

# SCIENTIFIC REPORTS

OPEN

## Radiogenomics-based cancer prognosis in colorectal cancer

Bogdan Badic<sup>1</sup>, Mathieu Hatt<sup>1</sup>, Stephanie Durand<sup>2</sup>, Catherine Le Jossic-Corcoc<sup>2</sup>, Brigitte Simon<sup>2</sup>, Dimitris Visvikis<sup>1</sup> & Laurent Corcos<sup>2</sup>

Received: 16 February 2019  
Accepted: 20 June 2019  
Published online: 05 July 2019

Radiogenomics aims at investigating the relationship between imaging radiomic features and gene expression alterations. This study addressed the potential prognostic complementary value of contrast enhanced computed tomography (CE-CT) radiomic features and gene expression data in primary colorectal cancers (CRC). Sixty-four patients underwent CT scans and radiomic features were extracted from the delineated tumor volume. Gene expression analysis of a small set of genes, previously identified as relevant for CRC, was conducted on surgical samples from the same tumors. The relationships between radiomic and gene expression data was assessed using the Kruskal–Wallis test. Multiple testing was not performed, as this was a pilot study. Cox regression was used to identify variables related to overall survival (OS) and progression free survival (PFS). *ABCC2* gene expression was correlated with N ( $p = 0.016$ ) and M stages ( $p = 0.022$ ). Expression changes of *ABCC2*, *CD166*, *CDKNV1* and *INHBB* genes exhibited significant correlations with some radiomic features. OS was associated with Ratio 3D Surface/volume ( $p = 0.022$ ) and *ALDH1A1* expression ( $p = 0.042$ ), whereas clinical stage ( $p = 0.004$ ), *ABCC2* expression ( $p = 0.035$ ), and Entropy<sub>GLCM\_E</sub> ( $p = 0.0031$ ), were prognostic factors for PFS. Combining CE-CT radiomics with gene expression analysis and histopathological examination of primary CRC could provide higher prognostic stratification power, leading to improved patient management.

Colorectal carcinoma (CRC) is characterized by a marked intra-tumor genetic heterogeneity<sup>1</sup>, which may be, at least partly, assessed on medical images using radiomics analysis to quantitatively characterize a delineated tumor volume using various metrics<sup>2</sup>. Recent studies have investigated the prognostic value of computed tomography (CT)-derived radiomic features of patients with colorectal cancers<sup>3–7</sup>. In addition, high-throughput platforms have been developed with the aim of obtaining genomic and transcriptomic data to better characterize cancer at the molecular level<sup>8</sup>. Current clinical applications of genomic information use mutational data from *KRAS*, *NRAS* and *BRAF* genes, which are effectors of the *MAPK* (mitogen-activated protein kinase) pathway, as prognostic and predictive biomarkers for standard treatment, and the *HER2* gene for targeted therapies<sup>9,10</sup>.

In a broader sense, radiogenomics aims at investigating the relationship between imaging features and gene expression alterations, as well as their potential complementary value in predictive modeling<sup>11</sup>. To date, no studies have examined either the relationships between gene expression and radiomics image features of CRC, or their complementary prognostic value. To investigate this question, we selected a small set of genes previously identified as relevant for CRC and conducted a pilot study to explore their relationship and complementary prognostic value for overall (OS) and progression-free survival (PFS), in a retrospectively recruited cohort of 64 patients.

### Results

**Relationships between clinical, histopathological characteristics and gene expression profiles.** The deregulation of gene expression observed in our cancer sample set showed good concordance with that observed in the TCGA samples (Table 1). Only *ABCC2* gene expression was correlated with UICC Stage ( $p = 0.028$ ), N stage ( $p = 0.016$ ) and M stage ( $p = 0.022$ ). Age or gender were not correlated with gene expression profiles.

