Cellular and Molecular Life Sciences (2025) 82:196
https://doi.org/10.1007/s00018-025-05725-6

ORIGINAL ARTICLE

Human-specific genomic evolution of a regulatory network enables fine-tuning of N-cadherin gene expression

Éva Erdmann[1] · Savera Agolli[1] · Simon Fix[1] · Félicie Cottard[2] · Christine Keyser[3] · Vincent Zvenigorosky[3] · Angéla Gonzalez[4] · Zakary Haili[1] · Bruno Kieffer[1] · Jocelyn Céraline[1,5,6,7]

Received: 15 January 2025 / Revised: 4 April 2025 / Accepted: 22 April 2025

Abstract

Androgen receptor (AR), a member of the nuclear receptor superfamily controls prostate epithelial cell plasticity by repressing a panel of genes involved in epithelial-mesenchymal transition (EMT), including the human *CDH2* gene encoding N-cadherin. At the opposite, pathological AR variants such as AR-V7 associated with prostate tumor progression upregulate those EMT genes. Here, focusing on the human *CDH2* gene, we show that this duality between AR and AR-V7 relies on a potential human accelerated region present in the intron 1. This fastest-evolving region of the human genome is actually a variable number tandem repeat (VNTR) comprising 24 repetitions of a DNA sequence that englobes binding sites for steroid hormone receptors, recombination signal binding protein for immunoglobulin kappa j region (RBPJ) an effector of the Notch pathway, and zinc finger e-box binding homeobox 1 (ZEB1). Genomic DNA sequencing, multiple sequence alignment, data mining, as well as protein-DNA interaction and gene expression analyses indicate that this VNTR constitutes a potential transcriptional hub for different transcription factors to control human *CDH2* expression. Also, our data suggest that prostate tumor cells may unlock an up to now unknown molecular mechanism associated with a fine-tuned control of human *CDH2* gene expression.

✉ Jocelyn Céraline
ceraline@unistra.fr

1 CNRS UMR 7104, INSERM U1258, IGBMC, Université de Strasbourg, Illkirch 67404, France

2 CNRS UMR 7242, ESBS, Université de Strasbourg, Illkirch 67404, France

3 CNRS UMR 8045 BABEL, Université de Paris, Paris 75012, France

4 Strasbourg Institute of Legal Medicine, Strasbourg 67085, France

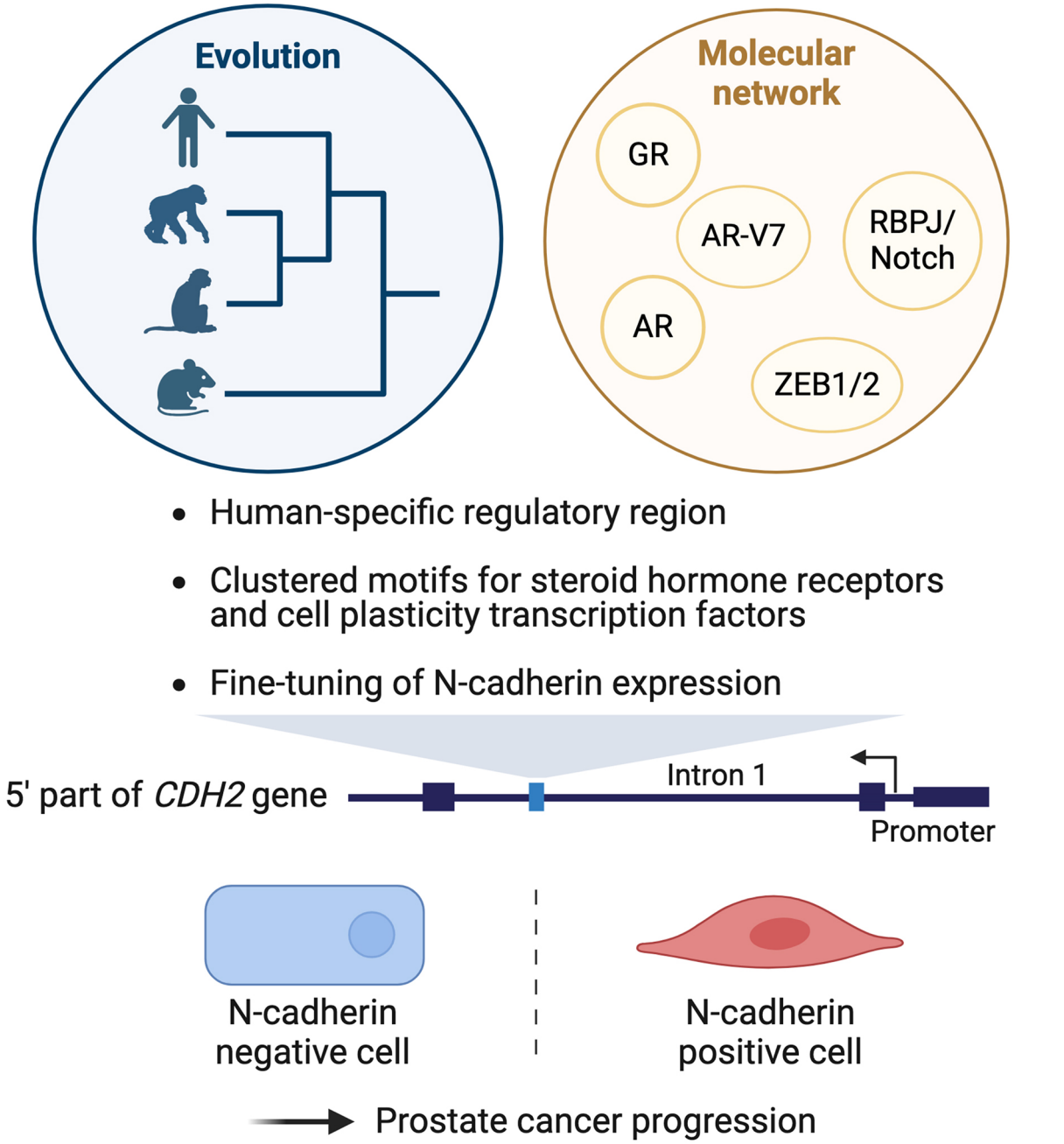
5 Hôpitaux Universitaires de Strasbourg, Strasbourg 67091, France

6 Fédération de Médecine Translationnelle de Strasbourg, FMTS, Université de Strasbourg, Strasbourg 67085, France

7 CNRS UMR 7104, INSERM U1258, IGBMC, 1, rue Laurent Fries, Illkirch 67404, France

Published online: 09 May 2025

Graphical Abstract



Keywords Androgen receptor · AR-V7 · CDH2 · N-cadherin · Cell plasticity · Prostate cancer · Transcriptional regulation

Introduction

Epithelial–mesenchymal plasticity (EMP) encompasses epithelial-to-mesenchymal transition (EMT) and its reverse counterpart mesenchymal-to-epithelial transition (MET), two complex and dynamic cellular programs that naturally occur during normal development, tissue repair and wound healing [1]. EMP, when fully completed, reversibly converts immobilized and polarized epithelial cells into mobile and

extracellular matrix (ECM) component-secreted mesenchymal cells. Mostly, a panel of intermediate cell states between the two extreme epithelial and mesenchymal states is generated according to the intracellular integration of the different signaling pathways that are supposed to fine-tune EMP processes [2]. Indeed, different extra-cellular ligands such as the transforming growth factor-β (TGFβ), growth factors, hormones, Hedgehog and Wnt act through their respective receptors to control EMP [3]. Also, ECM-cell interactions through the matrix-associated cluster of differentiation 44 (CD44) are connected to EMP [4]. All these pathways finally result in an interplay between transcription factors such as SNAI1/2, ZEB1/2, TWIST1, GRHL2, OVOL1/2, and PRRX1, post-transcriptional regulators (miRNAs), and epigenetic factors to control the expression of epithelial and mesenchymal specific genes [3, 5]. These pathways can be deregulated in tumors pushing EMP to generate and maintain cancer stem cells, to favor cell migration and metastasis and, ultimately, to sustain tumor resistance [1, 6, 7].

Together with others, we recently pinpointed a negative role of androgen receptor (AR) in EMP in prostate cancer (PCa) [8–13]. AR (NR3C4) is a ligand-dependent transcription factor belonging to the nuclear receptor superfamily [14]. AR regulates a wide range of physiological processes including reproduction, development and homeostasis, but is also involved in PCa cell proliferation and survival [15]. Upon ligand binding, AR displays genomic actions by binding to androgen response elements (AREs) located at promoter, enhancer and super-enhancer regions of target genes. Liganded-AR acquires the capacity to recruit specific pioneer factors and coactivators such as Forkhead Box A1 (FOXA1) and Nuclear Receptor Coactivator 1 (NCOA1) to form transcriptional activation complexes. Liganded-AR can also recruit corepressors such as Nuclear Receptor Repressor 1 (NCOR1) to inhibit gene expression [16].

AR genomic activities being linked to PCa cell proliferation and survival, AR inhibition through AR signaling inhibitors (ARSI) has emerged as the main therapeutic option for advanced PCa. Unfortunately, despite iterative ARSI innovation, advanced PCa systematically progresses to a lethal disease. In particular, many tumors show signs of lineage plasticity under the pressure of effective ARSI resulting in aggressive tumor growth and metastatic spread [8, 17, 18]. Recently, cross-analyses of different transcriptomic datasets show that, besides its well-known role in transcriptional activation, AR can repress certain genes involved in cellular plasticity, and that this property could be lost during ARSI therapy [13]. In addition, AR-V7, an AR splice variant not targeted by ARSI and frequently detected in metastatic lesions and ARSI-resistant PCa is shown to positively regulate genes involved in EMT [10, 11, 13]. So, these data highlight a troubling clinical situation during ARSI therapies where the protective role of AR as a repressor of EMP would be impaired on one side while the expression of EMT markers would be upregulated by pathological AR variants on the other side.

To further our understanding of the molecular mechanisms underpinning the pathological role of AR in the control of cell plasticity in PCa, we focus on the regulation of classical type 1 neural cadherin (CDH2, N-cadherin) expression. The human *CDH2* gene is localized in on chromosome 18q12.1 and is coding for a calcium-dependent transmembrane glycoprotein that mostly establishes homophilic cell adhesion by forming strong zipper-like punctate adhesions between apposing cell membranes. N-cadherin expression is retrieved from embryo to adult and is known to play a predominant role in cell migration during organogenesis [19]. As modulation of cell adhesion and cell migration properties is fundamental for a proper embryonic development, N-cadherin expression level needs to be spatiotemporally fine-tuned. Indeed, several signaling pathways, such as PDGFR and TGFβ, and transcription factors, including Twist, Snail, and Slug have been reported to control *CDH2* expression [19]. N-cadherin total levels can also be adjusted through FRMP-induced *CDH2* mRNAs degradation and Rab18-induced lysosomal degradation of the protein. Reduction of cell surface levels of N-cadherin can further be regulated through transport regulation to the Golgi and the plasma membrane, but also through internalization of surface CDH2 by a clathrin-dependent and independent endocytosis mechanism [20].

The ability of full-length AR to recognize AREs present in intron 1 of the human *CDH2* gene has been reported [21]. We demonstrated that AR-V7 also binds to these AREs, suggesting that different *CDH2* expression levels result from AR-mediated assembly of distinct regulatory complexes rather than difference in DNA recognition properties in this genomic region [10, 11]. Here, we describe the intronic region in human *CDH2* gene recognized by AR as a human-specific variable number tandem repeat (VNTR) containing several transcription factor binding sites. Different molecular approaches as well as in silico analyses were performed to decipher the different signaling pathways embedded within this intronic region, including AR, glucocorticoid receptor (GR), Notch and ZEB1, whose intricated actions result into fine-tuned expression of human *CDH2* gene. Further, our results suggest that, during progression of PCa, AR, GR, Notch and ZEB1 signaling may completely disorganize the fine-tuned regulation of N-cadherin, allowing consequently cell plasticity and PCa progression to a more aggressive disease.

Materials and methods

Cell culture

LNCaP cells (ATCC, #CRL-1740, RRID: CVCL_W668) were maintained in RMPI-1640 complete medium supplemented with 10% (v/v) of fetal bovine serum (FBS) (BD biosciences, Le Pont de Claix, France), 10 mM HEPES, 1 mM sodium pyruvate (Invitrogen, Thermo Fisher Scientific France, Illkirch-Graffenstaden, France), 100 U/mL penicillin, 100 µg/mL streptomycin (Sigma-Aldrich, Saint-Quentin-Fallavier, France). HEK293T cells (ATCC, #CRL-3316, RRID: CVCL_0063) were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% (v/v) FBS, 100 U/mL penicillin and 100 µg/mL streptomycin (Sigma-Aldrich, Saint-Quentin-Fallavier, France). These cell lines were authenticated by Short Tandem Repeat (STR) DNA profiling in the past three years. Stable transduced LNCaP cells were used to obtain doxycycline-inducible expression of enhanced green fluorescent protein (eGFP)-GR (eGFP-GR) as previously described [11, 13]. Inducible LNCaP-eGFP-GR cells were maintained in RMPI-1640 complete medium supplemented with 10% (v/v) of tetracycline-system-approved FBS (BD Biosciences, Le Pont de Claix, France), 10 mM HEPES, 1 mM sodium pyruvate (Invitrogen, Thermo Fisher Scientific France, Illkirch-Graffenstaden, France), 100 U/mL penicillin, 100 µg/mL streptomycin (Sigma-Aldrich, Saint-Quentin-Fallavier, France), 200 µg/ml geneticin, and 400 ng/ml puromycin (Life Technologies, Thermo Fisher Scientific France, Illkirch-Graffenstaden, France).

