BBA - Molecular Cell Research 1871 (2024) 119658

Contents lists available at ScienceDirect

BBA - Molecular Cell Research

journal homepage: www.elsevier.com/locate/bbamcr

Research paper

Loss of the epithelial marker CDX1 predicts poor prognosis in early-stage CRC patients

Laura Solé [a,1], Teresa Lobo-Jarne [a,1], Júlia-Jié Cabré-Romans [a], Antón González [b], Lierni Fernández [b], Laura Marruecos [a,c], Marta Guix [d], Miriam Cuatrecasas [e,f,g], Sandra López [e,f], Beatriz Bellosillo [b], Ferran Torres [h], Mar Iglesias [b,i], Anna Bigas [a,i,j], Lluís Espinosa [a,i,*]

[a] Program in Cancer Research, Institut Hospital del Mar d'Investigacions Mèdiques, Barcelona, Spain
[b] Pathology Department, Hospital del Mar, Barcelona, Spain
[c] The Walter and Eliza Hall Institute, Melbourne, Australia
[d] Oncology Department, Hospital del Mar, Barcelona, Spain
[e] Pathology Department, Centre of Biomedical Diagnosis (CDB), Hospital Clinic, 08036 Barcelona, Spain
[f] Centro de Investigacion Biomedica en Red en Enfermedades Hepáticas y Digestivas (CIBEREHD), 28029 Madrid, Spain
[g] Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
[h] Biostatistics Unit, Medical School, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain
[i] Centro de Investigacion Biomedica en Red Cancer (CIBERONC), Madrid, Spain
[j] Josep Carreras Leukemia Research Institute, Barcelona, Spain

ARTICLE INFO

Keywords:
Colorectal cancer
Fetal conversion
CDX1
Patient prognosis

ABSTRACT

Background: We have previously shown that non-curative chemotherapy imposes fetal conversion and high metastatic capacity to cancer cells. From the set of genes differentially expressed in Chemotherapy Resistant Cells, we obtained a characteristic fetal intestinal cell signature that is present in a group of untreated tumors and is sufficient to predict patient prognosis. A feature of this fetal signature is the loss of CDX1.
Methods: We have analyzed transcriptomic data in public datasets and performed immunohistochemistry analysis of paraffin embedded tumor samples from two cohorts of colorectal cancer patients.
Results: We demonstrated that low levels of *CDX1* are sufficient to identify patients with poorest outcome at the early tumor stages II and III. Presence tumor areas that are negative for CDX1 staining in stage I cancers is associated with tumor relapse.
Conclusions: Our results reveal the actual possibility of incorporating CDX1 immunostaining as a valuable biomarker for CRC patients.

1. Background

Colorectal cancer (CRC) is the third leading cause of death by cancer (reviewed in [1]). Although immunotherapy is acquiring increasing interest specifically in tumors with microsatellite instability [2], adjuvant chemotherapy after surgery is still the standard treatment in resected CRC. Despite an adequate therapeutic intervention, 35 % of all CRC patients (in average) relapse and/or metastasize and ultimately die (source; https://www.cancer.net/cancer-types/colorectal-cancer/statistics). The identification of markers that predict tumor relapse, other than tumor stage, and strategies to combat cancer cells surviving chemotherapy is an urgent clinical and social need.

In this direction, the establishment of a consensus molecular subtype classification of CRC tumors with implications in patient prognosis has represented an important advance towards understanding the high heterogeneity of the disease [3]. More recently, the identification of a single CRC cell population, characterized by the expression of a 99-genes signature, with high risk to metastasize has increased the options for early detection and eradication of tumor metastasis even at the time of primary tumor removal [4]. However, the use of extensive molecular signatures still represents a significant difficulty in the real clinical practice.

* Corresponding author at: Program in Cancer Research, Institut Hospital del Mar d'Investigacions Mèdiques, Barcelona, Spain.
E-mail address: lespinosa@imim.es (L. Espinosa).
[1] These authors contributed equally.

https://doi.org/10.1016/j.bbamcr.2024.119658
Received 17 October 2023; Received in revised form 18 December 2023; Accepted 3 January 2024
Available online 10 January 2024

Fetal/embryonic conversion of intestinal tumor cells is emerging as one of the main mechanisms used by cancer cells to persist upon chemotherapy treatment [5–8], and we have identified a specific 36-gene fetal signature composed of 28 upregulated and 8 downregulated genes (36FS) that is consistently acquired by therapy resistant cells. This signature is also present in about 20 % of naïve CRC tumors and is predictive of poor patient prognosis [9]. Whereas genes upregulated in 36FS include genes that are expressed in the stromal component, genes downregulated are mostly epithelial and include factors regulating intestinal identity such as *MYB* [10,11] and *CDX1* [12,13].

CDX1 and the closely-related factor CDX2 are members of the caudal homeobox-like family of proteins. *CDX1* and *CDX2* are expressed in a manner similar to HOX genes and are functionally exchangeable in the regulation of vertebral patterning [14]. In the intestine, CDX factors bind the chromatin to determine the timing of enhancer activation in postnatal duodenum. However, CDX1 and CDX2 function is not totally redundant and while *CDX2* knockout mice display peri-implantation lethality, *CDX1*-null mice are viable and fertile, and show no overt intestinal defects. Work from different groups indicate some gene selectivity for CDX-dependent gene transcription. For example, binding of CDX2 to two different enhancer elements regulates *HOXC8* expression, which is cell type specific [15]. Otherwise, CDX1 binds to TCF factors to allow full transcriptional activation of the Wnt target *AXIN2* [16]. CDX1 function is not restricted to the intestine but it also refines positional identity of the vertebrate hindbrain by directly repressing MAFB transcription [17].