**Relationships between radiomics and gene expression changes.** Some of the first, second and third order radiomic features were significantly correlated with changes in gene expression, e.g. *ABCC2*, *CD166* (Kruskal–Wallis analysis). The expression changes in four genes (*ABCC2*, *CD166*, *CDKNV1*, *INHBB*) exhibited a

<sup>1</sup>INSERM, UMR 1101, LaTIM, 22 rue Camille Desmoulins, 29238, Brest, France. <sup>2</sup>INSERM, UMR 1078, 22, avenue Camille Desmoulins, 29238, Brest, France. Dimitris Visvikis and Laurent Corcos contributed equally. Correspondence and requests for materials should be addressed to B.B. (email: [bogdan.badic@chu-brest.fr](mailto:bogdan.badic@chu-brest.fr))

Gene Symbol	Our cohort		TCGA-Microarray			TCGA-RNAseq		
	Mean	SEM	Mean	SEM	p adj*	Mean	SEM	p adj*
<i>ABCB1</i>	0.451	0.05	0.332	0.021	<0.0001	0.447	0.026	<0.0001
<i>ABCC2</i>	4.211	1.03	3.45	0.333	<0.0005	16.625	1.69	<0.0001
<i>ABCG2</i>	0.113	0.01	0.074	0.008	<0.0001	0.05	0.005	<0.0001
<i>ALDH1A1</i>	0.667	0.08	0.579	0.043	<0.0001	0.574	0.044	<0.0001
<i>CD166 (ALCAM)</i>	1.722	0.12	1.89	0.086	<0.0001	1.955	0.057	<0.0001
<i>CDKN1A</i>	0.592	0.04	0.584	0.023	<0.0001	0.457	0.017	<0.0001
<i>INHBB</i>	1.599	0.23	4.217	0.533	<0.0001	6.88	0.645	<0.0001

**Table 1.** Gene expression changes in colorectal carcinoma as compared to normal tissue in our cohort and similarly in 2 TCGA cohorts. \*Comparison of CRC *versus* NT (Normal Tissues) was performed by Welch's t-test and p values for TCGA data were corrected for multiple testing by Benjamini-Hochberg method (p.adjust function in R environment). SEM - Standard error of the mean.

	Covariate	HR	95% CI	p-value
<b>Progression free survival</b>				
Histopathological	N = 1	7.8	1.7 to 35.0	0.0076
	Stage = 3	11.1	2.1 to 57.9	0.0044
Gene expression	<i>ABCC2</i> (unchanged)	5.0	1.1 to 22.4	0.0354
Radiomic	Flatness	0.0003	0.0 to 0.3	0.0234
	SENT <sub>GLMCR</sub>	0.5	0.31 to 0.91	0.0205
	Entropy <sub>GLMCE</sub>	0.046	0.006 to 0.35	0.0031
	GLNU <sub>L</sub>	0.074	0.007 to 0.77	0.0298
<b>Overall survival</b>				
Histopathological	Stage = 4	3.1	1.1 to 8.6	0.0304
Gene expression	<i>ALDH1A1</i> (up-regulated)	2.8	1.0 to 7.3	0.0420
Radiomic	Ratio 3D Surface/volume	3.3	1.2 to 9.2	0.0219
	Flatness	0.012	0.0001 to 0.6	0.0275
	IDM <sub>R</sub>	270.1	1.3 to 55079.2	0.0401
	ID <sub>R</sub>	655.3	1.1 to 393927.0	0.0481

**Table 2.** Overall survival and progression free survival Cox univariate analysis. HR = hazard ratio. SENT (sum entropy). Entropy<sub>GLCM</sub> = entropy from grey-level-co-occurrence-matrix. GLNU = grey-level nonuniformity. IDM = Inverse difference moment. ID = Inverse difference.

significant correlation with radiomic features. Of note, the discretization method influenced the correlation with gene expression, e.g. *ABCC2* gene expression was correlated with Low gray level zone emphasis (LGLZE) and Small zone low gray emphasis (SZLGE) after **R** discretization but not after **L** or **E**.

**Overall relationships between clinical, histopathological, radiomics, gene expression profiles and outcome.** Cox univariate analysis revealed histopathological (stage IV), *ALDH1A1* expression and several radiomic features (Ratio 3D Surface/volume, Flatness, Inverse difference moment (IDM<sub>R</sub>), Inverse difference (ID<sub>R</sub>)) as predictive factors for overall survival (OS) (Table 2). Multivariate analysis identified Ratio 3D Surface/volume and *ALDH1A1* as independent prognostic factors. Their combination in Cox multivariate regression led to a HR of 8.4, compared to lower values of 2.8 and 3.3 alone, as well as for the stage alone (HR = 3.1). It should be emphasized however that significant overlap was observed between the 95% CI of these stratification results: [3.4–20.6] for the multivariate combination of Ratio and *ALDH1A1*, compared to [1–7.3], [1.2–9.2] and [1.1–8.6] for Ratio, *ALDH1A1* and stage respectively.

