

Proteogenomic analysis of Inhibitor of Differentiation 4 (ID4) in basal-like breast cancer

Laura A. Baker

Garvan Institute of Medical Research

Holly Holliday

Garvan Institute of Medical Research

Daniel Roden

Garvan Institute of Medical Research

Christoph Krisp

Macquarie University

Sunny Z. Wu

Garvan Institute of Medical Research

Simon Junankar

Garvan Institute of Medical Research

Aurelien A. Serandour

Cancer Research UK

Hisham Mohammed

Cancer Research UK

Radhika Nair

Rajiv Gandhi Centre for Biotechnology

Geetha Sankaranarayanan

Cancer Research UK

Andrew M.K. Law

Garvan Institute of Medical Research

Andrea McFarland

Garvan Institute of Medical Research

Peter T. Simpson

University of Queensland

Sunil Lakhani

University of Queensland

Eoin Dodson

Garvan Institute of Medical Research

Christina Selinger

Royal Prince Alfred Hospital

Lyndal Anderson

Royal Prince Alfred Hospital

Goli Samimi

National Cancer Institute

Neville F. Hacker

Royal Hospital for Women

Elgene Lim

Garvan Institute of Medical Research

Christopher J. Ormandy

Garvan Institute of Medical Research

Matthew J. Naylor

The University of Sydney School of Medicine

Kaylene Simpson

Peter MacCallum Cancer Centre

Iva Nikolic

Peter MacCallum Cancer Centre

Sandra O'Toole

Garvan Institute of Medical Research

Warren Kaplan

Garvan Institute of Medical Research

Mark J. Cowley

Garvan Institute of Medical Research

Jason S. Carroll

Cancer Research UK

Mark Molloy

Macquarie University

Alexander Swarbrick (✉ a.swarbrick@garvan.org.au)

Garvan Institute of Medical Research <https://orcid.org/0000-0002-3051-5676>

Research article

Keywords: Breast Cancer, Transcription factor, Helix-loop-helix, DNA damage

Posted Date: May 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-16182/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Breast Cancer Research on June 11th, 2020. See the published version at <https://doi.org/10.1186/s13058-020-01306-6>.

Abstract

Background Basal-like breast cancer (BLBC) is a poorly characterised, heterogeneous disease. Patients are diagnosed with aggressive, high-grade tumours and often relapse with chemotherapy resistance. Detailed understanding of the molecular underpinnings of this disease is essential to the development of personalised therapeutic strategies. Inhibitor of Differentiation 4 (ID4) is a helix-loop-helix transcriptional regulator required for mammary gland development. ID4 is overexpressed in a subset of BLBC patients, associating with a stem-like poor prognosis phenotype, and is necessary for the growth of cell line models of BLBC, through unknown mechanisms.

Methods Here, we have defined unique molecular insights into the function of ID4 in BLBC and the related disease high-grade serous ovarian cancer (HGSOC), by combining RIME proteomic analysis, ChIP-seq mapping of genomic binding sites and RNA-seq.

Results These studies reveal novel interactions with DNA damage response proteins, in particular, mediator of DNA damage checkpoint protein 1 (MDC1). Through MDC1, ID4 interacts with other DNA repair proteins (γ H2AX and BRCA1) at fragile chromatin sites. ID4 does not affect transcription at these sites, instead binding to chromatin following DNA damage. Analysis of clinical samples demonstrates that ID4 is amplified and overexpressed at a higher frequency in BRCA1 -mutant BLBC compared with sporadic BLBC, providing genetic evidence for an interaction between ID4 and DNA damage repair deficiency.

Conclusions These data link the interactions of ID4 with MDC1 to DNA damage repair in the aetiology of BLBC and HGSOC.

Background

Breast cancer is a heterogeneous disease. Gene expression signatures delineate five major subtypes with distinct pathologies, treatment strategies and clinical outcomes (1, 2). The basal-like subtype (BLBC), accounting for ~18% of diagnoses, is a subtype with particularly poor prognosis, largely due to the molecular and clinical heterogeneity of these tumours, and the corresponding lack of targeted therapeutics. The molecular drivers of BLBC are poorly understood, thus there are limited available targeted therapies.

One such molecular driver of a subset of BLBCs is mutation in the breast and ovarian cancer susceptibility gene (*BRCA1*) (3, 4). Mutations in *BRCA1* occur in approximately 0.25% of European women, predisposing them to breast and ovarian cancer, particularly to poor prognosis subtypes including BLBC and high-grade serous ovarian cancer (HGSOC) (3). These two subtypes of cancer are similar in terms of their gene expression profiles and genetic dependencies, and thus present similar sensitivity to therapeutic targeting (5-7). *BRCA1* has many cellular functions including transcription and

gene splicing, yet is best known for its role in mediating DNA damage repair (8, 9). BRCA1 coordinates efficient repair of double stranded DNA breaks through the Homologous Recombination (HR) pathway (9). In the absence of functional BRCA1, cells accumulate mutations, genomic instability and demonstrate an increased frequency of genomic rearrangements (10).

Another molecular driver of BLBC is the Helix-Loop-Helix (HLH) transcriptional regulator Inhibitor of differentiation 4 (ID4). We and others have previously shown ID4 to be important for both mammary gland development and also for the aetiology of BLBC (11-13). ID4 is overexpressed in a subset of BLBC patients, marking patients with poor survival outcome, and is necessary for the growth of BLBC cell lines (11, 14-21). Precisely how ID4 mediates this function in BLBC is unclear.

ID proteins (ID1-4) lack a basic DNA binding domain, and thus, their classical mechanism of action is believed to entail dominant-negative regulation of canonical binding partners; basic HLH (bHLH) transcription factors. ID proteins dimerise with bHLH proteins and prevent them from interacting with DNA, affecting the transcription of lineage-specific genes (22-26). Yet this model of ID protein function is largely based on evidence from studies of ID1-3 in non-transformed fibroblasts, neural and embryonic tissue. ID proteins are tissue specific in their expression and function and hence this model may not apply to all four ID proteins across various tissues and in disease (25, 27-32). Indeed, contrary mechanisms have been described for ID2 in liver regeneration, with ID2 interacting with chromatin at the c-Myc promoter as part of a multi-protein complex to repress c-Myc gene expression (33, 34), and ID4 has been shown to bind to and suppress activity of the ER α promoter and regions upstream of the ER α and FOXA1 genes in mouse mammary epithelial cells (12). These data suggest that despite lacking a DNA binding domain, ID proteins may interact with chromatin complexes under certain conditions. However, no studies have systematically mapped the protein or chromatin interactomes for any ID family member.

To this end, we applied chromatin immunoprecipitation-sequencing (ChIP-seq) to interrogate the ID4-chromatin binding sites, Rapid Immunoprecipitation and Mass spectrometry of Endogenous proteins (RIME) to determine the ID4 protein interactome. In addition, ID4 knockdown and RNA-sequencing analysis was used to determine transcriptional targets of ID4. These analyses reveal a novel link between ID4 with the DNA damage repair apparatus.

Methods

Mammalian cell culture growth conditions

Cell lines were obtained from American Type Culture Collection and verified using cell line fingerprinting. HCC70 and HCC1937 cell lines were cultured in RPMI 1640 (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS, 20 mM HEPES (Thermo Fisher Scientific) and 1 mM Sodium Pyruvate (Thermo Fisher Scientific). MDA-MB-468 and OVKATE cell lines were cultured in RPMI 1640 (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS, 20 mM HEPES (Thermo Fisher Scientific) and 0.25% (v/v) insulin (Human) (Clifford Hallam Healthcare). Cells stably-transduced with SMARTChoice lentiviral vectors were grown in the presence of 1 µg/mL puromycin.

Imaging

Immunohistochemistry images were obtained using an inverted epifluorescence microscope (Carl Zeiss, ICM-405, Oberkochen, Germany). Images were captured by the Leica DFC280 digital camera system (Leica Microsystems, Wetzlar, Germany). The Leica DM 5500 Microscope with monochrome camera (DFC310Fx) or Leica DMI SP8 Confocal with 4 lasers (405 nm, 488 nm, 552 nm and 638 nm) and two PMT detectors were used to capture standard fluorescent and confocal images.

SMARTChoice inducible lentiviral system

ID4 and control lentiviral shRNA constructs (SMARTchoice) were purchased commercially (Dharmacon, GE, Lafayette, CO, USA). Successfully transfected cells were selected using puromycin resistance (constitutive under the humanEF1a promoter). For ID4 knockdown analysis, HCC70 cells with SMARTChoice shID4 #1 #178657 (VSH6380-220912204), SMARTChoice shID4 #2 #703009 (VSH6380-221436556), SMARTChoice shID4 #3 #703033 (VSH6380-221436580), SMARTChoice shNon-targeting (VSC6572) and mock-infected cells. Cells were treated with vehicle control or with 1 µg/mL doxycycline for 72 hours before harvesting protein and RNA directly from adherent cells. The SMARTChoice shID4 #2 #703009 (VSH6380-221436556) produced the highest level of ID4 knockdown and was used for further analysis.

Non-lethal DNA damage induction with ionising radiation

Cells were seeded at 2.5×10^5 (HCC70) or 2.2×10^5 (MDA-MB-468) cells/well in a 6-well plate in normal growth medium. One day post seeding, cells were exposed to 2-5 Gy of ionising radiation using an X-RAD 320 Series Biological Irradiator (Precision X-Ray, CT, USA). Cells were returned to normal tissue culture incubation conditions and harvested at designated time points.

Gene expression analysis

Total RNA was prepared for using the miRNeasy RNA extraction kit (Qiagen), according to the manufacturer's instructions. cDNA was generated from 1000ng RNA using the Transcriptor First Strand cDNA Synthesis Kit (Roche) using oligo-dT primers and following the manufacturer's instructions. qPCR analysis was used to analyse mRNA expression levels using Taqman probes (Applied Biosystems/Life Technologies) as per manufactures specifications (Table 2) using an ABI PRISM 7900 HT machine. qPCR data was analysed using the $\Delta\Delta C_t$ method (35).

Gene	Taqman assay
ID4	Hs02912975_g1
B2M	Hs99999907_m1
GAPDH	Hs02758991_s1
NEAT1	Hs01008264_s1
MALAT1	Hs00273907_s1
ELF3	HS00963881_M1
GBA	HS00986836_G1
ZFP36L1	Hs00245183_m1
FAIM	HS00216756_M1

Genes analysed and the corresponding Taqman assay used to analyse their expression level

Protein analysis

Cells were lysed, unless specified, using RIPA [0.88% (w/v) Sodium Chloride, 1% (v/v) Triton X-100, 0.5% (w/v) Sodium Deoxycholate, 0.1% (w/v) SDS, 0.61% (v/v) Tris (Hydroxymethyl)Aminomethane and protease and phosphatase inhibitors (Roche)] or Lysis Buffer 5 (10 mM Tris pH 7.4, 1 mM EDTA, 150 mM NaCl, 1% Triton X-100 and protease and phosphatase inhibitors). If required, protein was quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions. Western blotting analysis was conducted as previously described (11). MDC1 protein was analysed using 3-8% tris/acetate gels and PVDF nitrocellulose membrane for MDC1 analysis (BioRad). All other proteins were analysed using the LiCor Odyssey system (Millenium Science, Mulgrave, VIC, Australia). Protein expression was analysed using antibodies targeting ID4 (Biocheck anti-ID4 rabbit monoclonal BCH-9/82-12, 1:40,000), β -Actin (Sigma anti-Actin mouse monoclonal A5441, 1:5,000) and MDC1 (Sigma anti-MDC1 mouse monoclonal M2444, 1:1000).

