European Journal of Cancer 177 (2022) 94–102

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Tumour break load is a biologically relevant feature of genomic instability with prognostic value in colorectal cancer

Soufyan Lakbir [a,b], Sara Lahoz [c], Miriam Cuatrecasas [d], Jordi Camps [c,e], Roel A. Glas [a,b], Jaap Heringa [a,f], Gerrit A. Meijer [b], Sanne Abeln [a,g,**], Remond J.A. Fijneman [b,*]

[a] *Bioinformatics Group, Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam 1081HV, the Netherlands*
[b] *Department of Pathology, Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam 1066CX, the Netherlands*
[c] *Translational Colorectal Cancer Genomics, Gastrointestinal and Pancreatic Oncology Team, Institut D'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), Hospital Clínic de Barcelona, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Barcelona, 08036, Spain*
[d] *Pathology Department, Biomedical Diagnostic Center (CDB), Hospital Clínic de Barcelona, Institut D'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Universitat de Barcelona (UB), Barcelona, 08036, Spain*
[e] *Department of Cell Biology, Physiology and Immunology, Faculty of Medicine, Autonomous University of Barcelona, Bellaterra, 08193, Spain*
[f] *AIMMS – Amsterdam Institute for Molecules Medicines and Systems, Vrije Universiteit Amsterdam, Amsterdam 1081HV, the Netherlands*
[g] *Life Sciences and Health Research Group, Centrum Wiskunde & Informatica (CWI), Science Park 123, Amsterdam 1098 XG, the Netherlands*

Received 6 July 2022; received in revised form 28 September 2022; accepted 30 September 2022
Available online 8 October 2022

KEYWORDS
Colorectal cancer;
Genomic instability;
Tumour break load;
Prognostic biomarker

Abstract ***Background:*** Clinically implemented prognostic biomarkers are lacking for the 80% of colorectal cancers (CRCs) that exhibit chromosomal instability (CIN). CIN is characterised by chromosome segregation errors and double-strand break repair defects that lead to somatic copy number aberrations (SCNAs) and chromosomal rearrangement-associated structural variants (SVs), respectively. We hypothesise that the number of SVs is a distinct feature of genomic instability and defined a new measure to quantify SVs: the tumour break

* *Corresponding author*: Plesmanlaan 121, 1066CX Amsterdam, the Netherlands.
** *Corresponding author*: De Boelelaan 1081A, 1081HV Amsterdam, the Netherlands.
E-mail address: s.abeln@vu.nl (S. Abeln), r.fijneman@nki.nl (R.J.A. Fijneman).

https://doi.org/10.1016/j.ejca.2022.09.034

load (TBL). The present study aimed to characterise the biological impact and clinical relevance of TBL in CRC.

Methods: Disease-free survival and SCNA data were obtained from The Cancer Genome Atlas and two independent CRC studies. TBL was defined as the sum of SCNA-associated SVs. RNA gene expression data of microsatellite stable (MSS) CRC samples were used to train an RNA-based TBL classifier. Dichotomised DNA-based TBL data were used for survival analysis.

Results: TBL shows large variation in CRC with poor correlation to tumour mutational burden and fraction of genome altered. TBL impact on tumour biology was illustrated by the high accuracy of classifying cancers in TBL-high and TBL-low (area under the receiver operating characteristic curve [AUC]: 0.88; $p < 0.01$). High TBL was associated with disease recurrence in 85 stages II–III MSS CRCs from The Cancer Genome Atlas (hazard ratio [HR]: 6.1; $p = 0.007$) and in two independent validation series of 57 untreated stages II–III (HR: 4.1; $p = 0.012$) and 74 untreated stage II MSS CRCs (HR: 2.4; $p = 0.01$).

Conclusion: TBL is a prognostic biomarker in patients with non-metastatic MSS CRC with great potential to be implemented in routine molecular diagnostics.

1. Introduction

Colorectal cancer (CRC) is the third most common and second leading cause of cancer-related deaths worldwide [1,2]. More than 80% of CRCs show chromosomal instability (CIN), whereas approximately 15%–20% display microsatellite instability (MSI) [1,2]. CIN is characterised by an increased rate of somatic copy number aberrations (SCNAs) and chromosomal rearrangement-associated structural variants (SVs) [3–6]. MSI is characterised by a hypermutable phenotype of short repetitive sequences in the genome due to an impaired DNA mismatch repair mechanism [4,7,8]. MSI cancers and cancers with mutations in DNA polymerase-ε (POLE) or DNA polymerase-δ catalytic subunit 1 (POLD1) have a high tumour mutational burden (TMB), a measure for the number of single/simple nucleotide variants (SNVs). This discerns MSI from microsatellite stable (MSS) CRCs [9,10]. MSI CRCs are associated with a relatively good prognosis for stage II localised disease but tend to have a worse prognosis when metastasised and not treated with immune checkpoint inhibitors [8,11–13].

