introduced directly downstream of the Gateway cassette by digesting pLD-T-IRES-Venus-WPRE with the restriction enzyme XbaI and blunted using DNA polymerase I, large (Klenow) fragment (New England Biolabs Inc., Beverly, Massachusetts, US).

The *piggyBac* PB-TGcMV-Neo plasmid used for the hit confirmation was derived from the PB-TET (AddGene, Cambridge, Massachusetts, US). The 'IRES-beta-Geo' fragment from PB-TET was replaced by the 'promoter PGK-neomycin resistance gene' fragment by homologous recombination in *Saccharomyces cerevisiae*. Briefly, PB-TET was digested with RsrII and ApaI to remove the beta-geo gene. The fragment coding for the promoter PGK and neomycin resistance gene, the URA3 cassette and the yeast 2 μ-origin of replication were PCR amplified and cloned into the digested PB-TET plasmid by homologous recombination in yeast. Yeast was then transformed with the p414-Cre plasmid in order to excise the URA3 cassette and the 2 μ-origin of replication after expression of the Cre recombinase.

All short hairpin RNAs (shRNAs) against the RHOXF2 and green fluorescent protein (GFP) mRNAs were derived from the RNAi Consortium (TRC) lentiviral libraries and obtained in the pLKO.1 vector. Five shRNAs against RHOXF2 were tested and were designed against the following regions: 5'-CGGGATGAGAGATGATTACTT (RHOXF2_shRNA_1); 5'-GACGAGAAAGAACAACAGGAT (RHOXF2_shRNA_2); 5'-AGAAGCATGATGTGACTGAA (RHOXF2_shRNA_3); 5'-GAGGGCATTAATGGCAAGAAA (RHOXF2_shRNA_4); 5'-GTCGCTTACTGAAGAGGTCAA (RHOXF2_shRNA_5).

Library preparation, virus production and lentiviral infection

The 12,212 hORF collection (representing 10,214 distinct genes) represents version 3.1 of an ongoing effort to create a complete human set of protein-encoding genes [11,12]. Initially based on the Mammalian Gene Collection cDNA collection, that work transfers full-length ORFs (excluding the 5' and 3' mRNA untranslated regions) into a Gateway system. Version 3.1 was obtained from Open Biosystems (Thermo Scientific, Waltham, Massachusetts, US) and re-groups non-fully sequenced verified and non-clonal ORFs (limitations corrected in the 8.1 version [14]). The collection was divided into 34 minipools of 376 hORFs; for each minipool, the hORFs were cloned *en masse* from the pDONR223 into the lentiviral expression vector pLD-T-IRES-Venus-WPRE-STOP by Gateway LR reaction. After electroporation, transformants were selected on Luria Broth (LB) plus ampicillin and the lentiviral vector was extracted.

Lentivirus was produced by normalizing the amount of DNA for the 34 hORF minipools and by co-transfection with the packaging plasmid psPAX2 and the envelope plasmid pMD2.G into the packaging HEK293T cells using FuGENE (Roche Mississauga, Ontario, Canada).

HEK293_M2 cells were then infected at a multiplicity of infection of 0.3. Doxycycline (2 µg/ml) was added to induce the expression of the Venus fluorescent protein and Venus-positive cells were sorted using a BD FACS Aria Cell Sorter (East Rutherford, New Jersey, US). Because the expression of the gene coding for Venus is linked to the expression of the hORFs, selection of the fluorescent cells allowed the selection for non-silent, stably integrated lentivirus as well as functional inducible hORFs.

Genomic DNA extraction, library preparation and data analysis for the hORFeome representation in HEK293_M2 cells

We thawed 6.3×10^6 HEK293_M2 cells harboring the virally integrated human ORFeome version 3.1 collection and these were cultured for 1 day to recover. Genomic DNA was recovered from 16×10^6 cells and 6 µg of genomic DNA (gDNA) were used to amplify the collection (three PCR reactions with 2 µg of gDNA). Following PCR, the amplified ORFs were used directly to prepare a Nextera sequencing library (Illumina Nextera DNA Sample Preparation Kit, San Diego, California, US) and sequenced on an Illumina MiSeq according to the manufacturer's protocol. Fifty-nucleotide paired-end reads were aligned against a reference database consisting of the ORF sequences in the hORFeome 3.1 collection using bwa (version 0.6.1). Coverage at each base position in each reference sequence was determined using the bedtools program genomeCoverageBed (v2.14.2). A custom Perl script was used to summarize the coverage levels to determine the mean coverage for each reference sequence and the proportion of each reference sequence with different levels of coverage. Raw data have been deposited in the ArrayExpress repository under accession number E-MTAB-2498.

hORFeome drug target screen

Aliquots of 500,000 HEK293_M2 cells were seeded in a T175 flask and grown for 15 hours before the induction of hORF expression by addition of doxycycline. After 18 hours of induction, cells formed micro-colonies that were cultured in the presence of drug or DMSO (as control) for 2 to 3 weeks until distinct colonies were detectable.

After each screen, surviving cells were harvested and gDNA was extracted as follows. Cells were lysed using SNET buffer (10 mM Tris pH 8.0, 0.1 M EDTA, 0.5% SDS, 0.1 mg/ml Proteinase K, 25 µg/ml RNase A). gDNA was isolated using one volume of phenol/chloroform/isoamyl alcohol (25:24:1) pH 8.0, precipitated by adding two

volumes of 95% ethanol (-20°C), washed with ice cold 70% ethanol and resuspended in Tris-HCl pH 8.0.

During the screen selection experiments, only a minority of hORFs protected HEK293_M2s against the cytotoxic effect of the drug. Compared to the initial cell population, HEK293_M2 cells harboring the most resistant genes should be over-represented. In order to further amplify the set of resistant genes from the resulting reduced population, hORFs were PCR amplified from only 80 ng of genomic DNA. The sequences of the forward and reverse primers specific for the inserted hORFs were 5'-CGGTACCCGGGGATCCTCTAGTCAGCTGAC and 5'-CCATTTGTCTCGAGGTTCGAGAATTCTAGCTAGAATC, respectively. Each reaction was carried out in a 50 µl volume containing 25 µl of 2× Phusion Flash High-Fidelity Master Mix (Finnzymes, Espoo, Finland), 200 nM of each primer and 80 ng of genomic DNA. The PCR profile was: 1 minute at 98°C for one cycle; 10 s at 98°C, 20 s at 65°C, 4 minutes at 72°C for 35 cycles; 10 minutes at 72°C for one cycle. PCR products were purified using QIAquick PCR Purification Kit (QIAGEN Venlo, Limburg, Netherlands), 150 ng of purified PCR product was biotinylated using the BioPrime DNA Labeling Kit (Invitrogen, Carlsbad, California, US) and unincorporated biotin-14-dCTP was removed by passing the samples through Sephadex G-50 columns (GE Healthcare Fairfield, Connecticut, US). Sample (150 ng) was added to the Affymetrix GeneChip Human Gene 1.0 ST array and hybridized at 45°C for 17 hours with a rotation of 60 rpm. Chips were washed and stained with SAPE (2× MES staining buffer, 20 mg/mL bovine serum albumin (BSA) and 1 mg/ml streptavidin-phycoerythrin), washed on an Affymetrix fluidics station and scanned.

Representation of the complete hORFeome collection was assessed using the same experimental conditions with some modifications. gDNA was extracted from at least 16×10^6 cells (approximately 1,300 cells per hORF), 3 PCR reactions were performed using 2 µg of gDNA each (approximately 75 genomes per hORFs), two biotinylation reactions were performed using 500 ng of combined PCR products and 3.5 µg of sample was added to an Affymetrix GeneChip Human Gene 1.0 ST array.

Microarray raw data have been deposited in the ArrayExpress repository [15] under accession number E-MTAB-2493.

Selection and individual validation of over-represented hORFs

The hORFs conferring resistance to the drug were identified in a step-wise procedure. Screen results were first visualized by plotting the log₂ of the signal intensity for each hORF retrieved via PCR from cells cultured in the presence of drug (on the x-axis) and by plotting the log₂ ratio of the signal intensity for each hORF of the cells

cultured in the presence of drug divided by the signal intensity for each hORF of the cells grown in the presence of DMSO (on the y-axis). Genes with a $\log_2(\text{drug}/\text{DMSO}) > 3$ and a $\log_2(\text{drug}) > 6$ enrichment were then listed and ranked. Primary hits selected for validation were those previously enriched genes in common among the drug screen replicates (at least two independent experiments).

For validation, 10,000 HEK293-M2 cells were seeded in 300 μl of DMEM plus 10% iFBS plus penicillin/streptomycin in a 48-well plate pretreated with poly-L-lysine (Sigma, P4832). After 12 hours of incubation, gene expression was induced by the addition of doxycycline (2 $\mu\text{g}/\text{ml}$ final concentration). After 18 hours of induction, drug was added and cell growth was assessed 3 days after drug addition by sulforhodamine B (SRB) assay as previously described [16].

Immunostaining

The HEK293_M2 + RHOXF2 stable cell line (PB-TGcMV-Neo) were seeded in an eight-well chamber slide (BD Falcon, East Ruherford, New Jersey, US) and cultivated overnight in the presence or absence of doxycyclin (2 $\mu\text{g}/\text{ml}$ final). Cells were treated with 50 nM of mitomycin C for 2 days before fixation with 2% paraformaldehyde and permeabilization with 0.3% Triton X-100. Cells were blocked for 30 minutes with blocking/dilution buffer (10% goat serum, 0.5% NP-40, 0.5% saponin, 1 \times phosphate-buffered saline (PBS)) and incubated with the anti-phospho-histone H2A.X (serine 139, clone JBW301; Millipore, Billerica, Massachusetts, US) overnight at 4°C. Cells were then incubated with an Alexa 546 anti-mouse antibody for 1 hour at room temperature, incubated with DAPI (0.8 $\mu\text{g}/\text{ml}$; Sigma) for 10 minutes and mounted using ProLong Gold (Invitrogen). Images were captured using a 40 \times dry objective. Nuclei and γ -H2A.X foci were quantified using CellProfiler [17].