Plasmids

pEGFP-AR and pEGFP-ARV7 plasmids were constructed as previously described [22]. The pEGFP-GR plasmid was purchased from Addgene (Addgene, #47504, Watertown, MA, USA). Expression plasmids without the eGFP tag, named pE-AR, pE-GR and pE-AR-V7, were created. Briefly, pEGFP-AR/GR/AR-V7 vectors were digested with restriction enzymes *Nhe* I and *Xho* I to remove the EGFP coding sequence. A fill-in reaction using DNA Polymerase I (New England Biolabs France Genopole, Evry, France) was performed to form blunt ends, which were subsequently ligated using T4 DNA ligase (New England Biolabs France Genopole, Evry, France). The same strategy was applied to construct an empty (control) plasmid, designated pE-C3, from the commercial vector pEGFP-C3 (Takara Europe SAS, Saint-Germain-en-Laye, France). To produce the expression plasmid pE-ZEB1, cDNA of ZEB1 was amplified from human ZEB1 ORF vector (Applied Biological Materials, Richmond, Canada) using ZEB1_ECORI_for and ZEB1_BAMHI_rev primers (Supplementary Table S10) and the PCR product was then cloned into the pE-C3 empty vector between *Eco*R I and *Bam*H I. To generate doxycycline-inducible LNCaP-eGFP-GR cells, the Lenti-X™ Tet-ON® 3G Inducible Expression System (Takara Europe SAS, Saint-Germain-en-Laye, France) was used. Briefly, cDNA of GR was amplified from the pEGFP-GR plasmid with LVX_GR_BSRGI_for and LVX_GR_MLUI_rev primers (Supplementary Table S10). The pLVX-TRE3G-EGFP-GR vector was then produced by inserting the amplicon into the previously described pLVX-TRE-eGFP vector between *Brs*G I/*Mlu* I [11]. The luciferase reporter plasmids HRE-M1 to HRE-M6 -luc were constructed by the insertion of one copy of the corresponding motif (M1 to M6) present in the enhancer of *CDH2* between the *Mlu* I and the *Bgl* II restriction sites in the pGL4 luciferase reporter vector (Promega, Charbonnières-les-Bains, France). The Notch signaling reporter plasmids 2x RBPJ and 2x mut-RBPJ were established in the same way from the pGL4 vector, inserting 2 repeats of the RBPJ motif "CTTCCC A" or 2 repeats of the mutated sequence "CTAGCGT" respectively. To produce the h*CDH2* Enhancer reporter plasmid, a 640 bp fragment corresponding to the complete sequence of the human *CDH2* enhancer (T2T CHM13v2.0/hs1, chr18:28,351,277−28,351,916) was amplified from the genomic DNA of LNCaP cells using the primers Enh_MLUI_for and Enh_BGLII_rev (Supplementary Table S10) and cloned between the *Mlu* I and the *Bgl* II restriction sites in the pGL4 luciferase vector. The enhancer sequence mutated for the 24 repeats of RBPJ motif was obtained from Eurofins Genomics (Eurofins Genomics, Ebersberg, Germany), amplified by PCR with the same Enh_MLUI_for and Enh_BGLII_rev primers and inserted as before into the pGL4 vector to generate the h*CDH2* mut-Enhancer plasmid. In luciferase assays, induction of Notch signaling was achieved using the expression plasmid pE-NICD1 encoding activated Notch1 (Notch 1 intracellular domain). PiggyBac vector pB-TAG-NICD was purchased from Addgene (Addgene, #130934, Watertown, MA, USA) and used to amplify NICD1 cDNA with NICD1_ECORI_for and NICD1_BAMHI_rev primers (Supplementary Table S10). The PCR product was then cloned into the pE-C3 empty vector described above between *Eco*R I and *Bam*H I. The pGL4.70[hRluc] vector, herein referred to as pRenilla-luc, was purchased from Promega (Promega, Charbonnières-les-Bains, France). All PCR were performed using the CloneAmp™ HiFi PCR Premix (Takara Europe SAS, Saint-Germain-en-Laye, France). All cloning reactions were achieved using the In-fusion® HD Cloning Kit (Takara Europe SAS, Saint-Germain-en-Laye, France). Sequences of oligonucleotides used for cloning and PCR are available in Supplementary Table S10.

Transient transfection and luciferase assays

LNCaP cells were seeded in 24-well plates (7×10^4 cells/well) and transfected with 1 µg of pEGFP-AR or pEGFP-GR using JetPEI transfection reagent (Polyplus Transfection, Illkirch-Graffenstaden, France). Twenty-four hours after, the growth medium was changed by phenol red free RPMI-1640 medium supplemented with 5% (v/v) charcoal-stripped fetal bovine serum (Sigma-Aldrich, Saint-Quentin-Fallavier, France) and with 10 nM of dihydrotestosterone (DHT) or 10 nM of dexamethasone (DEX) (Sigma-Aldrich, Saint-Quentin-Fallavier, France). For luciferase assays in HEK293T cells, the JetPEI transfection reagent (Polyplus Transfection, Illkirch-Graffenstaden, France) was used according to the Batch protocol provided by the manufacturer. HEK293T cells were plated and transfected the same day in 96-wells plates in phenol red free DMEM medium supplemented with 5% (v/v) charcoal-stripped fetal bovine serum (Sigma-Aldrich, Saint-Quentin-Fallavier, France) and with, as required, 1 nM or 10 nM of DHT or DEX or vehicle (ethanol). For nuclear receptor reporter assays, a quantity of 40,000 cells were transfected with 75 ng of pEGFP-AR, pEGFP-ARV7 or pEGFP-GR expression constructs in combination with 150 ng of HRE-M1 to HRE-M6 -luc, or h*CDH2* Enhancer-luc reporter constructs and 20 ng of pRenilla-luc. For Notch signaling reporter assays, 40,000 cells were transfected with 50 ng of pE-NICD1 or pE-GR expression constructs, or 50 ng of pE-C3 empty plasmid as control, in combination with 150 ng of 2x RBPJ, 2x mut-RBPJ, h*CDH2* Enhancer or h*CDH2* mut-Enhancer luciferase reporter plasmid and 50 ng of pRenilla-luc. In Notch luciferase assays, GR activities were assessed in the presence of 10 nM of dexamethasone. Luciferase assays were performed 24 h after transfection using Dual-Glo® Luciferase Assay System (Promega, Charbonnières-les-Bains, France) and following the manufacturer protocol. The relative luciferase unit (RLU) from individual transfections was normalized by measurement of Renilla luciferase activity. Transfection experiments were repeated three times in duplicate or triplicate and data are presented as mean +/- sem.

RNA isolation, reverse transcription and real-time PCR

Total RNA was isolated 3 days post-transfection or post-induction with doxycycline using the NucleoSpin® RNA plus kit (Macherey-Nagel, Hoerdt, France) and according to the manufacturer's protocol. This experiment was performed in three biological replicates for each condition. Reverse transcription was conducted from 500 ng of total RNA using iScript™ Reverse Transcription Supermix for RT-qPCR (Bio-Rad, Marnes-la-Coquette, France) as recommended by the manufacturer. Then, real-time PCR was performed with LightCycler® 480 SYBR Green I Master (Roche Diagnostics France SAS, Meylan, France) and with validated primers for *EGFP* (QT01171611, Qiagen, Courtaboeuf, France) and for *CDH2* isoform 1 (NM_001792) (QT00063196, Qiagen, Courtaboeuf, France). The housekeeping gene *HMBS* (QT00014462, Qiagen, Courtaboeuf, France) expression was used to normalize the results according to the 2DeltaCt method. *CDH2* isoform expression levels were similarly analyzed by RT-qPCR using primers specific to *CDH2* isoform 1 (NM_001792.5) and *CDH2* isoform 2 (NM_001308176.2) (Supplementary Table S10). Data are presented as mean +/- sem from 2 to 4 independent experiments in triplicate.

Protein isolation and western blot

Media were renewed 72 h after transfection or induction with doxycycline and protein extracts were prepared on day 6 in RIPA lysis buffer (Thermo Fisher Scientific France, Illkirch-Graffenstaden, France) supplemented with 1x of protease inhibitor cocktail (PIC) (Roche Diagnostics France SAS, Meylan, France), 1x phosphatase inhibitor cocktail (Phosph. IC) (Sigma-Aldrich, Saint-Quentin-Fallavier, France) and 25 units of benzonase (Merck-Millipore, Molsheim, France) during 30 min on ice. Cellular debris were removed from lysates by centrifugation at 14 000 g for 15 min at 4 °C. Protein concentration was quantified using Bio-Rad Protein Assay (Bio-Rad, Marnes-la-Coquette, France) according to the manufacturer's protocol. An equivalent quantity of total proteins was separated by 7.5% (w/v) SDS-PAGE and transferred onto a nitrocellulose membrane. Membranes were blocked with TBS/0.1% (v/v) Tween/4% (w/v) nonfat dry milk and probed with primary antibodies at 4 °C overnight. After incubation with horseradish peroxidase-conjugated secondary antibodies for 1 h, immunoreactive proteins were visualized by chemiluminescence (Immobilon Western, Millipore, Molsheim, France). Antibodies are detailed in Supplementary Table S9.

Nuclear extracts

A quantity of 6 million of HEK293T cells was transfected with 15 µg of pE-AR, pE-GR, pE-ARV7 or pE-ZEB1 expression constructs and plated the same day in 100 mm Petri dishes in phenol red free DMEM medium supplemented with 5% (v/v) charcoal-stripped fetal bovine serum and with, as required, 10 nM of DHT or DEX or vehicle (ethanol). Nuclear protein extracts were prepared 24 h post-transfection. Cells were collected in cold phosphate buffered saline (PBS) and centrifuged at 500 g for 15 min at 4 °C. Cell pellets were resuspended in 300 µl of cold buffer A {10 mM HEPES-KOH pH7.9, 1.5

mM MgCl2, 10 mM KCl, 0.5 mM DTT, 1x PIC, 1x Phosph. IC} and incubated on ice for 10 min. After centrifugation at 14 000 g for 30 s at 4 °C, supernatants (cytosolic fractions) were removed and pellets were resuspended in 70 µl of cold buffer B {20 mM HEPES-KOH pH7.9, 1.5 mM MgCl2, 420 mM KCl, 0.5 mM DTT, 10% (v/v) glycerol, 1x PIC, 1x Phosph.IC}. Lysates were incubated on ice for 30 min then centrifugated at 14 000 g for 10 min at 4 °C. Supernatants (nuclear fractions) were collected, diluted with an equal volume of cold buffer C {20 mM HEPES-KOH pH7.9, 1.5 mM MgCl2, 0.5 mM DTT, 30% (v/v) glycerol, 1x PIC, 1x Phosph.IC} and stored at -80 °C. Protein concentration was quantified using Bio-Rad Protein Assay (Bio-Rad, Marnes-la-Coquette, France) according to the manufacturer's protocol. Nuclear protein extracts from LNCaP and HEK293T (293T) cells were prepared in the same way.

Fluorescent electrophoretic mobility shift assay

Oligonucleotides with or without Cyanine 5 (CY5) fluorescent dye modification at the 5' terminus were purchased from Eurofins Genomics (Eurofins Genomics, Ebersberg, Germany) (Supplementary Table S10). For ZEB1 binding analyses, DNA probes containing one of the E2 box motifs identified in the *COL4A3* promoter regions and corresponding to the motifs found in human *CDH2* intron 1 were used [23] (Supplementary Table S10). Oligonucleotides were annealed at 20 µM in 10 mM Tris–HCl pH 8.0, 100 mM NaCl, and 1 mM EDTA as followed: samples were incubated 5 min at 95 °C, then allowed to cool to room temperature. A quantity of 10 µg of nuclear extract was incubated for 30 min on ice in binding buffer {10 mM Tris-HCl pH 7.5, 0.5 mM DTT, 0.5 mM EDTA, 1 mM β-mercaptoethanol, 1x PIC, 5% (v/v) glycerol, 50 µg/mL Poly-dIdC} supplemented, as required, with 100 nM DHT or DEX, and with 100 ng of specific antibody (Supplementary Table S9). Then 600 fmol of CY5-labeled DNA probe were added and reactions were incubated for 20 min on ice in the dark. Competition assays were performed in the presence of 600 fmol of labeled probe and a 40-fold molar excess (24000 fmol) of unlabeled DNA (cold probe). After pre-running the gel for 60 min at 4 °C, binding reactions were loaded on an 4% (w/v) native poly-acrylamide gel and ran in cold 0.5x Tris Borate EDTA (TBE) buffer supplemented with 2.5% (v/v) glycerol at 100 volts and at 4 °C for 1 h. Gels were imaged using the ImageQuant 800 imaging system (Cytiva, Saint-Germain-en-Laye, France).