Multiple reports indicated that loss of CDX2 is robustly associated with poor prognosis in CRC tumors [18–20]. Similarly, CDX1 inactivation has been linked to the undifferentiated histology of CRC tumors [21]. Re-expression of CDX1 or CDX2 in poorly differentiated CRC cells reduces cell growth and metastasis [22,23], and overexpression of miR-215, a direct transcriptional target of CDX1, decreases cell clonogenicity [24]. However, most of the impact of CDX1 loss in tumor malignancy has not definitely been established and it was suggested mostly in the context of the double CDX2 and CDX1 deficiency [25–27]. Further increasing uncertainty about the role of CDX1 in cancer, the human CDX1 gene is located on chromosome 5q, near the adenomatous polyposis coli (APC) gene, thus genetic rearrangements that frequently occur at that locus are commonly, but not always, involving both APC and CDX1 genes [28].

We have here investigated the possibility of using CDX1 mRNA or protein as prognostic biomarker in CRC. We demonstrate that loss or reduction in *CDX1* levels predicts poor prognosis at early tumor stages.

2. Methods

2.1. Patient gene expression datasets analysis

The transcriptome of tumor samples from GSE39582 [29] and GSE14333 [30] obtained by microarrays was downloaded from the Gene Expression Ommnibus (GEO) [31] and analyzed with the affy R package. Transcriptomic and available clinical data from COAD (colon) and rectum (READ) TCGA dataset was downloaded using TCGA Biolinks R package (v2.24.1) [32] as transcriptomic profiling category, gene expression quantification data type and STAR-Counts workflow type. Data from both COAD and READ datasets were merged, normalized and transformed using edgeR R package (v.3.38.1) [33].

Marisa (GSE39582) data set includes expression and clinical data for 566 patients with CRC, Jorissen (GSE14333) data set includes 226 CRC patients and TCGA dataset includes 539 CRC patients.

2.2. Statistical analysis

Survival analyses for relapse were described using Kaplan-Meier curves, the log-rank test was used to compare survival curves and the Cox proportional hazards models to estimate hazard ratios (HR) and 95%CIs. A standard log-rank test was applied to assess significance between groups. All the survival analyses and graphs were performed in R using the survival (v.3.3–1) and survimer packages and a p-value < 0.05 was considered statistically significant. Patients were classified according to the optimized cut-point of the designated genes, unless otherwise stated, and the optimized cutpoint was assessed using the surv_cutpoint function from the survminer R package (v.0.4.9).

2.3. Human colorectal tumors

Formalin-fixed, paraffin-embedded tissue blocks of gastrointestinal tumor samples were obtained from Parc de Salut Mar Biobank (MarBiobank, Barcelona) (discovery cohort and validation cohort 1) and the tumor tissue Biobank at Hospital Clinic-IDIBAPS (Barcelona) (validation cohort 2). Patient samples were selected based on clinical diagnosis of CRC without additional bias. For stage I samples, we specifically collected relapse cases and a comparable number of paired non-relapse cases (matched by age, gender and histopathologic parameters). Samples were retrieved under informed consent and approval of the Clinical Research Ethics Committee-Parc de Salut Mar (CEIC-PSMAR #2019/8595/I) according to Spanish ethical regulations and the guidelines of the Declaration of Helsinki. Patient identity for pathological specimens remained anonymous in the context of this study. Patient data was collected and treatment regimens were standard and adjusted to patient and tumor characteristics (Supplementary Tables S1, S2, S3 and S5).

2.4. Immunohistochemical staining

Tumor tissues were obtained after conventional fixation in 4 % formaldehyde overnight at room temperature, and subsequently paraffin-embedding. Paraffin-embedded tumor sections of 4 μm were deparaffinized, rehydrated and endogenous peroxidase activity was quenched (20 min, 1.5 % H_2O_2). Citrate-based antigen retrieval was used. Primary antibodies were diluted in PBS containing 0.05 % BSA (CDX1: 1:200, [Sigma, HPA055196, rabbit polyclonal] and CDX2: 1:5000 [Abcam, ab76541, rabbit monoclonal]), incubated overnight at 4 °C, and developed with the Envision+ System HRP Labeled Polymer anti-Rabbit or anti-Mouse and 3,3′-diaminobenzidine (DAB). Samples were mounted in DPX and the immunostaining was assessed by pathologists AG, LF and MI, using an Olympus BX61 microscope. H.

3. Results

3.1. Reduced CDX1 mRNA levels provide prognosis information in human CRC regardless of CDX2

CDX1 transcription factor regulates intestinal-specific gene expression and it is functionally related with the CRC biomarker CDX2 [18–20]. First, we performed a comparative analysis of the impact of *CDX1* and *CDX2* levels in patient prognosis in the Marisa, TCGA and Jorissen datasets. We found that both *CDX1* (Fig. 1A–C) and *CDX2* levels (Fig. 1D–F) were correlated with patient prognosis at all stages, but also when considering patients at stages II and II + III (stages B + C in Jorissen) (Fig. S1A–F) or the poorest prognosis Stage III + IV tumors (Fig. S1G–H).

We speculated that correlated expression of *CDX1* and *CDX2* in tumors [34,35], which were present in the Marisa, TCGA and Jorissen datasets (Fig. S1I–K) may affect the interpretation of our results. Importantly, in the TCGA and Jorissen datasets, low CDX1 levels identified patients with poor prognosis even in the presence of high CDX2 levels (compare red and green lanes), although tumors expressing low levels of both CDX1 and CDX2 imposed the worst prognosis (blue lanes) (Fig. 1G–I). In the Marisa dataset, that included only one patient with low *CDX1* and high *CDX2* levels, the CDX1-low and CDX2-low was again the group with poorest outcome (Fig. 1I). Moreover, in the poorest outcome Stage III + IV tumors CDX1 low levels were still a prognosis

factor.

3.2. CDX1 and CDX2 are independent prognosis factors in CRC independent of tumor stage, gender and oncogene mutations

To better characterize the prognosis value of CDX1 and CDX2, independently, we performed multivariate analysis including well-established prognostic factors such as *KRAS* or *BRAF* mutations, *TP53* status and tumor stage. We found that *CDX1* and *CDX2* were independent predictors of outcome in CRC patients, with CDX1 being significant in the three cohorts analyzed and CDX2 being significant only in the TCGA cohort (Fig. 2A–C). Mutations in *KRAS* and distal localization of the tumor were also identified as poor prognosis factors in the Marisa dataset (Fig. 2A).The lower prognostic value of *CDX2* as an independent biomarker in CRC could be related to its association with the subset of poor prognosis *BRAF* mutant tumors [36–39].