RNA extraction and library preparation for gene expression profiling of the RHOXF2 overexpressing cell line

The HEK293_M2 + RHOXF2 stable cell line (PB-TGcMV-Neo) was seeded in 10-cm dishes, cultivated 24 hours in the presence or absence of doxycycline (2 $\mu\text{g}/\text{ml}$ final) before being treated with 40 nM of mitomycin C or DMSO (control) during a 2-day period. Total RNA was extracted (RNeasy Mini Kit, QIAGEN), mRNA-focused libraries were generated from 1 μg of total RNA (Illumina TruSeq RNA Sample Preparation Kit V.2) and sequenced using an Illumina HiSeq2000 according to the Illumina protocol. Raw data have been deposited in the ArrayExpress repository under accession number E-MTAB-2497.

RNA-sequencing data analysis

Single-end reads, 51 nucleotides in length, were generated on the Illumina HiSeq. Sequence data were aligned

to the UCSC hg19 reference genome using TopHat (v2.0.0), provided with a RefSeq GTF file and instructed to align only across known junctions. Differential expression, as fragments per kilobase of exon per million fragments mapped (FPKM) differences, were generated by analysis with cuffdiff (v1.1.0) using the indicated comparisons. Based on their differential expression, potential activated and repressed isoforms were identified and related enriched biological processes were assessed using GOrilla [18].

RHOXF2 quantitative RT-PCR

Total RNA from 20 different normal human tissues (each pool from 3 donors) were obtained from the First-Choice Human Total RNA Survey Panel (Ambion, Austin, Texas, US). Total RNA (5 μg) from each tissue was used for oligo(dT)12-18-primed reverse transcription using the SuperScript II reverse transcriptase (Invitrogen). Quantitative PCR was performed using the LightCycler 480 SYBR Green I Master (Roche) and the LightCycler 480 System with the following primers:

QhRHOXF1_1F (5'-ACCGTGTCTACTGCCTGAGT GTA) and QhRHOXF1_1R (5'-TTCATGCCGTTCTCGT GGTTCACA) on RHOXF1 exon 1; QhRHOXF1_2F (5'-T GGAGGAGCTGGAAAGTGT) and QhRHOXF1_2.2R (5'-GGCCCTTTTATTCTTAAACC) spanning RHOXF1 exon 1 and exon 3; QhRHOXF2_1F (5'-CCGGACCAGT GTAGCCAGTA) and QhRHOXF2_1R (5'-TCTTTTCT TCTCCGCCTTG) spanning RHOXF2 exon 1 and exon 2; QhRHOXF2_2F (5'-ATGGTGCTGCTGCTTACTGA) and QhRHOXF2_2R (5'-TCGAGGTCTCCTTCCCATAG) on RHOXF2 exon 2; QhRHOXF2_4F (5'-CAGCGGGATGAG AGATGATT) and QhRHOXF2_4R (5'-TTGGGGAATGT GAAAGAAGG) on RHOXF2 exon 3.

Housekeeping/reference genes were: QhCYCG_F (5'-CT TGTCATGGCCAACAGAGG) and QhCYCG_R (5'-GC CCATCTAAATGAGGAGTTGGT) on CYCG (cyclophilin G); QhGUSB_F (5'-ACGCAGAAAATATGTGGTTGGA) and QhGUSB_R (5'-GCACTCTCGTCGGTACTGTT) on GUSB (beta-glucuronidase); QhActbF1 (5'-GAAGTC CCTTGCCATCCTAAAAG) and QhActbR1 (5'-AGGA CTGGGCCATTCTCCTTA) on ACTB (beta-actin); QhEef1a1F2 (5'-CTGAACCATCCAGGCCAAAT) and QhEef1a1R2 (5'-AGCCGTGTGGCAATCCA) on EEF1A1 (eukaryotic translation elongation factor 1 alpha 1); QhGAP DH_F1 (5'-CATGAGAAGTATGACAACAGCCT) and QhGAPDH_R1 (5'-AGTCCTTCCACGATACCAAAGT) on GAPDH (glyceraldehyde-3-phosphate dehydrogenase).

RHOXF2 immunohistochemistry

Formalin-fixed, paraffin-embedded sections of human normal testis were deparaffinized for 12 hours at 58°C and rehydrated. Heat-induced epitope retrieval was in sodium citrate buffer pH 6.0 using a pressure cooker

[19]. All antibodies were diluted in Tris-buffered Saline (TBS) with 1% BSA. Sections were blocked in 10% goat serum (Gibco) for 2 hours at room temperature. The anti-RHOXF2 primary antibody (Sigma, HPA003314) was used at a 1:150 dilution and incubated overnight at 4°C in a humid chamber. Endogenous peroxidase activity was suppressed by incubating the section in 0.3% H₂O₂ for 15 minutes. Horse radish peroxidase (HRP)-conjugated secondary antibody (anti-rabbit IgG-HRP; GE Healthcare) was used at a 1:250 dilution and incubated for 1 hour at room temperature. 3,3'-Diaminobenzidine (Abcam) was used as chromogen. Sections were counterstained with Mayer's hematoxylin (Sigma) before dehydration and mounting.

Western blotting

For the detection of endogenous RHOXF2, total cell lysates were prepared from 200,000 cells. Briefly, samples were lysed in 2× SB loading buffer (60 mM Tris-HCl pH6.8, 3% SDS, 10% glycerol, 0.05% bromophenol blue, 10% beta-mercaptoethanol) and subjected to SDS-PAGE (12%). The RHOXF2-specific mouse polyclonal antibody (Abcam, ab67811) and the tubulin specific rat antibody (Abcam, ab6160) were used at a dilution of 1:800 and 1:40,000 respectively. Immunological complexes were visualized by enhanced chemiluminescence using HRP-conjugated anti mouse or anti rat IgG.

Drug dose-response curves

K562 cells (100,000) were cultured in 600 μl of media in a 48-well plate for 24 hours before the addition of diluted drug. Two days after drug treatment, the percentage of drug inhibition was assessed by counting K562 using a Coulter particle count and cell analyser and compared to cells grown in the presence of DMSO only.

For the adherent HPAC, SW1353, and NCI-H1299 cell lines, 5,000 cells were seeded in 300 μl of media in 96-well plate and incubated for 24 hours prior to drug addition. Following two days of drug treatment, the percentage of drug inhibition was estimated by SRB assay [16] and compared to cells grown in DMSO as control.

Results

Development of the assay: DHFR and methotrexate

We developed an original phenotypic assay in which the mammalian cell line HEK293_M2 was used to identify hORFs capable of rescuing small molecule toxicity when overexpressed (Figure 1; Additional file 1). As a proof of concept, we rescued methotrexate toxicity by overexpression of its target, the gene encoding the enzyme dihydrofolate reductase (DHFR).

We first cloned the human *DHFR* gene into the Gateway-compatible lentiviral vector pLD-T-IRES-Venus-WPRE-STOP where gene expression is under the control of the Tet-On promoter (Additional file 2). After transduction

of the rtTA-expressing cell line HEK293_M2 by lentivirus, we detected a high level of protein in the presence of doxycycline by western blotting (Figure 2a; Additional file 3). No protein was detected in the absence of doxycycline, suggesting robust control of transgene expression. Cells were next exposed to increasing doses of methotrexate in the absence or presence of different doses of doxycycline. Increasing levels of DHFR overexpression in HEK293_M2 cells rescued methotrexate toxicity in a dose-dependent manner, demonstrating that the lentiviral construct was functional (Figure 2b; Additional file 3).

We implemented our approach on a larger scale by cloning, *en masse*, a pool of 376 random hORFs (including *DHFR*) into the lentiviral vector. Corresponding lentivirus were used to generate a population of HEK293_M2 cells with the stably integrated hORFs. Gene expression was induced as before by addition of doxycycline and cells were exposed to a lethal concentration of methotrexate. Ten days after treatment, surviving cells were harvested, genomic DNA was extracted, and hORFs were PCR amplified and hybridized on an Affymetrix GeneChip Human Gene 1.0 ST array to determine which hORFs conferred resistance to methotrexate. *DHFR* was found as the predominant gene in this pool of 376 hORFs whose overexpression provided resistance to methotrexate, validating our strategy at the scale of several hundred pooled clones (Additional file 4).