PCR amplification and sequencing strategy

The *primer3* tool [24] was used to design forward and reverse PCR primers to amplify the VNTR region present in intron 1 of the human *CDH2* taking the human genome (hg38, GRCh38.p13, NC_000018.10, position 28,156,372 to 28,156,865) as reference (Supplementary Table S10). PCR amplification was performed using the iPLEX Pro PCR reagent set (Agena Bioscience, Hamburg, Germany). The reaction mix was made in a final volume of 25 µL, containing 10X buffer, 25 µM dNTP mix, 1 µM primer mix, and 5 U/µL Taq polymerase. DNA was amplified with an initial denaturation at 95 °C for 2 min, followed by 45 cycles of 30 s denaturation at 95 °C, 30 s annealing at 56 °C, and 1 min extension at 72 °C, with a final 5 min extension at 72 °C. Samples were run on a 1.5% (w/v) agarose gel. A volume of 10 µL of PCR product, mixed with 2 µL loading buffer, was loaded per well. After gel running, band intensities were estimated to calculate the elution volume needed for a final concentration of 2 ng/µL. PCR products were purified using the NucleoSpin Gel and PCR Clean up kit (Machery-Nagel, Hoerdt, France). A second "BigDye" PCR was performed for Sanger sequencing. The BigDye™ Terminator v1.1 sequencing kit (Thermo Fisher Scientific France, Illkirch-Graffenstaden, France) was used. The reaction was made in a final volume of 10 µL, containing 5X reaction buffer, BigDye™ Terminator v1.1 enzyme, 0.8 µM primers, and 2 ng/µL purified PCR products. BigDye™ PCR reactions were performed with an initial denaturation at 96 °C for 60 s, followed by 25 cycles of 10 s denaturation at 96 °C, 5 s annealing at 53 °C, 4 min extension at 60 °C, and a final 7 min extension at 72 °C. BigDye™ PCR products were precipitated with 80% (v/v) and then 70% (v/v) ethanol. After centrifugation, the pellet was air-dried and re-suspended in 10 µL of H2O. A volume of 4 µL of the purified PCR products were sequenced using the 3500 Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific France, Illkirch-Graffenstaden, France). Obtained sequences were analyzed using the Sequencher™ 5.4.6 software (Gene Codes Corporation, Ann Arbor, MI, USA) to visualize both sense and anti-sense sequences for each sample. After visual inspection for ambiguities, further analysis was performed using in-house R scripts to finalize the analysis of repetitions. Genomic DNA from human prostate adenocarcinoma were obtained from Origene (CliniSciences, Nanterre, France). DNA sequencing of the VNTR region present in human prostate cancer cell lines and tissues was assessed after amplification with Platinum *Pfx* DNA polymerase (Thermo Fisher Scientific France, Illkirch-Graffenstaden, France) according to the manufacturer's protocol and with VNTR_region_2 forward and reverse primers (Supplementary Table S10).

Sequence analysis

Analyzing a genomic region with VNTRs using Sanger sequencing presents inherent challenges. In samples that

are homozygous for the number of repeats but have different motif polymorphisms on each chromosome, the inability of capillary sequencing to separate forward and reverse strands results in ambiguous nucleotides, represented here by IUPAC codes (e.g., W, R, Y, M, K, S). For heterozygosity in terms of the number of repeats, possible configurations of the motifs and numbers of repeats were iterated to determine the configuration explaining the IUPAC ambiguities caused by the overlap in the flanking region and within motifs. Full sequences with ambiguities are provided in the Supplementary Tables S1 and S2.

In silico analyses

Motif identification was revealed with JASPAR Transcription Factor Binding Sites (TFBS) predictions visualized directly in the UCSC genome browser as genomic tracks [25, 26] (https://jaspar.elixir.no/genome-tracks/) (https://genome.ucsc.edu/index.html). To determine the factors that bind our region of interest, ChIP-seq data were probed using the Cistrome DB Toolkit with "Transcription factor or Chromatin regulator" as Data type in Cistrome and "chr18:28156372–28156865" as Interval (hg38 genome coordinates) [27] (http://dbtoolkit.cistrome.org). The con servation of this genomic region within intron1 of *CDH2* during evolution was assessed by the "24 primates EPO-Extended" and "91 eutherian mammals EPO-Extended" multiple alignments in Ensembl release 113 [28] (https://www.ensembl.org/index.html). Full alignments and reference genomes are available at the following links http://www.ensembl.org/Homo_sapiens/Location/Compara_Alignments?align=2050_db=core_r=18%3A28156372-28156865 and http://www.ensembl.org/Homo_sapiens/Location/Compara_Alignments?align=2073_db=core_r=18%3A28156372-28156865.

ChIP-seq data analysis and visualization

For each sample of AR (GSE62492) and GR (GSE39879) ChIP-seq identified with an SRA accession number, the raw data in fastq format were downloaded to the Galaxy application (https://usegalaxy.fr/) from Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) databases of the National Center for Biotechnology Information (NCBI) (Supplementary Table S4). The reads were first aligned to the CHM13_T2T_v2.0 reference genome using the Bowtie2 program. Then, the bamCoverage tool was used to generate a signal coverage file in bigWig format for visualization on the Integrated Genome Viewer (IGV) genome browser. ChIP-seq data from murine Ar (GSE163145) were directly downloaded from GEO in bigWig format and visualized on the IGV genome browser.

Public transcriptomic datasets

The SU2C/PCF Dream Team 2019 cohort consists of whole exome sequencing of 444 castrate resistant PCa tumor/normal pairs [29]. Of these 444 samples, 266 show gene quantification data. The Fred Hutchinson CRC 2016 cohort concerns 63 patients with disseminated PCa for a total of 176 samples [30]; from this project, only samples from metastatic PCa ($n=149$) were considered for correlation analysis. The WCDT MCRPC project from the National Cancer Institute (NCI) at the National Institutes of Health (NIH) analyzes 101 patients with metastatic Castration Resistant PCa (mCRPC) and provides 99 samples with transcriptomic data [31]. GSE126078 data obtained from GEO database include the RNA sequencing of 43 patient-derived xenograft (PDX) models and 98 metastatic castration-resistant PCa (mCRPC) for a total of 141 samples [32]. For the MSK 2010 cohort from the Memorial Sloan Kettering Cancer Center (MSKCC) Prostate Oncogenome Project, only samples from primary PCa ($n=131$) were considered for analysis in this paper [33]. The TCGA PRAD project of the NCI provides 540 samples from 484 patients with prostate adenomas and adenocarcinomas and normal tissues [31] (TCGA Research Network: https://www.cancer.gov/tcga); from this project, only samples from primary PCa ($n=488$) were used for analysis. The GSE25183 study consists of the analysis of 21 prostate cell lines with variable number of replicates per sample for a total of 56 RNA-seq data [34]. All transcriptomic datasets are recapitulated in the Supplementary Table S5.

Correlation analysis

Published transcriptomic datasets were downloaded from databases in the format of processed RNA-seq data or mRNA expression values from microarray (Supplementary Table S5). The log2-transformed expression values were used to compute z-scores per gene. To explore the relationship between *CDH2* expression and AR activity in PCa cohorts, an AR activity score was generated based on the expression of AR target genes identified from the literature [35]. To refine the signature, only 20 genes that are upregulated by AR were included (Supplementary Table S6). The AR score was calculated by the summation of the expression z-scores of the 20 AR target genes. The same strategy was applied to compute a Notch activity score based on 20 genes that are overexpressed after activation of Notch pathway [36] (Supplementary Table S7). Analysis of the correlation between two variables was assessed by measuring Pearson's coefficients and associated *p-values* (Supplementary Table S8).

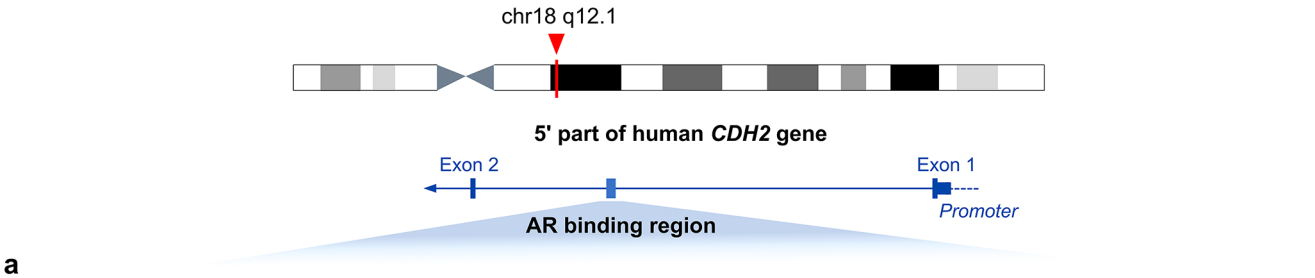
Statistics and graphic representation

All statistical analyses and data visualization were performed in python 3 using the pandas, matplotlib, numpy, scipy, and seaborn packages [37]. Unpaired two-tailed Student's t-test or Mann-Whitney test was used to determine the statistical difference between two sets of data. *P-value* < 0.05 was considered significant.

Results

Genomic variations within the androgen receptor binding region in intron 1 of the human *CDH2* gene

A cluster of AR binding motifs present in intron 1 of human *CDH2* gene was earlier reported (hg38, GRCh38.p13, NC_000018.10, position 28,156,372 to 28,156,865) [21]. We have previously shown that this genomic region was recognized by both full-length AR and AR-V7, but with diverged consequences on N-cadherin expression [10]. A deeper in silico analysis of this AR binding genomic region in intron 1 of the human *CDH2* uncovered two yet unreported peculiar features. First, according to the human reference genome considered, the number of AR binding sites diverged. Indeed, taking hg38 as reference genome, 21 AR binding sites were previously identified (Fig. 1a) while 24 AR binding sites are found within the novel human reference genome T2T-CHM13/hs1 (chromosome 18: position 28,351,299 to 28,351,858) (Fig. 1b). To further examine the variability in the number of potential AR binding sites within the intronic region of interest of human *CDH2* gene, we analyzed the genome of diverse present-day humans (Table 1). These samples included 5 men and 4 women from 3 populations, Yakuts or Sakha (North-Eastern Siberia), Beninese (West Africa) and Europeans. Variability among these human sequences was analyzed after amplification of the region of interest and Sanger sequencing. The number of AR binding sites ranged from 22 to 24 among these nine present-day human genomic DNA (Table 1; Supplementary Table S1). To assess any alteration of this genomic region in a tumor context, we further checked for any variation in the number and or the sequence of these AR binding motifs within intron 1 of the human *CDH2* gene in PCa cells and tissues. Genomic DNA was extracted from LNCaP and C4-2B cells and from two human PCa samples, before amplification and sequencing. For the two cell lines, the number of AR binding motifs and their sequences within intron 1 of the human *CDH2* gene were identical when considering T2T-CHM13/hs1 as reference genome. For the two prostate cancer tissues, we retrieved a number of 22 or 24 AR binding sites as observed for the nine present-day human genomic DNA. These data suggest that the enhancer sequence integrity was not necessarily affected in the context of prostate cancer (Supplementary Table S2).



a

Chr18 (hg38) : 28,156,372 - 28,156,865 (493 bp; 21 AR binding sites)

*ctgggtctgcatacaaaatgt*GGTACAGAATGTCACCTTCCCAGGTACAGAATGTCACCTTCCCAGGTATAGCATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTA
CAGAATGTCACCTTCCCAGGTATAGAATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTACAGAATGTCACCTTCCCA
GGTACAGCATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTGCAGCATGTCACCTTCCCAGGTGCAGAATGTCACCTTCCCAGGTACAGCATGTCACCT
TCCCAGGTACAGCATGTCACCTTCCCAGGTATAGCATGTCACCTTCCCAGGTATAGCATGTCACCTTCCCAGGTACAGAATGTCACCTTCCCAGGTATAGCATGTCA
CCTTCCCAGGTATAGCATGTCACCTTCCCAGGTATAGCATGTCACCTTCCCA*ggagtagagcc*

b

Chr18 (T2T-CHM13) : 28,351,299 - 28,351,858 (559 bp; 24 AR binding sites)

*ctgggtctgcatacaaaatgt*GGTACAGAATGTCACCTTCCCAGGTACAGAATGTCACCTTCCCAGGTATAGCATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTA
CAGAATGTCACCTTCCCAGGTACAGAATGTCACCTTCCCAGGTATAGAATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCA
GGTACAGAATGTCACCTTCCCAGGTACAGAATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTGCAGAATGTCACCT
TCCCAGGTATAGCATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTACAGCATGTCACCTTCCCAGGTACAGAATGTCACCTTCCCAGGTATAGCATGTC
ACCTTCCCAGGTATAGCATGTCACCTTCCCAGGTACAGAATGTCACCTTCCCAGGTATAGCATGTCACCTTCCCAGGTATAGCATGTCACCTTCCCAGGTATAGCAT
GTCACCTTCCCA*ggagtagagcc*

Fig. 1 Schematic view of the enhancer within intron 1 of human *CDH2* gene. Cluster of AR binding sites (in blue) present in *CDH2* enhancer according to the human reference genome (**a**) hg38 (GRCh38.p13; NC_000018.10; position 28,156,372 to 28,156,865) or (**b**) T2T-CHM13/hs1 (chromosome 18: position 28,351,299 to 28,351,858). Enhancer 5' and 3' flanking regions are indicated in italics and red letters. In black, a 7 bp DNA repeated sequence with no base change interspaces two AR binding sites evenly throughout the enhancer

Table 1 Characteristics of nine present-day human genomic samples and the number of androgen receptor (AR) binding motifs found within the intronic region in *CDH2* gene

Individual	Species	Population	Sex	Number of AR binding sites
YAK-M031	Homo sapiens	Sakha (Yakuts)	Male	24/24
YAK-BA11	Homo sapiens	Sakha (Yakuts)	Male	22/22
YAK-M01	Homo sapiens	Sakha (Yakuts)	Female	24/24
BEN-GOR162P	Homo sapiens	Beninese (West Africa)	Male	22/22
BEN-GOR145M1E	Homo sapiens	Beninese (West Africa)	Female	22/22
EU-KOV	Homo sapiens	European	Male	22/22
EU-KEN	Homo sapiens	European	Male	24/24
EU-CLEM	Homo sapiens	European	Female	24/24
EU-ROM	Homo sapiens	European	Female	22/22

The second interesting genomic feature relied on the level of conservation of this cluster within intron 1 of *CDH2* during evolution. This cluster of 24 AR binding sites is completely absent in genomes of other organisms (Fig. 2a and S1). However, only one divergent sequence is retrieved in the homolog genomic region of primates, excepted in tarsiers (Fig. 2b). Our analysis of this genomic region in primates revealed an evolutive history from sequences $_{5'}$GGTTTAgaaTGTCAT$_{3'}$ and $_{5'}$GGTTTAggaTGTAAC$_{3'}$, present respectively in loris and lemur to that one present in *Hominoidea* (Fig. 2b). Also, it is noteworthy to highlight modifications during evolution in the contiguous 7-bp sequence (Fig. 2b; sequence in black). While one copy of the invariable $_{5'}$CTTCCCA$_{3'}$ sequence that interspaces AR binding motifs in human was perfectly retrieved in great apes and lemurs, only an imperfect copy is observed in other primates. This supports the notion that these sequence variations are not random but follow phylogenetic patterns. Also, these data suggest that the sequence $_{5'}$GGTATAGAATGTCACCTTCCCA$_{3'}$, present in a unique copy in great apes and gibbon may constitute an ancestral motif that has been amplified in humans as part of a fast-evolving region in the human genome [38].