Some radiomic features, namely Flatness, Sum entropy (SENT<sub>R</sub>), entropy from Grey-level-co-occurrence-matrix (Entropy<sub>GLMCE</sub>), Grey-level non-uniformity (GLNU<sub>L</sub>), *ABCC2* mRNA level, Stage III and node status (N) were significantly associated with PFS on Cox univariate analysis. On multivariate analysis, stage III, *ABCC2* and Entropy<sub>GLMCE</sub> remained independent prognostic factors of PFS (p = 0.0001). Their combination through Cox modeling led to an HR above 22, compared to lower values when considered alone or two by two (Table 3; Fig. 1). Again, overlaps between the 95% CI of each HR were observed nonetheless.

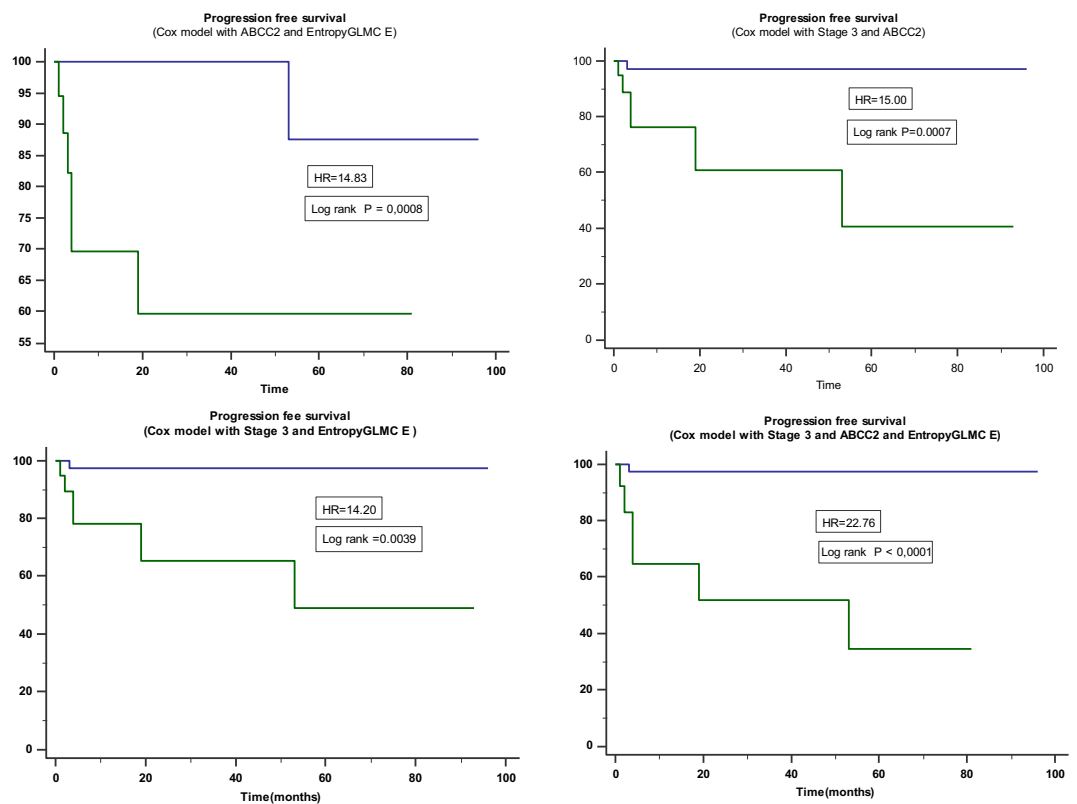
## Discussion

We report here a promising potential added value of combining some radiomic features derived from contrast-enhanced CT images with gene expression features for predicting disease outcome of patients with primary CRC.