The same approach was next applied to the complete collection of 12,212 hORFs. Here, we divided the collection into 34 minipools of 376 hORFs each, transferred them *en masse* into the pLD-T-IRES-Venus-WPRE-STOP lentivirus vector and then pooled 34 minipools together to produce sufficient quantities of lentivirus for the infection of HEK293_M2 cells at a multiplicity of infection of 0.3 (Figure 1). Because induction of the Venus marker is dependent on hORF expression, addition of doxycycline for 24 hours generated sufficient fluorescence to sort those cells with stable lentivirus integration (Additional file 5). The presence and relative abundance of the virally integrated hORFs were assessed using next-generation sequencing. Based on the identification of at least 95% of the sequence with a minimum of 10× coverage, 87% of the mappable hORFs were considered as present in the HEK293_M2 cells. Further analysis of the sequencing results demonstrated a size-dependent relative abundance of the hORFs, with smaller clones being more represented in the population (Additional file 6). An aliquot of 500,000 HEK293_M2 cells (an average of 41 cells per hORF) was seeded in a T175 flask, and grown for 15 hours before the induction of hORF expression by addition of doxycycline. After 18 hours of doxycycline induction, cells formed micro-colonies, which were cultured in the presence of methotrexate for 2 weeks. hORFs derived from the surviving cells were identified by PCR amplification

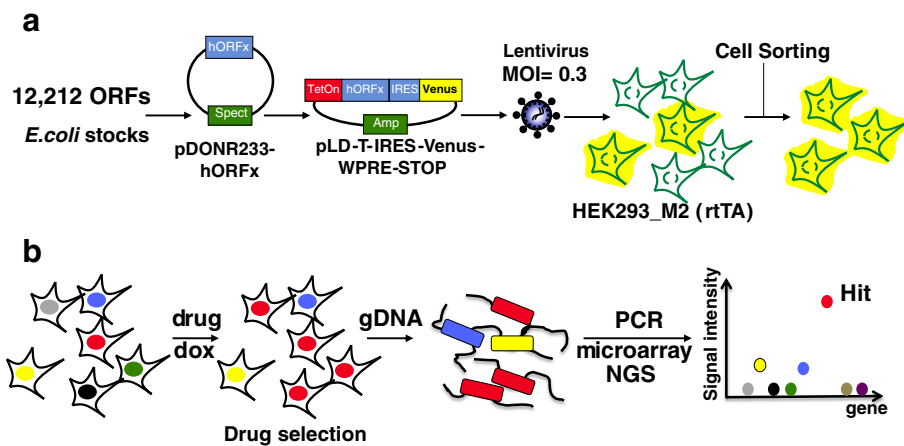


Figure 1 A gain-of-function cell-based assay to characterize the mode of action of small chemical compounds. **(a)** A collection of 12,212 hORFs was cloned *en masse* into the lentiviral expression vector pLD-IRES-Venus-WPRE-STOP. The resulting constructs were used to produce lentivirus and infect the rtTA-expressing cell line HEK293_M2 at a multiplicity of infection (MOI) of 0.3. Sorting of the Venus-positive cells resulted in the isolation of human cells with the functional integrated lentiviral constructs. **(b)** After seeding, doxycycline (dox) was added to the cells in order to induce the expression of the hORFs. After selection in the presence of a lethal dose of drug, surviving cells were harvested, gDNA extracted and the nature of the hORFs providing resistance to the chemical identified after hybridization on microarray. NGS, next-generation sequencing.

and microarray hybridization. As we found for the mini-pool experiment, *DHFR* was the top gene conferring resistance to methotrexate when overexpressed (Figure 2c). These results encouraged us to test the same strategy for drugs with diverse modes of action.

Screens with additional drugs

We tested six additional diverse drugs, including the antifolate aminopterin, the lipase inhibitor orlistat, the Hsp90 inhibitor 17-(dimethylaminoethylamino)-17-demethoxygeldanamycin (a geldanamycin analog) and

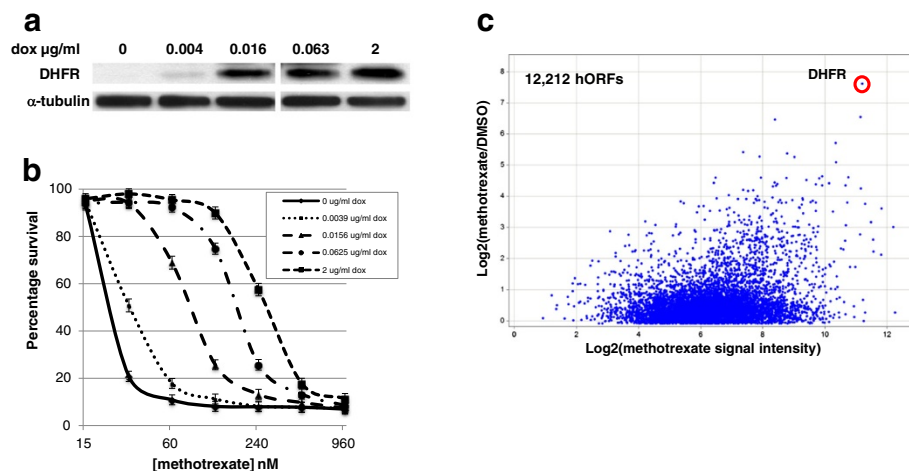


Figure 2 DHFR was identified as the top candidate during a large-scale screen against methotrexate. **(a)** Immunoblotting showing the conditional expression of DHFR in HEK293_M2 cells. Addition of 4 ng/ml, 16 ng/ml, 63 ng/ml and 2,000 ng/ml of doxycycline (dox) induced DHFR protein expression in the stable cell line HEK293_M2. Alpha tubulin was used as a loading control. For clarity, protein expression level at 0.0032 ng/ml of doxycycline has been omitted. **(b)** Single DHFR overexpression rescued methotrexate toxicity in HEK293_M2 cells. The gene coding for DHFR was cloned into the lentiviral vector and used to create stable HEK293_M2 cells able to conditionally express DHFR. Dose-response curves of the cells grown in the presence of an increasing concentration of doxycycline were calculated after 2 days of exposure to 15, 30, 60, 125, 250, 500 and 1,000 nM of methotrexate and compared with those for cells cultured in the presence of DMSO as control. Cell survival was quantified using a SRB assay. Error bars represent the standard deviation for triplicate assays. **(c)** Identification of genes conferring resistance to methotrexate when overexpressed. HEK293_M2 cells harboring the 12,212 hORF collection were grown in the presence of a lethal dose of methotrexate. The nature of the hORFs conferring resistance to the drug was identified by plotting the log₂ of the signal intensity for each hORF in the cells cultured in the presence of methotrexate on the x-axis; and by plotting the log₂ ratio of the signal intensity for each hORF of the cells cultured in the presence of the drug divided by the signal intensity for each hORF of the cells grown in the presence of DMSO on the y-axis.

three DNA damaging agents (bleomycin, cisplatin and mitomycin C) based on the diversity of their mode of action (Table 1).

Each drug was screened at least twice and the top hits (those hORFs most enriched in the final sample of cells; see Methods) were selected for validation. Of 123 candidates, we successfully cloned 120 hORFs into the modified Gateway-compatible *piggyBac* vector PB-TGcMV-Neo (Additional file 7) and generated 120 independent HEK293_M2 stable cell lines. Protection against the drug toxicity was assessed two days after culturing cells in the presence of four concentrations of each drug with or without doxycycline induction. Relative cell densities were compared with cells containing an empty vector control construct and were quantified by SRB assay. A total of 17 unique hits were confirmed to rescue their respective drug toxicity (Table 2). The number of validated hits varied considerably according to the compound, with bleomycin and 17-DMAG having low validation rates (7% and 3%, respectively) and, in contrast, cisplatin and mitomycin C having high validation levels (27% and 42%, respectively) (Table 2). Several scenarios, both technical and biological, may explain this modest confirmation rate. First, PCR amplification of the hORFs from gDNA and probe cross-hybridization on microarrays can be expected to generate a certain percentage of false positives, despite each screen having been performed in duplicate. Second, since this version of the hORFeome library is not fully

sequence verified, we speculate that some genes contain mutations that impair proper expression. We attempted to minimize the impact of mutations by restriction verifying four independent colonies from the original pDONR233 bacterial stock, cloning positive clones as a pool for subsequent cloning into the PB-TGcMV-Neo vector and transfection into HEK293_M2 cells. Finally, the biological mechanisms by which these agents work will influence the number of hits obtained from each screen, and therefore affect the number of hits that can be confirmed. For this proof-of-principle study we chose to be conservative, selecting more hits for validation at the cost of a high false-positive rate.

Screens for the antifolates methotrexate and aminopterin

For the antifolate screens, the single confirmed ORF was *DHFR*. This result highlights the fact that when a single hORF confers a strong growth advantage, it can dominate the other clones in the population and potentially mask the detection of other potential protective genes. One way to address this and potentially improve the dynamic range for this drug class would be to repeat these screens with a pool of hORFs lacking *DHFR*.