Therefore, this cluster in intron 1 of human *CDH2* could be one of these particular and recently evolved "enhancer" sequences, specifically in humans, that control mesenchymal identity [39], and should be considered as a key genomic region for cell-type specific chromatin topology and cell-type specific gene regulation of the *CDH2* gene.

A potential transcriptional hub to control human *CDH2* expression

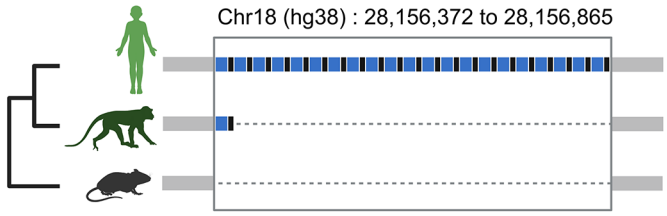
Canonical AR binding sites are known to be bipartite elements composed of two hexameric core half-site motifs forming an imperfect inverted repeat separated by a three base-pair spacer [40]. In this *CDH2* intron 1 region, the sequence of the 3' hexameric core ($_{5'}$TGTCAC$_{3'}$) remains invariable, whereas the 5' half-site motif and the three base-pair spacer present 3 polymorphic sites (*) ($_{5'}$GGT**Ag*aTGTCAC$_{3'}$). This leads us to suggest a motif ranking from M1 to M6 according to their order of appearance within the enhancer region, with motif 5 lacking in T2T-CHM13/hs1 (Table 2).

A comprehensive analysis of transcription factor binding sites within this enhancer sequence using JASPAR revealed unique features. While the first half-site $_{5'}$*GGTACA*$_{3'}$ of motif 1 and motif 2 was predicted to be recognized by AR, it was not the case for the corresponding half-site in other repeated motifs. Indeed, the $_{5'}$*GGTATA*$_{3'}$ found in motif 3 and motif 4 and the $_{5'}$*GGTGCA*$_{3'}$ present in motif 5 and motif 6 were not predicted to be recognized by AR (Fig. 3a; Table 2). Furthermore, the first half-site $_{5'}$*GGTACA*$_{3'}$ of motif 1 and motif 2 could also been recognized by the other three 3-ketosteroid receptors comprising glucocorticoid receptor (GR, NR3C1), mineralocorticoid receptor (MR, NR3C2) and progesterone receptor (PR, NR3C3) (Fig. 3a). Regarding the 3' half-site $_{5'}$*TGTCAC*$_{3'}$ retrieved in all the six motifs, the first three base pairs are part of an AR, GR, MR or PR binding site while the three last ones diverge from the consensus motif recognized by these steroid receptors, mainly in position 5 where a C is predominantly found. However, the 3' half-site $_{5'}$*TGTCAC*$_{3'}$ fits well with the first part of estrogen receptor 1 (ESR1, NR3A1) binding site, suggesting a potential involvement of ESR1 in the control of expression of human *CDH2* gene (Fig. 3a). These observations led us to consider motifs 1 to 6 as potential steroid hormone receptor binding sites, hereinafter named hormone response elements (HREs).

We further analyzed the invariable sequence $_{5'}$CTTCCCA$_{3'}$, which is repeated 24 times, interspersing two steroid hormone receptor binding sites. This sequence appeared to be the reverse complement of a potential binding site for the recombination signal binding protein for immunoglobulin kappa j region (RBPJ), a transcriptional effector of Notch signaling (Fig. 3a and b). We also highlighted 1 × $_{5'}$*CAGGTG*$_{3'}$ and 22 × $_{5'}$*CAGGTA*$_{3'}$ motifs overlapping both the RBPJ and HRE sequences (Fig. 3b). These motifs correspond to potential binding sites of the zinc finger E-box binding homeobox 1 (ZEB1) (Fig. 3a).

Available cistrome datasets were further questioned for experimental proofs of binding of any of these transcription factors to the enhancer of interest using CistromeDB Toolkit [27]. When aligned on hg38, peaks for all the steroid

a



b

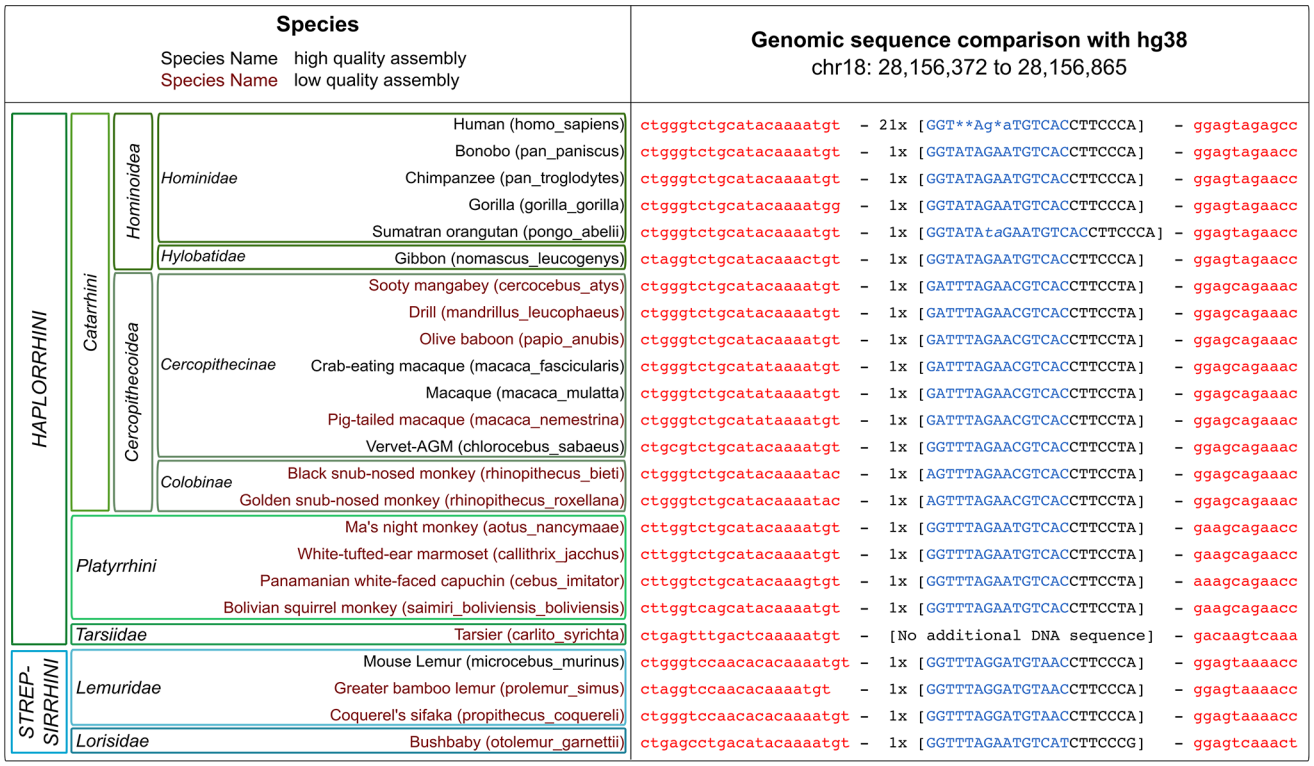


Fig. 2 Genome evolution of the region of interest in intron 1 of *CDH2* gene. (a) Illustration of the genomic comparison of the region of interest in intron 1 of *CDH2*. AR binding sites are represented in blue. The DNA sequence that interspaces two AR binding sites in human are represented in black. Dashed line corresponds to gap in genome alignments. Created in BioRender. https://BioRender.com/o5r8i2h. (b) 24 Primates EPO-Extended multiple alignment of the sequence representing the cluster present in intron 1 of human *CDH2* gene (chr18: 28,156,372 to 28,156,865; hg38) from Ensembl release 113 (https://www.ensembl.org/index.html). For full alignments and reference genomes, see methods. Enhancer 5' and 3' flanking regions are indicated in red letters. In blue, AR binding site sequence during primate evolution. Asterisks refer to polymorphic sites in the sequence of interest in human genome (See Figure 1 and Table 2). In black, the relative sequence that corresponds to the 7-bp DNA repeated sequence that interspaces two AR binding sites in human

Table 2 DNA sequence variability and number of potential AR binding sites within the intronic region of interest of human *CDH2* gene according to the considered reference genome, hg38 (chr 18: 28,156,372 to 28,156,865) or T2T-CHM13/hs1 (chr 18: 28,351,299 to 28,351,858). The three base-pair spacer is indicated in lowercase. Polymorphic sites are in bold

AR binding motif	DNA sequence	Number of motif repeats	
		hg38	**T2T-CHM13/hs1**
Motif 1	5' -GGT**A**C**A**g**a**aTGTCAC- 3'	5	8
Motif 2	5' -GGT**A**C**A**g**c**aTGTCAC- 3'	7	7
Motif 3	5' -GGT**AT**A**g**aaTGTCAC- 3'	1	1
Motif 4	5' -GGT**AT**A**g**caTGTCAC- 3'	6	7
Motif 5	5' -GGT**GC**A**g**caTGTCAC- 3'	1	0
Motif 6	5' -GGT**GC**A**g**aaTGTCAC- 3'	1	1