Co-immunoprecipitation

Co-immunoprecipitations (IP) were conducted using 10 μ L per IP Pierce Protein A/G magnetic beads (Thermo Fisher Scientific) with 2 μ g of antibody: IgG rabbit polyclonal (Santa Cruz sc-2027), ID4 (1:1 mix of rabbit polyclonal antibodies: Santa Cruz L-20: sc-491 and Santa Cruz H-70: sc-13047), MDC1 (rabbit polyclonal antibody Merck Millipore #ABC155). Beads and antibody were incubated at 4°C on a rotating platform for a minimum of 4 h. Beads were then washed three times in lysis buffer before cell lysate was added to the tube. Lysates were incubated with beads overnight at 4°C on a rotating platform. Beads were washed three times in lysis buffer and resuspended in 2 x NuPage

sample reducing buffer (Life Technologies) and 2 x NuPage sample running buffer (Life Technologies) and heated to 85°C for 10 min. Beads were separated on a magnetic rack and supernatant was analysed by western blotting as described above.

Rapid Immunoprecipitation and Mass Spectrometry of Endogenous Proteins (RIME)

Cells were fixed using paraformaldehyde (PFA) (ProSciTech, Townsville, QLD, Australia) and prepared for RIME (36) and ChIP-seq (37, 38) as previously described. Cross-linking was performed for 7 min for RIME experiments and 10 min for ChIP-seq, ChIP-exo and ChIP-qPCR experiments. Samples were sonicated using a Bioruptor Standard (Diagenode, Denville, NJ, USA) for 30-35 cycles of 30 sec ON/30 sec OFF (sonication equipment kindly provided by Prof. Merlin Crossley, UNSW). IP was conducted on 60 (ChIP-seq/ ChIP-exo) to 120 (RIME) million cells using 100 µL beads/20 µg antibody. Correct DNA fragment size of 100-500bp was determined using 2% agarose gel electrophoresis.

Patient-derived xenograft tumour models were cross-linked at 4°C for 20 min in a solution of 1% Formaldehyde (ProSciTech), 50 mM Hepes-KOH, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors (H. Mohammed, personal communication). Samples (0.5 mg of starting tumour weight) were dissociated using a Polytron PT 1200E tissue homogeniser (VWR) and sonicated using the Branson Digital Sonifer probe sonicator (Branson Ultrasonics, Danbury, CT, USA) with a microtip attachment for 3-4 cycles of 10 x [0.1 sec ON, 0.9 sec OFF].

Mass spectrometry analysis was conducted at the Australian Proteomic Analysis Facility (APAF) at Macquarie University (NSW, Australia (39)). Briefly, samples were denatured in 100 mM triethylammonium bicarbonate and 1% w/v sodium deoxycholate, disulfide bonds were reduced in 10 mM dithiothreitol, alkylated in 20 mM iodo acetamide, and proteins digested on the dynabeads using trypsin. After C18 reversed phase (RP) StageTip sample clean up, peptides were submitted to nano liquid chromatography coupled mass

spectrometry (MS) (nanoLC-MS/MS) characterisation. MS was performed using a TripleToF 6600 (SCIEX, MA, USA) coupled to a nanoLC Ultra 2D HPLC with cHipLC system (SCIEX). Peptides were separated using a 15 cm chip column (ChromXP C18, 3 μ m, 120 Å) (SCIEX). The mass spectrometer was operated in positive ion mode using a data dependent acquisition method (DDA) and data independent acquisition mode (DIA or SWATH) both using a 60 min acetonitrile gradient from 5 - 35%. DDA was performed of the top 20 most intense precursors with charge stages from 2+ - 4+ with a dynamic exclusion of 30 s. SWATH-MS was acquired using 100 variable precursor windows based on the precursor density distribution in data dependent mode. MS data files were processed using ProteinPilot v.5.0 (SCIEX) to generate mascot generic files. Processed files were searched against the reviewed human SwissProt reference database using the Mascot (Matrix Science, MA, USA) search engine version 2.4.0. Searches were conducted with tryptic specificity, carbamidomethylation of cysteine residues as static modification and the oxidation of methionine residues as a dynamic modification. Using a reversed decoy database, false discovery rate was set to less than 1% and above the Mascot specific peptide identity threshold. For SWATH-MS processing, ProteinPilot search outputs from DDA runs were used to generate a spectral library for targeted information extraction from SWATH-MS data files using PeakView v2.1 with SWATH MicroApp v2.0 (SCIEX). Protein areas, summed chromatographic area under the curve of peptides with extraction FDR \leq 1%, were calculated and used to compare protein abundances between bait and control IPs.

Chromatin immunoprecipitation-quantitative real-time PCR analysis

Chromatin immunoprecipitation (ChIP) was conducted as described previously (37); however, following overnight IP, the samples were processed using a previously described protocol (40).

DNA was purified then quantified using quantitative real-time PCR analysis. Control regions analysed and primers used are listed in Table S4.

Relative enrichment of each region/primer set was calculated by taking an average of each duplicate reaction. The input Ct value was subtracted from the sample Ct value and the Ct converted using the respective PE for each primer set. The relative ChIP enrichment is then calculated by dividing the gene region of interest by the specific control region that is negative for both ID4 and H3K4Me3 binding (IFF01/NOP2 #1 primer). The formula for this normalisation is below:

$$\Phi Ct = Ct_{\text{region of interest}} - Ct_{\text{input region}}$$

$$\text{ChIP enrichment} = PE^{[-\Phi Ct(\text{region of interest})]} - PE^{[-\Phi Ct(\text{IFF01})]}$$

A sample was considered to be enriched if the fold-change over IgG control for each region was >2.

Chromatin immunoprecipitation-sequencing

Chromatin Immunoprecipitation and sequencing (ChIP-seq) was conducted as previously described (37). Samples were prepared and sequenced at Cancer Research United Kingdom (CRUK), Cambridge, UK. Antibody conditions for ChIP are the same as those used for RIME, with the addition of antibodies targeting H3K4Me3 (Active Motif #39159) and γ H2AX (Ser139) (1:1 mix of Cell Signalling #2577 and Merck Millipore clone JBW301). Samples were sequenced at CRUK using an Illumina HiSeq 2500 single-end 50 bp sequencing. Quality control was conducted using FastQC (41) and sequencing adapters trimmed using cutadapt (42). Reads were aligned using Bowtie for Illumina v0.12.7 (43) followed by Sam-to-Bam conversion tool (44) and alignment using Bwa v0.705a (44). Alignment statistics were generated using samtools flagstat [version 0.1.18 (r982:295)]

(44). ChIP-seq peaks were called using the peak calling algorithm HOMER v4.0 and MACS v1.4.2 (45, 46).

Chromatin Immunoprecipitation-exonuclease sequencing (ChIP-exo) was conducted as previously described (38). Samples were prepared and sequenced at CRUK.

qPCR analysis of ChIP DNA

Publicly available H4K3Me3 ChIP-sequencing data and the ID4 ChIP-sequencing data generated in this project were visualised using UCSC Genome Browser (genome.ucsc.edu and (47)). Regions of positive and negative enrichment were selected and the 500-1,000 bp DNA sequence was imported into Primer3, a primer design interface, web version 4.0.0 (48). Primers were designed with a minimum primer amplicon length of 70 bp. Primers were confirmed to align with specific DNA segments by conducting an *in silico* PCR using UCSC Genome Browser (genome.ucsc.edu and (47)). Oligo primers were ordered from Integrated DNA Technologies (Singapore). Primers were tested to determine adequate primer efficiency (between 1.7-2.3). All assays were set up using an EPmotion 5070 robot (Eppendorf, AG, Germany) and run on an ABI PRISM 7900 HT machine (Life Technologies, Scoresby, VIC, Australia). Briefly, reactions were performed in triplicate in a 384-well plate. Each reaction consisted of 1 μ L 5 μ M Forward primer, 1 μ L 5 μ M Reverse primer, 5 μ L SYBR Green PCR Mastermix (Thermo Fisher) and 3 μ L DNA. A standard curve was created using unsonicated, purified DNA extracted from the HCC70 cell line in 10-fold dilutions (1, 0.1, 0.01, 0.001, 0.0001).

PCR cycling was as follows: 1 cycle at 50°C for 2 min, 1 cycle at 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. A dissociation step was conducted at 95°C for 15 sec and 60°C for 15 sec. Data was analysed and a standard curve created using SDS 2.3 software (Applied Biosystems). The slope was used to calculate the PE using the qPCR Primer Efficiency Calculator (Thermo Fisher Scientific, available at thermofisher.com).

Patient-derived xenograft and histology

All experiments involving mice were performed in accordance with the regulations of the Garvan Institute Animal Ethics Committee. NOD.CB17-Prkdc^{scid}/Arc mice were sourced from the Australian BioResources Ltd. (Moss Vale, NSW, Australia). Assoc. Prof Alana Welm (Oklahoma Medical Research Foundation) kindly donated the patient-derived xenografts (PDX) models used in this study. Models were maintained as described elsewhere (49). Tumour chunks were transplanted into the 4th mammary gland of 5-week old recipient NOD.CB17-Prkdc^{scid}/Arc mice. Tumours were harvested at ethical endpoint, defined as having a tumour approximately 1 mm³ in size or deterioration of the body condition score. At harvest, a cross-section sample of the tumour was fixed in 10% neutral buffered formalin (Australian Biostain, Traralgon, VIC, Australia) overnight before transfer to 70% ethanol for storage at 4°C before histopathological analysis. The formalin fixed paraffin embedded (FFPE) blocks were cut in 4 µm-thick sections and stained for ID4 (Biocheck BCH-9/82-12, 1:1,000 for 60 min following antigen retrieval using pressure cooker 1699 for 1 min, Envision Rabbit secondary for 30 min). Protein expression was scored by a pathologist using the H-score method (50).

Fluorescent In Situ Hybridisation

Tissue sections were analysed using Fluorescent In Situ Hybridisation (FISH) to examine the genomic region encoding *ID4* (6p22.3). *ID4* FISH Probe (Orange 552 nm-576 nm, Empire Genomics, NY, USA) was compared to the control probe *CEP6* (Chromosome 6, Green 5-Fluorescein dUTP). This *CEP6* probe marks a control region on the same chromosome as *ID4* and is used to normalise *ID4* copy number. Breast pathologist Dr Sandra O'Toole oversaw the FISH quantification for all samples.

Immunofluorescence and Proximity ligation assays

Immunofluorescence: Cells were seeded on glass coverslips (Coverglass, 13 mm, VITLAB, Germany). At harvest, media was removed, cells were washed twice with PBS without salts and fixed in 4% paraformaldehyde (PFA) (ProSciTech) for 10 min. Cells were again washed twice with PBS without salts (Thermo Fisher Scientific) before permeabilising for 15 min with 1% Triton-X (Sigma-Aldrich) in PBS and then blocking with 5% BSA in PBS without salts for 1 h at room temperature. Cells were washed twice with PBS without salts and antibodies were applied overnight at 4°C: ID4 (Biocheck BCH-9/82-12, 1:1,000), MDC1 (Sigma-Aldrich M2444, 1:1,000), BRCA1 (Merck Millipore (Ab-1), MS110, 1:250), γ H2AX (Ser139) (Merck Millipore clone JBW301 05-636, 1:300), FLAG (Sigma-Aldrich M2, 1:500) and V5 (Santa Cruz sc-58052, 1:500). Cells were washed twice with PBS without salts then secondary antibodies were applied for 1 h at room temperature. Cells were washed twice with PBS without salts, with the second wash containing DAPI (1:500 dilution) and phalloidin (1:1,000 dilution) (CytoPainter Phalloidin-iFluor 633 Reagent Abcam ab176758). Cells were then mounted on slides using 4 μ L of Prolong Diamond (Thermo Fisher Scientific).

Duolink Proximity ligation assay analysis (PLA): PLA was conducted using Duolink PLA technology with Orange mouse/ rabbit probes (Sigma-Aldrich, DUO92102) according to the manufacturers instructions. Images were captured using SP8 6000 confocal imaging with 0.4 μ m Z-stacks. Maximum projects were made for each image (100-200 cells) and quantified using FIJI by ImageJ (51) as described previously (52). Quantification was conducted on a minimum of 50 cells. Data is represented as number of interactions (dots) per cell.