The impact of chemotherapy in CRC patients is limited by the inter-individual variability in drug response and the acquisition of resistance in which ATP-binding cassette (ABC) transporters play a crucial role by increasing

	HR	95% CI	p-value
<b>Model Overall survival (OS)</b>			
Ratio 3D Surface/volume	2.8	1.0 to 7.3	0.0420
<i>ALDH1A1</i>	3.3	1.2 to 9.2	0.0219
Cox model with Ratio and <i>ALDH1A1</i>	8.4	3.4 to 20.6	0.0005
<b>Model Progression free survival (PFS)</b>			
Stage 3 only	11.1	2.1 to 57.9	0.0044
Cox model combining Stage 3 and Entropy <sub>GLMC E</sub>	14.2	2.4 to 85.3	0.0039
Cox model combining Stage 3 and <i>ABCC2</i>	15.0	2.9 to 77.4	0.0007
Cox model combining <i>ABCC2</i> and Entropy <sub>GLMC E</sub>	14.8	2.9 to 76.2	0.0008
Cox model combining Stage 3, <i>ABCC2</i> and Entropy <sub>GLMC E</sub>	22.8	3.7 to 141	<0.0001

**Table 3.** Kaplan-Meier analysis and resulting hazard ratios for stratifying patients for OS and PFS using single features identified by the univariate Cox modeling or their combination.



**Figure 1.** Kaplan-Meier analysis of PFS with Cox models combining Stage III, *ABCC2* expression and Entropy<sub>GLMC E</sub> led to increased stratification power.

the efflux of anticancer drugs outside of cancer cells<sup>12,13</sup>. Significantly, higher *ABCC2* mRNA levels were previously observed in adenomas with mild to moderate dysplasia, as well as in carcinoma, whereas *ABCG2* mRNA levels were decreased<sup>14</sup>, similarly to the data presented here. We also found correlations between *ABCC2* mRNA levels and PFS of CRC patients. The rise in *ABCC2* mRNA in CRC could reflect a functional defect. By contrast, the strong drop in *ABCG2* mRNA was not associated with any other parameter. However, it seems possible that some other member of the family of ABC transporters could have compensated for this reduction, leaving it relevant as a marker, but with low or unappreciable effect on the response to drugs. A similar situation occurred for the *ABCB1*, *CDKNV1* or *CDKNV2* genes. It remains, however, that these gene expression changes should be worth investigating further to determine how they contribute to characterizing, or possibly, participate in the establishment or maintenance of the cancerous phenotype.

*ALDH1A1* is a detoxifying enzyme that confers resistance to alkylating chemotherapeutic agents and protects against oxidative damage by catalyzing the irreversible oxidization of cellular aldehydes<sup>15</sup>. High *ALDH1* expression indicates a poor prognosis in CRC patients that correlates with the T stage, N stage and tumor

differentiation<sup>16</sup>. Overall survival was associated in our study with *ALDH1A1* gene expression on Cox univariate, but not multivariate analysis, probably due to the limited size of our cohort.

A number of CT-derived radiomic features have been linked to gene expression profiles, for example in lung cancer<sup>2</sup>. CE-CT features analysis used for preoperative tumor staging<sup>5</sup> correlated with lymph nodes metastasis in CRC<sup>17</sup>. CE-CT radiomics analysis could differentiate high-grade from low-grade CRC<sup>18</sup>. Features derived from intensity histograms, including entropy, standard deviation and skewness, were associated with *KRAS* mutation and tumor grade in CRC liver metastases<sup>4</sup>. Progression free survival, but not overall survival, was correlated with entropy from GLCM in our study. This difference can be explained by the fact that first order statistics features describe the distribution of individual voxel values without concern for spatial relationships<sup>19</sup>.

In our study, some second and third order texture features were significantly associated with PFS and OS. The repeatability of these metrics could be affected by the discretization choices, but the discretization value chosen for this study (64 bins) has been shown previously to provide a good compromise<sup>20</sup>. Geometric features such as flatness, minor axis length or sphericity are known already to be reliable shape descriptors that can be related to the high repeatability of segmentation<sup>21</sup>. Some shape-based features were significantly associated with OS and PFS in our study.

As shown here, combining an increasing number of features improved the stratification of patients resulting in a nomogram combining clinical stage, gene expression (*ALDH1A1*, *ABCC2*) and CE-CT radiomic features, suggesting that complementary prognostic value could be obtained not only from histopathological examination, but also using the pre-therapeutic CT and gene expression features. Our nomogram could thus be used to identify patients with poor prognosis who could then be offered alternative treatment options, such as targeted therapies or treatment intensification, using baseline CE-CT images that are acquired routinely for staging in clinical management.