Among the potential expected hits present in our collection was the enzyme gamma-glutamyl hydrolase (GGH), which catalyzes the removal of polyglutamates from methotrexate. Therefore, overexpression of GGH would be expected to decrease the intracellular presence

Table 1 Characteristics of screened compounds

Drug name	Screen concentration	Category*	Clinical use	Known target(s)	Reference
Methotrexate	60 nM	Antimetabolites/ immunosuppressants	Antineoplastic Antirheumatic Dermatologic agent	DHFR	[20]
Aminopterin	20 nM	Antimetabolites (discontinued)	Dermatologic agent (phase I) Antineoplastic agent (phase II)	DHFR, FPGS(?)	[21]
Orlistat	20 μM	Antiobesity preparations, excluding diet products	Anti-obesity agent	LPL PNLIP FASN	[22] [23]
Bleomycin	1 μM	Cytotoxic antibiotics and related substances	Antineoplastic agent	DNA LIG1 LIG3	[24]
Cisplatin	2 μM	Other antineoplastic agents	Antineoplastic agent	DNA	[25]
Mitomycin C	80 nM	Cytotoxic antibiotics and related substances	Antineoplastic agent	DNA	[26]
17_DMAG	70 nM	Experimental Hsp90 inhibitor	Antineoplastic agent (phases I and II)	Hsp90 HSP90AA1 HSP90AA2 HSP90B1 HSP90AB1	[27]

Table 2 Validated hits

Drug	Gene name	Accession number	Description	Number of validated/tested candidate genes
Methotrexate	<i>DHFR</i>	BC000192	Dihydrofolate reductase	1/3
Aminopterin	<i>DHFR</i>	BC000192	Dihydrofolate reductase	1/9
Bleomycin	<i>IMMT</i>	BC002412	Inner membrane protein, mitochondrial (mitofilin)	2/30
	<i>STX3</i>	BC007405	Syntaxin 3	
Cisplatin	<i>MAP2K1IP1</i>	BC026245	Mitogen-activated protein kinase kinase 1 interacting protein 1	4/15
	<i>MMD</i>	BC026324	Monocyte to macrophage differentiation-associated	
	<i>PTPN2</i>	BC016727	Protein tyrosine phosphatase, non-receptor type 2	
	<i>RHOXF2</i>	BC021719	Rhox homeobox family, member 2	
Mitomycin C	<i>C20orf54</i>	BC009750	Chromosome 20 open reading frame 54	8/19
	<i>CCDC45</i>	BC009518	Coiled-coil domain containing 45	
	<i>ELF5</i>	BC029743	E74-like factor 5 (ets domain transcription factor)	
	<i>PTPN2</i>	BC008244	Protein tyrosine phosphatase, non-receptor type 2	
	<i>RHOXF2</i>	BC021719	Rhox homeobox family, member 2	
	<i>TMEM150</i>	BC050466	Transmembrane protein 150	
	<i>USPL1</i>	BC038103	Ubiquitin specific peptidase like 1	
	<i>ZFP64</i>	BC012759	Zinc finger protein 64 homolog (mouse)	
17-DMAG	<i>EIF4B</i>	BC073139	Eukaryotic translation initiation factor 4B	1/29
Orlistat	<i>MGLL</i>	BC006230	Monoglyceride lipase	2/18
	<i>PON3</i>	BC070374	Paraoxonase 3	

of methotrexate-polyglutamate, thereby increasing the efflux of methotrexate out of the cell and decreasing cytotoxicity. However, GGH overexpression was demonstrated to be insufficient to produce resistance to the drug in a human fibrosarcoma (HT-1080) and a human breast carcinoma (MCF-7) cell line [28].

DHFRL1 (dihydrofolate reductase like-1) consistently appeared as a top hit for screens with both methotrexate and aminopterin but failed to confer resistance when individually expressed in HEK293_M2 cells. Based on the strong similarity between DHFR and DHFRL1 (a total of 14 non-synonymous changes), we hypothesized that the observed unspecific signal was probably due to cross-hybridization. Interestingly, a similar problem occurred when the same strategy using the same hORFeome library was tested in the model organism *S. cerevisiae*. Both DHFR and DHFRL1 were enriched in the screens against methotrexate (and their expression confirmed at the protein level) but only DHFR conferred methotrexate resistance (data not shown).

Orlistat screens

Orlistat is a lipase inhibitor that reportedly targets LPL (lipoprotein lipase), PNLIP (pancreatic lipase) and FASN (fatty acid synthase). Although LPL is absent from our collection, PNLIP and FASN are present. However, genes for the latter two did not confer overexpression resistance.

Our screen identified the gene coding for monoglyceride lipase (MGLL) as a resistance hit. MGLL hydrolyzes intracellular triglyceride stores in adipocytes and other cells to fatty acids and glycerol. The enzyme might also be involved in the hydrolysis of monoglycerides [29]. *In vitro* experiments suggest that orlistat can inhibit the MGLL-like activity in rat-cerebellar membranes and rat cerebellar homogenate [30], rendering the human MGLL a potential direct target for orlistat.

Another validated hit, *PON3*, is also involved in lipid metabolism. The gene encodes paraoxonase 3, which is secreted into the bloodstream and associates with high-density lipoprotein. The protein can also rapidly hydrolyze lactones and inhibit the oxidation of low-density lipoprotein.

Bleomycin screens

Although the exact mechanism of action of bleomycin is unknown, available evidence indicates that its main mode of action is via the inhibition of DNA synthesis, with additional evidence for its inhibition of RNA and, to a lesser extent, protein synthesis [31].

We show that overexpression of *IMMT* (inner membrane protein, mitochondrial) partially rescues bleomycin toxicity. Bleomycin treatment induces damage of both nuclear and mitochondrial DNA [32]. A recent study demonstrated that the mitochondrial localization of PARP-1

requires interaction with IMMT and that it is involved in the maintenance of mitochondrial DNA integrity [33]. The authors also suggest that IMMT overexpression results in an increase of PARP-1 in the extracellular compartment. We therefore speculate that IMMT overexpression recruits sufficient PARP-1 into the mitochondrion to maintain mitochondrial DNA integrity and promote cell survival.

Cisplatin and mitomycin C screens

RHOXF2 and PTPN2 were found to suppress the toxicity of cisplatin and mitomycin C. Together with the antifolates, these DNA damaging agents were the only screens yielding identical hits. Also, as for methotrexate and aminopterin, both these alkylating agents have the same mechanism of action.

Another validated hit for mitomycin C, ZFP64 (zinc finger protein 64 homolog), is noteworthy because it has been shown to be potentially phosphorylated by ATM and ATR, the major signal transducing kinases of the DNA damage response in response to ionizing radiation [34]. This observation is consistent with a potential role for ZFP64 in the repair of DNA damage generated by both alkylating agents and ionizing radiation.

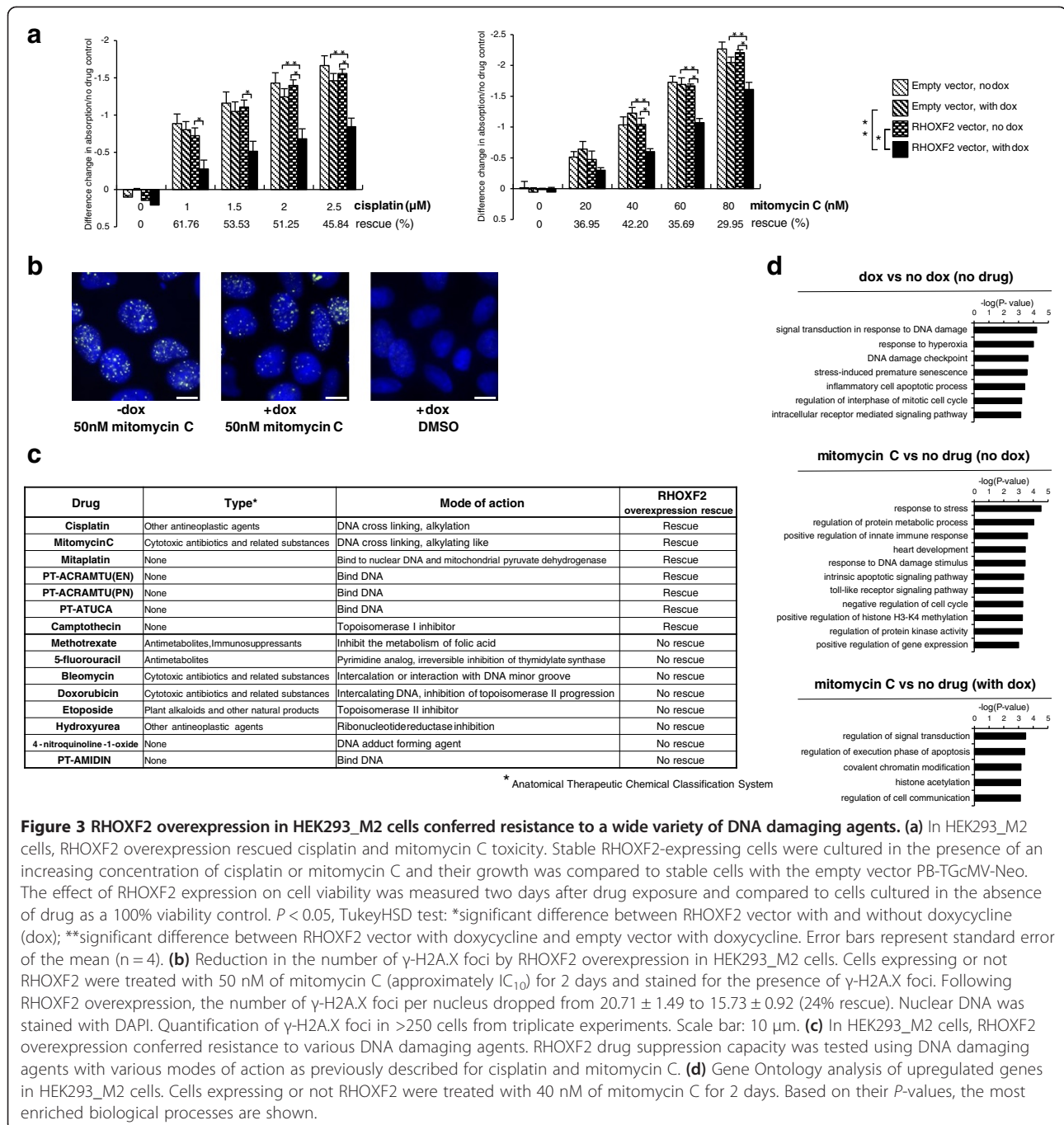
RHOXF2 overexpression confers resistance to several DNA damaging agents

Among the top candidates from the screens and confirmations, we focused on RHOXF2, a relatively uncharacterized member of the homeobox gene family, because this gene, when overexpressed, provided resistance to both cisplatin and mitomycin C. RHOXF2 clearly reduced drug-induced toxicities when overexpressed in the HEK293_M2 cell line treated with cisplatin (1 to 2.5 μ M) or with mitomycin C (20 to 80 nM). Rescue efficiency ranged from 45 to 60% and 30 to 42%, respectively (Figure 3a). This observation was confirmed using an independent cell biological assay in which overexpression of this transcription factor significantly reduced the formation of γ -H2A.X foci in cells treated with 50 nM of mitomycin C for 2 days (Figure 3b).