Quantification of DNA damage foci: Image quantification was conducted using FIJI v2.0.0 image processing software (Fiji is just ImageJ, available at Fiji.sc, (51)) as previously described (52). Four to five images were taken of each sample. The DAPI channel was supervised to enable accurate gating of cell nuclei for application to other channels. Size selection (pixel size 2,000 to 15,000) and circularity (0.30-1.00) cut-offs were used. Cells on the edge of the image were excluded from the analysis. The number of DNA damage foci

per cell nucleus was calculated for approximately 100-200 cells. The information for individual samples was then collated and analysed using the Pandas package in Python 3.5.

Clinical Cohorts

Basal-like breast cancer: Samples were stratified into groups as follows: 42 BLBC (negative for ER α , PR, HER2 and positive for CK5/6, CK14 or EGFR), 14 triple negative non-BLBC (negative for ER α , PR, HER2, CK5/6, CK14 and EGFR) and 26 HER2-Enriched (negative for ER α and PR, positive for HER2). BRCA1-mutation status in this cohort is unknown, however it is expected to occur in approximately 6.5% of BLBC patients (53). Samples were obtained under the Garvan Institute ethical approval number HREC 08/145.

Kathleen Cuningham Foundation Consortium for research into familial breast cancer (KConfab): BRCA1-mutant BLBC was sourced from KConfab. A total of 97 BRCA1-mutant BLBC cases were obtained under the Garvan Institute ethical approval number HREC 08/145.

Ovarian Cancer: A total of 97 HGSOC cases were obtained under the Human Research Ethics Committee of the Sydney South East Area Hospital Service Northern Section (00/115) (54).

Results

ID4 interacts with chromatin without regulating transcription

To investigate the molecular function of ID4, we first examined ID4 protein expression and cellular localisation across a panel of breast and ovarian cancer cell lines. Ovarian cancer cell lines were included due to the established molecular similarities between BLBC and HGSOC (5-7). Furthermore, evidence suggests ID4 may play an important role in the aetiology of both BLBC and HGSOC (11, 55). As described previously (11), ID4 is predominantly expressed by cell lines of the BLBC subtype (Fig. S1A). Variable ID4 protein expression was observed across the ovarian cell lines, and unlike in BLBC ID4 did not associate specifically with HGSOC, the more aggressive ovarian cancer subtype. Four ID4 expressing models were selected for further analysis: MDA-MB-468 (BLBC), HCC70 (BLBC), HCC1954 (HER2-Enriched) and OVKATE (HGSOC), because these models represent different biological systems with similar levels of ID4 expression.

ID4 predominantly localised to chromatin, as evidenced by distinct punctate staining in the nucleus, with a proportion of ID4 staining overlapping with DAPI-low nuclear regions (Fig. 1A), typical of uncondensed euchromatin. Due to this localisation, and a previous report of ID4 associating with enhancer regions in normal mammary epithelial cells (12), chromatin immunoprecipitation and sequencing (ChIP-seq) was used to interrogate whether, and how, ID4 binds chromatin in BLBC. ID4 ChIP-seq was conducted on asynchronous HCC70 cells in three biological replicates and compared with IgG ChIP and an input control. Western blot analysis confirmed successful immunoprecipitation of ID4 (Fig. S1B). ChIP-seq data was integrated by removing signal identified in the input and IgG controls and examining the overlap between the three ID4 ChIP-seq replicates. Remarkably, ID4 reproducibly associated with seven sites across the genome (Fig. S2A and Table S1). Some of these were focussed, small binding events typical of transcription factor binding. For example, ID4 binding to the transcription start site of *FAIM* and *GBA* (140-230 bp) (Fig. S2C) was typical of transcription factor binding peaks. In other locations, ID4 binding was spread over very large regions of DNA, up to 10kb in length, which is reminiscent of some histone marks, and not transcription factor binding profiles. ID4 bound primarily to gene bodies of the protein-coding genes *ELF3* and *ZFP36L1*, the long non-coding RNAs *NEAT1* and *MALAT1* and the microRNA host-gene *MIRLET7BHG*. The MACS peak-calling algorithm identified multiple peaks per gene for these genes, due to wide spread binding across the region (*ELF3*, *NEAT1*, *MALAT1*, *MIRLET7BHG* and *ZFP36L1*) (Fig. 1B and Fig. S2C). This binding is absent from intergenic regions (Fig. 1B).

Due to the large expanses of ID4 binding, ChIP-exonuclease (ChIP-exo) was employed to provide higher resolution mapping of ID4 binding. ChIP followed by exonuclease digestion and DNA sequencing identifies precise transcription factor binding sites with higher resolution than traditional ChIP-seq (38). ID4 ChIP-exo was conducted in biological duplicate in HCC70 and HCC1954 breast cancer cell lines (Fig. S3A and Table S1). Consistent with the initial ChIP-seq experiment, we identified significant enrichment of ID4 binding to the genomic regions encoding *NEAT1*, *MALAT1* and *GBA* (Fig. S3B). Signal was observed at *ELF3* and *ZFP36L1*, however these regions were not identified through MACS analysis (Fig. S3C). Additional ID4 peaks were identified in several transfer RNAs (tRNA) (Fig. 1C) and within the gene bodies of *KDM4C* and *ERRFI1* (Fig. S3D). This method was unable to further resolve ID4-chromatin binding information and the data resembled the ChIP-seq data. This suggests that ID4 forms large contiguous and uninterrupted interactions with expanses of chromatin. The ID4-bound chromatin aligned with regions of high transcriptional activity, co-localising with DNaseI hypersensitivity clusters, active histone marks, and *BRCA1* and *POLR2A* transcription factor binding sites (Fig. 1D).

ChIP-qPCR was used to validate ID4 binding in HCC70, HCC1954 and MDA-MB-468 cell lines. Primers to the genes of interest were designed to tile across multiple sites in each gene, with several regions enriched for ID4 binding compared to input and IgG controls (Fig. 1E and Fig. S4). ChIP-qPCR validation from independent ChIP experiments validated the findings made from the unbiased ChIP-seq approaches.

To address the specificity of ID4 binding, ChIP-qPCR analysis was conducted on HCC70 cell line modified with the SMARTChoice doxycycline-inducible ID4 short hairpin. Doxycycline treatment resulted in an 87%

reduction in ID4 protein expression at 72 h (Fig. S5). ID4 knockdown reduced ID4-chromatin binding at a majority of sites (Fig. S6b-c), confirming the specific binding of ID4 to these loci.

Although ID proteins are classically thought to be transcriptional regulators, to date, no studies have reported a systematic analysis of ID4 transcriptional targets. Hence, to determine whether ID4 affects transcription in BLBC, we analysed changes in gene expression using RNA-sequencing (RNA-seq) following ID4 depletion using the SMARTChoice model described above. This analysis was also used to identify whether ID4 regulates the expression of genes adjacent to ID4 chromatin interactions. Differential expression analysis revealed limited changes in gene expression following loss of ID4, with only six genes identified as being differentially expressed (Table S2), indicating ID4 does not function as a transcriptional regulator in this model. Loss of ID4 did not affect the expression of the ID4-bound genes identified through ChIP-seq and ChIP-exo analysis, as determined through RNA-seq and qPCR (Fig. 1F). This suggests that ID4 has a transcription-independent role on chromatin.

ID4 interacts with DNA damage response proteins

To understand how ID4 interacts with chromatin in the absence of a known DNA interaction domain, endogenous ID4 was purified from cells using the proteomic technique Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME) (36). Anti-ID4 and control IgG were used for immunoprecipitations in biological triplicate (each in technical duplicate or triplicate) followed by mass spectrometry in the HCC70, MDA-MB-468, HCC1954 and OVKATE cell lines. Proteins were considered bona fide ID4-binding targets if they appeared in two or more technical and biological replicates and were not present in any IgG controls (Fig. 2A).

RIME analysis identified cell line-unique and common ID4-binding candidates (Fig. 2B and Table S3). In total 1,106 unique proteins were identified across the four cell lines. Gene Set Enrichment Analysis (GSEA) of the putative ID4 binding partners showed enrichment for chromatin (DNA damage and transcription proteins and histones) and RNA-associated gene sets (including RNA splicing and processing factors) (Fig. 2C and Table S3). This compliments the above ChIP-seq data, demonstrating that ID4 interacts with chromatin and nuclear machinery. The BRCA1-PCC network was the most significantly enriched gene set in the ID4 purified proteome (PCC: Pearson Correlation Coefficient, 101/852 ID4-associated genes were present in the set of 1652 BRCA1-associated genes; $p=6.47E-59$) (Fig. 2C). This gene set comprises networks controlling breast cancer susceptibility (56). Other gene sets originating from this previous study (56), including the CHEK2-PCC and ATM-PCC, were also identified as highly enriched, suggesting an association between ID4 and DNA damage repair factors.

Across the entire RIME dataset, only one peptide mapping to a bHLH protein, HEB (also known as TCF12) was identified, in one of nine technical replicates in the OVKATE cell line. We therefore concluded that

HLH proteins are not significant ID4 interactors in BLBC and HGSOC.

Six proteins were commonly identified in all four-cell lines: ID4, mediator of DNA damage checkpoint 1 (MDC1), disintegrin and metalloprotease domain 9 protein (ADAM9), Histidine Rich Glycoprotein (HRG), splicing factor 3a subunit 2 (SF3A2) and spectrin repeat containing nuclear envelope family member 3 (SYNE3). The most highly enriched protein, MDC1, was the only protein to be identified in all RIME technical and biological replicates in all four cell lines. MDC1 has many cellular functions, including cell cycle control and transcription (57, 58), yet is best characterised for its role in the DNA damage response, where it associates with DNA (59). Of the other proteins identified, ADAM9, and HRG affect cell motility, SYNE3 is important in cytoskeletal and nuclear organisation (60) and SF3A2 is required for pre-mRNA splicing and the generation of alternative transcripts (61). ID4 was previously suggested to bind to mutant p53 and SRSF1 (62, 63), yet no evidence was found of these associations in our analysis.

To validate these findings in a more clinically relevant setting, we conducted ID4 RIME in Patient derived xenograft (PDX) models of triple negative breast cancer (49). Three ID4-expressing models (HCI001, HCI002, HCI009) were selected for RIME analysis based on detectable expression of ID4 (Fig. 2D). ID4 and MDC1 were the only targets identified in ID4 RIME in all three xenograft models and absent in the IgG RIME (see Table S3 for full list of proteins identified in PDX models). This discovery proteomics was supported by quantitative mass spectrometry analysis of the models using the sensitive label-free quantitative proteomic method termed SWATH (Sequential Window Acquisition of all THEoretical mass spectra) (64). This analysis showed a high abundance of both ID4 and MDC1 in all three PDX models (Fig. 2D). Due to the robust identification of MDC1 in the four cell lines and the subsequent validation of this interaction in three PDX models, we focused further studies on the functional importance of the ID4-MDC1 relationship.

ID4-MDC1 binding was validated using independent methods including immunoprecipitation with an anti-ID4 antibody and anti-MDC1 antibody, followed by western blotting with an independent monoclonal ID4 or a MDC1 antibody (Fig. 2E). This co-immunoprecipitation experiment confirmed the ID4-MDC1 interaction in MDA-MB-468 cells.

ID4 interacts with DNA repair proteins at fragile chromatin sites in a DNA damage-dependent manner and regulates DNA damage signalling

To further explore the interaction between ID4 and MDC1 we used proximity ligation assays (PLA), a quantitative measure of protein interaction *in situ*, and confirmed binding of ID4 with MDC1 in MDA-MB-468 cell line. This interaction occurred predominantly in the cell nucleus (Fig. 3A). Depletion of MDC1 using siRNA resulted in reduction of the PLA signal to control levels (Fig. 3B). PLA analysis was also used to test the association of ID4 with γ H2AX, as it is required for signalling double-stranded DNA breaks and

with BRCA1, as it is a critical mediator of the homologous recombination (HR) repair pathway. PLA analysis identified enrichment of ID4 binding with MDC1, γ H2AX and BRCA1 above control levels in HCC70 and with MDC1 in OVKATE cell lines (Fig. S7).