For the large proportion of patients with non-metastatic (stages I–III) MSS CRC, the risk of disease recurrence is currently estimated based on stage of disease. Although staging is helpful to stratify groups of patients in general, within each stage, it remains unclear for individual patients who have a relatively good prognosis and who are at risk of disease recurrence [1,14]. The vast majority of these cancers are characterised by CIN, a consequence of mitotic chromosome segregation errors and DNA double-strand break (DSB) repair defects. Segregation errors result in numerical alterations, such as whole-chromosome gains and losses [10,15]. DSB repair defects result in SVs comprising of deletions, insertions, duplications, inversions, and chromosomal translocations [10,15,16]. The fraction of genome altered (FGA) is a measure of the total amount of SCNAs, describing the extent of numerical alterations [17–21]. A high FGA has been reported to be associated with poor prognosis in stage II colon cancer [20] and metastatic CRC [19,21], indicating the potential prognostic impact of FGA. However, although FGA is representative of the extent of aneuploidy, it fails to capture the amount of structural variation.

The detection of all SVs in a given tumour sample requires high-resolution data such as deep-coverage whole genome sequencing (WGS) data. Although the use of WGS is gaining track in cancer diagnostics, the availability of WGS data from clinically well-defined cohorts of patients with non-metastatic CRC is still lacking [22]. In contrast, SCNA profiling is a widely established approach to characterise chromosomal alterations in cancer [23]. Computational methods detecting SCNAs also yield genomic information about the shift in DNA copy number levels within chromosome arms, which are indicative of SCNA-associated (i.e. unbalanced) SVs [23,24]. In the present study, we defined the tumour break load (TBL) as a measure of the number of SCNA-associated SVs and aimed to characterise the biological and clinical impact of the TBL in MSS CRC.

2. Materials and methods

2.1. Patient cohorts and data

Public data from The Cancer Genome Atlas (TCGA) [25] and segmented copy number and clinical data from two additional validation series of CRC patients were used in this study [20,26]. Specifically, processed

Affymetrix SNP 6.0 array (n = 633), whole-exome sequencing (n = 521), RNA sequencing (n = 462) and clinicopathological data including MSI status were retrieved for 633 colorectal adenocarcinoma (COADREAD) samples [9]. Updated clinical data for survival analysis [27], consensus molecular subtype (CMS) classification data [28] and curated oncogenic pathway alteration states [29] were retrieved from other publications that analysed TCGA COADREAD data. In the present study, the subset of MSS CRC cases with a low TMB (<300 SNVs; i.e. also wildtype for *POLE* and *POLD1*) will be further referred to as 'MSS CRC' (see also Supplementary Materials and Methods). An overview of the data flow is shown in Fig. S1.

2.2. Calculation of the TBL

The TBL was defined as the sum of SCNA-associated SVs, that is, the sum of unbalanced somatic chromosomal breaks per tumour sample (Table S1). The detection and filtering of somatic chromosomal breaks and subsequent calculation of the TBL are described in the Supplementary Materials and Methods.

2.3. RNA-based classification models

TCGA RNA gene expression data were used to classify genomic instability states, for example, MSI status, TBL status, and FGA status. As the TBL is a continuous value, a selection of the upper and lower quantiles of the TBL distribution, denoted as 'high TBL' (>75%) and 'low TBL' (<25%), was made as predefined class labels. The TBL model was trained with a random forest model using a 65%/35% train-test split with a ten-time repeated fivefold CV loop. To assess the performance of the model, a permutation test was performed. A Wilcoxon signed-rank test was used to assess the statistical significance of TBL between subgroups. A full description of the model setup and MSI and FGA models is provided in the Supplementary Materials and Methods.

2.4. Survival analysis

To assess the prognostic impact of the TBL, a threshold to divide tumour samples of clinical series into 'TBL-high' and 'TBL-low' was determined. Kaplan–Meier analysis and Cox proportional hazards regression models were applied to assess the association of the predefined dichotomised TBL states (Table S1), the TBL expression profiles, CMS classes and the altered oncogenic pathway states with disease-free survival or time to recurrence data (see also Supplementary Materials and Methods).

2.5. Data availability

The data generated in this study are available within this article, its supplementary data files, and the code and data repositories. All code and models used for the analysis described in this study are available at https://codeocean.com/capsule/1605813/tree/ and https://github.com/ibivu/Tumor_Break_Load.

3. Results

3.1. TBL is a distinct feature of genomic instability in CRC

The TBL is a measure for the number of SCNA-associated SVs, that is, unbalanced chromosomal breaks. As a first step, we aimed to obtain insight into the variability of TBL among CRC samples. Therefore, the prevalence of SCNA-associated SVs was determined in 633 colorectal adenocarcinoma cases (COADREAD). Segmented DNA copy number data from patient normal and tumour tissue were used to determine cutoff values that allowed to remove the far majority of germline events, background noise and artefacts while retaining most somatic events (Fig. S2). In this way, the prevalence of (unbalanced) SVs could be estimated per sample, the sum of which is expressed as the TBL (Table S1). Example SCNA profiles with a low TBL (TBL: 19) and high TBL (TBL: 151) are shown in Fig. 1A. The median TBL for the 633 TCGA COADREAD samples was 47 and ranged from 0 to 337 SCNA-associated SVs per sample (Fig. 1B). CRC samples that were MSI or hypermutators due to mutations in *POLE* or *POLD1* showed a high TMB (TMB: median 3.05 [1.83–4.02]) and a relatively low TBL (TBL: median 19 [0–138]). In contrast, non-hypermutated MSS CRC samples showed large variation in TBL (TBL: median 54 [1–337]; Fig. 1C). These data show that the TBL is a highly variable feature in MSS CRC.

