Journal of Pathology
J Pathol 2023
Published online 15 November 2023 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.6228

ORIGINAL ARTICLE

Genetic and immune landscape evolution in MMR-deficient colorectal cancer

Benjamin R Challoner[1], Andrew Woolston[2], David Lau[3], Marta Buzzetti[2], Caroline Fong[3], Louise J Barber[2], Gayathri Anandappa[3], Richard Crux[3], Ioannis Assiotis[1], Kerry Fenwick[1], Ruwaida Begum[3], Dipa Begum[1,3], Tom Lund[1,3], Nanna Sivamanoharan[1,3], Harold B Sansano[3], Melissa Domingo-Arada[3], Amina Tran[3], Hardev Pandha[4], David Church[5], Bryony Eccles[6], Richard Ellis[7], Stephen Falk[8], Mark Hill[9], Daniel Krell[10], Nirupa Murugaesu[11,12], Luke Nolan[13], Vanessa Potter[14], Mark Saunders[15], Kai-Keen Shiu[16], Sebastian Guettler[1], James L Alexander[17], Héctor Lázare-Iglesias[18], James Kinross[17], Jamie Murphy[17], Katharina von Loga[1,3], David Cunningham[3], Ian Chau[3], Naureen Starling[3], Juan Ruiz-Bañobre[18,19], Tony Dhillon[20] and Marco Gerlinger[2,21]*

[1] *The Institute of Cancer Research, London, UK*
[2] *Barts Cancer Institute, Queen Mary University of London, London, UK*
[3] *The Royal Marsden NHS Foundation Trust, London, UK*
[4] *University of Surrey, Guildford, UK*
[5] *Oxford University Hospitals NHS Foundation Trust, Oxford, UK*
[6] *University Hospitals Dorset NHS Foundation Trust, Bournemouth, UK*
[7] *Royal Cornwall Hospitals NHS Trust, Truro, UK*
[8] *University Hospitals Bristol NHS Foundation Trust, Bristol, UK*
[9] *Maidstone and Tunbridge Wells NHS Trust, Maidstone, UK*
[10] *Royal Free London NHS Foundation Trust, London, UK*
[11] *St George's University Hospitals NHS Foundation Trust, London, UK*
[12] *Genomics England, London, UK*
[13] *Hampshire Hospitals NHS Foundation Trust, Winchester, UK*
[14] *University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK*
[15] *The Christie NHS Foundation Trust, Manchester, UK*
[16] *University College London Hospitals NHS Foundation Trust, London, UK*
[17] *Imperial College Healthcare NHS Trust, London, UK*
[18] *University Clinical Hospital of Santiago de Compostela, Santiago de Compostela, Spain*
[19] *University of Santiago de Compostela, Santiago de Compostela, Spain*
[20] *Royal Surrey Hospital NHS Foundation Trust, Guildford, UK*
[21] *St Bartholomew's Hospital Cancer Centre, London, UK*

*Correspondence to: M Gerlinger, Barts Cancer Institute, Queen Mary University of London, St Bartholomew's Hospital Cancer Centre, London, UK.
E-mail: m.gerlinger@qmul.ac.uk

Abstract

Mismatch repair-deficient (MMRd) colorectal cancers (CRCs) have high mutation burdens, which make these tumours immunogenic and many respond to immune checkpoint inhibitors. The MMRd hypermutator phenotype may also promote intratumour heterogeneity (ITH) and cancer evolution. We applied multiregion sequencing and CD8 and programmed death ligand 1 (PD-L1) immunostaining to systematically investigate ITH and how genetic and immune landscapes coevolve. All cases had high truncal mutation burdens. Despite pervasive ITH, driver aberrations showed a clear hierarchy. Those in WNT/β-catenin, mitogen-activated protein kinase, and TGF-β receptor family genes were almost always truncal. Immune evasion (IE) drivers, such as inactivation of genes involved in antigen presentation or IFN-γ signalling, were predominantly subclonal and showed parallel evolution. These IE drivers have been implicated in immune checkpoint inhibitor resistance or sensitivity. Clonality assessments are therefore important for the development of predictive immunotherapy biomarkers in MMRd CRCs. Phylogenetic analysis identified three distinct patterns of IE driver evolution: pan-tumour evolution, subclonal evolution, and evolutionary stasis. These, but neither mutation burdens nor heterogeneity metrics, significantly correlated with T-cell densities, which were used as a surrogate marker of tumour immunogenicity. Furthermore, this revealed that genetic and T-cell infiltrates coevolve in MMRd CRCs. Low T-cell densities in the subgroup without any known IE drivers may indicate an, as yet unknown, IE mechanism. PD-L1 was expressed in the tumour microenvironment in most samples and correlated with T-cell densities. However, PD-L1 expression in cancer cells was independent of T-cell densities but strongly associated with loss of the intestinal homeobox transcription factor CDX2. This explains infrequent PD-L1 expression by cancer cells and may contribute to a higher recurrence risk of MMRd CRCs with impaired CDX2 expression.