To more closely examine the correlation between DNA damage and ID4 complex association with chromatin, we used ionising radiation (IR) to induce DNA damage in cells. Irradiation markedly increased ID4 chromatin binding, measured by ChIP-qPCR; up to 160-fold compared to undamaged controls at the NEAT1 gene, and 5-70-fold at ELF3, MALAT1, MIRLET7BHG, ZFP36L1, GBA and various tRNAs (Fig. 3C and Fig. S8). This indicates that ID4 binding is specific and induced by DNA damage. ChIP-qPCR for MDC1 similarly showed induced binding to a majority of these sites in damaged cells (Fig. S8). Taken together, these results suggest a positive correlation between DNA damage and ID4-chromatin binding.

To investigate the functional role for ID4 in the DDR, DNA damage was induced using IR and damage sensing was monitored over time. Cells were treated with doxycycline to deplete ID4 for 72h (point of maximal ID4 knockdown) prior to treatment with ionising radiation. Cells were fixed at 0, 15min and 8h following ionising radiation, to examine the DNA damage response over time. As previously reported (65), the number of γ H2AX foci in control cultures increased at 15min post-IR and remained high at 8h (Fig. 3D), while the number of 53BP1 foci increased at 8h following ionising radiation (Fig. 3E). Upon ID4 depletion, formation of γ H2AX and 53BP1 foci was significantly reduced at 8h post-IR. This suggests that in the absence of ID4, the DNA was either repaired more efficiently, or alternatively that the DNA damage response was not triggered properly.

ID4 genetically interacts with *BRCA1* mutation

The data presented earlier suggests an interaction between ID4 function and DNA damage pathways. Patients with germline *BRCA1* mutations are at elevated risk of BLBC and HGSOB (3). While BRCA1-associated cancers often undergo *BRCA1* loss of heterozygosity (LOH), recent evidence suggests that mutation and copy number change of additional genes, such as *TP53*, often occur prior to, and may be required for, loss of the wildtype *BRCA1* allele (66). We therefore assessed whether there is evidence for copy number change of *ID4* in (i) sporadic disease or (ii) in the context of familial breast cancer driven by germline *BRCA1* mutation.

ID4 is reportedly amplified in sporadic BLBC (5, 6, 55), however these results are based on small cohorts and imprecise array Comparative Genomic Hybridisation (aCGH) techniques, both of which affect accurate determination of amplification frequency. To definitively quantify *ID4* amplification frequencies, we used fluorescent *in situ* hybridisation (FISH) in a clinical diagnostic laboratory, a highly specific approach for analysing copy number. FISH was applied to a discovery cohort of 82 sporadic oestrogen receptor-negative invasive ductal carcinomas: composed of 42 BLBC (negative for ER α , PR, HER2 and positive for CK5/6, CK14 or EGFR), 14 triple negative non-BLBC (negative for ER α , PR, HER2, CK5/6, CK14

and EGFR) and 26 HER2-Enriched (negative for Era and PR, positive for HER2) samples (11). *ID4* amplification was exclusive to the BLBC subtype with an amplification frequency of ~10% (4/42 cases), and amplification correlated with ID4 protein expression ($r = 0.265$, $p = 0.0088$) (Fig. 4A and Fig. S9).

We also examined whether ID4 is amplified in sporadic HGSOC. Analysis of 97 sporadic HGSOC cases (54) showed a comparable rate of *ID4* amplification relative to sporadic BLBC of ~10% (10/97 in cases of HGSOC) (Fig. 4E,G). However, these frequencies are lower than that published using Array CGH (5, 6), potentially suggesting that aCGH methods overestimate amplification frequencies or that these are distinct clinical cohorts. ID4 protein expression in HGSOC correlated with amplification ($r = 0.457$ and $p < 0.0001$) (Fig. 4F-G).

To determine potential genetic interactions between *ID4* and *BRCA1*, *ID4* copy number and protein expression was evaluated in 97 familial BLBCs from patients with known germline *BRCA1* mutations and compared to the 42 unselected BLBC (which are not expected to frequently carry *BRCA1* germline mutations) described above. *ID4* was amplified at twice the frequency in *BRCA1*-mutant BLBC (~22%, 21/97 cases) compared to sporadic BLBC (~10%, 4/42 cases), indicating a selection for *ID4* amplification in cancers arising in patients carrying a mutant *BRCA1* allele (Fig. 4B-C). ID4 protein expression was also significantly higher in the *BRCA1*-mutant BLBC compared to the sporadic BLBC cohort (Fig. 4D). These data provide further evidence for an important role for ID4 in BLBC and HGSOC aetiology.

Discussion

By elucidating the drivers and dependencies of BLBC, we aim to improve our understanding of this complex, heterogeneous disease, potentially leading to the identification of novel targets and therapeutics. Previous work from our group and others has shown that ID4 acts as a proto-oncogene in BLBC (11, 14-21). ID4-positive BLBC have a very poor prognosis and depletion of ID4 reduced BLBC cell line growth *in vitro* and *in vivo* (11), suggesting that ID4 controls essential, yet unknown intrinsic pathways in BLBC.

We have conducted the first systematic mapping of the chromatin interactome of ID4 in mammalian cells. Using ChIP-seq we have identified novel ID4 binding sites within the BLBC genome. ID4 bound to large regions of chromatin, up to 10kb in length, at a very small number of loci, suggesting that ID4 is not binding as part of a conventional transcriptional regulatory complex. This conclusion is supported by the observation that ID4 knockdown did not affect gene expression at these loci. The regions identified through ChIP-seq were typically the bodies of genes that are highly transcribed and mutated in cancer (67, 68), characteristics of fragile, DNA damage-prone sites. Interestingly, these sites primarily encode non-coding RNA, including lncRNA, microRNA precursors and tRNA. The lncRNA NEAT1 and MALAT1 are some of the most abundant cellular RNAs and the genes encoding them undergo recurrent mutation in breast cancer (69). Furthermore, they are upregulated in ovarian tumour cells and are associated with higher tumour grade and stage and in metastases (70). The genomic loci encoding tRNAs are also highly transcriptionally active, co-localising with DNaseI hypersensitivity clusters and transcription factor

binding sites, including BRCA1 and POLR2A (71). Transcriptional activity is highly stressful and associated with genomic stress and DNA damage (67). Binding of ID4 to these sites was increased upon DNA damage. Together these data suggest that ID4 binds preferentially to sites of active transcription and DNA damage, consistent with its interaction with MDC1.

We have conducted the most systematic and unbiased proteomic analysis of binding partners for any ID family member. RIME revealed hundreds of ID4 interacting proteins, which were highly enriched for BRCA1-associated proteins. Five novel proteins were found to interact with ID4 with high confidence in all 4 models examined, namely MDC1, ADAM9, HRG, SF3A2 and SYNE3. These warrant further investigation to examine the role they may have in the mechanism of action of ID4. Interestingly, no bHLH proteins were reproducibly found in complex with ID4, although they are the canonical ID binding partners in certain non-transformed cells (22-26). This is unlikely to be a technical artefact, as ID4-bHLH interactions are readily detected in non-transformed mammary epithelial cells using the same method (H. Holliday *et al*; Unpublished data). Rather, ID4 may have alternate binding partners as a consequence of downregulation of bHLH proteins in BLBC (H. Holliday *et al*; Unpublished data).

ID4 has been demonstrated to be a member of a ribonucleoprotein complex along with mutant p53, SRSF1 and lncRNA MALAT1 (63). This complex promotes splicing of VEGFA pre-mRNA, which signals in a paracrine manner to macrophages and ultimately results in tumour angiogenesis (62, 63, 72). p53 and SRSF1 were not identified in our RIME analysis. This disparity may be the consequence of methodological differences such as fixation, cell lysis conditions and detection techniques. It is possible that ID4 is a member of a large complex encompassing several splicing factors (including SF3A1 and SRSF1). The observations that ID4 interacts with both the MALAT1 gene, as well as MALAT1 lncRNA itself (63), suggests that ID4 and MALAT1 function are intimately linked, and warrants further investigation.

MDC1, the most reproducible binding partner of ID4 identified through RIME analysis, is recruited to sites of DNA damage to amplify the phosphorylation of H2AX (to form γ H2AX) and recruit downstream signalling proteins (59). Deficiency in MDC1, much like BRCA1, results in hypersensitivity to double-stranded DNA breaks (73). MDC1 interacted with many of the sites of chromatin interaction by ID4 in a DNA damage dependent manner. We also find ID4 in close proximity of known MDC1 interactors, including BRCA1 and γ H2AX. These data suggest a model in which ID4 associates with the DNA damage repair apparatus at sites of genome instability or damage via its interaction with MDC1. MDC1 was also recently found to associate with ID3, suggesting that this is a conserved feature of ID proteins (74). How MDC1 binds ID4 is unknown, however a quasi HLH domain (75) structure within MDC1 may enable interaction with the HLH domain of ID4. As complex feedback mechanisms govern the DNA damage response, further investigation is required to determine whether ID4 association with MDC1 ultimately promotes or impedes DNA repair.

BRCA1 has proposed functions as a transcription factor controlling differentiation in non-transformed cells (71, 76), but a primary role in DNA repair in cancer (9). Similarly, ID4 is an important regulator of

stemness in the developing mammary gland, acting to inhibit differentiation (11, 12). However, the results presented here have uncovered an unexpected role for ID4 in the DNA damage response in BLBC, suggesting a similar dichotomy of function to BRCA1. That is, primarily regulating transcription during development whilst predominantly regulating the DNA damage response in cancer. Transcription is a stressful cellular process causing significant DNA damage and repair (67). Thus, a role for transcription factors in localising DNA damage machinery to chromatin may be an important cellular capability.

Mutations in BRCA1 predispose carriers predominantly to cancers of the breast and ovaries (mostly BLBC and HGSOC), though the mechanism driving tumorigenesis in these patients is still unclear. While many BRCA1-associated cancers undergo LOH for the wildtype BRCA1 allele, many acquire other genomic 'hits' prior to this event, which may be required for subsequent LOH (77). In addition to reporting a biochemical interaction between ID4 and MDC1, we also show a novel genetic interaction with BRCA1, in that *ID4* is amplified at twice the frequency in *BRCA1*-mutant BLBC compared to sporadic BLBC, making it one of the most frequently amplified genes in that disease. A caveat of this finding is that other cancer-associated genes, such as *E2F3*, are located adjacent to *ID4* at Chr6q22 and so may also be a target of the amplification event.

Further work is required to understand the drivers of ID4 amplification and its contribution to DNA repair, however at least 2 scenarios are possible. In the first, ID4 acts to suppress DNA repair proficiency and so cooperates with *BRCA1* haploinsufficiency to promote genomic instability and tumourigenesis. This is consistent with the positive correlation between ID4 expression and 'BRCAness' (15), a defect in BRCA1 function in the absence of germline *BRCA1* mutations (78). However perhaps more likely is the opposing scenario, that *ID4* amplification and overexpression promote DNA damage repair, consistent with our observations of ID4 association with MDC1 at fragile sites and the ongoing requirement for ID4 in BLBC proliferation that we previously reported (11). In the case of BRCA1-associated breast cancer, *ID4* amplification, like *TP53* mutation, may be permissive for subsequent *BRCA1* LOH which is otherwise lethal (66), explaining the high frequency of *ID4* amplification in familial cancers. Further resolving the function of ID4 in BLBC will require detailed biochemical analysis of DNA repair functional assays, genetic studies with ID4 knockout cells or animals and access to a large cohort of familial breast cancers with detailed gene copy number, BRCA1 mutation and methylation data.

Conclusions

In summary, in this integrated analysis of ID4 function in BLBC and HGSOC, we show that ID4 localises to DNA at sites of active transcription and DNA damage, bridged through its biochemical interaction with the DNA damage response machinery; namely MDC1. Rather than regulating transcription at these sites, our data points to a role of ID4 in the DNA damage response. Finally, we establish a genetic interaction between *BRCA1* and *ID4*, where *ID4* is amplified at twice the frequency in *BRCA1*-mutant BLBC compared to sporadic disease. Though further studies are required to dissect the precise involvement of ID4 in the DNA damage response, these novel insights expand upon our current understanding of the mechanism by which ID4 drives malignancy in breast and ovarian cancer.