Next, we examined the relationship between TBL and other features of genomic instability. TBL was compared with TMB as a measure for the number of SNVs and to FGA as a measure for the abundance of SCNAs. When focussing on the subset of non-hypermutated MSS cases (Fig. 1C), a poor correlation between the TBL and TMB was observed (R^2: 0.057; Fig. 2A). Interestingly, a poor correlation was also observed between the TBL and FGA (R^2: 0.072; Fig. 2B). These data indicate that TBL and FGA, while both characteristics of CIN, each represent distinct features of genomic instability.

3.2. TBL has a large impact on CRC biology

Having identified TBL as a distinct feature of genomic instability, we aimed to investigate its impact on CRC

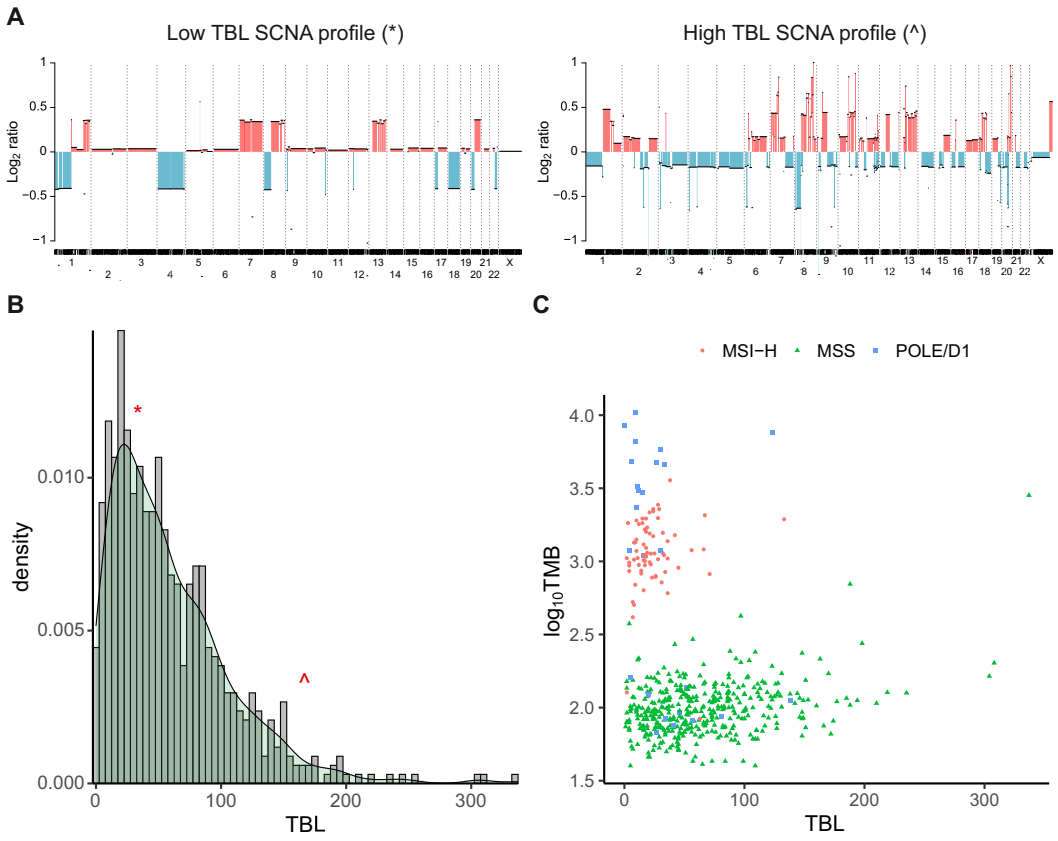

Fig. 1. **Characterisation of the tumour break load (TBL) in colorectal cancer.** (A) Segmented SNP 6.0 array DNA copy number profiles for a sample with a low TBL (TBL: 19; ∗) and a high TBL (TBL: 151; ^). DNA copy number represented as the Log_2 ratio. Gain in copy number indicated as red and loss in copy number indicated as blue. (B) Density histogram depicting the TBL distribution of 633 TCGA COADREAD cancer samples. The TBL value of the high and low copy number profiles depicted in panel A are marked by ^ and ∗, respectively. (C) Scatter plot of the TBL against the log_{10} tumour mutational burden (TMB) from 521 colorectal cancer samples for which both SNV and SCNA data were available. The dots indicate CRC samples that are microsatellite instable (MSI-H; red), CRC samples with a mutation in POLE or POLD1 (POLE/D1; blue) and non-hypermutated CRC samples that are microsatellite stable (MSS; green). CRC, colorectal cancer; POLD, polymerase-δ; POLE, polymerase-ε; SCNA, somatic copy number aberration; SNV, single/simple nucleotide variant; TCGA, The Cancer Genome Atlas; TMB, tumour mutational burden. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