We next asked if RHOXF2's protective effect was restricted to cisplatin and mitomycin C or if this gene, when overexpressed, could also suppress the toxicity of DNA damaging agents with different mechanisms of action. Accordingly, we assessed the cytotoxicity of antimetabolites (methotrexate, 5-fluorouracil), a radiomimetic (bleomycin), replication inhibitors (camptothecin, doxorubicin, etoposide), mitaplatin, 4-nitroquinoline-1-oxide and several experimental platinum anticancer agents (PT-ACRAMTU (EN), PT-ACRAMTU(PN), PT-ATUCA, PT-AMIDIN) on HEK293-M2 cells expressing or not the transcription factor (Figure 3c; Additional file 8). With the exception of the topoisomerase I inhibitor camptothecin, all drugs whose toxicity was suppressed by RHOXF2 overexpression directly bind to DNA to exert their effects.

These observations motivated us to ask (i) if and how RHOXF2, an uncharacterized transcription factor, modulates gene expression in HEK293_M2 cells and (ii) if the nature of the regulated genes can illuminate the protective response to DNA damaging agents. We cultured HEK293_M2 cells with or without overexpression of RHOXF2 in the presence of 40 nM of mitomycin C or DMSO (control) and profiled their transcriptional response by RNA-seq. The sequencing data confirmed the RHOXF2 transcript was absent in the uninduced HEK293_M2 samples and verified its strong induction in the presence of doxycycline regardless of the drug treatment (Additional file 9). Neither RHOXF1 nor RHOXF2B expression was affected by RHOXF2 overexpression (Additional file 9). Interestingly, and despite their limited number, Gene Ontology analysis of the genes whose expression was induced during RHOXF2 overexpression showed that the most significantly enriched pathways are related to DNA damage processes ($P < 7 \times 10^{-5}$ and $P < 3 \times 10^{-4}$), stress response ($P < 3 \times 10^{-4}$), apoptosis ($P < 4 \times 10^{-4}$) and cell cycle ($P < 7 \times 10^{-4}$) (Figure 3d; Additional file 9). As anticipated, in the absence of RHOXF2 and the presence of mitomycin C, we found a strong enrichment for genes related to the response to stress ($P < 3 \times 10^{-5}$) and regulation of protein metabolic process ($P < 9.5 \times 10^{-5}$). Finally, in the presence of RHOXF2, genes involved in the regulation of signal transduction ($P < 4 \times 10^{-4}$) were induced during drug treatment. Therefore, we suggest that when treated with DNA damaging agents, the protective effect conferred by overexpression of RHOXF2 is due, in part, to its activation of several key genes involved in the response to DNA damage.

Finally, we asked if the RHOXF2-dependent protective effect observed in HEK293_M2 cells was restricted to one cell type. We then tested if RHOXF2 overexpression could confer resistance to the rtTA-expressing human breast adenocarcinoma cell line MCF7_M2 and the adenocarcinomic human alveolar basal epithelial cell line A549_M2. As quantified by both a two-fold shift in IC_{50} (44.91 μ M to 88.17 μ M) and by the 21% decrease in toxicity at 3 μ M, RHOXF2 overexpression protected MCF7_M2 cells against cisplatin-mediated toxicity (Additional file 10a). In the absence of RHOXF2, addition of doxycycline resulted in a modest increase in IC_{50} (41.06 μ M to 58.59 μ M; empty vector). The effect of the vector alone, while detectable, is less than that observed in the presence of RHOXF2. In contrast, although RHOXF2 overexpression in A549_M2 cells provided some apparent resistance (shift in IC_{50} from 448 nM to 1,099 nM), the measured effect was, in fact, largely due to doxycycline (shift in IC_{50} from 718 nM to 1,454 nM with empty vector) (Additional file 10b). In conclusion, our experiments suggest that the protective effect of RHOXF2 is not confined to a single genetic background



and expression of the transcription factor is sufficient to protect both HEK293 and MCF7 cells against the DNA damaging agent cisplatin. As with all large-scale screens, the biological significance of this observation should be further confirmed, for example, by interrogating a larger panel of cell lines.

Based on its viability and transcriptional phenotypes in HEK293_M2 cells, we further investigated the potential role for RHOXF2 in neoplastic cells.

In normal tissue, RHOXF2 expression is confined to testis but RHOXF2 is expressed in several cancer cell lines

In a small-scale study performed by northern blot analysis on normal tissues, expression of the transcription factor RHOXF2 was initially reported to be restricted to testis [35]. We performed a large-scale systematic survey interrogating a more representative panel of 20 different normal human tissues (3 donors per tissue type) by quantitative RT-PCR (Figure 4a). We first confirmed RHOXF2's testis-

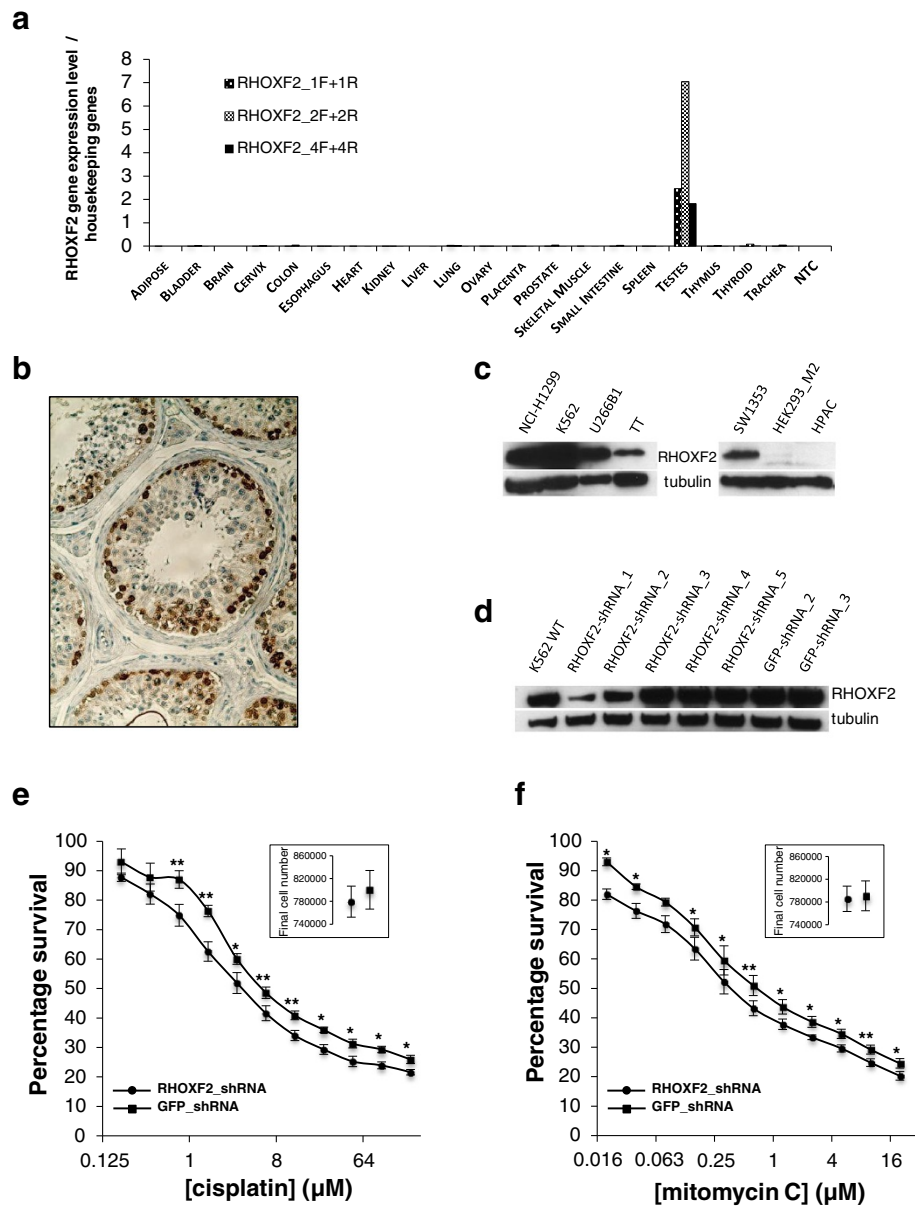


Figure 4 RHOXF2 is expressed in various cancer cell lines and modulates K562 resistance to DNA damaging agents. **(a)** In normal human tissues, RHOXF2 was exclusively detected in testis. Total RNA from 20 distinct tissues (mix of 3 different donors) was reverse transcribed into first strand cDNA and used as template for quantitative PCR. Three distinct pairs of primers covering exons 1, 2 and 4 were used. RHOXF2 expression level was calculated relative to the signal obtained for the amplification of housekeeping genes (CYCG, GUSB, ACTB, EEF1A1, GAPDH). **(b)** Representative example of immunohistochemical staining for RHOXF2 in normal testis. Section has been counterstained with Mayer's hematoxylin. **(c)** RHOXF2 was detected in various tumor cell lines by western blotting. **(d)** RHOXF2 shRNA validation in the human chronic myelogenous leukemia cell line K562. WT, wild type. **(e, f)** Partial depletion of RHOXF2 increased K562 sensitivity to cisplatin and mitomycin C. Final numbers of viable cells were calculated in the absence of drug (inset). Student's *t*-test; **P* < 0.05; ***P* < 0.01; means \pm standard error of the mean for *n* = 4 experiments.

specificity and then investigated which specific testis cell types expressed RHOXF2 protein by immunohistochemistry on normal testis samples. We detected a strong signal for RHOXF2 in spermatogonia and primary spermatocytes but not in spermatids (Figure 4b). Because RHOXF2 gene expression was detected in early stage spermatogenic cells

and not in later stages, we hypothesized that RHOXF2 might be involved in the self-renewal or maintenance of the undifferentiated state of the testis germ cells.