List Of Abbreviations

BLBC: Basal-like breast cancer

BRCA1: Breast and ovarian cancer susceptibility gene

ChIP-seq: Chromatin Immunoprecipitation followed by DNA Sequencing

FISH: Fluorescent in situ hybridisation

GSEA: Gene Set Enrichment Analysis

HGSOC: High-grade serous ovarian cancer

HLH: Helix-loop-helix

HR: Homologous Recombination

ID4: Inhibitor of Differentiation 4

IR: Ionising Radiation

LOH: Loss of heterozygosity

MDC1: Mediator of DNA damage checkpoint 1

PDX: Patient Derived Xenograft

PLA: Proximity Ligation Assay

RIME: Rapid Immunoprecipitation and Mass spectrometry of Endogenous proteins

SWATH: Sequential Window Acquisition of all THEoretical mass spectra

Declarations

Ethics approval and consent to participate

Mouse PDX experiments were conducted under the Garvan Institute ethical approval number 12/11. Breast cancer cohorts were obtained under the Garvan Institute ethical approval number HREC 08/145. Ovarian cancer cohort was obtained under the Human Research Ethics Committee of the Sydney South East Area Hospital Service Northern Section (00/115)

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are included as supplementary information files.

Competing interests

The authors declare no competing interests.

Funding

Australia Postgraduate Award (to LAB and HH), Boehringer Ingelheim Fonds, Cancer Council NSW, The National Health and Medical Research Council of Australia, The McMurtrie family, Beth Yarrow Memorial Award, Castle Harlan Award, LH Ainsworth Award and Research conducted at APAF was supported by the Australian Government's National Collaborative Research Infrastructure Scheme.

Authors' contributions

L.A.B designed and performed the experiments and wrote the manuscript. H.H assisted with experiments, discussions and revised the manuscript. D.R performed CHIP-seq/exo and RNA-seq analysis. C.K and M.M performed the mass spectrometry and analysis. A.A.S, H.M and G.S advised and contributed to the RIME and CHIP experiments. A.M performed mouse experiments. A.M.K.L assisted with image analysis. P.T.S, S.L, C.S, L.A, G.A, N.F.H and E.L contributed to analysis of clinical cohorts. S.Z.W, S.J, R.N, E.D, K.S and I.N contributed to experiments and discussions. M.J.N, C.J.O, W.K, M.J.C and J.S.C helped conceptualise the project. A.S conceived the study and was in charge of overall direction and planning.

Acknowledgements

We thankfully acknowledge the support of the above funding bodies and Grant Houghton for assistance with data analysis.

References

1. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-52.
2. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*. 2001;98(19):10869-74.
3. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. 1994;266(5182):66-71.
4. Turner N, Reis-Filho J, Russell A, Springall R, Ryder K, Steele D, et al. BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene*. 2007;26(14):2126-32.

5. Network CGA. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.
6. Network CGAR. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011;474(7353):609-15.
7. Marcotte R, Sayad A, Brown KR, Sanchez-Garcia F, Reimand J, Haider M, et al. Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. *Cell*. 2016;164(1):293-309.
8. Mullan P, Quinn J, Harkin D. The role of BRCA1 in transcriptional regulation and cell cycle control. *Oncogene*. 2006;25(43):5854.
9. Scully R, Xie A, Nagaraju G. Molecular functions of BRCA1 in the DNA damage response. *Cancer biology & therapy*. 2004;3(6):521-7.
10. Vollebergh M, Lips E, Nederlof P, Wessels L, Schmidt M, Van Beers E, et al. An aCGH classifier derived from BRCA1-mutated breast cancer and benefit of high-dose platinum-based chemotherapy in HER2-negative breast cancer patients. *Annals of Oncology*. 2010;22(7):1561-70.
11. Junankar S, Baker LA, Roden DL, Nair R, Elsworth B, Gallego-Ortega D, et al. ID4 controls mammary stem cells and marks breast cancers with a stem cell-like phenotype. *Nature communications*. 2015;6.
12. Best SA, Hutt KJ, Fu NY, Vaillant F, Liew SH, Hartley L, et al. Dual roles for Id4 in the regulation of estrogen signaling in the mammary gland and ovary. *Development*. 2014;141(16):3159-64.
13. Dong J, Huang S, Caikovski M, Ji S, McGrath A, Custorio MG, et al. ID4 regulates mammary gland development by suppressing p38MAPK activity. *Development*. 2011;138(23):5247-56.
14. Beger C, Pierce LN, Krüger M, Marcusson EG, Robbins JM, Welcsh P, et al. Identification of Id4 as a regulator of BRCA1 expression by using a ribozyme-library-based inverse genomics approach. *Proceedings of the National Academy of Sciences*. 2001;98(1):130-5.
15. Branham M, Campoy E, Laurito S, Branham R, Urrutia G, Orozco J, et al. Epigenetic regulation of ID4 in the determination of the BRCAness phenotype in breast cancer. *Breast cancer research and treatment*. 2016;155(1):13-23.
16. Crippa E, Lusa L, De Cecco L, Marchesi E, Calin GA, Radice P, et al. miR-342 regulates BRCA1 expression through modulation of ID4 in breast cancer. *PloS one*. 2014;9(1):e87039.
17. de Candia P, Akram M, Benezra R, Brogi E. Id4 messenger RNA and estrogen receptor expression: inverse correlation in human normal breast epithelium and carcinoma. *Human pathology*. 2006;37(8):1032-41.
18. Roldán G, Delgado L, Musé IM. Tumoral expression of BRCA1, estrogen receptor alpha and ID4 protein in patients with sporadic breast cancer. *Cancer biology & therapy*. 2006;5(5):505-10.
19. Shan L, Yu M, Qiu C, Snyderwine EG. Id4 regulates mammary epithelial cell growth and differentiation and is overexpressed in rat mammary gland carcinomas. *The American journal of pathology*. 2003;163(6):2495-502.
20. Thike AA, Tan PH, Ikeda M, Iqbal J. Increased ID4 expression, accompanied by mutant p53 accumulation and loss of BRCA1/2 proteins in triple-negative breast cancer, adversely affects

- survival. *Histopathology*. 2015.
21. Wen YH, Ho A, Patil S, Akram M, Catalano J, Eaton A, et al. Id4 protein is highly expressed in triple-negative breast carcinomas: possible implications for BRCA1 downregulation. *Breast cancer research and treatment*. 2012;135(1):93-102.
 22. Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell*. 1990;61(1):49-59.
 23. Jen Y, Weintraub H, Benezra R. Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. *Genes & development*. 1992;6(8):1466-79.
 24. Loveys DA, Streiff MB, Kato GJ. E2A basic-helix-loop-helix transcription factors are negatively regulated by serum growth factors and by the Id3 protein. *Nucleic acids research*. 1996;24(14):2813-20.
 25. Langlands K, Yin X, Anand G, Prochownik EV. Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *Journal of Biological Chemistry*. 1997;272(32):19785-93.
 26. Roberts EC, Deed RW, Inoue T, Norton JD, Sharrocks AD. Id helix-loop-helix proteins antagonize pax transcription factor activity by inhibiting DNA binding. *Molecular and cellular biology*. 2001;21(2):524-33.
 27. Jen Y, Manova K, Benezra R. Expression patterns of Id1, Id2, and Id3 are highly related but distinct from that of Id4 during mouse embryogenesis. *Developmental Dynamics*. 1996;207(3):235-52.
 28. Norton JD. ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis. *Journal of cell science*. 2000;113(22):3897-905.
 29. Chaudhary J, Johnson J, Kim G, Skinner MK. Hormonal Regulation and Differential Actions of the Helix-Loop-Helix Transcriptional Inhibitors of Differentiation (Id1, Id2, Id3, and Id4) in Sertoli Cells 1. *Endocrinology*. 2001;142(5):1727-36.
 30. Engel I, Murre C. The function of E-and Id proteins in lymphocyte development. *Nature reviews Immunology*. 2001;1(3):193.
 31. Ruzinova MB, Benezra R. Id proteins in development, cell cycle and cancer. *Trends in cell biology*. 2003;13(8):410-8.
 32. Asirvatham AJ, Schmidt MA, Chaudhary J. Non-redundant inhibitor of differentiation (Id) gene expression and function in human prostate epithelial cells. *The Prostate*. 2006;66(9):921-35.
 33. Ferrer-Vicens I, Riffo-Campos ÁL, Zaragoza R, García C, López-Rodas G, Viña JR, et al. In vivo genome-wide binding of Id2 to E2F4 target genes as part of a reversible program in mice liver. *Cellular and Molecular Life Sciences*. 2014;71(18):3583-97.
 34. Rodríguez JL, Sandoval J, Serviddio G, Sastre J, Morante M, Perrelli M-G, et al. Id2 leaves the chromatin of the E2F4–p130-controlled c-myc promoter during hepatocyte priming for liver regeneration. *Biochemical Journal*. 2006;398(3):431-7.
 35. Schefe JH, Lehmann KE, Buschmann IR, Unger T, Funke-Kaiser H. Quantitative real-time RT-PCR data analysis: current concepts and the novel “gene expression’s C T difference” formula. *Journal of*

- molecular medicine. 2006;84(11):901-10.
36. Mohammed H, D'Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell reports*. 2013;3(2):342-9.
 37. Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. ChIP-seq: using high-throughput sequencing to discover protein–DNA interactions. *Methods*. 2009;48(3):240-8.
 38. Serandour AA, Brown GD, Cohen JD, Carroll JS. Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. *Genome biology*. 2013;14(12):1.
 39. Huang Q, Yang L, Luo J, Guo L, Wang Z, Yang X, et al. SWATH enables precise label-free quantification on proteome scale. *Proteomics*. 2015;15(7):1215-23.
 40. Taberlay PC, Kelly TK, Liu C-C, You JS, De Carvalho DD, Miranda TB, et al. Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell*. 2011;147(6):1283-94.
 41. Andrews S. FastQC: A quality control tool for high throughput sequence data. Reference Source. 2010.
 42. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011;17(1):pp. 10-2.
 43. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357.
 44. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
 45. Shi W, Ma Z, Willers H, Akhtar K, Scott SP, Zhang J, et al. Disassembly of MDC1 foci is controlled by ubiquitin-proteasome-dependent degradation. *Journal of Biological Chemistry*. 2008;283(46):31608-16.
 46. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome biology*. 2008;9(9):R137.
 47. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome research*. 2002;12(6):996-1006.
 48. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007;23(10):1289-91.
 49. DeRose YS, Gligorich KM, Wang G, Georgelas A, Bowman P, Courdy SJ, et al. Patient-Derived Models of Human Breast Cancer: Protocols for In Vitro and In Vivo Applications in Tumor Biology and Translational Medicine. *Current protocols in pharmacology*. 2013;14.23. 1-14.23. 43.
 50. Budwit-Novotny DA, McCarty KS, Cox EB, Soper JT, Mutch DG, Creasman WT, et al. Immunohistochemical analyses of estrogen receptor in endometrial adenocarcinoma using a monoclonal antibody. *Cancer research*. 1986;46(10):5419-25.
 51. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*. 2012;9(7):676-82.