Our study has several limitations. First, our cohort is retrospective and from a single center. It is also limited by the number of patients and events, due to the fact that tumor tissue samples were available only for a fraction of the patients, and that funds for the expensive transcriptomic analysis were limited. This limited number of patients and events led to high 95% confidence intervals around the estimated HR values which limit the reliability of their comparison. Second, repeatability of radiomics data is important to guarantee reliable results and several texture features have already shown their repeatability in other cohorts and types of cancers<sup>22,23</sup>. Since only routine CE-CT acquisitions were available for our cohort, we did not perform a repeatability evaluation of texture features and their prognostic value. Nevertheless, we performed texture analysis on the entire tumor volume, despite the time-consuming nature of whole tumor analysis. By contrast, the use of the largest cross-sectional area rather than the whole tumor to extract texture features could be a limitation in some tumors given the overall aim to quantify heterogeneity<sup>3</sup>. Finally, only radiomics extracted from portal enhanced CT scans were used for this study. Some recent observations showed that features extracted from non-enhanced and contrast-enhanced CT provide complementary prognostic information, thus suggesting that analysis of this two modalities could better characterize CRC tumors<sup>24</sup>.

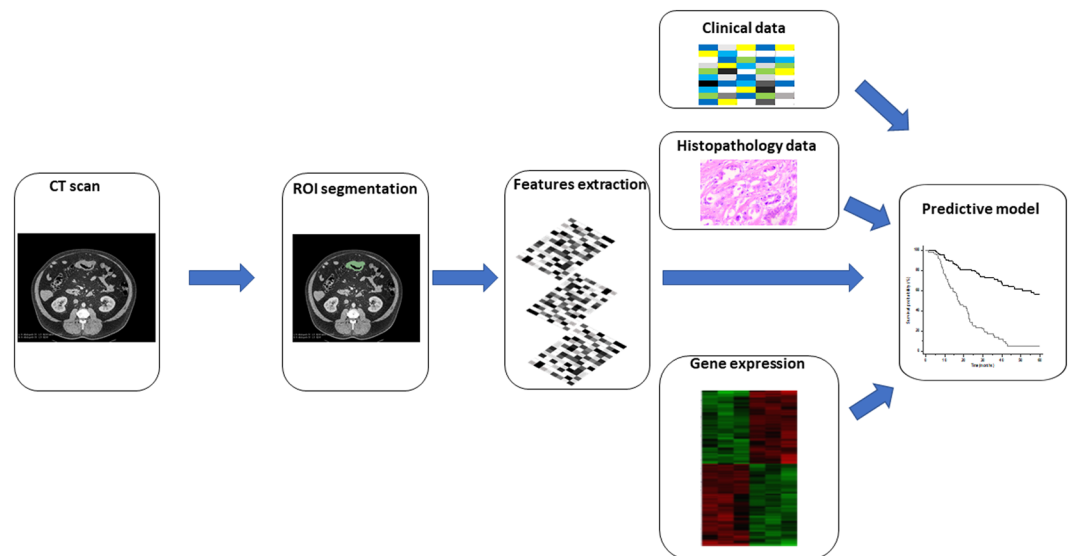
Combining CE-CT radiomics, gene expression analysis and histopathological examination of primary CRC demonstrated an association with survival, suggesting that this strategy could provide complementary prognostic value, which might be beneficial for therapeutic decisions. Further analyses of genomic mutations or rearrangements could also strengthen such multi-parameter approaches, provided all of this patient information can be obtained readily and be cost-efficient.

## Patients and Methods

**Patients.** Between January 2008 and May 2017, 653 colorectal resections were performed at our institution. For 64 of them (36 M, 28 F; mean age 71 (range 26–93)), 27 in the right colon, 3 in the transverse colon, 20 in the left colon and 14 in the rectum) all the data and material required for our retrospective investigation were available: clinical and histopathological data, CT scans, outcome information (date of diagnosis, surgery, progression or last follow-up and death) and tissue samples (from our local registered tumor tissue collection). According to the UICC staging, there were 12 stage I, 25 stage II, 15 stage III and 12 stage IV tumors. Informed consent was obtained from each participant included in the study. Tumor localization was determined by colonoscopy and the tumor was visible on CE-CT images. We excluded patients with previous chemotherapy or radiotherapy or tumors not confirmed by anatomopathological examination. We also excluded patients with tumors located in the medium and lower rectum, because of lower CT scan resolution preventing appropriate delineation of rectal tumors, and because of different treatment modalities between colon and rectum tumors. Additional patient characteristics, such as potential comorbidity conditions at time of diagnosis (drug absorption, tobacco exposure or alcohol consumption, etc.), were not available. This study was conducted after approval by Brest University Hospital institutional ethics committee.