Computational analysis predicts RHOXF2 to be a testicular cancer candidate gene [36] and, in fact, the transcript was detected in several types of human testicular

cancers [37]. Furthermore, in the present study, we show that RHOXF2 overexpression conferred resistance to cisplatin (Figure 3a) and that RHOXF2 is found in several other cancer cell lines, the non-small cell lung carcinoma NCI-H1299, the blast phase chronic myelogenous leukemia K562, the multiple myeloma U266B1, the thyroid gland medullary carcinoma TT (CRL-1803) and the chondrosarcoma SW1353 (Figure 4c; Additional file 11).

In the chronic myelogenous leukemia K562 cell line, RHOXF2 depletion modulates the response to antineoplastic agents

Our study showed that RHOXF2 overexpression confers resistance to cisplatin and mitomycin C in HEK293_M2 cells, in which the protein is not endogenously expressed (Figure 4c; Additional file 9). We first generalized this observation by positing that overexpression or ectopic expression of certain genes greatly reduces some detrimental drug effects on the growth of particular cancer cell lines. We then reasoned that if this is truly drug-gene specific, the opposite mechanism should experimentally verify it. More specifically, depletion of the endogenously expressed protein should confer sensitivity to the cell line during drug treatment. The most direct translation of this gene-drug specificity could then be exploited to understand drug resistance in cancer cell lines. Therefore, we asked if sensitivity to DNA damaging agents was observed when RHOXF2 is knocked down in cancer cells where the protein is normally expressed. We generated several stable cell lines to deplete RHOXF2 (NCI-H1299, K562, SW1353, and the human pancreatic HPAC as negative control) with five shRNAs that target the mRNA. Two shRNAs against GFP were used as a negative control. Two shRNAs showed a strong knock-down of the protein in all three RHOXF2-positive cell lines (Figure 4d; Additional file 12e,h).

K562 showed a significant increase in sensitivity for both cisplatin (IC_{50} : 3.231 μ M to 2.052 μ M) and mitomycin C (IC_{50} : 377.8 nM to 250.4 nM) when the transcription factor was depleted, demonstrating the potential role of RHOXF2 in response to DNA damaging agents and particularly in response to DNA crosslinking agents (Figure 4e,f). Independent K562 transduction resulted in identical sensitivity to the drugs, showing the phenotype is independent of the lentiviral integration site. We ruled out the possibility that the calculated drug sensitivity simply reflected a growth difference among cell lines by demonstrating that the RHOXF2_shRNA and GFP_shRNA control cell lines divided at the same rate (Figure 4e,f, insets). Finally, we assessed the robustness of our knock-downs in the presence of the drugs by confirming that the modulation of RHOXF2 protein level is maintained even when the K562 cells were exposed to cisplatin or mitomycin C at IC_{70} (Additional file 13a,b).

As expected, no change was measured for the RHOXF2-negative HPAC cells (Additional file 12a,b). Interestingly, no increase in sensitivity to cisplatin and mitomycin C was observed for NCI-H1299 and SW1353 knock-down cell lines (Additional file 12c,d,f,g).

The modest increase in sensitivity of K562 cells to DNA damaging agents can be explained by different factors. First, transcription factors are generally present in limited quantities in cells [38-40]. Here, although clearly depleted in the presence of the shRNAs, RHOXF2 is still present at a relatively high level in K562 cells (Figure 4d). Therefore, we speculate that despite the efficient knock-down, the limited shift in IC_{50} is in part due to the presence of enough protein to sustain a certain level of resistance during drug treatment. In order to test our first hypothesis, a complete knock-out of the transcription factor would be necessary (for example, using CRISPR technology). Secondly, a recent study in mouse embryonic stem cells reports that knock-down of a few transcription factors is associated with substantial transcriptome change [41]. These data not only indicate the robustness of the pluripotency gene network but also suggest that perhaps RHOXF2 knock-down hardly perturbs the transcription factor network in the K562 cellular context.

Discussion

We have reduced to practice an elegant method to simultaneously evaluate thousands of genes for their capacity to provide a survival advantage in the presence of a toxic dose of drug using cells expressing the entire hORF collection and a microarray or sequencing-based readout.

The availability of hORF collections represents great progress compared to traditional approaches (random mutagenesis, cDNA libraries, and so on) because it permits the systematic interrogation of a gene's biological function in a targeted manner. Such gain-of-function approaches would be bolstered when combined with the reciprocal loss-of-function approach [42]. To date, several studies have already explored the availability of such collections. Most of the studies tested the hORFs function in an array (well-based) format [43-45] that simplifies the readout and analysis of the results. Here, we tested the function of thousands of ORFs simultaneously in a competitive manner. Our pool approach greatly reduced the tedious and expensive upstream steps of cloning, transduction and testing of the genes. The primary challenges to the pool approach are (i) potential loss of ORFs during cloning and infection (leading to representation variability), (ii) the potential masking effect of overrepresented clones in a population during selection (for example, DHFR in the case of methotrexate) and (iii) the necessity for dedicated analysis of the final results (hits prioritization). In addition to future enhancements of the human ORFeome collection, other promising gain-of-

function approaches to assess drug mode of action are the development of functional variomics technology [46] and the exploration of human isoform space in human cell lines.

In developing the assay, we first determined the optimal period of time between the beginning of the gene expression and the addition of the chemical compound. To address that question, we quantified the number of Venus-positive cells 12 hours, 24 hours and 48 hours after induction of expression of the hORF collection. Because the human gene and the Venus marker are co-transcribed, quantifying the level of Venus expression by flow cytometry serves as a proxy for the level of human gene expression. Because the number of Venus-positive cells peaks at 24 hours after doxycycline addition (Additional file 5), we postulated that 18 hours post-induction would be sufficient time after which to add drug to test the human gene for any protective effects.

We then addressed the sensitivity of the technique by improving the ratio between the number of seeded cells, the growth area and the time of drug addition. Seeding 500,000 cells in a T175 flask 33 hours before drug addition provided enough time and density for isolated cells to form separate micro-colonies that should over-express only one or a few genes. This 'clonogenic assay approach' was more sensitive to lower drug concentrations and provided a better selective environment for the surviving cells, presumably by minimizing signaling from a large population of dying cells that would arise at higher plating densities. Similarly, adding fresh medium and drug every 48 hours for the duration of the screen improved the selection process by washing away the dead and dying adherent cells that could have interfered with the subsequent screen readout.

During the course of the screen, the cell population followed a stereotypical pattern of growth, starting with a dramatic decrease in cell number and disappearance of the majority of cells, followed by a stabilization of the population and concluding with the development of medium to large size colonies originating from the initial surviving cells. Daily visual inspection of the cell population was crucial to determine when to stop the drug selection and harvest the cells, typically after 20 to 30 doubling times.

Several refinements may enhance this assay. First, improving the coverage and expression quality of the human genes using a more complete collection of fully sequenced ORFs as such collections become available [14]. Moreover, as illustrated in Additional file 6, relative abundance of the virally integrated hORFs is size-dependent. Therefore, to correct for that variation, preparation of the normalized amount of plasmid DNA for lentivirus production could be done as follows: i) reduce the pool size (less than 376 hORFs per pool as in our

study); ii) hORFs should be grouped by gene length. One can also match the particular drug under interrogation to disease-appropriate cell types and environmental conditions. Whenever possible, experiments should be performed with an initial high number of cells to guarantee the full representation of the collection during drug target screening. Although DNA damaging resistance was observed in the RHOXF2-overexpressing HEK293_M2 cells, it would be informative to further study if that observation is dependent on RHOXF2 expression status. Finally, next-generation sequencing will facilitate the throughput of hit detection and reduce the number of false positives arising from microarray cross-hybridization.

Our study revealed several genes whose functions relate to the action of the drugs tested. RHOXF2 is of a particular interest because of (i) its strong rescue phenotype in HEK293_M2 cells, (ii) its capacity to suppress both cisplatin and mitomycin C toxicity, and (iii) the dearth of functional information on this transcription factor. Homeobox genes encode transcription factors that play a central role during embryogenesis. *RHOXF2* belongs to the *Rhox* family of genes, which are expressed not only during embryogenesis but also in adult reproductive tissue [47]. In normal tissue, RHOXF2 expression is likely restricted to testis but its function remains largely unknown [35,48]. Here, we demonstrated that (i) the presence of RHOXF2 was sufficient to lower both cisplatin and mitomycin C toxicity in HEK293_M2 cells, and (ii) the transcription factor might exert its effect via the activation of genes involved in response to DNA damage. Furthermore, RHOXF2 protein expression was detected in several cancer cell lines. In one of these lines (chronic myelogenous leukemia/K562) knocking down the protein increased the toxicity for both DNA damaging agents. Interestingly, although no growth disadvantage was observed in K562 cells partially depleted for RHOXF2 (Figure 4e,f), its depletion via shRNA increased the sensitivity of this leukemia cell line to several pharmacological compounds. We therefore speculate that although not a driver of tumorigenesis, RHOXF2 could confer on the neoplastic cells a selective advantage during drug treatment.