3.3. Detection of CDX1 protein from paraffin-embedded CRC samples informs on patient prognosis

To bring our study closer to clinical routine, we analyzed whether immunohistochemistry (IHC) analysis of CDX1 protein was efficient in predicting patient prognosis in a discovery cohort ($n = 39$) of paraffin-embedded CRC samples. We initially determined the intensity levels of CDX1 (from 0 = negative to 3 = high, comparable to normal tissue) (Fig. 3A) and the tumor area positive to generate an H-score (intensity = 0–3 * % positive tumor area; 0–100). We then categorized tumors in quartiles (Q1-Q4) depending on the H-score (H-score 0; H-score 1–75 = Q1; H-score 76–150 = Q2; H-score 151–225 = Q3; H-score 226–300 = Q4) (Supplementary Table S1). Analysis of patient survival according to the H-score demonstrated that H-scores of 0 and >0 (Q1 to Q4) represented the best cutoff for patient stratification (Fig. S2A), with loss of CDX1 protein being suggestive of reduced overall survival (Fig. 3B). Disease-free survival data from this cohort of patients was not available. We then performed IHC analysis of CDX1 from two different CRC validation cohorts, one containing 191 samples obtained from MarBiobank (Supplementary Table S2) and the other containing 125 samples from

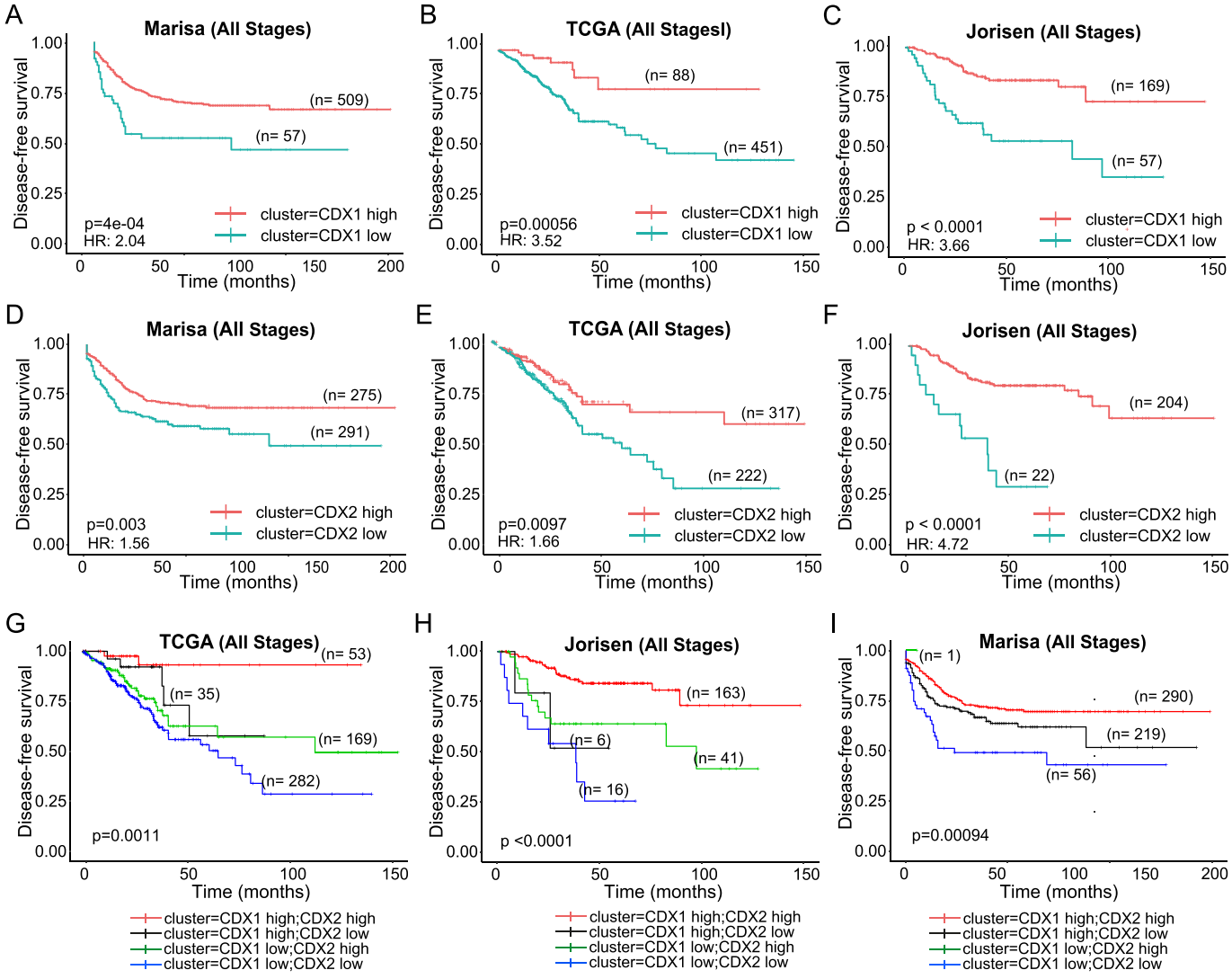


Fig. 1. Reduced *CDX1* mRNA levels provide prognosis information in human CRC regardless of *CDX2*. (A–F) Kaplan–Meier representation of disease-free survival over time for patients at all stages from Marisa (A, D), TCGA (B, E) and Jorissen (C, F) colorectal cancer databases, according to *CDX2* or *CDX1* expression levels as indicated. (G-I) Kaplan–Meier representation of disease-free survival over time for patients from TCGA (G), Jorissen (H) and Marisa (I) colorectal cancer databases, according to the indicated combination of CDX1 and CDX2 expression levels. Hazard ratios (HRs) and 95 % CIs are estimated from Cox proportional hazards models, and *p*-values are derived from the log-rank test. *$P < 0.05$; **$P < 0.01$. ***$P < 0.001$. All patients were considered.