52. Law AM, Yin JX, Castillo L, Young AI, Piggin C, Rogers S, et al. Andy's Algorithms: new automated digital image analysis pipelines for FIJI. *Scientific reports*. 2017;7(1):15717.
53. Hartman AR, Kaldate RR, Sailer LM, Painter L, Grier CE, Endsley RR, et al. Prevalence of BRCA mutations in an unselected population of triple-negative breast cancer. *Cancer*. 2012;118(11):2787-95.
54. Montavon C, Gloss BS, Warton K, Barton CA, Statham AL, Scurry JP, et al. Prognostic and diagnostic significance of DNA methylation patterns in high grade serous ovarian cancer. *Gynecologic oncology*. 2012;124(3):582-8.
55. Ren Y, Cheung HW, von Maltzhan G, Agrawal A, Cowley GS, Weir BA, et al. Targeted tumor-penetrating siRNA nanocomplexes for credentialing the ovarian cancer oncogene ID4. *Science translational medicine*. 2012;4(147):147ra12-ra12.
56. Pujana MA, Han J-DJ, Starita LM, Stevens KN, Tewari M, Ahn JS, et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nature genetics*. 2007;39(11):1338-49.
57. Townsend K, Mason H, Blackford AN, Miller ES, Chapman JR, Sedgwick GG, et al. Mediator of DNA damage checkpoint 1 (MDC1) regulates mitotic progression. *Journal of Biological Chemistry*. 2009;284(49):33939-48.
58. Wilson KA, Colavito SA, Schulz V, Wakefield PH, Sessa W, Tuck D, et al. NFB1/MDC1 regulates Cav1 and Cav2 independently of DNA damage and p53. *Molecular Cancer Research*. 2011;9(6):766-81.
59. Stewart GS, Wang B, Bignell CR, Taylor AMR, Elledge SJ. MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature*. 2003;421(6926):961-6.
60. Wilhelmsen K, Litjens SH, Kuikman I, Tshimbalanga N, Janssen H, van den Bout I, et al. Nesprin-3, a novel outer nuclear membrane protein, associates with the cytoskeletal linker protein plectin. *The Journal of cell biology*. 2005;171(5):799-810.
61. Takenaka K, Nakagawa H, Miyamoto S, Hiroaki M. The pre-mRNA-splicing factor SF3a66 functions as a microtubule-binding and-bundling protein. *Biochemical Journal*. 2004;382(1):223-30.
62. Fontemaggi G, Dell'Orso S, Trisciuglio D, Shay T, Melucci E, Fazi F, et al. The execution of the transcriptional axis mutant p53, E2F1 and ID4 promotes tumor neo-angiogenesis. *Nature structural & molecular biology*. 2009;16(10):1086-93.
63. Pruszko M, Milano E, Forcato M, Donzelli S, Ganci F, Di Agostino S, et al. The mutant p53-ID4 complex controls VEGFA isoforms by recruiting lncRNA MALAT1. *EMBO reports*. 2017;18(8):1331-51.
64. Liebler DC, Zimmerman LJ. Targeted quantitation of proteins by mass spectrometry. *Biochemistry*. 2013;52(22):3797-806.
65. Bekker-Jensen S, Lukas C, Melander F, Bartek J, Lukas J. Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFB1. *The Journal of cell biology*. 2005;170(2):201-11.
66. Konishi H, Mohseni M, Tamaki A, Garay JP, Croessmann S, Karnan S, et al. Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells.

- Proceedings of the National Academy of Sciences of the United States of America. 2011;108(43):17773-8.
67. Seo J, Kim SC, Lee H-S, Kim JK, Shon HJ, Salleh NLM, et al. Genome-wide profiles of H2AX and γ -H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells. *Nucleic acids research*. 2012:gks287.
 68. Natale F, Rapp A, Yu W, Durante M, Taucher-Scholz G, Cardoso MC. Genome-wide multi-parametric analysis of H2AX or γ H2AX distributions during ionizing radiation-induced DNA damage response. *Epigenetics & Chromatin*. 2013;6(1):1.
 69. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534(7605):47-54.
 70. Vafae F, Colvin EK, Mok SC, Howell VM, Samimi G. Functional prediction of long non-coding RNAs in ovarian cancer-associated fibroblasts indicate a potential role in metastasis. *Scientific reports*. 2017;7(1):10374.
 71. Gardini A, Baillat D, Cesaroni M, Shiekhattar R. Genome-wide analysis reveals a role for BRCA1 and PALB2 in transcriptional co-activation. *The EMBO journal*. 2014:e201385567.
 72. Donzelli S, Milano E, Pruszko M, Sacconi A, Masciarelli S, Iosue I, et al. Expression of ID4 protein in breast cancer cells induces reprogramming of tumour-associated macrophages. *Breast Cancer Research*. 2018;20(1):59.
 73. Lou Z, Minter-Dykhouse K, Franco S, Gostissa M, Rivera MA, Celeste A, et al. MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Molecular cell*. 2006;21(2):187-200.
 74. Lee J-H, Park S-J, Hariharasudhan G, Kim M-J, Jung SM, Jeong S-Y, et al. ID3 regulates the MDC1-mediated DNA damage response in order to maintain genome stability. *Nature communications*. 2017;8(1):903.
 75. Stucki M, Clapperton JA, Mohammad D, Yaffe MB, Smerdon SJ, Jackson SP. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell*. 2005;123(7):1213-26.
 76. Furuta S, Jiang X, Gu B, Cheng E, Chen P-L, Lee W-H. Depletion of BRCA1 impairs differentiation but enhances proliferation of mammary epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(26):9176-81.
 77. Martins FC, De S, Almendro V, Gönen M, Park SY, Blum JL, et al. Evolutionary pathways in BRCA1-associated breast tumors. *Cancer Discov*. 2012;2(6):503-11.
 78. Konstantinopoulos PA, Spentzos D, Karlan BY, Taniguchi T, Fountzilas E, Francoeur N, et al. Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. *Journal of Clinical Oncology*. 2010;28(22):3555-61.
 79. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nature biotechnology*. 2011;29(1):24-6.

80. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*. 2013;14(2):178-92.
81. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. 2007.
82. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer cell*. 2006;10(6):515-27.
83. Anglesio MS, Wiegand KC, Melnyk N, Chow C, Salamanca C, Prentice LM, et al. Type-specific cell line models for type-specific ovarian cancer research. *PloS one*. 2013;8(9):e72162.

Supplementary Information

Figure S1. Analysis of ID4 protein expression across a panel of breast and ovarian cell lines

(A) Western blotting analysis of ID4 protein expression across a panel of breast (top) and ovarian (bottom) normal and cancer cell lines. A rabbit monoclonal antibody specific to ID4 (Biocheck, BCH-9/82-12) was used for detection (11). Two isoforms of ID4 are detectable across the panel. β -Actin is shown as a loading control. Modifications to images indicated with vertical black line. Cell lines selected for further analysis are indicated with a black square. Breast cancer subtypes and ovarian cancer histotypes indicated (82, 83). (B) Western blot validation of RIME protocol. MDA-MB-468 cells were processed using the RIME protocol. Antibodies targeting ID4 (pooled polyclonal antibodies) and IgG (rabbit species matched control) control were used for immunoprecipitation. Western blot analysis for ID4 protein expression, using an independent ID4 monoclonal antibody, following immunoprecipitation.

Figure S2. ID4 ChIP-seq analysis of HCC70 cell line identifies reproducible ID4-chromatin binding sites

(A) Table summarising three biological replicates of ID4 ChIP-seq analysis in HCC70 cell line. (B) 10 000bp resolution image of ID4 ChIP-seq technical replicate #1 binding to MALAT1 gene. Red sequencing reads are aligned to the positive strand (5' - 3'), and blue to the negative strand of DNA (3' - 5'). ID4 binding to (C) left to right: GBA, FAIM, MIRLET7BHG, NEAT1 and ZFP36L1. The chromosomal location, size of the gene and Refseq, human reference genome, are displayed at the top of the image. Reads have been aligned to the human reference genome Hg19 and peaks called using MACs peak calling algorithm (v2.0.9) (38). Images contain ChIP-seq coverage data and the peaks called for each ID4 technical replicate and the consensus peaks called for all three ID4 ChIP-seq biological replicate for selected gene regions. ID4 binding is shown in comparison to IgG and Input data for the same region. Data visualised using IGV (79, 80). Transcription Start Site (TSS) indicated with black arrow.

Figure S3. ID4 ChIP-exonuclease sequencing analysis of HCC70 and HCC1954 cell lines reproduces ChIP-seq analysis

(A) Table summarising ChIP-exonuclease sequencing analysis of ID4 and IgG binding events in HCC70 and HCC1954 breast cancer cell lines. ID4 binding normalised as for ChIP-seq analysis. ChIP-exo analysis of the HCC70 and HCC1954 breast cancer cell lines showing ID4 binding to NEAT1, MALAT1 and GBA (B), ZFP36L1 and ELF3 (C), and KDM4C and ERFFI1 (D). Reads have been aligned to the human reference genome Hg19 and peaks called using MACs peak calling algorithm (v2.0.9) (38). Images contain ChIP-exo coverage data and the peaks called for each ID4 technical replicate and the consensus peaks called for both ID4 ChIP-exo technical replicate for selected gene regions. ID4 binding is shown in comparison to IgG and input data for the same region. Data visualised using IGV (79, 80). The chromosomal location and size of the gene are displayed at the top of the image. Below this, Refseq, human reference genome, displays the gene corresponding to particular genomic loci. Transcription Start Site (TSS) indicated with black arrow.

Figure S4. Validation of ID4 binding to specific loci in HCC70, MDA-MB-468 and HCC1954 cell lines.

(A) Schematic of primer binding across ELF3 gene region. Primers 1-5 are scattered along the length of the ELF3 gene ID4. ChIP-qPCR analysis in (B) HCC70 (n=2-8), (C) MDA-MB-468 (n=2) and (D) HCC1954 (n=2) cells. Multiple primers were designed to tile across the large ID4 binding sites. ID4 binding normalised to input DNA and to a region not bound by ID4 (negative region) and represented as fold-change over IgG control. 1-Way-ANOVA of ID4 binding across all primers compared to negative region: HCC70 p=0.0001, MDA-MB-468 p=0.549, HCC1954 p=0.519.

Figure S5. ID4 knockdown using SMARTChoice inducible lentiviral knockdown system.

Representative western blot showing ID4 and β -Actin protein expression in wild-type HCC70 cell line and HCC70 cells stably transfected with SMARTChoice vectors containing doxycycline-inducible shRNAs targeting ID4 (shID4 #1, #2 and #3) or non-targeting control region (shNTC). Cells treated with and without DOX for 72 h. (B) Western blotting densitometry quantification of ID4 protein expression across six biological replicates of ID4 knockdown following treatment with and without DOX for 72 h. ID4 expression normalised to β -ACTIN and no DOX control. Error bars measure standard error, n=4-6. (C) qRT-PCR analysis of ID4 RNA expression matched to cells in (B). Data normalised to B2M housekeeping gene and no DOX control. Error bars measure standard error, n=4-6. Student's t-test compares shNTC with the other modified cell lines. ** p<0.01, **** p<0.001.

Figure S6. Validation of specificity of ID4-chromatin binding using the SMARTChoice ID4 knockdown model.

(A) Western blot showing ID4 and β -Actin protein expression in HCC70 cells stably transfected with shID4 #2 SMARTChoice vector treated with and without doxycycline for 72 h. (B) ID4 ChIP-qPCR analysis of cells from (A). ID4 binding normalised to input DNA, to a region not bound by ID4 (negative region) and to the IgG control. Data represented as fold-change over untreated cells. (C) replicate of ID4 ChIP-qPCR analysis of HCC70 cells stably transfected with shID4 #2 SMARTChoice vector treated with and without doxycycline for 72 h.

Figure S7. Proximity ligation assay analysis identifies constitutive and DNA damage-inducible binding of ID4 to DNA damage proteins.

PLA analysis of ID4 association with FLAG, MDC1, BRCA1 and γ H2AX in (A) HCC70 and (B) OVKATE cell line. Analysis is representative of approximately 150-200 cells of pooled from two independent experiments. 1-way Anova $p=0.098$. B, right; example image of ID4 and MDC1 PLA. Quantification of interactions presented on right (number of dots/ cell). 10 μ m scale.

Figure S8. Ionising radiation increases DNA damage foci formation and localisation of MDC1 to chromatin

Induction of DNA damage using ionising radiation measured using immunofluorescence analysis in MDA-MB-468 cells for γ H2AX (top) and MDC1 (bottom) foci formation (5Gy with 5 h recovery time prior to fixation). Example images on left, quantification of foci on right. Four to five images were taken for each condition; 50-100 cells. 20 μ m scale (B) Second replicate of HCC70 ID4 ChIP-qPCR. (C) ChIP-qPCR analysis of MDC1 positive control binding following ionising radiation DNA damage induction. Data normalised as for Fig. 3C, with data shown as fold-change compared to no-IR, $n=1$.