Conclusions

We demonstrate the usefulness of our approach in order to understand drug mechanisms of action by combining genetic perturbation with drug treatment. Here we recapitulated examples of drug resistance and uncovered several unanticipated drug-target interactions by virtue of the suppression of chemical toxicity observed upon gene overexpression. We also provide a step-by-step protocol for performing such screens and for analyzing the data (Additional file 1). Moreover, our innovative

overexpression approach could also be applied to identify genes whose up-regulation could be toxic or enhance cell proliferation in different cell types. For example, the assay could be useful to uncover tumor-specific activity in the context of genome engineering as well as for the discovery of cell-type-specific oncogenes.

Additional files

Additional file 1: Step-by-step protocol for a genome-wide mammalian overexpression drug screen.

Additional file 2: Plasmid map of the lentiviral vector pLD-T-IRES-Venus-WPRE-stop.

Additional file 3: Original immunoblotting showing the conditional expression of DHFR in HEK293_M2 cells. Addition of 4 ng/ml, 16 ng/ml, 32 ng/ml, 63 ng/ml and 2,000 ng/ml of doxycycline induced DHFR protein expression in the stable cell line HEK293_M2. Expression level was compared to alpha tubulin as a loading control. The asterisk marks a human protein that cross-reacts with the anti-DHFR antibody.

Additional file 4: DHFR was identified as the predominant gene in this pool of 376 hORFs whose overexpression provided resistance to methotrexate. HEK293_M2 cells harboring a minipool collection of 376 hORFs (including *DHFR*) were grown in the presence of a lethal dose of methotrexate. The nature of the hORFs conferring resistance to the drug was identified by plotting the log₂ of the signal intensity for each hORF in the cells cultured in the presence of methotrexate on the x-axis; and by plotting the log₂ ratio of the signal intensity for each hORF of the cells cultured in presence of the drug divided by the signal intensity for each hORF of the cells grown in presence of DMSO on the y-axis.

Additional file 5: Detection of Venus-expressing HEK293_M2 cells after 12, 24 and 48 hours of induction. (a,b) HEK293_M2 cells stably transduced with the hORFeome collection (b) or not (a) were cultured in the presence of doxycycline (2 µg/ml) for 12, 24 and 48 hours. After gene induction, cell number and intensity of the Venus fluorescence were measured by flow cytometry. The percentage of Venus positive cells was compared to the total number of living cells measured.

Additional file 6: Human ORFeome relative abundance in HEK293_M2s at T0.

Additional file 7: Plasmid map of the transposon-based vector PB-TGcMV-Neo.

Additional file 8: In HEK293_M2 cells, RHOXF2 overexpression conferred resistance to various DNA damaging agents. Stable RHOXF2 cells was cultured in the presence of an increasing concentration of DNA damaging agents and their growth was compared to the stable cells with the empty vector PB-TGcMV-Neo. The effect of RHOXF2 expression on cell viability was measured two days after drug exposure and compared to cells cultured in the absence of drug as a 100% viability control. TukeyHSD test, $P < 0.05$; *significant difference between RHOXF2 vector with and without doxycycline; **significant difference between RHOXF2 vector with doxycycline and empty vector with doxycycline. Error bars represent standard error of the mean ($n = 4$).

Additional file 9: RHOXF2 raw FPKM values and Gene Ontology enrichment.

Additional file 10: RHOXF2 overexpression rescued cisplatin toxicity in HEK293_M2 and MCF7_M2 but not A549_M2 cells. Stable RHOXF2 cells were cultured in the presence of an increasing concentration of cisplatin and their growth was compared to the stable cells with the empty vector PB-TGcMV-Neo. The effect of RHOXF2 expression on cell viability was measured three days after drug exposure and compared to cell cultured in the absence of drug as a 100% viability control. Lines show the nonlinear fit of a variable-slope dose-response model for (a) MCF7_M2 cells, (b) A549_M2 cells and (c) HEK293_M2 cells as positive control; (d) resulting IC₅₀s. Data analysis was performed using the 'drc' package in R. Experiments were done in triplicate.

Additional file 11: RHOXF2 was detected in various tumor cell lines by western blotting.

Additional file 12: RHOXF2 knocked-down in HPAC, NCI-H1299 and SW1353 cells does not modulate cisplatin and mitomycin C sensitivities. (a-g) Dose-response curves for HPAC (a,b), NCI-H1299 (c,d) and SW1353 (f,g) were established after 2 days of exposure to cisplatin (a,c,f) or mitomycin C (b,d,g). Percentage of growth inhibition was calculated by comparing the number of cells treated with drug to the number of cells cultured in media with DMSO as control. Error bars represent standard error of the mean ($n = 4$). RHOXF2 depletion in NCI-H1299 (e) and SW1353 cells (h). Total cell extracts for each cell line were used to detect the presence of RHOXF2 by western blotting. Expression level was compared to the alpha tubulin as loading control.

Additional file 13: RHOXF2 knocked down in K562 cells. (a,b) Cell lines in which RHOXF2 and GFP were knocked down were grown in the presence of 35 µM cisplatin (a) or 7 µM of mitomycin C (b) for 2 days. Total cell extracts for each cell line were used to detect the presence of RHOXF2 by western blotting. Expression level was compared to alpha tubulin as loading control. (c) Total cell extracts from a RHOXF2-expressing cell line and a cell line with GFP knocked down but overexpressing RHOXF2 were used to detect the presence of RHOXF2 by western blotting.

Abbreviations

BSA: bovine serum albumin; DHFR: dihydrofolate reductase; DMEM: Dulbecco's modified Eagle's medium; gDNA: genomic DNA; GGH: gamma-glutamyl hydrolase; hORF: human ORF; HRP: horse radish peroxidase; MGLL: monoglyceride lipase; ORF: open reading frame; PBS: phosphate-buffered saline; PCR: polymerase chain reaction; RNAi: RNA interference; shRNA: short hairpin RNA; SRB: sulforhodamine B.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AA, SK, GG and CN contributed to the conception and design of this project. AA and CN developed the methodology. AA and SK performed experiments. ABM constructed the pLD-T-IRES-Venus-WPRE-STOP vector. AIS performed quantitative PCR experiments. DT performed the Illumina RNA sequencing. AA and LEH analyzed and interpreted the data. AA, SK, LEH, AIS, JM, GG and CN wrote the manuscript. CN supervised the study. All authors read and approved the final manuscript.

Acknowledgements

CN and GG were supported by a grant from the Canadian Cancer Society (#20830) and the NHGRI.

Author details

¹Department of Molecular Genetics, University of Toronto, Toronto, M5S 3E1, Canada. ²Banting and Best Department of Medical Research, University of Toronto, Toronto, M5S 3E1, Canada. ³Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. ⁴Donnelly Sequencing Center, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. ⁵Institute of Biomaterials and Biomedical Engineering, University of Toronto, 170 College Street, Toronto M5S 3E3, Canada. ⁶Department of Pharmaceutical Sciences, University of Toronto, 144 College Street, Toronto, Ontario M5S 3M2, Canada. ⁷Department of Pharmaceutical Sciences, University of British Columbia, 6619-2405 Wesbrook Mall, Vancouver, BC V6T 1Z3, Canada.

Received: 3 September 2013 Accepted: 22 April 2014

Published: 29 April 2014

References

- Jarosz DF, Taipale M, Lindquist S: Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. *Annu Rev Genet* 2010, **44**:189-216.
- Nobeli I, Favia AD, Thornton JM: Protein promiscuity and its implications for biotechnology. *Nat Biotechnol* 2009, **27**:157-167.
- Hopkins AL: Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008, **4**:682-690.