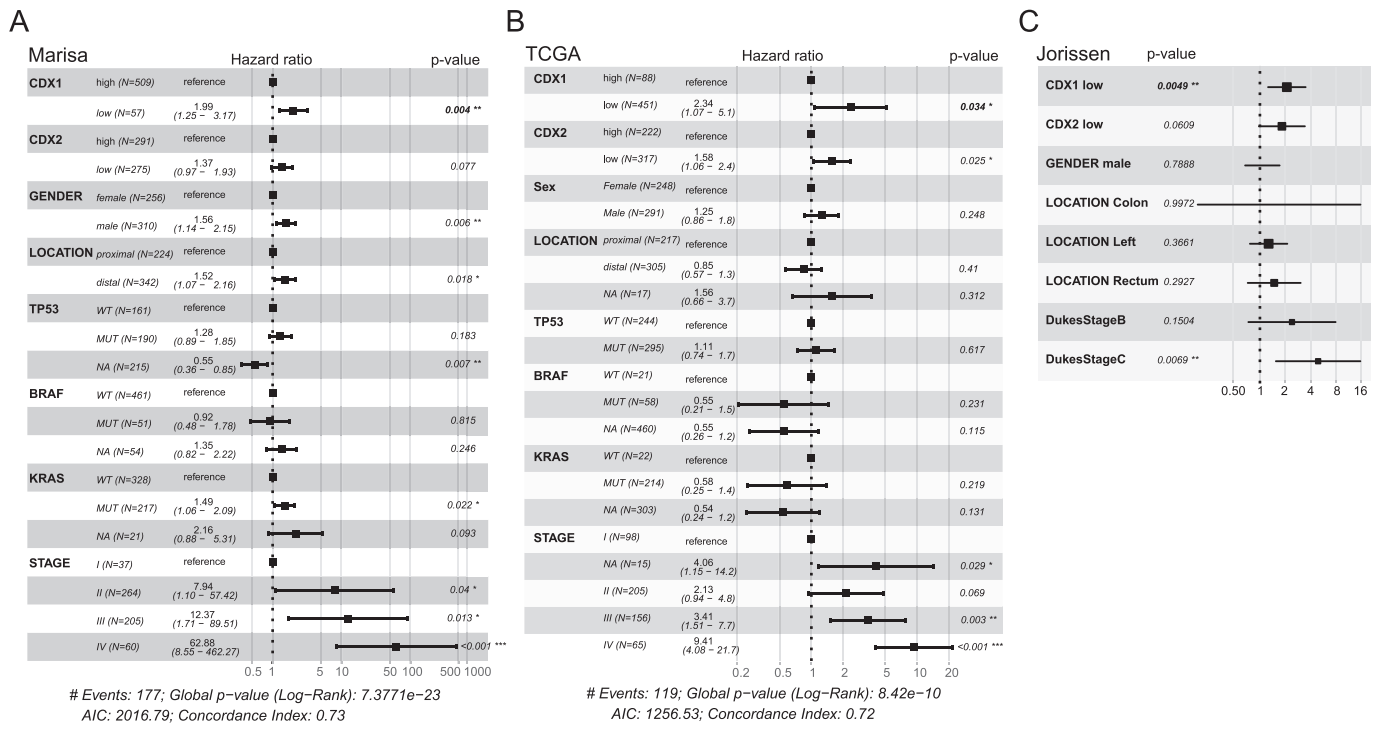


Fig. 2. CDX1 and CDX2 are independent prognosis factors in CRC independent of tumor stage, gender and oncogene mutations. Forest plot for the multivariate Cox proportional hazard regression model showing hazard ratio estimates and 95 % confidence intervals from the *CDX1* and *CDX2* genes in in Marisa (A), TCGA (B) and Jorissen (C) databases. Other parameters included in the analysis are indicated. WT: wildtype, MT: mutated, NA: not ascribed.

the Biobank from Hospital Clinic in Barcelona (Supplementary Table S3). When analyzed independently, CDX1-negative tumors stratified patients with lower disease-free survival in both cohorts with different statistical power, and marginal impact in overall survival (Fig. 3C). Pooling both cohorts together increased the significance of the results and demonstrated that loss of CDX1 is predictive of tumor relapse, although it did not reach statistical significance in relation with overall survival (Fig. 3D). In the same CRC cohort, we found that loss of CDX2 identified patients with very low overall and disease-free survival, however the number of patients with CDX2-negative tumors was extremely low (Fig. S2B and S2C).

3.4. Presence of CDX1 negative tumor areas in stage I CRC is linked to relapse

Stage I CRC tumors are considered to have a good prognosis and most of them are treated exclusively with surgery. However, a very small numbers of these patients relapse and there are no biomarkers to anticipate this event. We studied whether CDX1 and CDX2 levels predicted patient outcome at specific tumor stages, and in particular at stage I. In the 3 cohorts analyzed, low levels of CDX1, but not CDX2, characterize the poorest prognosis stage I (stage A in Jorissen) patients (Supplementary Table S4). We visualized these results in the Kaplan Meier curves of TCGA and Jorissen cohorts (Fig. 4A and B), since the Marisa cohort only included 4 CDX1-high patients. Addition of CDX2 levels in the analysis of stage I tumors did nor add any benefit in terms of significance (Fig. 4C and D). However, the significance and hazard ratio (HR) of CDX1 (Fig. 4E and F) or CDX1 plus CDX2 (Fig. 4G and H) were highly improved when considering stages I + II (or A + B for Jorissen) patients.