Figure S9. ID4 is overexpressed and amplified at a higher frequency in BRCA1-mutant BLBC

(A) Graph of ID4 protein expression measured by H-score. Contains 97 BRCA1-mutant BLBC patients. (B) ID4 protein expression measured by immunohistochemistry H-score compared with corresponding *ID4:CEP6* FISH ratio. Assuming non-Gaussian distribution, H-score and FISH correlated with a value of $r = 0.265$ and Spearman correlation p value of <0.00881 .

Table S1. ChIP-seq and ChIP-exo MACS peaks

Table S2. ID4 RNA-seq differentially expressed genes

Table S3. Putative ID4 interaction proteins identified by RIME

Table S4. ChIP-qPCR primers

Figures

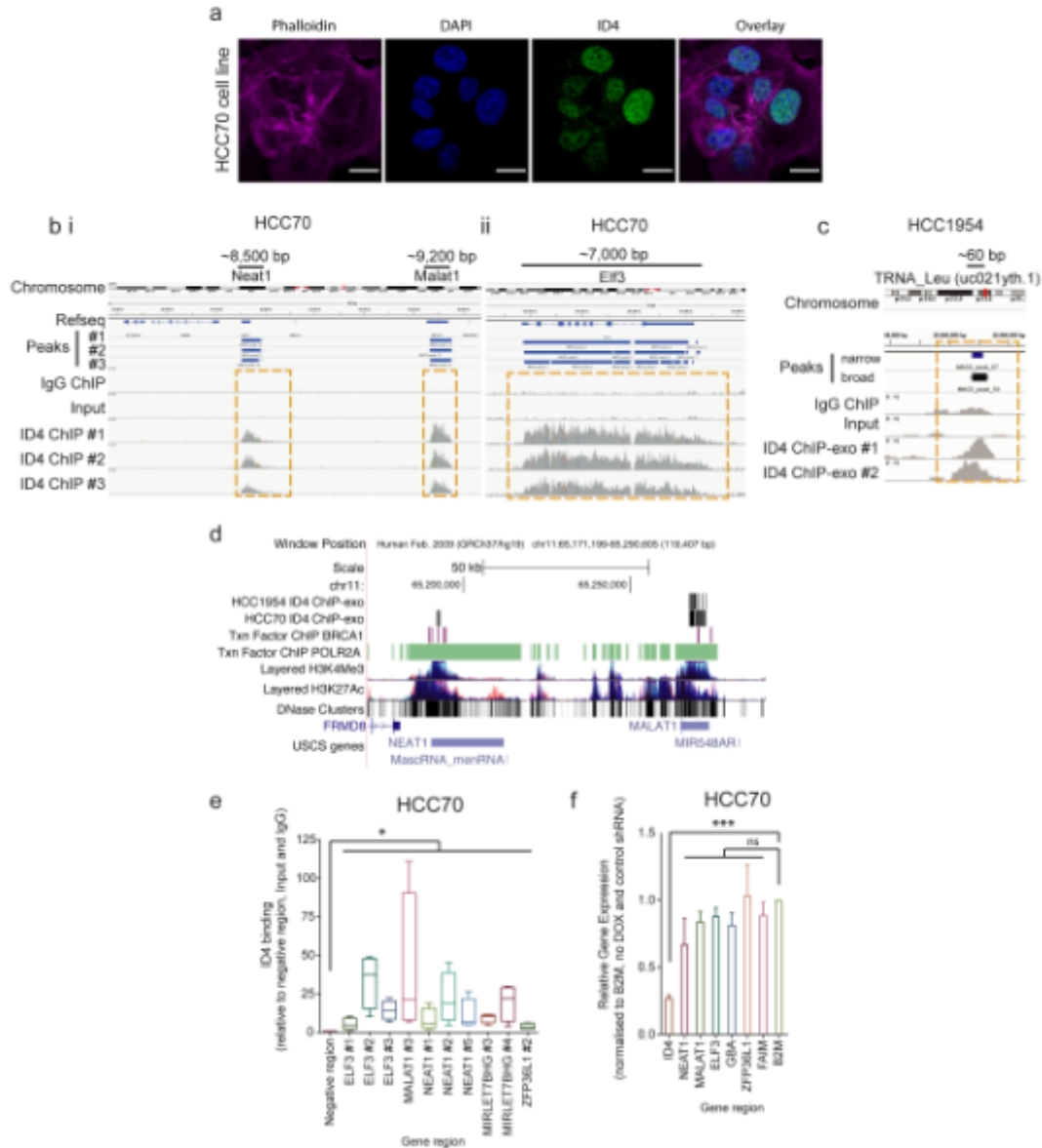
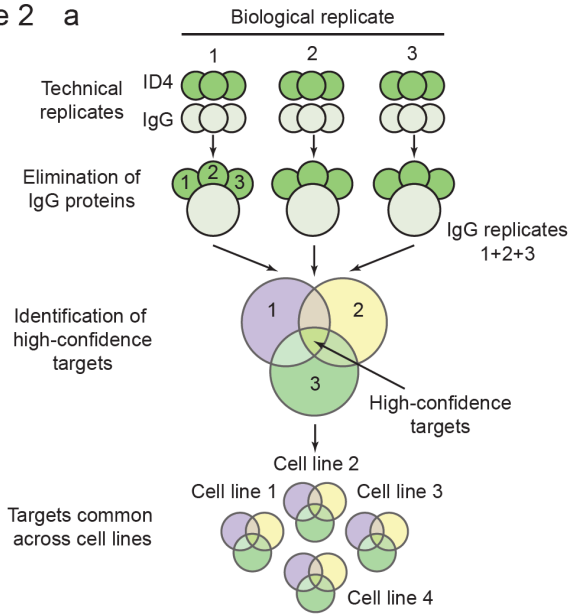


Figure 1

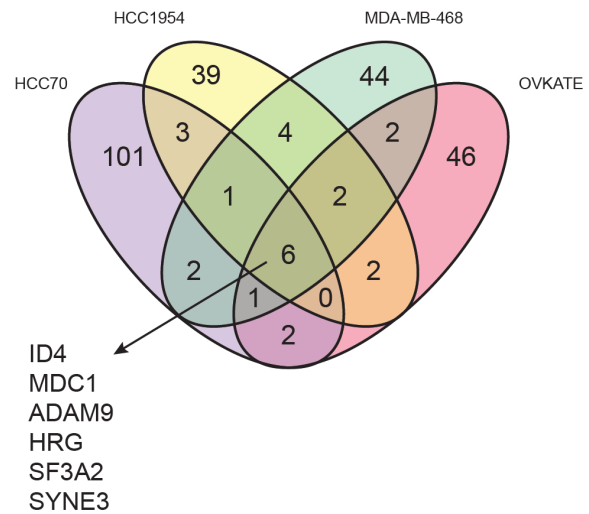
ID4 binds to chromatin in BLBC cell lines and PDX models and does not regulate transcription. (A) Immunofluorescence analysis of ID4 protein expression in the HCC70 BLBC cell line, blue; nuclear marker DAPI, magenta; cytoskeletal marker phalloidin, green; ID4. 20 μ m scale. (B) Three technical replicates of ID4 ChIP-seq analysis in HCC70 cell line. IgG ChIP-seq and Input controls are shown for comparison. ID4 binding to i) the genomic region encoding the long non-coding RNAs NEAT1 and MALAT1 and ii) the protein coding gene ELF3. (C) ID4 binding to tRNA_Leu (uc021yth.1) in HCC1954 cell line measured by ChIP-exo. Identification of ID4 binding enrichment identified by MACS peak-calling algorithm.

Chromosome location, transcription start site (TSS) and Refseq information tracks displayed. Reads have been aligned to the human reference genome Hg19 and peaks called using MACs peak calling algorithm (v2.0.9) (38). Images contain ChIP-seq coverage data and the peaks called for each ID4 technical replicate and the consensus peaks called for all three ID4 ChIP-seq technical replicate for selected gene regions. ID4 binding is shown in comparison to IgG and Input data for the same region. Data visualised using IGV (79, 80). Transcription Start Site (TSS) indicated with black arrow. (D) Alignment of ID4 ChIP-exo peaks with DNase hypersensitivity clusters, Transcription Factor ChIP and histone marks H3K27Ac and H3K4Me3 ChIP-seq at NEAT1 and MALAT1, UCSC Genome Browser. (E) Box and whisker plot of ID4 ChIP-qPCR analysis in HCC70 cells. Multiple primers were designed to tile across the large ID4 binding sites. ID4 binding normalised to input DNA and to a region not bound by ID4 (negative region) and represented as fold-change over IgG control. Ratio paired t-test, * $p < 0.05$. Whiskers indicate min to max. $n = 4$ (F) qRT PCR analysis of mRNA transcript expression of ID4-bound genomic regions following depletion of ID4 in HCC70 cells using a lentiviral, doxycycline-inducible short hairpin RNA #2 (SMARTChoice). Data normalised to B2M housekeeping gene and to HCC70 cells treated with vehicle control. One-way ANOVA, multiple comparisons test of control primer B2M with each primer set. *** $p < 0.001$, ns = non-significant. Error bars represent standard error. $n = 4$.

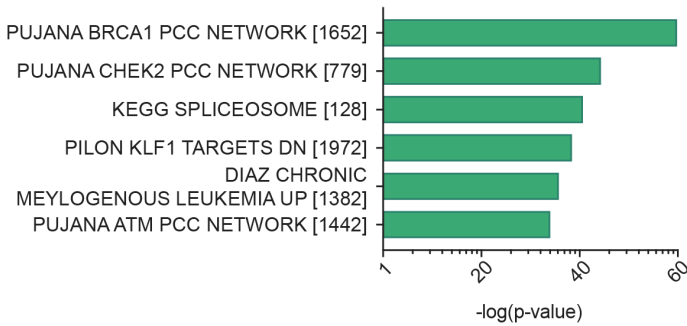
Figure 2 a



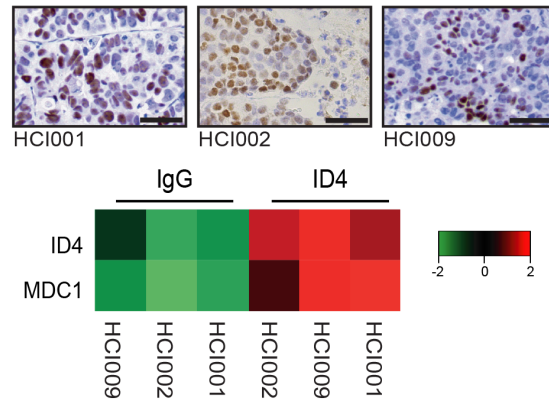
b



c



d



e

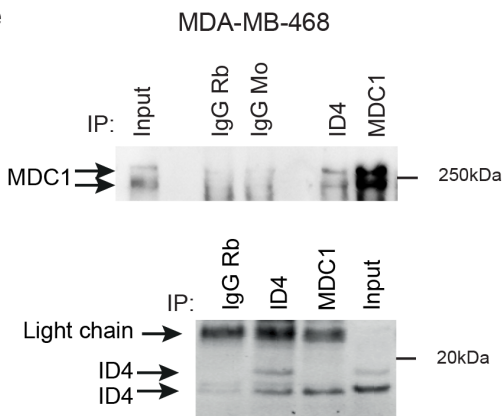


Figure 2

ID4 binds to MDC1 and interacts with the BRCA1 network. (A) Schematic of ID4 and IgG RIME data analysis; ID4 and IgG immunoprecipitations were conducted in technical duplicate or triplicate, and in biological triplicate. All proteins identified in the IgG controls (for each biological replicate) were removed from each of the ID4 IPs as non-specific, background proteins. The remaining proteins were compared across technical replicates to generate a list of medium-confidence proteins that were robustly identified