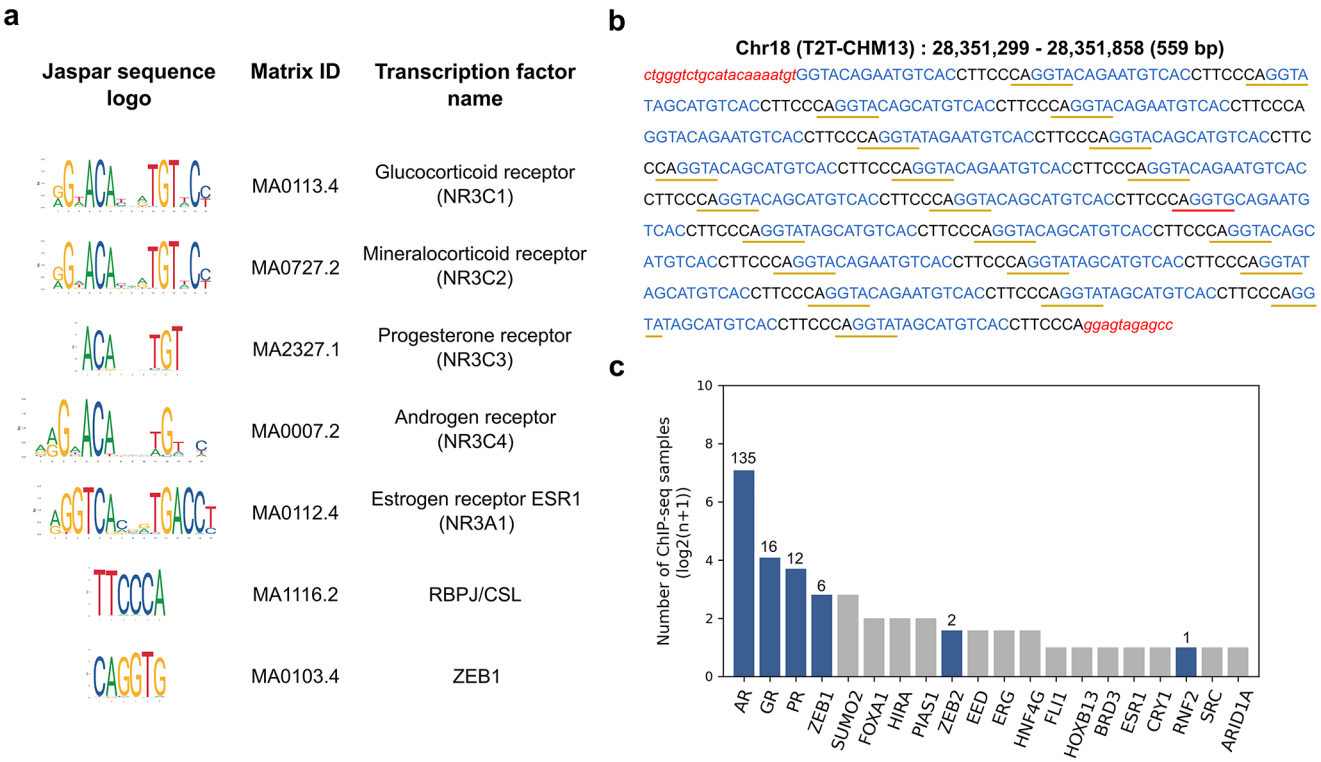


Fig. 3 Putative transcription factor binding sites within the enhancer present in intron 1 of human *CDH2* gene. (a) Comprehensive analysis of transcription factor binding sites using JASPAR within human genomic region corresponding to chr 18: 28,351,299 to 28,351,858 (T2T-CHM13/hs1). Homo sapiens sequence logos and matrix ID were obtained from the CORE collection of the 10th release (2024) of JASPAR (https://jaspar.elixir.no/). RBPJ and ZEB1 logos are presented as reverse complement. (b) Position of putative transcription factor binding sites within the enhancer present in intron 1 of human *CDH2* gene. Enhancer 5' and 3' flanking regions are indicated in italics and red letters. The 24 HREs present in the enhancer are indicated in blue. In black, 24 repeats of the RBPJ/CSL motif intersperse the HREs. Overlapping RBPJ and HRE motifs, 23 sequences corresponding to the ZEB1 motif are underlined in red (1 × 5'−CAGGTG−3') and gold (22 × 5'−CAGGTA−3'). (c) Top 20 transcription factors and chromatin regulators that bind within the human genomic region corresponding to chr 18: 28,156,372 to 28,156,865 (hg38). Y axis represents the number of ChIP-seq samples (log2 scale) exhibiting one or more peaks of the respective factor within the genomic region of interest (Supplementary Table S3). Color indicates transcription factors whose motif has been identified in the sequence of interest. For these factors, number of ChIP-seq samples is indicated at the top of the bars. Source http://dbtoolkit.cistrome.org

receptors, AR, GR, PR, ESR1, excepted MR were retrieved from various experimental models within the human genomic region encompassing chr 18: 28,156,372 to 28,156,865 (Fig. 3c; Supplementary Table S3). Peaks were also retrieved for ZEB1, ZEB2 and others chromatin regulators including SRC, FOXA1, BRD3 but not for RBPJ (Fig. 3c).

These data indicate either co-binding and/or competitive binding of these different transcription factors events to their overlapping cognate sequence motifs within this enhancer, suggesting a potential crosstalk between different pathways and transcription factors to control human *CDH2* expression.

Recognition of the different hormone response elements found in human *CDH2* enhancer by androgen receptor and glucocorticoid receptor

Due to the well-known role of GR in PCa progression associated with its capacity to bypass AR blockade [41], we hypothesized that GR might play a role in *CDH2* regulation in castration-resistant PCa. From this, we chose to focus our study on the impact of AR and GR in controlling N-cadherin expression.

First, we sought to confirm the recognition of the particular enhancer in intron 1 of human *CDH2* by AR and GR in prostate samples. To visualize the binding of these two nuclear receptors on the genomic region of interest, we analyzed two entries of the GEO database: (i) the GSE39879 dataset that reports on a GR cistrome performed in GR expressing LNCaP-1F5 cells in the presence of dexamethasone (Dex), a GR agonist [42], (ii) the GSE62492 dataset corresponding to AR cistrome in LNCaP cells in the presence of dihydrotestosterone (DHT) [43] (Supplementary Table S4). Alignment of raw reads to the human genome T2T-CHM13/hs1 illustrated AR and GR peaks in the genomic region of interest (chr18: 28,351,299 to 28,351,858) (Fig. 4a). Noteworthy, the absence of Ar binding at the murine *Cdh2* gene

observed from the GSE163145 dataset (Ar cistrome in Mus musculus ventral prostatic tissues) confirmed the specificity of this recognition to genus Homo [44] (Fig. S2; Supplementary Table S4).

According to Jaspar core, only the first half-site of motif M1 and M2 are predicted to be recognized by AR and GR while no data are available to assess the binding ability to the four remaining motifs in the context of the *CDH2* intron 1 enhancer sequence. Aiming at characterizing the induction activity of AR, GR and AR-V7 upon binding on each of the six motifs present in the enhancer, we performed luciferase assays in HEK293T cells using reporter plasmids containing only one copy of either M1, M2, M3, M4, M5 or M6, placed upstream a minimal promoter and the firefly luciferase gene. After co-transfection with a receptor expression plasmid, a luciferase activity was observed for AR or GR mainly for M1 and M2 reporter constructs (Fig. 4b), suggesting a preferential binding of these two transcription factors to M1 and M2 in agreement with Jaspar predictions. A similar profile was obtained for AR-V7 (Fig. S3). This preferential binding to M1 and M2 was also confirmed with fluorescent electrophoretic mobility shift assay (fEMSA) performed from protein extract of AR, AR-V7 or GR expressing HEK293T cells (Fig. S4). To further characterize AR, AR-V7 and GR transcriptional activities from this genomic region of interest, we performed a luciferase assay in the presence of the h*CDH2* Enhancer reporter plasmid containing the entire 640-bp full-length enhancer

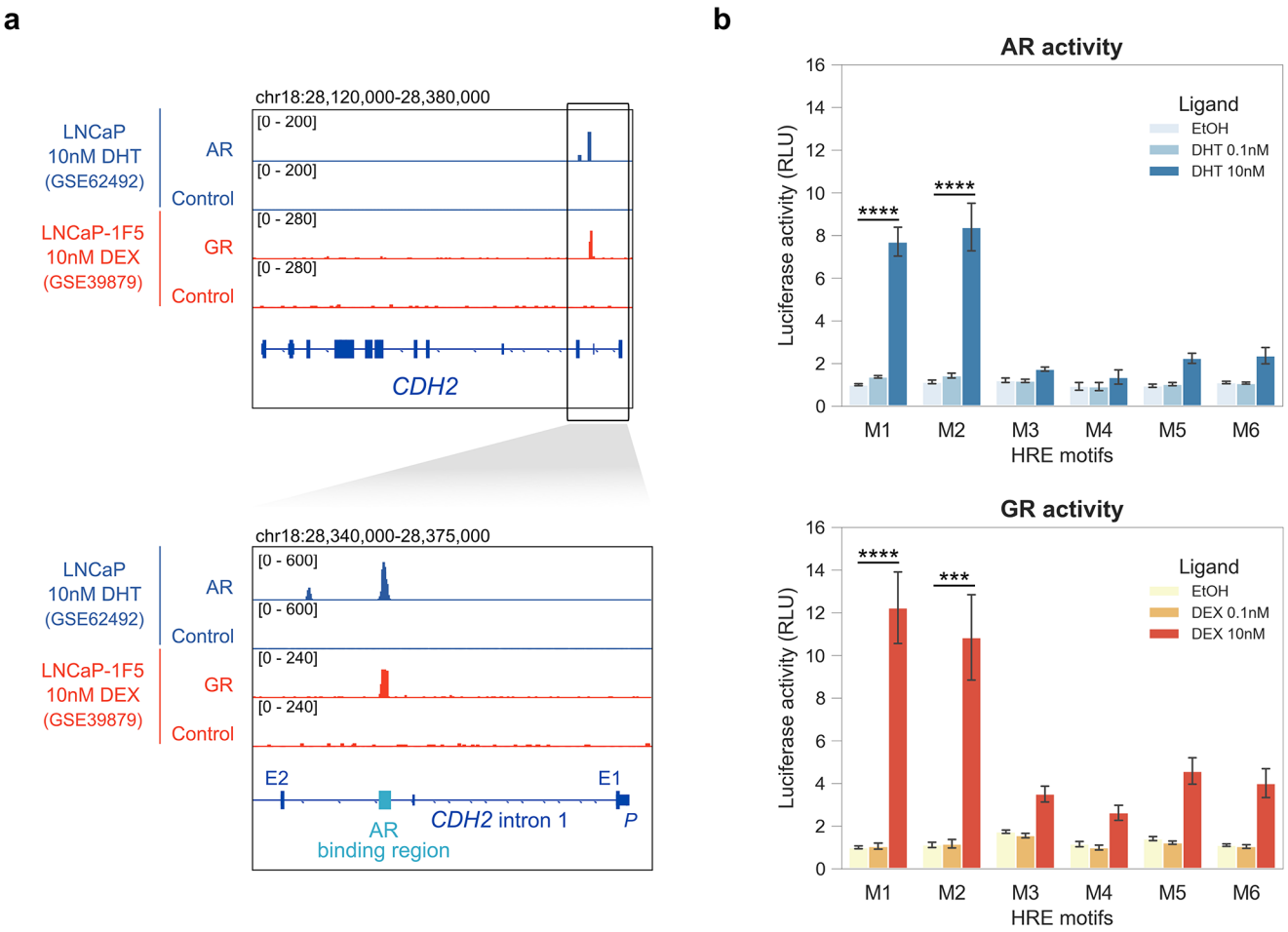


Fig. 4 Recognition of the HRE present in intron 1 of the *CDH2* gene by AR and GR. (**a**) Androgen receptor (AR) (GSE62492) and glucocorticoid receptor (GR) (GSE39879) binding to the enhancer of interest within intron 1 of the human *CDH2* gene. IGV views of AR and GR binding at the intron 1 of the *CDH2* gene on chromosome 18, positions 28,351,299 to 28,351,858 (AR binding region) on T2T-CHM13/hs1 genome. Control AR, input; Control GR, IgG mouse gamma immunoglobulin; E1/E2, exon 1 and exon 2; P, promoter. (**b**) Luciferase assays showing the activities of AR and GR on motifs HRE-M1 to HRE-M6 in the presence of indicated ligand. HEK293T cells were co-transfected with the indicated reporter plasmid HRE-M1 (M1) to HRE-M6-luc (M6), the internal control Renilla-luc, and the expression plasmid pEGFP-AR or pEGFP-GR in the presence (0.1 nM or 10 nM) or absence of the indicated ligand. Luciferase assays were performed 24 h after transfection. For each sample, HRE-mediated luciferase activities (Firefly) were normalized to Renilla activities and reported as relative value to the M1 EtOH conditions. Values are presented as the mean +/- SEM ($n=9$). Two-tailed Student's t-test (***, $pvalue<0.001$; ****, $pvalue<0.0001$). EtOH, Ethanol; DEX, dexamethasone; DHT, dihydrotestosterone

(T2T CHM13v2.0/hs1, chr18:28,351,277−28,351,916) present in intron 1 of human *CDH2* gene. Compared to AR, a 75-fold and 90-fold increased luciferase activity was observed with GR in the presence of 10 nM Dex and with AR-V7 respectively (Fig. S5). Altogether, these data led us to conclude that the respective action of AR, AR-V7 or GR on the expression of the human *CDH2* gene could mainly rely on their potential binding to 15 (8 M1 and 7 M2) out of 24 motifs present in the enhancer of interest.

Competition between androgen receptor and glucocorticoid receptor in N-cadherin regulation

To determine the functional impact of GR on *CDH2* expression, RT-qPCR and western-blot were performed in LNCaP cells at 3 and 6 days respectively after transient transfection with eGFP-AR or eGFP-GR expression plasmids. A significant increase in *CDH2* expression was obtained at mRNA and protein levels in GR expressing LNCaP in the presence of 10 nM Dex compared with control AR expressing LNCaP cells (Fig. 5a). *hCDH2* isoform-specific qRT-PCR further indicated that isoform 1 (NM_001792.5), but not isoform 2 (NM_001308176.2) was upregulated in the presence of GR (Fig. S6). These data suggest that as AR and AR-V7, GR could bind to the enhancer of interest in intron 1 of human *CDH2* gene, but unlike AR, GR activation leads to an increase of N-cadherin expression, an effect that is reminiscent of the ligand independent activation observed for AR-V7 in the same sequence context [11].