We investigated the possibility that IHC analysis of CDX1 provides information on the outcome of patients with stage I tumors. For this analysis, we specifically selected 11 samples of stage I tumors that relapsed after surgery and a comparable number of non-relapsed tumors (Supplementary Table S5). CDX1 expression was primarily preserved in all stage I tumors at different levels. Unexpectedly, we detected the presence in some of the tumors of circumscribed CDX1-negative areas with no significant morphological differences with the adjacent positive areas (Fig. 4I). CDX1-negative areas were almost exclusively present in samples from tumors that subsequently relapsed (5 of 11 relapsed tumors compared with 1 of 11 non-relapsed) (Fig. 4J). We then performed CDX2 IHC on these same samples and found that all but 3 of the stage I tumors analyzed were CDX2 positive with the presence of scattered negative cells (Fig. S3A and Supplementary Table S5). Similar to that found in the majority of stage I tumor, we detected loss of nuclear CDX2 in few cells of the CDX1-negative areas (Fig. S3B).

These results indicate that loss of CDX1, either complete or focal, could represent a powerful biomarker for high-risk stage I tumors at the time of diagnosis.

4. Discussion

We have now identified CDX1, one of the genes downregulated in the poor-prognosis fetal-type tumors [6], as a new biomarker for patient stratification at early-stage CRC. This finding has extremely important clinical implications since there are few (if any) molecular biomarkers that are translated to clinical practice to predict relapse at specific tumor stages and in particular at Stage I (reviewed in [40]). From a clinical perspective, the relevance of our findings relies in the identification of a useful biomarker that could be easily implemented in the pathology departments at the time of diagnosis. It is important to remark that CDX1-negative areas that are present in stage I tumors with higher probability to relapse are indistinguishable from the CDX1-positive areas at histopathological level, thus making essential IHC evaluation of CDX1. However, the number of samples analyzed in this study is relatively low and a more extensive analysis of this subset of patients with good prognosis is clearly required to obtain more conclusive results. Association of CDX1 loss with intestinal development, and CRC initiation and progression has previously been described, but mainly associated to CDX2 loss [12,13,26,28,34,35,41]. Although we are aware

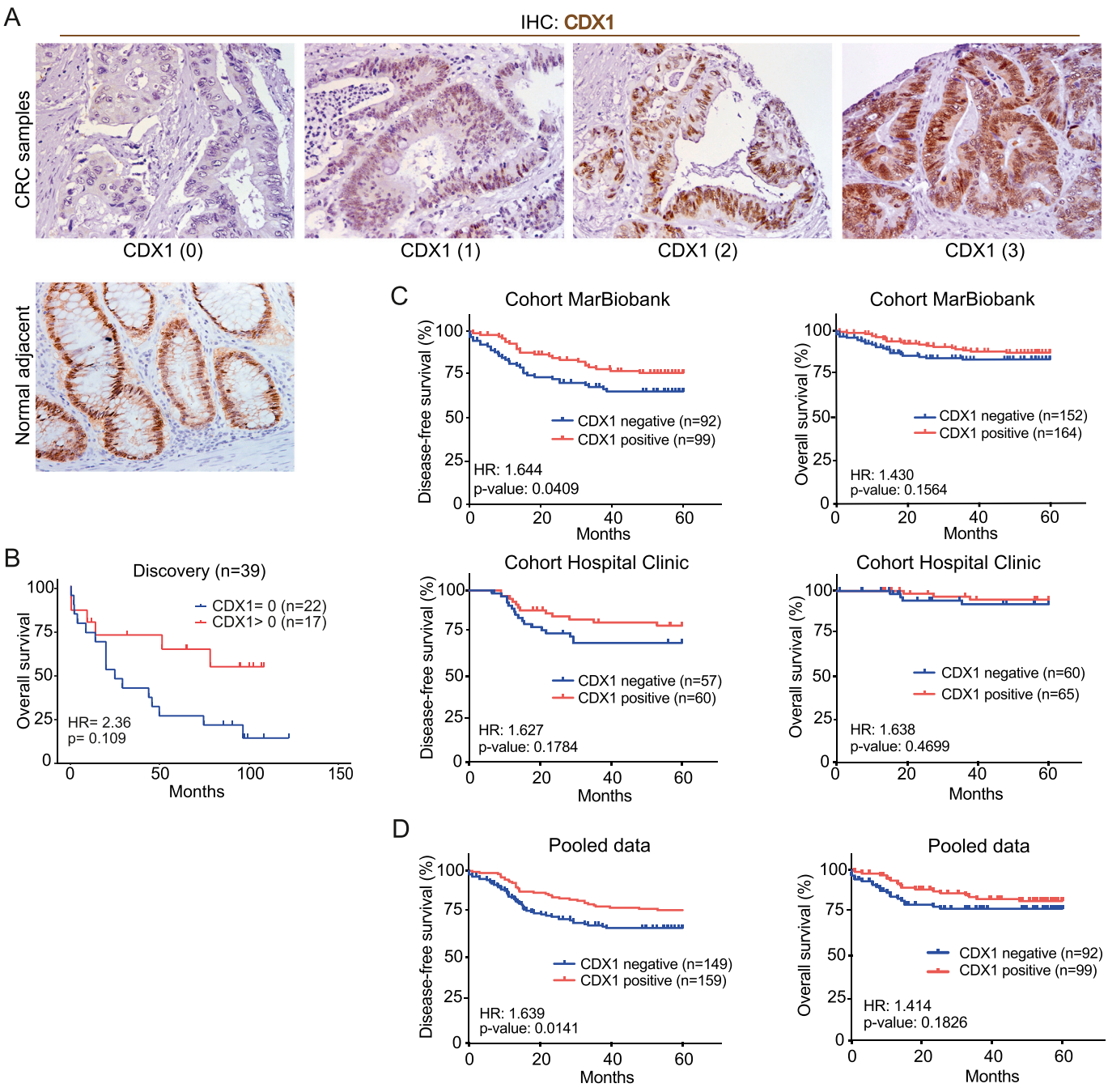
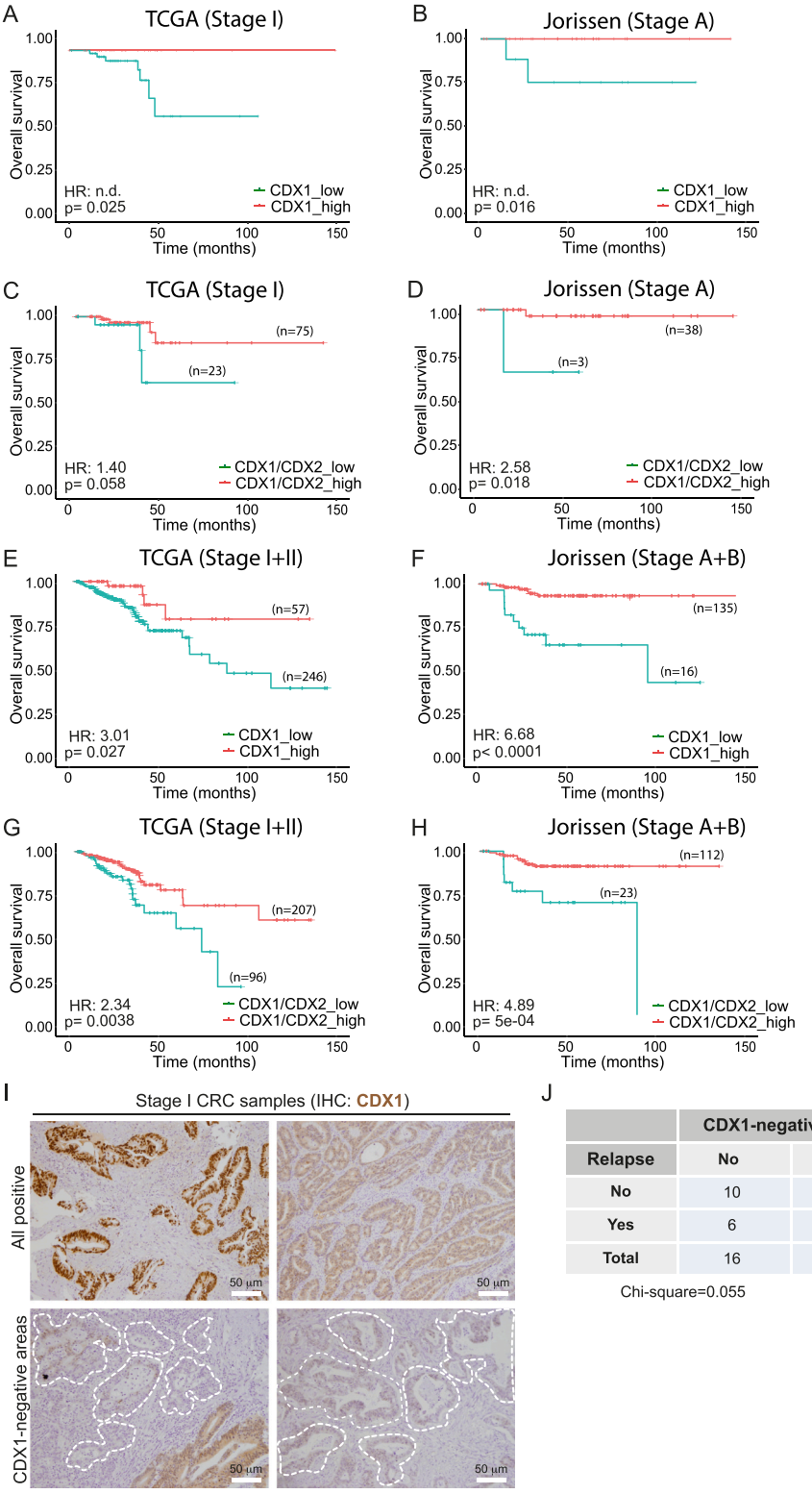