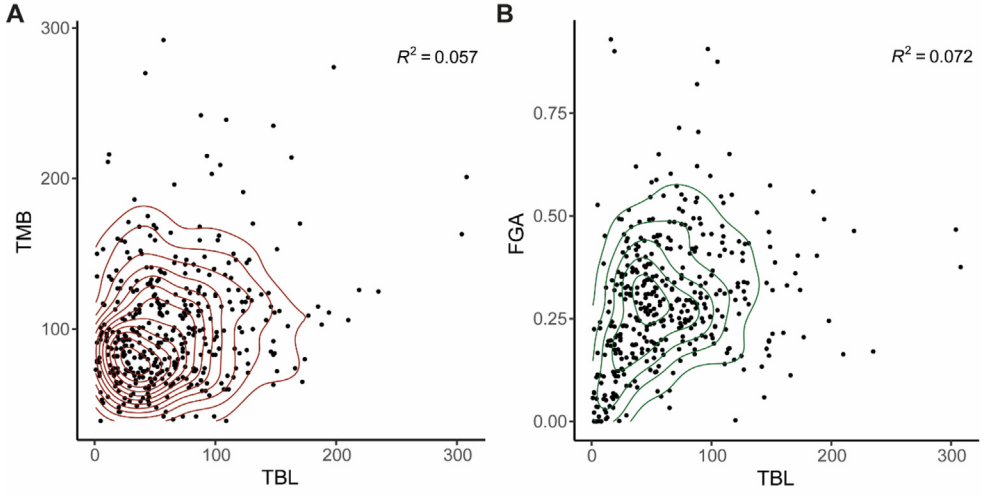

Fig. 2. **Poor correlation of TBL with other genomic instability features.** Density scatterplot of the tumour break load (TBL) plotted against tumour mutational burden (TMB) (A) and fraction genome altered (FGA) (B) in 401 non-hypermutated MSS CRC samples. The correlation was reported with the Pearson correlation coefficient (R^2). CRC, colorectal cancer; MSS, microsatellite stable.

biology. We hypothesise that if the observed variation in TBL is associated with distinct CRC biology, gene expression profiles could be used to predict which CRC cases have a high TBL (TBL-high) or a low TBL (TBL-low). TCGA RNA gene expression data were used to create a classification model that differentiates between TBL-high and TBL-low samples. First, this strategy was verified by demonstrating that MSI status could be accurately classified (Fig. S4A). To apply the RNA gene expression classification strategy on a continuous label such as the TBL, TBL-high and TBL-low groups were preselected using the upper 75% and lower 25% quantiles of the TBL distribution. These quantiles were selected as predefined classes to represent the most extreme TBL phenotypes. The trained TBL classifier could classify an independent test set of tumours in TBL-high and TBL-low expression profiles with high accuracy (AUC of 0.88; $p < 0.01$; Fig. 3A and B), an accuracy that was higher than observed for the classification of the FGA status (AUC of 0.83; $p < 0.01$; Fig. S4B). When the TBL classification model was applied to all MSS CRC samples (336 cases), there was a significant difference in TBL between TBL-low (median: 20) and TBL-high (median: 74; $p < 0.0001$; Fig. 3B). When the classification model was applied to the previously excluded MSI and POLE/D1 mutant samples, 86% of these samples were classified in the TBL-low group (Figs. S4C and D), verifying that the model captures the variation associated with the TBL. The predictive value of the TBL expression profiles to classify TBL-low from TBL-high samples shows that there is a clear biological distinction between CRC samples.

To understand the underlying biological processes associated with the TBL classification, an enrichment analysis was performed between the predicted TBL-high and TBL-low states and reported curated oncogenic pathway alteration states [29]. The TBL-high expression profile is associated with an altered TP53 pathway affected by mutations or SCNAs ($p < 0.05$), whereas the TBL-low expression profile is associated with the altered rat sarcoma virus (RAS), transforming growth factor-β (TGF-β) and phosphatidylinositol-3-kinase (PI3K) pathways ($p < 0.05$; Fig. 3C). The predicted TBL-high and TBL-low states did not show significant variation in immune cell composition (Fig. S5, Table S2).

Clinically, the TBL expression profiles are distinct from pathological stages ($p < 0.001$; Fig. S6A) and the CMS classification ($p < 0.0001$; Table S3). The TBL-high expression profile shows a strong association with disease recurrence in stages II–III MSS CRC (hazard ratio [HR] = 5.19; $p < 0.02$; Fig. S6B), which is in this data set more profound than the differences observed for CMS classification (CMS: HR = 1.29; $p = 0.7$; Fig. S6C–D) [28].