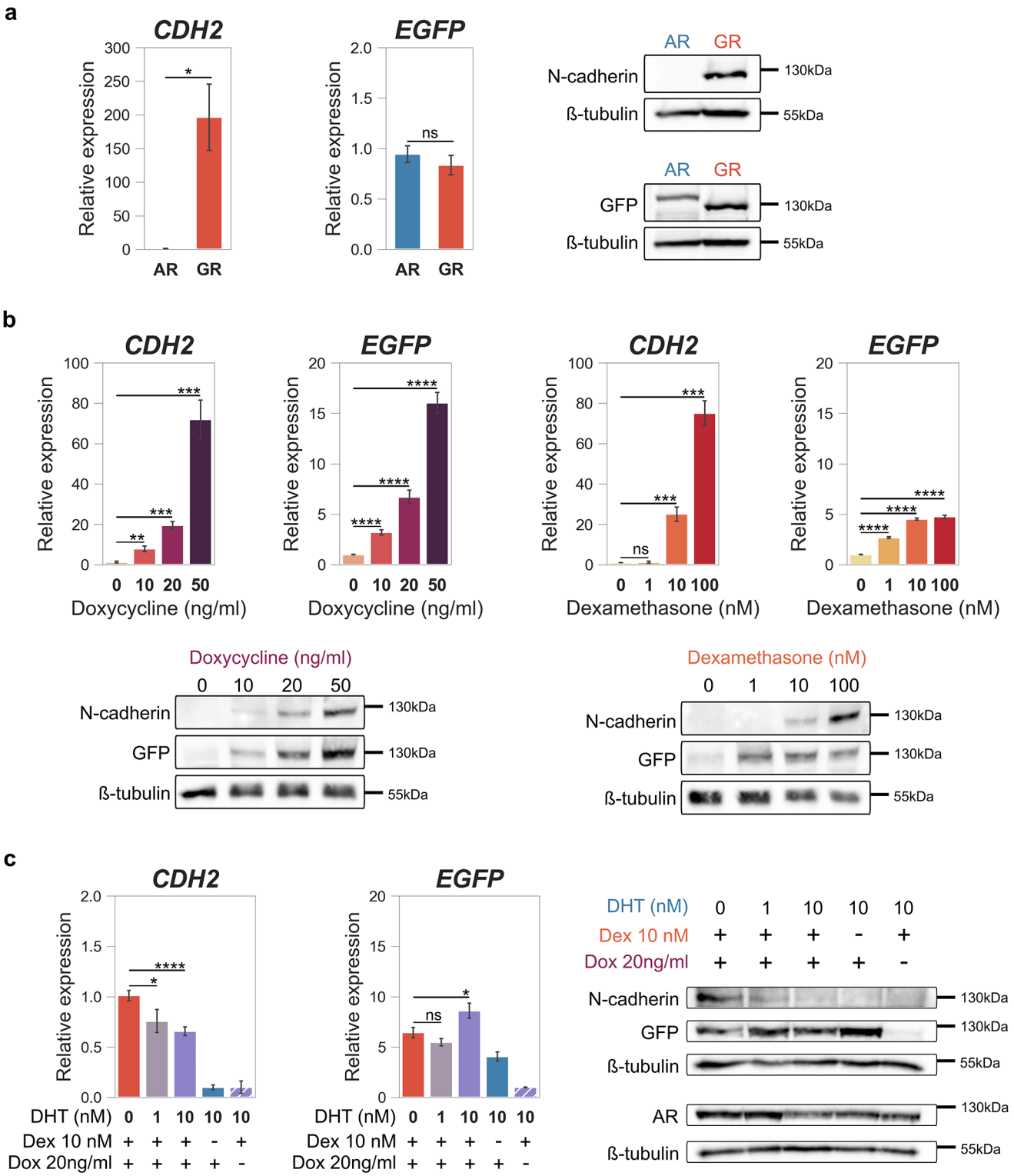
To investigate more deeply the potential induction of *CDH2* by GR in PCa cells, we established stable transduced LNCaP cells with a doxycycline inducible expression of eGFP-GR. In this system, increasing doses of doxycycline (Dox) in the presence of 10 nM Dex, or increasing doses of dexamethasone in the presence of 20 ng/ml Dox, were used to induce or activate GR, respectively. N-cadherin expression was then monitored by RT-qPCR and western blot at respectively 3 and 6 days of treatment. The results showed a dose-dependent increase in *CDH2* expression upon both GR induced expression by doxycycline and its activation by dexamethasone (Fig. 5b). Considering this opposite impact of AR and GR on the expression of N-cadherin, we next investigated whether AR and GR could compete to regulate N-cadherin expression in LNCaP cells. We treated our inducible LNCaP-GR expressing cells with DHT to activate the endogenous androgen receptor of these cells. Interestingly, in the presence of 20 ng/ml Dox and 10 nM Dex, a significant decrease in the expression of the *CDH2* gene was already obtained at the low dose of 1 nM DHT (Fig. 5c). This reduction was even more pronounced in the presence of 10 nM DHT. These data confirm the opposite regulation of AR and GR on N-cadherin expression in PCa cells and highlight the competition between these two nuclear receptors at the HRE motifs within the particular enhancer of the human *CDH2* gene.

Signaling pathways controlling *CDH2* expression in human prostate cancer

We have seen that both AR and GR are able to bind the *CDH2* enhancer and play a direct and opposite role in N-cadherin expression in prostate tumor cells. Considering the presence of motifs for ZEB1 and RBPJ (Notch pathway) in this same region of intron 1 of human *CDH2* and the role of these factors in cell plasticity, we investigated their possible implication in the control of N-cadherin expression.

Publicly available cistrome data indicate binding of ZEB1 to this locus, suggesting a role of this factor in N-cadherin regulation. However, two distinct E-box motifs are retrieved within the enhancer, with $1\times$,_*CAGGTG*_ , and $22\times$,_*CAGGTA*_ , sequences. Binding of ZEB1 to these motifs analyzed by fluorescent EMSA showed protein-DNA complex formation for both sequences (Fig. S7). Interestingly a stronger signal was observed for the ,_*CAGGTA*_ , motif predominantly found in the enhancer (Fig. S7). Regarding Notch signaling, no experimental data indicate potential binding of RBPJ to the human *CDH2* enhancer. Aiming at filling this gap and determining whether this enhancer can respond to the Notch signaling pathway, reporter assays were performed using two types of luciferase constructs, containing either two RBPJ motifs (2x RBPJ) or the entire enhancer comprising 24 repeats of the RBPJ motif (h*CDH2* Enhancer) (Fig. S8a). Furthermore, to validate the specificity of the Notch response, sequences mutated for the RBPJ motif were used, resulting in the 2x mut-RBPJ and h*CDH2* mut-Enhancer reporter plasmids (Fig. S8a). Luciferase assays showed a clear and significant response of both 2x RBPJ and h*CDH2* Enhancer reporters in the presence of activated Notch NICD (Notch Intracellular Domain) (Fig. S8b and S8c). Moreover, these Notch responses were completely abrogated for the sequence mutated for the RBPJ sites, whereas these mutations did not affect the responses observed in the presence of GR (Fig. S8c). These results support a possible impact of the Notch pathway on N-cadherin expression.

Then we aimed to evaluate the role of these transcription factors on the regulation of N-cadherin at a clinical level by exploiting publicly available transcriptomic datasets from patient cohorts and in vitro studies (Supplementary Table S5). We first established an AR activity score based on the expression of well-known AR target genes by summing the z-scores of 20 genes over a set of 30 genes previously described as being under AR control [35] (Supplementary Table S6; Fig. S9). The Notch signaling pathway is



activated via receptor-ligand interactions leading to Notch receptor cleavage, nuclear translocation of the NICD moiety, and expression of the NICD-transactivated target genes. Therefore, to assess the impact of the Notch pathway on *CDH2* gene expression, we similarly generated a Notch activity score using 20 genes that are overexpressed in the presence of an activated Notch pathway [36] (Supplementary Table S7).

We next performed correlation analyses in seven human PCa datasets to explore the relationship between the

Fig. 5 Upregulation of N-cadherin by GR in prostate cancer cells. (**a**) *CDH2* expression levels analyzed by qRT-PCR and western blot after transfection of LNCaP cells with the expression plasmid pEGFP-AR in the presence of 10 nM DHT or pEGFP-GR in the presence of 10 nM Dex. For each sample in qPCR, expression levels were normalized to the housekeeping gene *HMBS* and reported as relative value to the AR condition. (**b**) Q-RT-PCR and western blot showing *CDH2* expression levels in the doxycycline inducible LNCaP-eGFP-GR expressing cells after addition of increasing doses of Dox in the presence of 10 nM Dex, or after addition of increasing doses of Dex in the presence of 20 ng/ml Dox. For qPCR, expression levels of each sample were normalized to *HMBS* and reported as relative value to the dose 0 condition. (**c**) Competition between endogenous AR and eGFP-GR for N-cadherin regulation in doxycycline inducible LNCaP-eGFP-GR cells. *CDH2* expression was analyzed by qRT-PCR and western blot in the presence or absence of DHT to activate AR and Dex to activate GR. For each sample, expression levels were normalized to *HMBS* and reported as relative value to the DHT-free condition (*CDH2*), or to the control condition without Dox (*EGFP*). All experiments were performed at day-3 (qRT-QPCR) or day-6 (western blot) after transfection or doxycycline induction. *EGFP* relative expression (qPCR) and GFP immunodetection (western blot) referred to the level of expression of AR and GR transgenes. Values are presented as the mean +/- SEM. Mann-Whitney test (ns, non-significant; *, *pvalue*<0.05; **, *pvalue*<0.01; ***, *pvalue*<0.001; ****, *pvalue*<0.0001). In immunoblot, 50 to 70 µg of total protein extracts were used, and ß-tubulin was used as loading control

expression or the activity of the different transcription factors potentially binding *CDH2* enhancer and the expression of the *CDH2* gene. As a control, we also added the *CDH1* gene encoding E-cadherin to our analyses. Overall, the correlation heatmaps allowed us to visualize two gene clusters whose expressions were correlated (Fig. 6; Supplementary Table S8). In the first cluster, expression of the epithelial marker *CDH1* was mainly positively correlated with *AR* expression, as well as with the AR activity score (Fig. 6; Supplementary Table S8). A second cluster gathered *CDH2*, *NR3C1* (GR), *ZEB1*, *ZEB2* and *NOTCH1* expression and the Notch activity score (Fig. 6; Supplementary Table S8). Within this group, a positive correlation was particularly observed between *CDH2* expression and that of *ZEB1* and *ZEB2* in almost all datasets (Fig. 6; Supplementary Table S8). As these three genes mark the mesenchymal state of the cells, this association was expected and supports our hypothesis. Moreover, *CDH2* gene expression was also significantly positively correlated with the Notch activity score in six out of seven datasets, suggesting a possible positive impact of Notch signaling pathway on N-cadherin regulation (Fig. 6; Supplementary Table S8). Considering GR, three datasets showed a significant positive correlation between *CDH2* and *NR3C1* expression, while in the other four, no significant correlation was observed (Fig. 6; Supplementary Table S8). This result could be due to the possible co-expression of AR and GR in PCa samples and the competition between the two nuclear receptors for the control of *CDH2*. Interestingly, AR expression levels exhibit a different distribution depending on whether the dataset shows a significant correlation between *CDH2* and *NR3C1* expression (Fig. S10, D-F) or not (Fig. S10, A-C). Thus, the predominance of AR in some datasets could mask the impact of GR on *CDH2* expression.

Also, we observed a clear and significant negative correlation between *CDH2* expression and AR activity in almost all datasets, validating the repression function of AR on N-cadherin expression (Fig. 6; Supplementary Table S8).

Discussion

Aberrant expression of N-cadherin in cancer and its role in the migratory and invasive behavior of cells in the context of EMT have been widely described [45, 46]. However, the precise mechanisms by which N-cadherin expression is regulated are still poorly understood. In this study, we characterize a remarkable enhancer region within intron 1 of the human *CDH2* gene that enables a fine-tuned regulation of N-cadherin expression. This genomic sequence features a highly original composition, with a perfect juxtaposition over 528 bp of tandem repeats of recognition motifs for various transcription factors, including AR. The cluster that can be considered as a variable number tandem repeat (VNTR) comprises 24 HREs potentially recognized by nuclear steroid receptors, separated by 24 motifs for RBPJ, effector of the Notch pathway. Moreover, within these repeats, 23 motifs of the ZEB1 transcription factor overlap both HRE and RBPJ motifs.

Another special feature of this VNTR is its specificity to the Homo genus as no similar VNTR is observed in other species according to our phylogenetic analysis. Considering primates, a divergent sequence barely related to one motif example could be observed, except in tarsiers. This divergent sequence seems to evolve towards a perfect motif 3 $_{5'}$GGTATAGAATGTCAC$_{3'}$ juxtaposed to a complete copy of RBPJ binding site $_{5'}$CTTCCCA$_{3'}$ late during primate evolution. Indeed, motif 3 together with the RBPJ binding site was only present in gibbon and great apes. Also, it is noteworthy to highlight that the related sequence retrieved in loris, lemurs and in new world monkeys (*Platyrrhini*) were more closely related to motif 3 found in great apes, and that was not the case for old world monkey (*Cercopithecoidea*). From these observations, we postulated that motif 3 together with the RBPJ binding site as observed in great apes could be at the origin of the VNTR present in Homo sapiens. This implies a unique evolutionary divergence of intron 1 of *CDH2* gene between humans and other primates marked by human-specific sequence changes.