Fig. 3. Detection of CDX1 protein from paraffin-embedded CRC samples informs on patient prognosis in three different cohorts analyzed. (A) Representative images of CDX1 classified according to protein intensity levels (0–3) in the TMA of CRC tumors from the discovery cohort. CDX1 staining of a normal adjacent colonic tissue is shown. (B) Kaplan-Meier representation of overall survival of CRC patients from the discovery cohort classified according to presence (>0) or absence (=0) of nuclear CDX1. (C–D) Kaplan-Meier representation of disease-free and overall survival of CRC patients from the MarBiobank and Hospital Clinic validation cohorts (C) and pooled data from the two cohorts (D) classified according to presence (>0) or absence (=0) of nuclear CDX1. Hazard ratios (HRs) and 95 % CIs are estimated from Cox proportional hazards models, and p-values are derived from the log-rank test. *P < 0.05; **P < 0.01. ***P < 0.001. All patients were considered.

that the predictive capacity of CDX1 is significantly lower than CDX2, in contrast, it incorporates a higher number of patients, which is not less relevant. Moreover, CDX2 does not inform on patient outcome when determined at Stage I, a group of patients that would crucially benefit from a simple test capable to refine their diagnosis and therapeutic management. Since, the CDX1 gene is located near APC in the human genome, whether low CDX1 levels are linked to genetic APC deletion in a percent of cases, remains unstudied. However, the high correlation between CDX1 (cytogenetic band 5q32) and CDX2 (cytogenetic band 13q12.2) expression in the tumors suggest some kind of transcriptional coregulation more than a specific deletion of the CDX1 locus.