3.3. TBL is a prognostic biomarker in non-metastatic CRC

High levels of CIN are associated with tumour relapse and metastasis [6,30,31]. As we have demonstrated that TBL is a distinct feature of CIN with an impact on tumour biology, we aimed to assess whether the TBL itself is a prognostic biomarker for disease recurrence in patients with MSS CRC. Following dichotomisation of CRC samples into TBL-high and TBL-low groups, TCGA CRC patients that were TBL-high showed a shorter disease-free survival compared with patients in the TBL-low group (HR = 6.1; $p < 0.01$; Fig. 4A). Although these data imply that TBL has prognostic value, the TCGA patient data are heterogeneous and

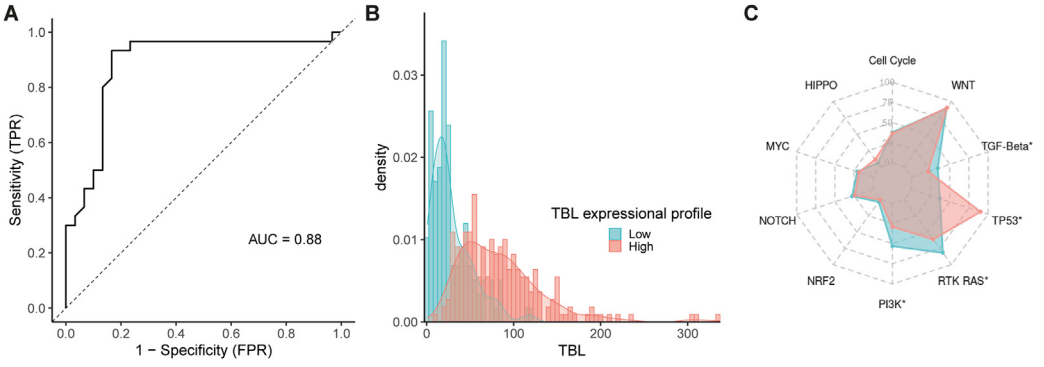

Fig. 3. **Biological impact of the tumour break load (TBL).** (A) Receiver operating characteristic (ROC) curve of the 35% test set of the TBL-high (75% quantile) and TBL-low (25% quantile) classifier trained on 168 TCGA MSS COADREAD samples. Area under the ROC curve (AUC) of the TBL classifier is 0.88, permutation p-value < 0.01. (B) The TBL distribution of the predicted TBL-high (red) and TBL-low (blue) expression profiles of 336 MSS CRC cases. The TBL distributions are significantly different ($p < 0.0001$) assessed with a Wilcoxon signed-rank test. (C) Spider plot of the percentage of predicted TBL-high (red) and TBL-low (blue) expression profile samples associated with an altered oncogenic pathway affected by SNVs or SCNAs. Statistical significance of the enrichment of TBL expression profiles with the altered oncogenic pathways was assessed with a Fisher's exact test. $*p \leq 0.05$. CRC, colorectal cancer; MSS, microsatellite stable; SCNA, somatic copy number aberration; SNV, single/simple nucleotide variant; TCGA, The Cancer Genome Atlas. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

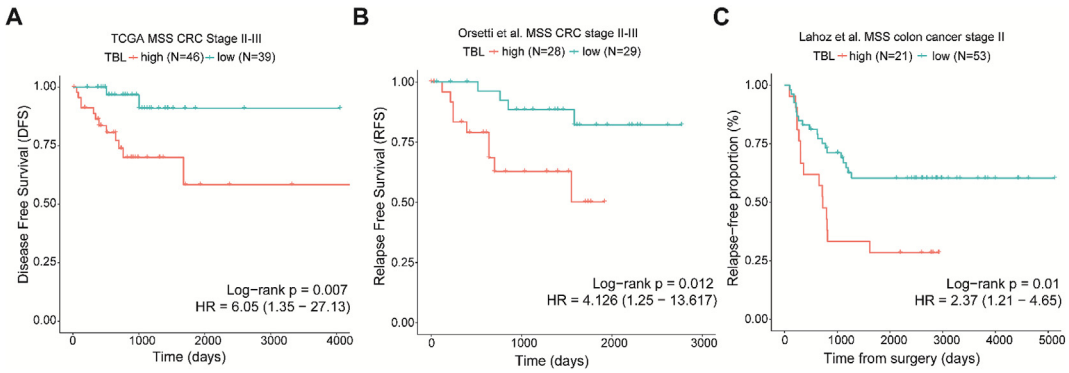

Fig. 4. **The prognostic impact of the TBL in non-metastatic CRC.** (A) Kaplan–Meier DFS curve for 85 TCGA stage II and stage III MSS CRC samples stratified in high and low TBL. (B) Kaplan–Meier curve of the relapse-free survival (RFS) for 57 untreated stages II–III MSS CRC samples from Orsetti B et al. [26] stratified in high and low TBL. (C) Kaplan–Meier curve of the relapse-free proportion for 74 stage II MSS colon cancer patients from Lahoz S et al. [20] stratified in high and low TBL. Statistical significance for the difference in survival has been assessed with a log-rank test. The hazard ratio (HR) has been assessed with a univariate Cox regression model. CRC, colorectal cancer; MSS, microsatellite stable; TBL, the tumour break load; TCGA, The Cancer Genome Atlas.

include patients who received adjuvant treatment. Therefore, to validate this finding, independent cohorts of untreated patients with stage II and stage III MSS CRC were analysed. In one cohort of 57 stage II and stage III untreated MSS CRC patients [26], TBL-high was associated with poor relapse-free survival (HR = 4.1; $p = 0.012$; Fig. 4B). Similarly, in a second cohort consisting of 74 chemotherapy-naïve stage II MSS colon cancer patients [20], TBL-high was associated with poor relapse-free proportion (HR = 2.4; $p = 0.01$; Fig. 4C). The findings were retained in a multivariate analysis including clinicopathological features for the TCGA and Orsetti et al. cohort (Fig. S7). In summary, these results show that TBL is a prognostic biomarker for disease recurrence in non-metastatic MSS CRC.