A workflow of our study is displayed in Fig. 2.

**Imaging.** Abdominal CE-CT assessment was performed as part of a routine imaging protocol at our institution before surgery or systemic treatment (radiotherapy/chemotherapy). Images were collected from the archive and communication system (PACS). The portal venous phase CT images were exploited for the present analysis because they allowed to clearly distinguish tumor tissue from adjacent normal bowel wall. Scans were performed on a Siemens Definition AS64 (Siemens medical, Erlangen, Germany). Standard acquisition settings were 120 kVp tube voltage, automatic tube current-exposure time product, 0.5 s rotation time, 1.25 mm slice thickness, matrix size of 512 × 512 pixels and inspiratory breath hold. Intravenous contrast (Xenetix 350, Guebert, Roissy, France) was administered at a 3 mL/sec injection rate with a pump injector (Medrad Stellant, Bayer, NY).



**Figure 2.** Workflow of prognosis model construction. A colorectal tumor is delineated in every slice and validated by an experienced physician. This allows creation of a 3D representation of the tumor. Radiomic features (intensity, shape, texture) are extracted from this delineated tumor, and integrated with clinical, histopathology and genomic/gene expression data.

**Radiomics analysis.** Primary tumors were semi-automatically delineated by one experienced physician using the 3D Slicer™ software<sup>25</sup> on the portal phase CT images. Imaging Biomarkers Standardization Initiative (IBSI) compliant radiomic features were extracted from each segmented volume using an in-house software that was validated with the most up-to-date IBSI reference document and benchmark values<sup>26</sup>. For textural features, matrices were constructed in 3D according to the merging technique (Supplementary Material 1). No filter-based analysis (either textural features on wavelet decompositions<sup>2</sup> or histogram analysis in sub-volumes identified by log of Laplacian filters<sup>27</sup>) was included to reduce the number of considered features and because these have not yet been included in the IBSI<sup>24</sup>.

The intensities in the original images were discretized for second and third order textural matrices' calculations<sup>28</sup>. It has been shown that this step can have a significant impact on the texture values and distributions<sup>28</sup>. Three different sets of textural features were thus generated with different methods. The first two consisted of a discretization into 64 bins using either the fixed bin number approach (denoted from here onwards L) or histogram equalization, (denoted from here onwards E)<sup>26,28</sup>. The third one consisted in resampling the original grey-levels into a variable number of bins of fixed width (in our case 10 Hounsfield units (HU), denoted from here onwards R)<sup>24,28</sup> (see Supplementary Material 1 for radiomics analysis).

**Tissue samples processing.** Tissue samples were obtained from our institution tissue bank. H/E staining was performed for all samples, and the tumor content within cancer samples was above 80%. Repartition of low grade (stages 0, I and II) and high-grade carcinoma samples (stages III and IV) was homogeneous. The tissue fragments were stored in RNAlater® stabilization solution (Ambion, France), a reagent that prevents mRNA from degradation. Total RNA was extracted with the AllPrep DNA/RNA Mini kit (Qiagen, Courtaboeuf, France) from homogenized tissue samples (20 mg). RNA purity and integrity were evaluated by measuring the optical density ratio (A260/A280) and the RNA Integrity Number (RIN) using the RNA 6000 Nano LabChip and the 2100 BioAnalyzer (Agilent, Massy, France), respectively. Only RNA samples with a 28S/18S ratio >1.0 and RIN >5.0 were used. Reverse transcription and real-time PCR amplification were performed using conventional methods with reagents from Applied Biosystems (Applied Biosystems, France) and a StepOnePlus real time PCR system. The RNA levels were standardized to those of beta2-microglobulin, which was invariant between cancer and healthy tissue. The engineered PCR array plates were prepared at Eurogentec (Eurogentec, Belgium). We selected genes that have been previously reported for their expression changes in CRC (*ABCB1*, *ABCC2*, *ALDH1A1*, *CDKNV1*, *CDKNV2*, *INHBB*, *CD166* and *ABCG2*) as compared to healthy tissue<sup>14,29–31</sup>. The gene characteristics are presented in Supplementary Table 1. To establish an independent validation set, we downloaded expression microarray and RNA-sequencing data from 221 and 358 CRC patients, respectively, from The Cancer Genome Atlas (TCGA) portal<sup>32</sup> and the Broad GDAC Firehose<sup>33</sup> by selecting the colorectal adenocarcinoma cohort (COADREAD). Patient characteristics from this cohort are presented in Supplementary Table 2.