Still, we are aware that the prognosis capacity of the entire fetal signature previously described [6] is significantly superior than CDX1 or any other fetal markers alone. For this reason, we are currently working on the incorporation of additional fetal markers to perform multiplex IHC analysis on paraffin-embedded samples. In addition, because several of the genes and proteins that are over-represented in the fetal-type tumor cells are also expressed in the cancer-associated fibroblasts and macrophages (not depicted), we are studying the possibility to setup a qPRC test from liquid biopsy, thus avoiding the artifacts derived from



J

	CDX1-negative areas		
Relapse	**No**	**Yes**	**Total**
No	10	1	11
Yes	6	5	11
Total	16	6	22

Chi-square=0.055

Fig. 4. Low levels of CDX1 RNA or the presence of CDX1 negative tumor areas in stage I CRC are linked to relapse. (A–B) Kaplan–Meier representation of overall survival over time for patients from TCGA stage I (A) and Jorissen stage A (B) according to *CDX1* expression levels. (C–D) Kaplan–Meier representation of overall survival over time for patients from TCGA stage I (C) and Jorissen stage A (D) according to *CDX1 and CDX2* expression levels. (E–F) Kaplan–Meier representation of overall survival over time for patients from TCGA stage I + II (E) and Jorissen stage A + B (F) according to *CDX1* expression levels. (G-H) Kaplan–Meier representation of overall survival over time for patients from TCGA stage I + II (G) and Jorissen stage A + B (H) according to *CDX1 and CDX2* expression levels. (I) Representative images of CDX1 classification according to the presence or absence of CDX1-negative areas in stage I CRC specimens. CDX1-negative areas are shown delimited by dashed white lines. (J) Number of patients displaying CDX1-negative areas inside the stage I CRC tumors and their relapse status. Different subgroups were compared with Pearson's chi-square test. In the Kaplan-Meier analysis, p-values are derived from the log-rank test.

stromal-derived transcripts.

5. Conclusions

Together, our results identified CDX1 as a potential prognostic biomarker at difficult-to-predict stages I and II + III. We propose that CDX1 loss should be assessed in stage I tumors at diagnosis to identify patients who are candidates for a closer follow-up. Implementation of CDX1 immunostaining could have a relevant clinical impact in therapeutical management of early-stage CRC patients in the near future.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbamcr.2024.119658.

Abbreviations

EMT	epithelial-to-mesenchymal transition
CRC	colorectal cancer
APC	adenomatous polyposis coli
DAB	3,3′-diaminobenzidine

Ethics approval and consent to participate

Samples from patients were kindly provided by MarBiobank and the biobank of Hospital Clinic de Barcelona integrated in the Spanish Hospital Biobanks Network (RetBioH; www.redbiobancos.es). Informed consent was obtained from all participants and protocols were approved by the institutional ethical committee of Hospital del Mar (Ethical Committee approval #2019/8595/I).

Consent for publication

Not applicable.

Funding

This work was funded by grants from Instituto de Salud Carlos III cofunded by the EU (PI22/00069 and DTS23/00005), Generalitat de Catalunya 2021 SGR 00039 and Instituto de Salud Carlos III-Fondo Europeo de Desarrollo Regional (CIBERONC; CB16/12/00244 and CB16/12/00241). TL-J is a recipient of the AECC postdoctoral grant POSTD21975. LS is a postdoctoral researcher supported by AGAUR (Programa Investigo 2022 2022; INV-1 00005/100005ID5).

CRediT autorship contribution statement

TL-J, LS, LM and J-JC-R performed experiments and analyzed data. AG, MG, BB, LF, SL, MC and MI selected the tumor samples, obtained patient-derived material and performed the clinicopathological characterization of human tumors. MG provided clinical advice and revised patient data. TL-J and FT performed statistical analysis of data. TL-J, LS, AB and LE design and supervised the project, prepared figures and wrote the manuscript.

All authors revised the final text and figures.

CRediT authorship contribution statement

Laura Solé: Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – review & editing. **Teresa Lobo-Jarne:** Data curation, Formal analysis, Investigation, Methodology. **Júlia-Jié Cabré-Romans:** Investigation, Methodology. **Antón González:** Data curation. **Lierni Fernandez:** Investigation, Methodology. **Laura Marruecos:** Investigation, Methodology. **Marta Guix:** Data curation, Formal analysis, Supervision. **Miriam Cuatrecasas:** Data curation, Formal analysis. **Sandra López:** Investigation, Methodology. **Beatriz Bellosillo:** Data curation, Formal analysis, Investigation, Methodology. **Ferran Torres:** Data curation, Formal analysis. **Mar Iglesias:** Formal analysis, Investigation, Supervision. **Anna Bigas:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Lluís Espinosa:** Conceptualization, Funding acquisition, Investigation, Resources, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Public datasets used in this study are referenced in the text. Data from our in-house cohort of samples are included in the supplementary tables.

Acknowledgments

We want to thank the Bigas' and Espinosa's lab members for constructive discussions and suggestions and technical support. We thank our patients for their generosity and to MarBiobank integrated in the Spanish Hospital Biobanks Network (RetBioH; www.redbiobancos.es).