4. Strebhardt K, Ullrich A: **Paul Ehrlich's magic bullet concept: 100 years of progress.** *Nat Rev Cancer* 2008, **8**:473–480.
5. Agarwal R, Kaye SB: **Ovarian cancer: strategies for overcoming resistance to chemotherapy.** *Nat Rev Cancer* 2003, **3**:502–516.
6. Seruga B, Ocana A, Tannock IF: **Drug resistance in metastatic castration-resistant prostate cancer.** *Nat Rev Clin Oncol* 2011, **8**:12–23.
7. Curt GA, Carney DN, Cowan KH, Jolivet J, Bailey BD, Drake JC, Chien Song KS, Minna JD, Chabner BA: **Unstable methotrexate resistance in human small-cell carcinoma associated with double minute chromosomes.** *N Engl J Med* 1983, **308**:199–202.
8. Cohen SN, Chang AC, Boyer HW, Helling RB: **Construction of biologically functional bacterial plasmids in vitro.** *Proc Natl Acad Sci U S A* 1973, **70**:3240–3244.
9. Chang AC, Nunberg JH, Kaufman RJ, Erlich HA, Schimke RT, Cohen SN: **Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase.** *Nature* 1978, **275**:617–624.
10. Rine J, Hansen W, Hardeman E, Davis RW: **Targeted selection of recombinant clones through gene dosage effects.** *Proc Natl Acad Sci U S A* 1983, **80**:6750–6754.
11. Rual JF, Hirozane-Kishikawa T, Hao T, Bertin N, Li S, Dricot A, Li N, Rosenberg J, Lamesch P, Vidalain PO, Clingsmith TR, Hartley JL, Esposito D, Cheo D, Moore T, Simmons B, Sequerra R, Bosak S, Doucette-Stamm L, Le Peuch C, Vandenhoute J, Cusick ME, Albala JS, Hill DE, Vidal M: **Human ORFeome version 1.1: a platform for reverse proteomics.** *Genome Res* 2004, **14**:2128–2135.
12. Lamesch P, Li N, Milstein S, Fan C, Hao T, Szabo G, Hu Z, Venkatesan K, Bethel G, Martin P, Rogers J, Lawlor S, McLaren S, Dricot A, Borick H, Cusick ME, Vandenhoute J, Dunham I, Hill DE, Vidal M: **hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes.** *Genomics* 2007, **89**:307–315.
13. Shin KJ, Wall EA, Zavzavadjian JR, Santat LA, Liu J, Hwang JI, Rebres R, Roach T, Seaman W, Simon MI, Fraser ID: **A single lentiviral vector platform for microRNA-based conditional RNA interference and coordinated transgene expression.** *Proc Natl Acad Sci U S A* 2006, **103**:13759–13764.
14. Yang X, Boehm JS, Salehi-Ashtiani K, Hao T, Shen Y, Lubonja R, Thomas SR, Alkan O, Bhimdi T, Green TM, Johannessen CM, Silver SJ, Nguyen C, Murray RR, Hieronymus H, Balcha D, Fan C, Lin C, Ghamsari L, Vidal M, Hahn WC, Hill DE, Root DE: **A public genome-scale lentiviral expression library of human ORFs.** *Nat Methods* 2011, **8**:659–661.
15. **ArrayExpress.** [<http://www.ebi.ac.uk/arrayexpress/>]
16. Vichai V, Kirtikara K: **Sulforhodamine B colorimetric assay for cytotoxicity screening.** *Nat Protoc* 2006, **1**:1112–1116.
17. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, Golland P, Sabatini DM: **Cell Profiler: image analysis software for identifying and quantifying cell phenotypes.** *Genome Biol* 2006, **7**:R100.
18. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z: **GORilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.** *BMC Bioinformatics* 2009, **10**:48.
19. **Abcam Immunohistochemistry Procedure.** [http://docs.abcam.com/pdf/protocols/Short_IHC-P_protocol.pdf]
20. Vander Heiden MG: **Targeting cancer metabolism: a therapeutic window opens.** *Nat Rev Drug Discov* 2011, **10**:671–684.
21. Farber S, Diamond LK, Mercer RD, Sylvester RF, Wolff JA: **Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid (aminopterin).** *N Engl J Med* 1948, **238**:787–793.
22. Shi Y, Burn P: **Lipid metabolic enzymes: emerging drug targets for the treatment of obesity.** *Nat Rev Drug Discov* 2004, **3**:695–710.
23. Hogan S, Fleury A, Hadvary P, Lengsfeld H, Meier MK, Triscari J, Sullivan AC: **Studies on the antiobesity activity of tetrahydrolipstatin, a potent and selective inhibitor of pancreatic lipase.** *Int J Obes* 1987, **11**:35–42.
24. Rose JL, Reeves KC, Likhovorik RI, Hoyt DG: **Base excision repair proteins are required for integrin-mediated suppression of bleomycin-induced DNA breakage in murine lung endothelial cells.** *J Pharmacol Exp Ther* 2007, **321**:318–326.
25. Kelland LR: **Emerging drugs for ovarian cancer.** *Expert Opin Emerg Drugs* 2005, **10**:413–424.
26. Iyer VN, Szybalski W: **A molecular mechanism of mitomycin action: linking of complementary DNA strands.** *Proc Natl Acad Sci U S A* 1963, **50**:355–362.
27. Egorin MJ, Lagattuta TF, Hamburger DR, Covey JM, White KD, Musser SM, Eiseman JL: **Pharmacokinetics, tissue distribution, and metabolism of 17-(dimethylaminoethylamino)-17-demethoxygeldanamycin (NSC 707545) in CD2F1 mice and Fischer 344 rats.** *Cancer Chemother Pharmacol* 2002, **49**:7–19.
28. Cole PD, Kamen BA, Gorlick R, Banerjee D, Smith AK, Magill E, Bertino JR: **Effects of overexpression of gamma-Glutamyl hydrolase on methotrexate metabolism and resistance.** *Cancer Res* 2001, **61**:4599–4604.
29. Karlsson M, Reue K, Xia YR, Lusic AJ, Langin D, Tornqvist H, Holm C: **Exon-intron organization and chromosomal localization of the mouse monoglyceride lipase gene.** *Gene* 2001, **272**:11–18.
30. Saario SM, Laitinen JT: **Monoglyceride lipase as an enzyme hydrolyzing 2-arachidonoylglycerol.** *Chem Biodivers* 2007, **4**:1903–1913.
31. Chen J, Stubbe J: **Bleomycins: towards better therapeutics.** *Nat Rev Cancer* 2005, **5**:102–112.
32. Shen CC, Wertelecki W, Driggers WJ, LeDoux SP, Wilson GL: **Repair of mitochondrial DNA damage induced by bleomycin in human cells.** *Mutat Res* 1995, **337**:19–23.
33. Rossi MN, Carbone M, Mostocotto C, Mancone C, Tripodi M, Maione R, Amati P: **Mitochondrial localization of PARP-1 requires interaction with mitofilin and is involved in the maintenance of mitochondrial DNA integrity.** *J Biol Chem* 2009, **284**:31616–31624.
34. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER 3rd, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y, Shiloh Y, Gygi SP, Elledge SJ: **ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.** *Science* 2007, **316**:1160–1166.
35. Wayne CM, MacLean JA, Cornwall G, Wilkinson MF: **Two novel human X-linked homeobox genes, hPEPP1 and hPEPP2, selectively expressed in the testis.** *Gene* 2002, **301**:1–11.
36. Hofmann O, Caballero OL, Stevenson BJ, Chen YT, Cohen T, Chua R, Maher CA, Panji S, Schaefer U, Kruger A, Lehvaslaiho M, Carninci P, Hayashizaki Y, Jongeneel CV, Simpson AJ, Old LJ, Hide W: **Genome-wide analysis of cancer/testis gene expression.** *Proc Natl Acad Sci U S A* 2008, **105**:20422–20427.
37. Korkola JE, Houldsworth J, Chadalavada RS, Olshen AB, Dobrzynski D, Reuter VE, Bosl GJ, Chaganti RS: **Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with in vivo differentiation of human male germ cell tumors.** *Cancer Res* 2006, **66**:820–827.
38. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM: **A census of human transcription factors: function, expression and evolution.** *Nat Rev Genet* 2009, **10**:252–263.
39. Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, Frishman D: **Protein abundance profiling of the Escherichia coli cytosol.** *BMC Genomics* 2008, **9**:102.
40. Yang VW: **Eukaryotic transcription factors: identification, characterization and functions.** *J Nutr* 1998, **128**:2045–2051.
41. Nishiyama A, Sharov AA, Piao Y, Amano M, Amano T, Hoang HG, Binder BY, Tapnio R, Basseu U, Malinou JN, Correa-Cerro LS, Yu H, Xin L, Meyers E, Zalzman M, Nakatake Y, Stagg C, Sharova L, Qian Y, Dudekula D, Sheer S, Cadet JS, Hirata T, Yang HT, Goldberg J, Evans MK, Longo DL, Schlessinger D, Ko MS: **Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells.** *Sci Rep* 2013, **3**:1390.
42. Hoon S, Smith AM, Wallace IM, Suresh S, Miranda M, Fung E, Fung E, Proctor M, Shokat KM, Zhang C, Davis RW, Giaever G, St Onge RP, Nislow C: **An integrated platform of genomic assays reveals small-molecule bioactivities.** *Nat Chem Biol* 2008, **4**:498–506.
43. Skalamera D, Dahmer M, Purdon AS, Wilson BM, Ranall MV, Blumenthal A, Gabrielli B, Gonda TJ: **Generation of a genome scale lentiviral vector library for EF1alpha promoter-driven expression of human ORFs and identification of human genes affecting viral titer.** *PLoS One* 2012, **7**:e51733.
44. Skalamera D, Ranall MV, Wilson BM, Leo P, Purdon AS, Hyde C, Nourbakhsh E, Grimmond SM, Barry SC, Gabrielli B, Gonda TJ: **A high-throughput platform for lentiviral overexpression screening of the human ORFeome.** *PLoS One* 2011, **6**:e20057.
45. Reece-Hoyes JS, Barutcu AR, McCord RP, Jeong JS, Jiang L, MacWilliams A, Yang X, Salehi-Ashtiani K, Hill DE, Blackshaw S, Zhu H, Dekker J, Walhout AJ: **Yeast one-hybrid assays for gene-centered human gene regulatory network mapping.** *Nat Methods* 2011, **8**:1050–1052.
46. Huang Z, Chen K, Zhang J, Li Y, Wang H, Cui D, Tang J, Liu Y, Shi X, Li W, Liu D, Chen R, Sucgang RS, Pan X: **A functional variomics tool for**

discovering drug-resistance genes and drug targets. *Cell Rep* 2013, **3**:577–585.

47. Maclean JA 2nd, Chen MA, Wayne CM, Bruce SR, Rao M, Meistrich ML, Macleod C, Wilkinson MF: **Rhox: a new homeobox gene cluster**. *Cell* 2005, **120**:369–382.
48. MacLean JA 2nd, Wilkinson MF: **The Rhox genes**. *Reproduction* 2010, **140**:195–213.

doi:10.1186/gm549

Cite this article as: Arnoldo *et al.*: A genome scale overexpression screen to reveal drug activity in human cells. *Genome Medicine* 2014 **6**:32.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