**Statistical analysis.** Statistical analyses were performed with MedCalc Software version 14.8.1. Comparisons between gene expression profiles and continuous variables were done using the non-parametric Kruskal-Wallis test. Comparisons between gene expression profiles and discrete variables were performed using the Chi2 test or Fisher's exact test. Progression free survival (PFS) was defined as the time from diagnosis to first event (local or metastatic failure or death). Patients with no events were censored at the time of last follow-up. Overall survival (OS)



was defined as the time from diagnosis to death from any disease-related cause or last follow-up. The Kaplan-Meier method and log-rank test with cut-off thresholds determined by receiver operating characteristic curve (ROC) analysis (according to Youden's index) were used to determine the prognostic value of each feature. For each variable, relative risks were estimated using a univariate Cox model and expressed with their 95% confidence interval. Cox multivariate analysis was carried out using characteristics found significant in the univariate analysis. Statistical significance was set as  $p < 0.05$  without correction for multiple testing as this is a hypothesis-generating study.

**Approval, accordance and informed consent.** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. All samples were included in a registered tumor tissue collection and the present study was conducted after approval by the CHU Brest institutional ethics committee<sup>34</sup>.

## Data Availability

CT images used in this study can be made available on request for specific research purposes.

## References

- Losi, L. Evolution of intratumoral genetic heterogeneity during colorectal cancer progression. *26* (2005).
- Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5**, 4006 (2014).
- Ng, F., Kozarski, R., Ganeshan, B. & Goh, V. Assessment of tumor heterogeneity by CT texture analysis: Can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur. J. Radiol.* **82**, 342–348 (2013).
- Lubner, M. G. *et al.* CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom. Imaging* **40**, 2331–2337 (2015).
- Liang, C. *et al.* The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. *Oncotarget* **7**, 31401–12 (2016).
- Dinapoli, N. *et al.* Radiomics for rectal cancer. *5*, 424–431 (2016).
- Huang, Y. *et al.* Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. <https://doi.org/10.1200/JCO.2015.65.9128> (2016).
- Sala, E. *et al.* HHS Public Access. **72**, 3–10 (2018).
- Therkildsen, C., Bergmann, T. K., Henrichsen-Schnack, T., Ladellund, S. & Nilbert, M. The predictive value of KRAS, NRAS, BRAF, PIK3CA and PTEN for anti-EGFR treatment in metastatic colorectal cancer: A systematic review and meta-analysis. *Acta Oncol. (Madr)*. **53**, 852–864 (2014).
- Mondaca, S. & Yaeger, R. Colorectal cancer genomics and designing rational trials. **6**, 1–10 (2018).
- Incoronato, M. *et al.* Radiogenomic Analysis of Oncological Data: A Technical Survey. 1–28. <https://doi.org/10.3390/ijms18040805> (2017).
- Pesson, M. *et al.* A gene expression and pre-mRNA splicing signature that marks the adenoma-adenocarcinoma progression in colorectal cancer. *PLoS One* **9** (2014).
- Theile, D., Grebhardt, S., Haefeli, W. E. & Weiss, J. Involvement of drug transporters in the synergistic action of FOLFOX combination chemotherapy. *Biochem. Pharmacol.* **78**, 1366–1373 (2009).
- Andersen, V., Vogel, L. K., Kopp, T. I. & Sæb, M. High ABCG2 and Low ABCG2 Gene Expression Are Early Events in the Colorectal Adenoma-Carcinoma Sequence. 1–13. <https://doi.org/10.1371/journal.pone.0119255> (2015).
- Leng, Z. *et al.* Krüppel-Like Factor 4 Acts as an Oncogene in Colon Cancer Stem Cell-Enriched Spheroid Cells. *PLoS One* **8** (2013).
- Chen, J. *et al.* Prognostic value of cancer stem cell marker ALDH1 expression in colorectal cancer: A systematic review and meta-analysis. *PLoS One* **10**, 1–15 (2015).