References

[1] F. Baidoun, et al., Colorectal cancer epidemiology: recent trends and impact on outcomes, Curr. Drug Targets 22 (2021) 998–1009.
[2] A. Fan, et al., Immunotherapy in colorectal cancer: current achievements and future perspective, Int. J. Biol. Sci. 17 (2021) 3837–3849.
[3] J. Guinney, et al., The consensus molecular subtypes of colorectal cancer, Nat. Med. 21 (2015) 1350–1356.
[4] A. Cañellas-Socias, et al., Metastatic recurrence in colorectal cancer arises from residual EMP1+ cells, Nature 611 (2022) 603–613.
[5] A. Álvarez-Varela, et al., Mex3a marks drug-tolerant persister colorectal cancer cells that mediate relapse after chemotherapy, Nat. Cancer 3 (2022) 1052–1070.
[6] L. Solé, et al., p53 wild-type colorectal cancer cells that express a fetal gene signature are associated with metastasis and poor prognosis, Nat. Commun. 13 (2022) 2866.
[7] E. Dhimolea, et al., An embryonic diapause-like adaptation with suppressed Myc activity enables tumor treatment persistence, Cancer Cell 39 (2021) 240–256 (e11).
[8] S.K. Rehman, et al., Colorectal cancer cells enter a diapause-like DTP state to survive chemotherapy, Cell (2021), https://doi.org/10.1016/j.cell.2020.11.018.
[9] C. Colomer, et al., IKKα kinase regulates the DNA damage response and drives chemo-resistance in cancer, Mol. Cell 75 (2019) 669–682.e5.
[10] S. Sampurno, et al., The Myb-p300-CREB axis modulates intestine home- ostasis, radiosensitivity and tumorigenesis, Cell Death Dis. 4 (2013) e605.
[11] D. Cheasley, et al., Myb controls intestinal stem cell genes and self-renewal, Stem Cells 29 (2011) 2042–2050.
[12] A. Hryniuk, S. Grainger, J.G.A. Savory, D. Lohnes, Cdx function is required for maintenance of intestinal identity in the adult, Dev. Biol. 363 (2012) 426–437.
[13] D.G. Silberg, G.P. Swain, E.R. Suh, P.G. Traber, Cdx1 and Cdx2 expression during intestinal development, Gastroenterology 119 (2000) 961–971.
[14] J.G.A. Savory, et al., Cdx1 and Cdx2 are functionally equivalent in vertebral patterning, Dev. Biol. 330 (2009) 114–122.
[15] J.K. Taylor, T. Levy, E.R. Suh, P.G. Traber, Activation of enhancer elements by the homeobox gene Cdx2 is cell line specific, Nucleic Acids Res. 25 (1997) 2293–2300.
[16] A.-B. Ramakrishnan, L. Chen, P.E. Burby, K.M. Cadigan, Wnt target enhancer regulation by a CDX/TCF transcription factor collective and a novel DNA motif, Nucleic Acids Res. 49 (2021) 8625–8641.
[17] K. Sturgeon, et al., Cdx1 refines positional identity of the vertebrate hind- brain by directly repressing Mafb expression, Development 138 (2011) 65–74.
[18] C. Pilati, et al., CDX2 prognostic value in stage II/III resected colon cancer is related to CMS classification, Ann. Oncol. 28 (2017) 1032–1035.
[19] S.H. den Uil, et al., Quantitative analysis of CDX2 protein expression improves its clinical utility as a prognostic biomarker in stage II and III colon cancer, Eur. J. Cancer 144 (2021) 91–100.
[20] P. Dalerba, et al., CDX2 as a prognostic biomarker in stage II and stage III colon cancer, N. Engl. J. Med. 374 (2016) 211–222.
[21] N.A.C.S. Wong, et al., Loss of CDX1 expression in colorectal carcinoma: promoter methylation, mutation, and loss of heterozygosity analyses of 37 cell lines, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 574–579.
[22] I.Y. Luk, et al., Epithelial de-differentiation triggered by co-ordinate epige- netic inactivation of the EHF and CDX1 transcription factors drives colorectal cancer progression, Cell Death Differ. 29 (2022) 2288–2302.

[23] C. Nakayama, et al., Transduced caudal-type homeobox (CDX) 2/CDX1 can induce growth inhibition on CDX-deficient gastric cancer by rapid intestinal differentiation, Cancer Sci. 109 (2018) 3853–3864.
[24] M.F. Jones, et al., The CDX1-microRNA-215 axis regulates colorectal cancer stem cell differentiation, Proc. Natl. Acad. Sci. U. S. A. 112 (E15) (2015) 50–58.
[25] D. Qualtrough, T. Hinoi, E. Fearon, C. Paraskeva, Expression of CDX2 in normal and neoplastic human colon tissue and during differentia- tion of an in vitro model system, Gut 51 (2002) 184–190.
[26] A. Hryniuk, S. Grainger, J.G.A. Savory, D. Lohnes, Cdx1 and Cdx2 function as tumor suppressors, J. Biol. Chem. 289 (2014) 33343–33354.
[27] G.V. Mallo, et al., Expression of the Cdx1 and Cdx2 homeotic genes leads to reduced malignancy in colon cancer-derived cells, J. Biol. Chem. 273 (1998) 14030–14036.
[28] C. Domon-Dell, et al., Cdx1 homeobox gene during human colon cancer progression, Oncogene 22 (2003) 7913–7921.
[29] L. Marisa, et al., Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value, PLoS Med. (2013), https://doi.org/10.1371/journal.pmed.1001453.
[30] R.N. Jorissen, et al., Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes stage B and C colorectal cancer, Clin. Cancer Res. (2009), https://doi.org/10.1158/1078-0432.CCR-09-1431.
[31] T. Barrett, R. Edgar, Gene expression omnibus: microarray data storage, submission, retrieval, and analysis, Methods Enzymol. 411 (2006) 352–369.
[32] A. Colaprico, et al., TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data, Nucleic Acids Res. 44 (2016) e71.
[33] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2009) 139–140.
[34] G.H. Kim, et al., Relationships between the expressions of CDX1 and CDX2 mRNA and clinicopathologic features in colorectal cancers, Korean J. Intern. Med. 20 (2005) 317–324.
[35] T. Tagawa, et al., Multiple microRNAs induced by Cdx1 suppress Cdx2 in human colorectal tumour cells, Biochem. J. 447 (2012) 449–455.
[36] K. Aasebø, et al., CDX2: a prognostic marker in metastatic colorectal cancer defining a better BRAF mutated and a worse KRAS mutated subgroup, Front. Oncol. 10 (2020) 8.
[37] M.S. Landau, S.-F. Kuan, S. Chiosea, R.K. Pai, BRAF-mutated microsatellite stable colorectal carcinoma: an aggressive adenocarcinoma with reduced CDX2 and increased cytokeratin 7 immunohistochemical expression, Hum. Pathol. 45 (2014) 1704–1712.
[38] I.V. Lundberg, et al., SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer, BMC Cancer 16 (2016) 471.
[39] N. Sakamoto, et al., BRAFV600E cooperates with CDX2 inactivation to promote serrated colorectal tumorigenesis, Elife 6 (2017).
[40] O.O. Ogunwobi, F. Mahmood, A. Akingboye, Biomarkers in colorectal cancer: current research and future prospects, Int. J. Mol. Sci. 21 (2020).
[41] E. Pilozzi, M.R. Onelli, V. Ziparo, P. Mercantini, L. Ruco, CDX1 expression is reduced in colorectal carcinoma and is associated with promoter hypermethylation, J. Pathol. 204 (2004) 289–295.