in >1 ID4 RIME technical replicate. The biological replicates were then compared to generate a list of targets present in >1 biological replicate in an individual cell line (i.e >6 technical replicates conducted over three biological replicates). This resulted in the identification of 22 (HCC70), 21 (HCC1954), 21 (MDA-MB-468) and 30 (OVKATE) proteins. Targets from the different cell lines were compared, identifying a list of ID4 interactors present across multiple cell lines. (B) Venn diagram showing comparison of the high-confidence targets identified in all four cell lines; HCC70, HCC1954, MDA-MB-468 and OVKATE. Common targets across the four cell lines are indicated. Overlap generated using Venny (81). (C) Top six highest enriched gene sets identified through Gene Set Enrichment Analysis (GSEA) of the ID4 proteome (1,106 proteins) identified in (B) (all proteins identified in ID4 RIME and not in IgG RIME), compared to C2 (curated gene sets), C4 (computational gene sets), C6 (oncogenic signatures), C7 (immunologic signatures) and H (hallmark gene sets) gene sets. Proteins identified in >1 technical replicate of ID4 RIME (and not in IgG RIME) were considered in GSEA analysis. D, top; Immunohistochemistry analysis of ID4 protein expression across HCI001, HCI002 and HCI009 triple negative PDX models, 100 μ m scale, bottom; SWATH proteomic analysis of ID4 and IgG RIME conducted on HCI001, HCI002, HCI009 PDX models. Heatmap showing quantification of ID4 and MDC1 abundance, both significantly differentially expressed proteins in ID4 RIME compared with control IgG RIME (p-value < 0.05 and fold change > 2). RIME and SWATH were conducted on the same sample; one biological replicate per tumour. 100 μ m scale. (E) Immunoprecipitation was conducted on MDA-MB-468 cells prepared using the RIME protocol. Top: Input control and IgG (mouse and rabbit) compared to IP with anti-ID4 antibodies and MDC1. Western blotting is shown using an independent monoclonal MDC1 antibody. Bottom: Input control and IgG rabbit compared to IP with anti-ID4 antibodies and MDC1. Western blotting is shown using an independent monoclonal ID4 antibody.

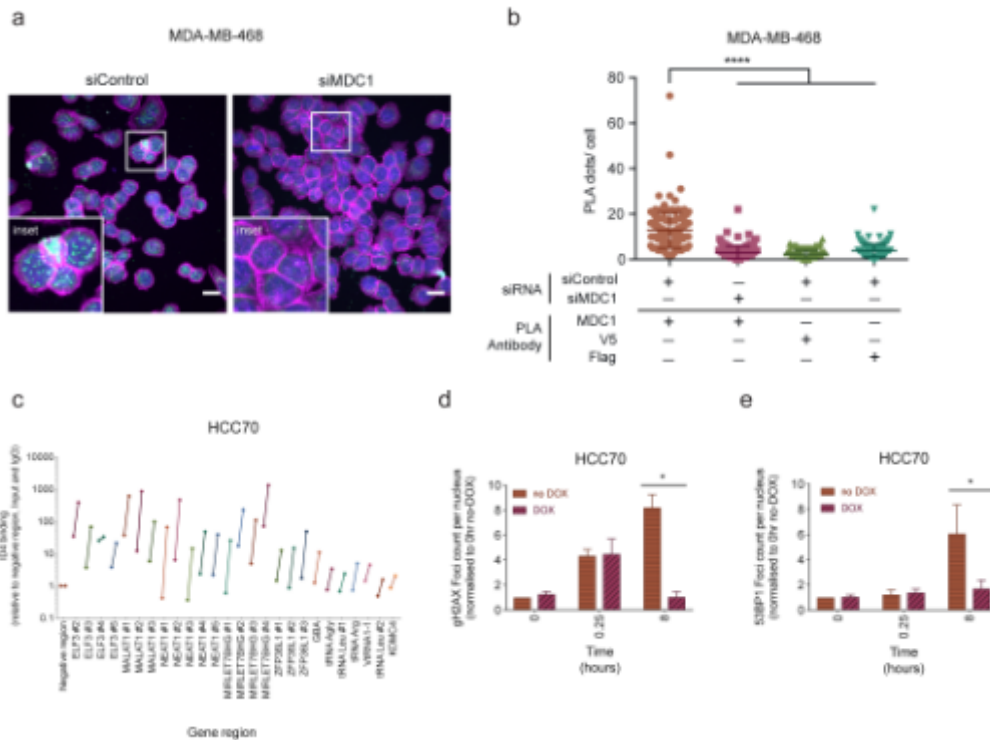


Figure 3

ID4 interacts with DNA damage proteins at fragile chromatin sites in a DNA damage-dependent manner. (A) Representative proximity ligation assay (PLA) confocal images of ID4 and MDC1 interactions in MDA-MB-468 cells. siRNAs targeting MDC1 (siMDC1) and a scrambled control (siControl) used to show specificity of the assay. PLA foci (green), DAPI (blue) and phalloidin (magenta). 20 μ m scale. (B) Graph showing quantification of PLA interactions between ID4 and MDC1 in cells treated with siControl and siMDC1. Data shown in comparison to the interaction between ID4:V5 and ID4:FLAG treated with siControl. 50-100 cells captured per condition. Quantification of interactions (number of dots/ cell). One-way ANOVA, multiple comparisons test. **** $p < 0.0001$. Error bars represent standard deviation. (C) ChIP-qPCR analysis of ID4 binding in untreated HCC70 cells compared to cells treated with ionising radiation (5Gy with 5 h recovery time prior to fixation). ID4 binding normalised to negative control region, input DNA and IgG control. Data for each gene region is shown for both untreated (left) and ionising radiation-treated (right) cells, connected by a line. Representative of two independent experiments. Quantification of a time course of D: γ H2AX and E: 53BP1 DNA damage foci formation following ionising radiation. ID4 was depleted from cells using the SMARTChoice inducible shRNA system following treatment with doxycycline for 72 h prior to treatment with 5 Gy ionising radiation at time 0 and allowed to recover for 0.25 and 8 hours prior to analysis. Four to five images were taken for each condition; 100-200 cells in total. Number of foci per cell nucleus was calculated using FIJI by ImageJ (Schindelin et al., 2012) and samples were then collated and analysed using the Pandas package in Python 3.5. Data is normalised to 0 h, no DOX, no IR time point. $n = 3-5$, Student's t-test, * $p < 0.05$. Error bars represent standard error.

Figure 4

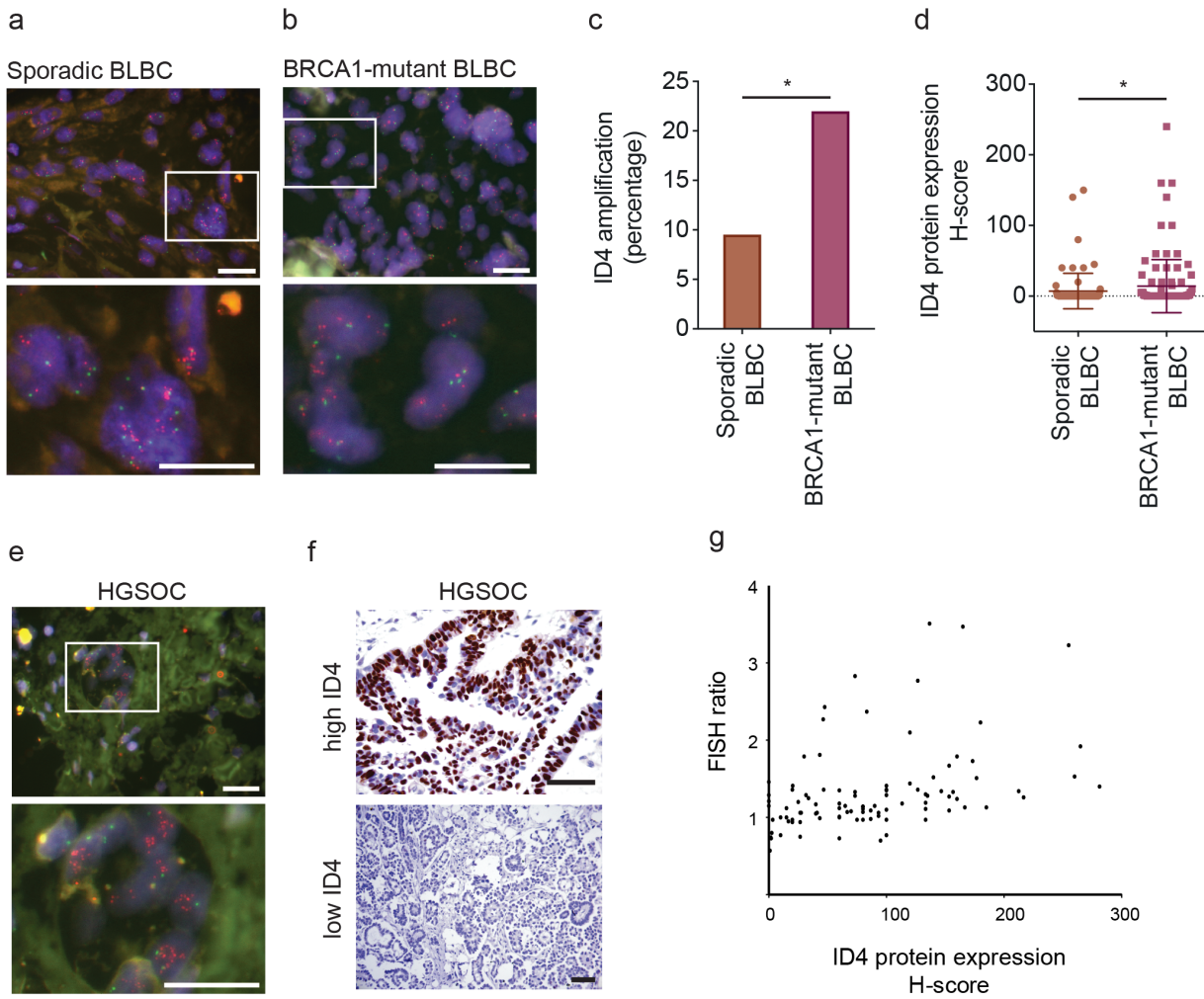


Figure 4

ID4 is selectively amplified in cancers arising in germline BRCA1 mutation carriers (A) ID4 amplification in sporadic and familial BLBC. Example images of fluorescent in situ hybridisation (FISH) analysis of ID4 (red) and CEP6 centromeric control (green) copy number aberrations in a cohort of 42 sporadic BLBCs. 10 µm scale. (B) Example images of FISH analysis of ID4 (red) and CEP6 centromeric control (green) copy number aberrations in a cohort of 97 BRCA1-mutant BLBCs. 10 µm scale. (C) Percentage of ID4 amplified

cases in sporadic and BRCA1-mutant BLBCs presented in (A) and (B). * $p < 0.05$ (Two sample t-test). (D) Immunohistochemical analysis of ID4 protein expression in the sporadic and BRCA1-mutant BLBCs presented in (A) and (B). ID4 protein quantified using the H-score method. H-score = intensity of staining (on a scale from 0-3) x percentage of cells positive. $p = 0.007$ (Mann-Whitney two-tailed test). (E) Example images of FISH analysis of ID4 (red) and CEP6 centromeric control (green) copy number aberrations in a cohort of 97 high-grade serous ovarian cancers. 10 μm scale. (F) Immunohistochemistry staining for ID4 in HGSOC. Example images of ID4 high expressing (top) and ID4 low expressing (bottom) HGSOC. (G) ID4 H-score compared with corresponding ID4:CEP6 FISH ratio. Assuming non-Gaussian distribution, H-score and FISH correlated with a value of $r = 0.457$ and Spearman correlation $p = < 0.0001$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS2.xlsx](#)
- [Supplementaryfigures.pdf](#)
- [TableS3.xlsx](#)
- [TableS1.xlsx](#)
- [TableS4.xlsx](#)