This VNTR that is crucial for the control of human *CDH2* gene expression can be considered as one of the human accelerated regions (HARs) playing a determining

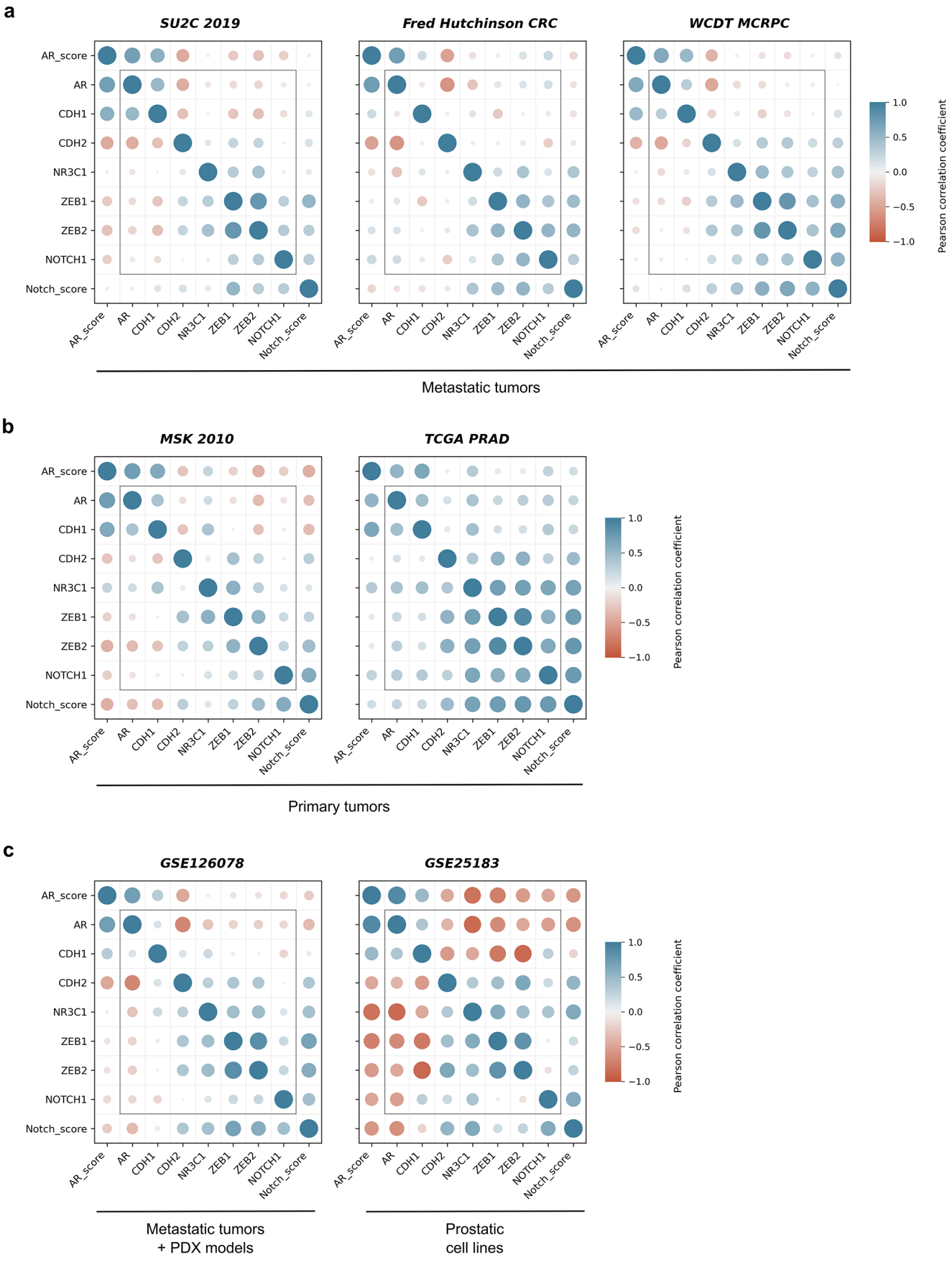

a
SU2C 2019
Fred Hutchinson CRC
WCDT MCRPC
AR_score
AR
CDH1
CDH2
NR3C1
ZEB1
ZEB2
NOTCH1
Notch_score
Pearson correlation coefficient
1.0
0.5
0.0
−0.5
−1.0
Metastatic tumors
b
MSK 2010
TCGA PRAD
Primary tumors
c
GSE126078
GSE25183
Metastatic tumors
+ PDX models
Prostatic
cell lines

Fig. 6 Gene expression correlation analyses in PCa cohorts and PCa cell lines. Heatmaps showing Pearson correlation coefficients between *AR*, *CDH1*, *CDH2*, *NR3C1*, *ZEB1*, *ZEB2 and NOTCH1* gene expression (gray square), AR activity score and Notch activity score. Conditions are ordered to allow visualization of correlation groups. (**a**) Metastatic tumor datasets: SU2C 2019 ($n=266$), Fred Hutchinson CRC ($n=149$) and WCDT MCRPC ($n=99$). (**b**) Primary tumor cohorts: MSK 2010 ($n=131$) and TCGA PRAD ($n=488$). (**c**) GEO series GSE126078 ($n=141$) and GSE25183 ($n=56$). The color and size of the circles represent the correlation coefficient between two variables. Pearson coefficients and associated *p-values* are indicated in Supplementary Table S8

role in mesenchymal gene expression [39]. These HARs are evolutionarily conserved sequences that exhibit a large number of nucleotide changes in the human lineage [38]. Most HARs are located in non-coding areas of the genome and are significantly enriched near genes involved in transcription, cell adhesion and development, and preferentially active in the brain [47]. Thus, it is now hypothesized that HARs may function as gene regulatory enhancers with particular implications for neurodevelopment [38, 48]. Interestingly, genomic sequences of tandem repeats containing highly clustered ZEB1 motifs have also recently been described within human epithelial genes [39]. To our knowledge, the enhancer region within the human *CDH2* gene described in this study appears to be the first example of a HAR containing repeats of HRE motifs. It now remains to be explored whether other HARs might include HRE clusters and identify associated cellular functions.

Also, further sequence analysis of this human *CDH2* enhancer revealed a significant variability in the AR binding sites, with the presence of 6 distinct HRE motifs within the cluster. We have previously described the capacity of AR and AR-V7 to bind to this enhancer region in prostate tumor cells [11]. In this study, we have shown that, similarly, the glucocorticoid receptor GR is also able to bind to this genomic region, and that the binding of these three nuclear receptors to the *CDH2* enhancer involves the preferential recognition of motifs 1 and 2 among the 6 HRE motifs present in the cluster. Our results show that, like AR-V7, GR can trigger the expression of N-cadherin in PCa cells in agreement with cumulative evidence of a crucial role of GR in tumor progression in prostate cancer [11, 41, 49, 50]. Moreover, as AR represses *CDH2* transcription, the presence of both AR and GR in PCa cells leads to a competition between these two factors to control *CDH2* expression. Our analysis of correlated gene expression using available transcriptome data from PCa samples further illustrated the interplay between AR and GR in this context.

In a hypothetical mechanism, in prostate epithelial cells AR binding to this enhancer locks transcriptional activation of the human *CDH2* gene. In the context of prostate cancer evolution with aberrant AR-V7 and GR expression, we can postulate that AR/AR-V7 or AR/GR competition displaces AR binding and unlocks *CDH2* expression.

Meanwhile, the presence of recognition motifs for different signaling pathways within the *CDH2* enhancer suggests a crosstalk between several transcription factors for the control of N-cadherin expression in humans. Querying public cistrome data highlighted AR and ZEB1 as key transcription factors linking this genomic region. Moreover, the overlap of recognition sequences for AR and ZEB1 within the enhancer strengthens the hypothesis of a competition between these two factors for N-cadherin regulation. Interestingly, the results of correlation analyses clearly support an opposite role of AR and ZEB1 on *CDH2* expression. Significantly, in addition to its known role as a transcriptional repressor of epithelial genes and master regulator of EMT, ZEB1 was also recently described as a transcriptional activator, thereby inducing the expression of mesenchymal cell-specific genes such as collagen or smooth muscle actin [51]. Our observations are consistent with these data. Although the precise mechanism involving ZEB1 in the control of *CDH2* remains to be understood, we should pay attention to the kinetics of ZEB1 expression in our model. Indeed, a late increase of ZEB1 is observed after expression of AR-V7 in our PCa cells [10], suggesting that after AR displacement from the enhancer, ZEB1 emerges to lock the mesenchymal state.

Despite the current lack of experimental evidence showing RBPJ bound to the enhancer, the presence of 24 RBPJ binding motifs in human *CDH2* enhancer and the results of our luciferase reporter assays suggest that the Notch pathway could possibly be involved in N-cadherin regulation. Furthermore, the ability of Notch signaling, an evolutionarily conserved pathway involved in the development and homeostasis of multiple tissues and organs via cell–cell contact, to induce partial EMT and maintain stemness has been described [52]. Consistently, exploration of transcriptome datasets showed a positive correlation of Notch activity with *CDH2* expression. Nevertheless, the role of Notch signaling in cell plasticity may be context-dependent, as it can also suppress neuroendocrine differentiation of PCa [53]. Additionally, a complex interplay between Notch, Notch ligands and EMT factors such as ZEB and Snail occurs during the EMT process [52]. Further studies are therefore needed to fully understand how these transcription factors act in the context of *CDH2* regulation.

N-cadherin is known to play a predominant role in cell migration during organogenesis and particularly during neurogenesis [19]. As modulation of cell adhesion and cell migration properties is fundamental for a proper embryonic development, N-cadherin expression level needs to be spatiotemporally fine-tuned. Interestingly, steroid hormones and their receptors as well as Notch and ZEB1 are known

as critical regulators of human neurogenesis [54–60]. Altogether, our work pinpointed a highly sophisticated enhancer region that gathers binding sites for these different transcription factors to probably enable an elaborated specific regulation of N-cadherin during neurogenesis in humans. It remains to know how these different transcription factors could intervene and what could be the link with human cortex development. Intriguingly, our data suggest that tumor cells may evolve to hijack this fine-tuned control of *CDH2* expression to gain cellular plasticity and move towards metastatic progression.

Abbreviations

AR	Androgen Receptor
ARE	Androgen Response Element
DHT	Dihydrotestosterone
EMP	Epithelial-Mesenchymal Plasticity
EMT	Epithelial-Mesenchymal Transition
GR	Glucocorticoid Receptor
GRE	Glucocorticoid Response Element
HRE	Hormone Response Element
MET	Mesenchymal-Epithelial Transition
NICD	Notch Intracellular Domain
PCa	Prostate Cancer
VNTR	Variable Number of Tandem Repeats

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00018-025-05725-6.

Acknowledgements Special thanks to Pluricell East platform (IGBMC, Illkirch) for their help in cell culture management, to the antibody engineering platform (IGBMC, Illkirch), and to Émilie Delorme, Silan Algul and Noémie Sacksteder for their remarkable contribution.

Author contributions JC, EE, CK and VZ designed the experiment. CK, VZ, AG and ZH performed the genomic analysis of the nine present-day humans. EE, SF and SA realized the molecular and cellular experiments. EE analyzed publicly available ChIP-seq data and realized correlation studies from patient cohorts. JC, EE, FC and BK cowrote the publication.

Funding This work was supported by the Association pour la Recherche sur les Tumeurs Prostatiques (ARTP), the association Alsace Contre le Cancer and the Agence Nationale de la Recherche (ANR-20-CE11-0004 ARCHAP).

Data availability The following previously published data sets were used: Sahu B (2013) FoxA1 specifies unique androgen and glucocorticoid receptor binding events in prostate cancer cells. NCBI Gene Expression Omnibus. ID GSE39879. https://www-ncbi-nlm-nih-gov.proxy.insermbiblio.inist.fr/geo/query/acc.cgi?acc=GSE39879 Takayama K (2015) Genome wide analysis of AR binding sites and histone modifications in prostate cancer. NCBI Gene Expression Omnibus. ID GSE62492. https://www-ncbi-nlm-nih-gov.proxy.insermbiblio.inist.fr/geo/query/acc.cgi?acc=GSE62492 Qiu X (2022) MYC drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets at androgen receptor targets. NCBI Gene Expression Omnibus. ID GSE163146. https://www-ncbi-nlm-nih-gov.proxy.insermbiblio.inist.fr/geo/query/acc.cgi?acc=GSE163146 Labrecque MP (2019) Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. NCBI Gene Expression Omnibus. ID GSE126078. https://www-ncbi-nlm-nih-gov.proxy.insermbiblio.inist.fr/geo/query/acc.cgi?acc=GSE126078 Prensner JR (2011) Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. NCBI Gene Expression Omnibus. ID GSE25183. https://www-ncbi-nlm-nih-gov.proxy.insermbiblio.inist.fr/geo/query/acc.cgi?acc=GSE25183 Taylor BS (2010) Comprehensive profiling of 218 prostate tumors (181 primaries, 37 metastases) and 14 prostate cancer cell lines and xenografts. MSKCC Prostate Oncogenome Project. cBioPortal for Cancer Genomics. ID Prostate Adenocarcinoma (MSK, Cancer Cell 2010). https://www.cbioportal.org/study/summary?id=prad_mskcc Abida W (2019) Whole exome sequencing of 444 castrate resistant prostate cancer tumor/normal pairs. SU2C-PCF Dream Team: Precision Therapy for Advanced Prostate Cancer. cBioPortal for Cancer Genomics. ID Metastatic Prostate Adenocarcinoma (SU2C/PCF Dream Team, PNAS 2019). https://www.cbioportal.org/study/summary?id=prad_su2c_2019 Kumar A (2016) Whole-exome sequencing, array comparative genomic hybridization (CGH) and RNA transcript profiling of 176 tumors from 63 patients with disseminated prostate cancer with matched normals. cBioPortal for Cancer Genomics. ID Prostate Adenocarcinoma (Fred Hutchinson CRC, Nat Med 2016). https://www.cbioportal.org/study/summary?id=prad_fhcrc Genomic Characterization of Metastatic Castration Resistant Prostate Cancer (2020). National Cancer Institute. Genomic Data Commons. ID WCDT-MCRPC. https://portal.gdc.cancer.gov/projects/WCDT-MCRPC National Institutes of Health, The Cancer Genome Atlas (TCGA): Prostate Adenocarcinoma (2019) National Cancer Institute. Genomic Data Commons. ID TCGA-PRAD. https://portal.gdc.cancer.gov/projects/TCGA-PRAD.

Declarations

Ethical approval No approval of research ethics committees was required to accomplish the goals of this study because experimental work was conducted with cell line models.

Consent to participate No consent to participate was required to accomplish the goals of this study because all patient datasets were acquired from published sources.

Consent for publication All authors read and approved the final manuscript.