4. Discussion

At present, there are several classification methods to define CRC subtypes, for example, MSI/MSS status [7], CMS classification [28], TMB and FGA [20,21] and clinicopathological features [32]. However, none of these methods can make an optimal clinical distinction for the risk of disease recurrence in the large proportion of patients with non-metastatic MSS CRC [1,14].

In the present study, we introduced and assessed the novel feature TBL as a measure for the number of (SCNA-associated) SVs and aimed to characterise its biological and clinical impact in CRC. We showed that the TBL is a distinct feature of genomic instability, reflected by its poor correlation with the FGA and TMB. We demonstrated that the TBL has a strong impact on tumour biology by showing that the TBL status can be predicted with high accuracy from RNA gene expression data. Finally, we showed that a high TBL is associated with poor prognosis in untreated stages II–III CRC, implying that the TBL is a prognostic biomarker that may improve the estimation of an individual's risk of disease recurrence and support decision-making whether or not to receive adjuvant treatment.

TBL was determined as the sum of SCNA-associated SVs, which implies there is a dependency between the TBL and the abundance of SCNAs represented by the FGA. Nevertheless, when focussing on the non-hypermutated MSS CRCs, there was a poor correlation between the TBL and the FGA, indicating that DSB repair deficiency problems and chromosome segregation errors can and should be examined as separate features of CIN [10,15,16]. Using an RNA gene expression-based classification strategy, TBL-high CRCs could be discriminated from TBL-low CRCs, demonstrating their distinct biology within MSS CRC. A comparison of the TBL classification to the well-known CMS classification showed that the TBL-high expression profile is enriched among CMS2 and CMS4, the two CMS subtypes that are characterised by high levels of SCNAs. CMS4 has been reported to have a worse prognosis than CMS2 [28]. In the TCGA MSS stage II–III cohort analysed in this study, the difference in disease recurrence between TBL-high and TBL-low in non-metastatic MSS CRC was more pronounced. Overall, these results prompt further investigation to validate whether the TBL captures the risk for disease recurrence in untreated stages II–III MSS CRC better than the currently frequently used approach of CMS classification.

The present study has several limitations. This is a retrospective study with limited availability of clinicopathological data. The sample size of each of the cohorts analysed is relatively small. It is therefore important that we were able to explore the prognostic value of the TBL in three independent cohorts and observed strong prognostic value in each of these cohorts. Moreover, we were able to perform a multivariate analysis for the

TCGA cohort, for which multiple clinicopathological variables were available. In this multivariate analysis, we observed that the TBL was the feature with the strongest prognostic value.

The strong prognostic effect of TBL on disease recurrence suggests an important role for double-strand breaks in the biological and clinical behaviour of CRC. One molecular process that may be involved is the mediation of the cGAS-STING antiviral pathway. The cGAS-STING pathway might be induced by the high number of SCNA-associated double-strand DNA breaks in TBL-high samples. Activation of the cGAS-STING pathway subsequently activates the nuclear factor kappa B (NF-κB) pathway downstream of STING, which can facilitate migration, invasion and metastasis [6,30,31]. In this way, a high TBL could intrinsically be associated with an increased likelihood of the development of (micro) metastases and early relapse in patients with resected primary CRC [33].

Currently, the prognosis of patients with non-metastatic CRC is mostly based on clinicopathological features. There is an increasing interest in the use of genomic information for determining the risk of disease recurrence; however, due to the lack of robust procedures for classification based on gene expression signatures, its clinical implementation is lagging behind [32]. In addition, in recent years, tremendous progress has been made in the detection of post-surgery cell-free circulating tumour DNA (ctDNA) as a biomarker for minimal residual disease in stages II and III CRC [33–36]. Patients with a post-surgery ctDNA-positive test result who do not receive adjuvant treatment have a very high chance of disease recurrence of more than 80%. However, the sensitivity of ctDNA-testing is currently not sufficient to detect all patients who develop a recurrence. For future investigations, it will be very interesting to explore the added value of tumour tissue prognostic biomarkers in patients with a post-surgery ctDNA-negative test who developed recurrences [35].

We here demonstrated that a high TBL is associated with poor prognosis in patients with resected stages II and III MSS CRC. The TBL is a DNA-derived feature that can be obtained from SCNA profiling data from various platforms with different resolutions, such as array comparative genomic hybridisation, SNP6 arrays and low-coverage WGS. For clinical applicability of the TBL as a prognostic biomarker, a relative distinction between TBL-high and TBL-low is expected to be sufficient, for which a precise determination of the exact number and location of chromosomal breaks is not required. We therefore expect that TBL classification can be performed in a robust manner using a wide range of platforms that are currently used to generate SCNA data.

In conclusion, the TBL is a prognostic biomarker that has the potential to be implemented for patients with non-metastatic CRC, pending further validation. Further assay development and prospective validation studies are required to develop the TBL measure into a robust biomarker that can be applied in a routine molecular diagnostic setting.

Funding

This collaboration project is co-funded by PPP Allowance (grants LSHM19027 and LSHM21018) made available by Health ~ Holland, Top Sector Life Sciences & Health, to stimulate public–private partnerships.