- Huang, Y. Q. *et al.* Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J. Clin. Oncol.* **34**, 2157–64 (2016).
- Huang, X. *et al.* CT-based Radiomics Signature to Discriminate High-grade From Low-grade Colorectal Adenocarcinoma. *Acad. Radiol.*, <https://doi.org/10.1016/j.acra.2018.01.020> (2018).
- Tixier, F. *et al.* Reproducibility of Tumor Uptake Heterogeneity Characterization Through Textural Feature Analysis in 18F-FDG PET. *J. Nucl. Med.* **53**, 693–700 (2012).
- Leijenaar, R. T. H. *et al.* The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci. Rep.* **5**, 11075 (2015).
- Desseroit, M. C. *et al.* Development of a nomogram combining clinical staging with 18F-FDG PET/CT image features in non-small-cell lung cancer stage I-III. *Eur. J. Nucl. Med. Mol. Imaging* **43**, 1477–1485 (2016).
- Yang, J. *et al.* Uncertainty analysis of quantitative imaging features extracted from contrast-enhanced CT in lung tumors. *Comput. Med. Imaging Graph.* 1–8. <https://doi.org/10.1016/j.compmedimag.2015.12.001> (2016).
- Images, C. T., Hawkins, S., Kim, J., Goldgof, D. B. & Gillies, R. J. *Translational Oncology Reproducibility and Prognosis of Quantitative Features*. **7**, 72–87 (2014).
- Badic, B., Desseroit, M. C., Mathieu Hatt, D. V. Potential Complementary Value of Noncontrast and Contrast Enhanced. *Acad. Radiol.* 1–11. <https://doi.org/10.1016/j.acra.2018.06.004> (2018).
- Parmar, C. *et al.* Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One* **9**, 1–8 (2014).
- Zwanenburg, A. & Leger, S. Vallières, Martin; Löck, Steffen; Image Biomarker Standardisation Initiative, for the. Image biomarker standardisation initiative. *eprint arXiv:1612.07003*.
- Ganeshan, B., Skogen, K., Pressney, I., Coutroubis, D. & Miles, K. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: Preliminary evidence of an association with tumour metabolism, stage, and survival. *Clin. Radiol.* **67**, 157–164 (2012).
- Hatt, M. *et al.* Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging*, <https://doi.org/10.1007/s00259-016-3427-0>.
- Wang, X. *et al.* Membranous ABCG2 expression in colorectal cancer independently correlates with shortened patient survival. *Cancer Biomarkers* **13**, 81–88 (2013).
- Candeil, L. *et al.* ABCG2 overexpression in colon cancer cells resistant to SN38 and in irinotecan-treated metastases. *Int. J. Cancer* **109**, 848–854 (2004).
- Kahlert, C. *et al.* Expression analysis of aldehyde dehydrogenase 1A1 (ALDH1A1) in colon and rectal cancer in association with prognosis and response to chemotherapy. *Ann. Surg. Oncol.* **19**, 4193–4201 (2012).
- No Title. Available at, <https://portal.gdc.cancer.gov>.
- No Title. Available at, <https://gdac.broadinstitute.org>.
- Badic, B. Caractérisation multiparamétrique des cancers colorectaux (2018).

## Acknowledgements

The manuscript is part of a PhD thesis “Caractérisation multiparamétrique des cancers colorectaux” by Bogdan Badic, Université de Bretagne Occidentale (HAL Id: tel-02078409). We thank Dr Laurent Doucet, for providing colon cancer tissue, and Dr Marie-Cécile Nicot for extracting tissue fragments from Brest Tumor tissue Bank. We thank the Ligue contre le cancer, comité du Finistère and the INSERM for funding the project. This study was funded in part by the Ligue contre le Cancer, comité du Finistère, and by INSERM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Contributions

B.B., D.V. and L.C. designed the study, B.B. and M.H. performed image analysis and statistical analysis. S.D, C.L.J.C., B.S., L.C. performed gene expression analyses in colon and colon cancer-derived stem cells and bioinformatics treatment of the data. All authors contributed to writing and reviewed/approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-46286-6>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019