Competing interests The authors declare no competing interest.

References

1. Nieto MA, Huang RY-J, Jackson RA, Thiery JP (2016) EMT: 2016. Cell 166:21–45. https://doi.org/10.1016/j.cell.2016.06.028
2. Zhang Y, Weinberg RA (2018) Epithelial-to-mesenchymal transition in cancer: complexity and opportunities. Front Med 12:361–373. https://doi.org/10.1007/s11684-018-0656-6
3. Lamouille S, Xu J, Derynck R (2014) Molecular mechanisms of epithelial–mesenchymal transition. Nat Rev Mol Cell Biol 15:178–196. https://doi.org/10.1038/nrm3758
4. Scott LE, Weinberg SH, Lemmon CA (2019) Mechanochemical signaling of the extracellular matrix in Epithelial-Mesenchymal transition. Front Cell Dev Biol 7:135. https://doi.org/10.3389/fcell.2019.00135
5. Tam WL, Weinberg RA (2013) The epigenetics of epithelial-mesenchymal plasticity in cancer. Nat Med 19:1438–1449. https://doi.org/10.1038/nm.3336
6. Shibue T, Weinberg RA (2017) EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat Rev Clin Oncol 14:611–629. https://doi.org/10.1038/nrclinonc.2017.44
7. Williams ED, Gao D, Redfern A, Thompson EW (2019) Controversies around epithelial–mesenchymal plasticity in cancer metastasis. Nat Rev Cancer 19:716–732. https://doi.org/10.1038/s41568-019-0213-x
8. Tanaka H, Kono E, Tran CP et al (2010) Monoclonal antibody targeting of N-cadherin inhibits prostate cancer growth, metastasis and castration resistance. Nat Med 16:1414–1420. https://doi.org/10.1038/nm.2236
9. Sun Y, Wang B-E, Leong KG et al (2012) Androgen deprivation causes epithelial-mesenchymal transition in the prostate: implications for androgen-deprivation therapy. Cancer Res 72:527–536. https://doi.org/10.1158/0008-5472.CAN-11-3004
10. Cottard F, Asmane I, Erdmann E et al (2013) Constitutively active androgen receptor variants upregulate expression of mesenchymal markers in prostate cancer cells. PLoS ONE 8:e63466. https://doi.org/10.1371/journal.pone.0063466
11. Cottard F, Madi-Berthélémy PO, Erdmann E et al (2017) Dual effects of constitutively active androgen receptor and full-length androgen receptor for N-cadherin regulation in prostate cancer. Oncotarget 8:72008–72020. https://doi.org/10.18632/oncotarget.18270
12. Miao L, Yang L, Li R et al (2017) Disrupting androgen receptor signaling induces Snail-Mediated Epithelial-Mesenchymal plasticity in prostate Cancer. Cancer Res 77:3101–3112. https://doi.org/10.1158/0008-5472.CAN-16-2169
13. Erdmann É, Ould Madi Berthélémy P, Cottard F et al (2022) Androgen receptor-mediated transcriptional repression targets cell plasticity in prostate cancer. Mol Oncol 16:2518–2536. https://doi.org/10.1002/1878-0261.13164
14. Germain P, Staels B, Dacquet C et al (2006) Overview of nomenclature of nuclear receptors. Pharmacol Rev 58:685–704. https://doi.org/10.1124/pr.58.4.2
15. Banerjee PP, Banerjee S, Brown TR, Zirkin BR (2018) Androgen action in prostate function and disease. Am J Clin Exp Urol 6:62–77
16. Gritsina G, Gao W-Q, Yu J (2019) Transcriptional repression by androgen receptor: roles in castration-resistant prostate cancer. Asian J Androl 21:215. https://doi.org/10.4103/aja.aja_19_19
17. De Velasco MA, Tanaka M, Yamamoto Y et al (2014) Androgen deprivation induces phenotypic plasticity and promotes resistance to molecular targeted therapy in a PTEN-deficient mouse model of prostate cancer. Carcinogenesis 35:2142–2153. https://doi.org/10.1093/carcin/bgu143
18. Nouri M, Ratther E, Stylianou N et al (2014) Androgen-targeted therapy-induced epithelial mesenchymal plasticity and neuroendocrine transdifferentiation in prostate cancer: an opportunity for intervention. Front Oncol 4:370. https://doi.org/10.3389/fonc.2014.00370
19. Derycke LDM, Bracke ME (2004) N-cadherin in the spotlight of cell-cell adhesion, differentiation, embryogenesis, invasion and signalling. Int J Dev Biol 48:463–476. https://doi.org/10.1387/ijdb.041793ld
20. Martinez-Garay I (2020) Molecular mechanisms of Cadherin function during cortical migration. Front Cell Dev Biol 8. https://doi.org/10.3389/fcell.2020.588152
21. Takayama K, Kaneshiro K, Tsutsumi S et al (2007) Identification of novel androgen response genes in prostate cancer cells by coupling chromatin Immunoprecipitation and genomic microarray analysis. Oncogene 26:4453–4463. https://doi.org/10.1038/sj.onc.1210229
22. Marcias G, Erdmann E, Lapouge G et al (2010) Identification of novel truncated androgen receptor (AR) mutants including unreported pre-mRNA splicing variants in the 22Rv1 hormone-refractory prostate cancer (PCa) cell line. Hum Mutat 31:74–80. https://doi.org/10.1002/humu.21138
23. Chung D-WD, Frausto RF, Chiu S et al (2016) Investigating the molecular basis of PPCD3: characterization of ZEB1 regulation of COL4A3 expression. Invest Ophthalmol Vis Sci 57:4136–4143. https://doi.org/10.1167/iovs.16-19533
24. Kõressaar T, Lepamets M, Kaplinski L et al (2018) Primer3_masker: integrating masking of template sequence with primer design software. Bioinformatics 34:1937–1938. https://doi.org/10.1093/bioinformatics/bty036
25. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R et al (2024) JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. Nucleic Acids Res 52:D174–D182. https://doi.org/10.1093/nar/gkad1059
26. Raney BJ, Barber GP, Benet-Pagès A et al (2024) The UCSC genome browser database: 2024 update. Nucleic Acids Res 52:D1082–D1088. https://doi.org/10.1093/nar/gkad987
27. Zheng R, Wan C, Mei S et al (2019) Cistrome data browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res 47:D729–D735. https://doi.org/10.1093/nar/gky1094
28. Harrison PW, Amode MR, Austine-Orimoloye O et al (2024) Ensembl 2024. Nucleic Acids Res 52:D891–D899. https://doi.org/10.1093/nar/gkad1049
29. Abida W, Cyrta J, Heller G et al (2019) Genomic correlates of clinical outcome in advanced prostate cancer. Proc Natl Acad Sci U S A 116:11428–11436. https://doi.org/10.1073/pnas.1902651116
30. Kumar A, Coleman I, Morrissey C et al (2016) Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med 22:369–378. https://doi.org/10.1038/nm.4053
31. Grossman RL, Heath AP, Ferretti V et al (2016) Toward a shared vision for Cancer genomic data. N Engl J Med 375:1109–1112. https://doi.org/10.1056/NEJMp1607591
32. Labrecque MP, Coleman IM, Brown LG et al (2019) Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. J Clin Invest 129:4492–4505. https://doi.org/10.1172/JCI128212
33. Taylor BS, Schultz N, Hieronymus H et al (2010) Integrative genomic profiling of human prostate Cancer. Cancer Cell 18:11–22. https://doi.org/10.1016/j.ccr.2010.05.026
34. Prensner JR, Iyer MK, Balbin OA et al (2011) Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1,

an unannotated LincRNA implicated in disease progression. Nat Biotechnol 29:742–749. https://doi.org/10.1038/nbt.1914
35. Nelson PS, Clegg N, Arnold H et al (2002) The program of androgen-responsive genes in neoplastic prostate epithelium. Proceedings of the National Academy of Sciences 99:11890–11895. https://doi.org/10.1073/pnas.182376299
36. Braune E-B, Geist F, Tang X et al (2024) Identification of a Notch transcriptomic signature for breast cancer. Breast Cancer Res 26:4. https://doi.org/10.1186/s13058-023-01757-7
37. Oliphant TE (2007) Python for scientific computing. Comput Sci Engg 9:10–20. https://doi.org/10.1109/MCSE.2007.58
38. Whalen S, Pollard KS (2022) Enhancer function and evolutionary roles of human accelerated regions. Annu Rev Genet 56:423–439. https://doi.org/10.1146/annurev-genet-071819-103933
39. Balestrieri C, Alfarano G, Milan M et al (2018) Co-optation of tandem DNA repeats for the maintenance of mesenchymal identity. Cell 173:1150–1164e14. https://doi.org/10.1016/j.cell.2018.03.081
40. Claessens F, Denayer S, Van Tilborgh N et al (2008) Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. Nucl Recept Signal 6. https://doi.org/10.1621/nrs.06008
41. Arora VK, Schenkein E, Murali R et al (2013) Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor Blockade. Cell 155:1309–1322. https://doi.org/10.1016/j.cell.2013.11.012
42. Sahu B, Laakso M, Pihlajamaa P et al (2013) FoxA1 specifies unique androgen and glucocorticoid receptor binding events in prostate cancer cells. Cancer Res 73:1570–1580. https://doi.org/10.1158/0008-5472.CAN-12-2350
43. Takayama K, Suzuki T, Tsutsumi S et al (2014) RUNX1, an androgen- and EZH2-regulated gene, has differential roles in AR-dependent and -independent prostate cancer. Oncotarget 6:2263–2276. https://doi.org/10.18632/oncotarget.2949
44. Qiu X, Boufaied N, Hallal T et al (2022) MYC drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets. Nat Commun 13:2559. https://doi.org/10.1038/s41467-022-30257-z
45. Mrozik KM, Blaschuk OW, Cheong CM et al (2018) N-cadherin in cancer metastasis, its emerging role in haematological malignancies and potential as a therapeutic target in cancer. BMC Cancer 18:939. https://doi.org/10.1186/s12885-018-4845-0
46. Cao Z-Q, Wang Z, Leng P (2019) Aberrant N-cadherin expression in cancer. Biomed Pharmacother 118:109320. https://doi.org/10.1016/j.biopha.2019.109320
47. Guardiola-Ripoll M, Fatjó-Vilas M (2023) A systematic review of the human accelerated regions in schizophrenia and related disorders: where the evolutionary and neurodevelopmental hypotheses converge. Int J Mol Sci 24:3597. https://doi.org/10.3390/ijms24043597
48. Capra JA, Erwin GD, McKinsey G et al (2013) Many human accelerated regions are developmental enhancers. Philos Trans R Soc Lond B Biol Sci 368:20130025. https://doi.org/10.1098/rstb.2013.0025
49. Virtanen V, Paunu K, Kukkula A et al (2023) Glucocorticoid receptor-induced non-muscle Caldesmon regulates metastasis in castration-resistant prostate cancer. Oncogenesis 12:1–12. https://doi.org/10.1038/s41389-023-00485-z
50. Eigentler A, Handle F, Schanung S et al (2024) Glucocorticoid treatment influences prostate cancer cell growth and the tumor microenvironment via altered glucocorticoid receptor signaling in prostate fibroblasts. Oncogene 43:235–247. https://doi.org/10.1038/s41388-023-02901-5
51. Debnath P, Huirem RS, Dutta P, Palchaudhuri S (2021) Epithelial–mesenchymal transition and its transcription factors. Biosci Rep 42:BSR20211754. https://doi.org/10.1042/BSR20211754
52. Jolly MK, Boareto M, Huang B et al (2015) Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. Front Oncol 5:155. https://doi.org/10.3389/fonc.2015.00155
53. Ku S-Y, Wang Y, Garcia MM et al (2024) Notch signaling suppresses neuroendocrine differentiation and alters the immune microenvironment in advanced prostate cancer. J Clin Investig 134:e175217. https://doi.org/10.1172/JCI175217
54. Liston C, Gan W-B (2011) Glucocorticoids are critical regulators of dendritic spine development and plasticity in vivo. Proc Natl Acad Sci 108:16074–16079. https://doi.org/10.1073/pnas.1110444108
55. Kuwahara N, Nicholson K, Isaacs L, MacLusky NJ (2021) Androgen effects on neural plasticity. Androgens: Clin Res Ther 2:216–230. https://doi.org/10.1089/andro.2021.0022
56. Eachus H, Ryu S (2024) Glucocorticoid effects on the brain: from adaptive developmental plasticity to allostatic overload. J Exp Biol 227:jeb246128. https://doi.org/10.1242/jeb.246128
57. Krontira AC, Cruceanu C, Dony L et al (2024) Human cortical neurogenesis is altered via glucocorticoid-mediated regulation of ZBTB16 expression. Neuron 112:1426–1443e11. https://doi.org/10.1016/j.neuron.2024.02.005
58. Kelava I, Chiaradia I, Pellegrini L et al (2022) Androgens increase excitatory neurogenic potential in human brain organoids. Nature 602:112–116. https://doi.org/10.1038/s41586-021-04330-4
59. Fiddes IT, Lodewijk GA, Mooring M et al (2018) Human-Specific NOTCH2NL genes affect Notch signaling and cortical neurogenesis. Cell 173:1356–1369e22. https://doi.org/10.1016/j.cell.2018.03.051
60. Noisa P, Lund C, Kanduri K et al (2014) Notch signaling regulates the differentiation of neural crest from human pluripotent stem cells. Development 141:e1106–e1106. https://doi.org/10.1242/dev.112482