Authors' contributions

S.Lakbir:. contributed to conceptualisation, formal analysis, validation, investigation, visualisation, methodology, writing the original draft, and software. S.Lahoz:. contributed to resources, validation, visualisation and reviewing the article. M.C. contributed to resources and reviewing the article. J.C. contributed to resources, validation, and reviewing the article. R.A.G. contributed to methodology and reviewing the article. J.H. and G.A.M. contributed to conceptualisation, supervision, reviewing the article, and funding acquisition. S.A. contributed to conceptualisation, supervision, funding acquisition, methodology, writing and reviewing the article, and formal analysis. R.J.A.F. contributed to conceptualisation, supervision, funding acquisition, writing and reviewing the article, and methodology.

Conflict of interest statement

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

R.J.A.F. reports grants and non-financial support from Personal Genome Diagnostics, non-financial support from Delfi Diagnostics, grants from MERCK BV, grants and non-financial support from Cergentis BV, outside the submitted work; In addition, R.J.A.F. has several patents pending. S.A. reports grants and non-financial support from Cergentis BV and a patent pending outside the submitted work. S.Lakbir. reports non-financial support from Cergentis BV and a patent pending outside the submitted work. J.H. reports a patent pending outside the submitted work. G.A.M. reports non-financial support from Exact Sciences, non-financial support from Sysmex, non-financial support from Sentinel CH. SpA, non-financial support from Personal Genome Diagnostics (PGDX), other from Hartwig Medical Foundation, grants from CZ (OWM Centrale Zorgverzekeraars groep Zorgverzekeraar u.a),

other from Royal Philips, other from GlaxoSmithKline, other from Keosys SARL, other from OpenClinica LLC, other from Roche Diagnostics Nederland BV, other from The Hyve BV, other from Open Text, other from SURFSara BV, other from Vancis BV, other from CSC Computer Sciences BV, outside the submitted work; In addition, G.A.M. has several patents pending. The other authors declare no potential conflicts of interest.

Acknowledgements

The results shown here are in part based upon data generated by the Cancer Genome Atlas Research Network: https://www.cancer.gov/tcga.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2022.09.034.

References

[1] Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet 2019;394:1467–80. https://doi.org/10.1016/S0140-6736(19)32319-0.
[2] Granados-Romero JJ, Valderrama-Treviño AI, Contreras-Flores EH, Barrera-Mera B, Herrera Enríquez M, Uriarte-Ruíz K, et al. Colorectal cancer: a review. Int J Res Med Sci 2017;5:4667. https://doi.org/10.18203/2320-6012.ijrms20174914.
[3] Shen Z. Genomic instability and cancer: an introduction. J Mol Cell Biol 2011;3:1–3. https://doi.org/10.1093/jmcb/mjq057.
[4] Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature 2013;501:338–45. https://doi.org/10.1038/nature12625.
[5] Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. Gastrointest Cancer Res 2012;5:19–27.
[6] Bakhoum SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and its microenvironment. Cell 2018;174:1347–60. https://doi.org/10.1016/j.cell.2018.08.027.
[7] Yamamoto H, Watanabe Y, Maehata T, Imai K, Itoh F. Microsatellite instability in cancer: a novel landscape for diagnostic and therapeutic approach. Arch Toxicol 2020;94:3349–57. https://doi.org/10.1007/s00204-020-02833-z.
[8] Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. Nat Commun 2017;8:15180. https://doi.org/10.1038/ncomms15180.
[9] Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, et al. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–7. https://doi.org/10.1038/nature11252.
[10] Sansregret L, Vanhaesebroeck B, Swanton C. Determinants and clinical implications of chromosomal instability in cancer. Nat Rev Clin Oncol 2018;15:139–50. https://doi.org/10.1038/nrclinonc.2017.198.
[11] Taieb J, Shi Q, Pederson L, Alberts S, Wolmark N, Van Cutsem E, et al. Prognosis of microsatellite instability and/or mismatch repair deficiency stage III colon cancer patients after disease recurrence following adjuvant treatment: results of an ACCENT pooled analysis of seven studies. Ann Oncol Off J Eur Soc Med Oncol 2019;30:1466–71. https://doi.org/10.1093/annonc/mdz208.
[12] McNamara MG, Jacobs T, Lamarca A, Hubner RA, Valle JW, Amir E. Impact of high tumor mutational burden in solid tumors and challenges for biomarker application. Cancer Treat Rev 2020;89:102084. https://doi.org/10.1016/j.ctrv.2020.102084.
[13] André T, Shiu K-K, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. N Engl J Med 2020;383:2207–18. https://doi.org/10.1056/NEJMoa2017699.
[14] Xu W, He Y, Wang Y, Li X, Young J, Ioannidis JPA, et al. Risk factors and risk prediction models for colorectal cancer metastasis and recurrence: an umbrella review of systematic reviews and meta-analyses of observational studies. BMC Med 2020;18:172. https://doi.org/10.1186/s12916-020-01618-6.
[15] Jin N, Burkard ME. MACROD2, an original cause of CIN? Cancer Discov 2018;8:921–3. https://doi.org/10.1158/2159-8290.CD-18-0674.
[16] Venkatesan S, Natarajan AT, Hande MP. Chromosomal instability—mechanisms and consequences. Mutat Res Toxicol Environ Mutagen 2015;793:176–84. https://doi.org/10.1016/j.mrgentox.2015.08.008.
[17] Nguyen B, Sanchez-Vega F, Fong CJ, Chatila WK, Boroujeni AM, Pareja F, et al. The genomic landscape of carcinomas with mucinous differentiation. Sci Rep 2021;11:9478. https://doi.org/10.1038/s41598-021-89099-2.
[18] Pikor L, Thu K, Vucic E, Lam W. The detection and implication of genome instability in cancer. Cancer Metastasis Rev 2013;32:341–52. https://doi.org/10.1007/s10555-013-9429-5.
[19] Mehta KR, Nakao K, Zuraek MB, Ruan DT, Bergsland EK, Venook AP, et al. Fractional genomic alteration detected by array-based comparative genomic hybridization independently predicts survival after hepatic resection for metastatic colorectal cancer. Clin Cancer Res 2005;11:1791–7. https://doi.org/10.1158/1078-0432.CCR-04-1418.
[20] Lahoz S, Archilla I, Asensio E, Hernández-Illán E, Ferrer Q, López-Prades S, et al. Copy-number intratumor heterogeneity increases the risk of relapse in chemotherapy-naive stage II colon cancer. J Pathol 2022;257(1). https://doi.org/10.1002/path.5870.
[21] Smeets D, Miller IS, O'Connor DP, Das S, Moran B, Boeckx B, et al. Copy number load predicts outcome of metastatic colorectal cancer patients receiving bevacizumab combination therapy. Nat Commun 2018;9:4112. https://doi.org/10.1038/s41467-018-06567-6.
[22] Roepman P, de Bruijn E, van Lieshout S, Schoenmaker L, Boelens MC, Dubbink HJ, et al. Clinical validation of whole genome sequencing for cancer diagnostics. J Mol Diagn 2021;23:816–33. https://doi.org/10.1016/j.jmoldx.2021.04.011.
[23] van den Broek E, van Lieshout S, Rausch C, Ylstra B, van de Wiel MA, Meijer GA, et al. GeneBreak: detection of recurrent DNA copy number aberration-associated chromosomal breakpoints within genes. F1000Research 2016;5:2340. https://doi.org/10.12688/f1000research.9259.1.
[24] Van Den Broek E, Dijkstra MJJ, Krijgsman O, Sie D, Haan JC, Traets JJH, et al. High prevalence and clinical relevance of genes affected by chromosomal breaks in colorectal cancer. PLoS One 2015;10. https://doi.org/10.1371/journal.pone.0138141.
[25] Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N Engl J Med 2016;375:1109–12. https://doi.org/10.1056/NEJMp1607591.
[26] Orsetti B, Selves J, Bascoul-Mollevi C, Lasorsa L, Gordien K, Bibeau F, et al. Impact of chromosomal instability on colorectal cancer progression and outcome. BMC Cancer 2014;14:121. https://doi.org/10.1186/1471-2407-14-121.
[27] Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell 2018;173:400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052.

[28] Guinney J, Dienstmann R, Wang X, De Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350–6. https://doi.org/10.1038/nm.3967.

[29] Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling pathways in the cancer genome Atlas. Cell 2018;173:321–337.e10. https://doi.org/10.1016/j.cell.2018.03.035.

[30] Motwani M, Pesiridis S, Fitzgerald KA. DNA sensing by the cGAS–STING pathway in health and disease. Nat Rev Genet 2019;20:657–74. https://doi.org/10.1038/s41576-019-0151-1.

[31] Bakhoum SF, Ngo B, Laughney AM, Cavallo J-A, Murphy CJ, Ly P, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. Nature 2018;553:467–72. https://doi.org/10.1038/nature25432.

[32] Koncina E, Haan S, Rauh S, Letellier E. Prognostic and predictive molecular biomarkers for colorectal cancer: updates and challenges. Cancers 2020;12:319. https://doi.org/10.3390/cancers12020319.

[33] Henriksen TV, Tarazona N, Reinert T, Carbonell-Asins JA, Renner D, Sharma S, et al. Circulating tumor DNA analysis for assessment of recurrence risk, benefit of adjuvant therapy, and early relapse detection after treatment in colorectal cancer patients. J Clin Oncol 2021;39. https://doi.org/10.1200/JCO.2021.39.3_suppl.11. 11–11.

[34] Marcuello M, Vymetalkova V, Neves RPL, Duran-Sanchon S, Vedeld HM, Tham E, et al. Circulating biomarkers for early detection and clinical management of colorectal cancer. Mol Asp Med 2019;69:107–22. https://doi.org/10.1016/j.mam.2019.06.002.

[35] Tie J, Cohen JD, Lahouel K, Lo SN, Wang Y, Kosmider S, et al. Circulating tumor DNA analysis guiding adjuvant therapy in stage II colon cancer. N Engl J Med 2022;386:2261–72. https://doi.org/10.1056/NEJMoa2200075.

[36] Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med 2016;8. https://doi.org/10.1126/scitranslmed.aaf6219. 346ra92–346ra92.