Keywords: cancer evolution; colorectal cancer; mismatch repair deficiency; biomarkers; PD-L1; immunotherapy

Received 23 June 2023; Revised 17 September 2023; Accepted 10 October 2023

Conflict of interest statement: MG receives research funding and speaker/advisory board fees from Merck KG, Bristol Myers Squibb, and Roche. DCu receives research funding from Astra Zeneca, Clovis, Eli Lilly, 4SC, Bayer, Celegene, and Roche and is on the advisory board for OVIBIO. DL is the recipient of the Australasian Gastro-Intestinal Trials Group/Merck Clinical Research Fellowship. NM is on the advisory board for Bristol Myers Squibb, Novartis, Pfizer, and Roche and the Speakers' Bureau for Merck, Pfizer, and Roche. JR-B receives travel and accommodation expenses from Bristol Myers Squibb, Merck Sharp & Dohme, Ipsen, PharmaMar, and Roche; honoraria for educational activities from Pfizer and Roche; honoraria for consultancy from Boehringer Ingelheim; and institutional research funding from Roche. KS receives travel, accommodation, national/international meetings registration, and consultancy renumeration from Bristol Myers Squibb, Daiichi-Sankyo, Guardant Health, Innovent Biologics, Merck KG, Mirati Therapeutics, Merck Sharp & Dohme, Roche, Servier, and Institutional funding for trials and research at UCLH from Adaptimmune Therapeutics, Bristol Myers Squibb, Merck KG, Merck Sharp & Dohme, and Roche.

Introduction

Loss of DNA mismatch repair (MMR) proteins (MLH1, PMS2, MSH2, or MSH6) in MMR-deficient (MMRd) colorectal cancers (CRCs) establishes a hypermutator phenotype and high mutation burdens. The large number of mutation-encoded neoantigens make these cancers highly immunogenic [1]. Following surgical resection, stage 1/2 MMRd CRCs have a lower recurrence risk compared with stage-matched MMR-proficient (MMRp) CRCs, probably through higher immunogenicity. The favourable prognosis of MMRd is largely lost in tumours with lymph node metastases (Stage 3) [2]. Yet, the contribution of immune evasion (IE) mechanisms to disease progression remains uncertain [3]. Furthermore, immunotherapy with programmed cell death protein 1/programmed death ligand 1 (PD1/PD-L1) checkpoint inhibitors (CPIs) is ineffective in metastatic (stage 4) MMRp CRCs [4], but has high response rates in MMRd CRCs (43.8%) [5], confirming an important role for the PD1/PD-L1 immune checkpoint in IE. Combined PD1 and CTLA4 CPIs achieved even higher response rates in stage 4 [6] and early-stage MMRd CRCs [7]. Loss of the intestinal homeobox transcription factor CDX2 is enriched in MMRd CRCs, associated with high-grade tumours that progress and with worse survival [8,9]. There remains a major need to assess current and identify novel biomarkers to predict who benefits from CPI.

Biomarkers for CPIs in other tumour types include mutation burden, the number of insertions and deletions (indels) [10], as well as cytotoxic CD8 T-cell infiltrates and expression of PD-L1 [11]. Inactivation of genes in the IFN-γ signalling pathway or in HLA class I antigen presentation, such as *Beta-2-Microglobulin* (*B2M*), have also been associated with resistance to CPIs in several tumour types [12–15]. Data in MMRd CRCs are less clear: neither CD8 T-cells, PD-L1 expression, nor mutation numbers correlated with CPI responses [4,16]. Moreover, disrupting mutation of *B2M* was not associated with CPI resistance [17]. Recent data even showed higher response rates in MMRd CRCs with *B2M* mutations and that γδ T-cells mediate CPI responses in these [18].

Importantly, this hypermutator phenotype and the large number of small indels with high immunogenic potential [10] distinguish MMRd tumours from other CPI-sensitive cancers, such as melanoma, lung, or renal cancer. We have shown in MMRd gastro-oesophageal cancers that this promotes extreme intratumour heterogeneity (ITH) and subclonal driver evolution [19]. ITH hinders biomarker identification, as a single sample may not be representative of the entire cancer cell population [20,21]. ITH may thus explain the slow progress in identifying predictive biomarkers in MMRd CRCs, but it has not been systematically assessed. A clear understanding of which genetic drivers are commonly truncal and which biomarkers are unlikely to show ITH, and can therefore be assessed reliably from single samples, and which ones require assays competent to address ITH, may be important for biomarker development. We applied multiregion multiomics and immunohistochemistry (IHC) analysis to reveal ITH of driver aberrations, immune infiltrates, and PD-L1 expression.

Materials and methods

Fifty-five primary tumour regions, 15 lymph nodes, and one distant metastasis from 20 CRCs were sequenced with a panel of 194 genes (supplementary material, Table S1) that are recurrently mutated in MMRd or MMRp CRCs or implicated in IE and may confer CPI resistance [12,15,19,22–24]. 3'-RNA sequencing was performed on 60 samples with sufficient RNA yield and IHC for CD8; PD-L1 and STING immunostaining was performed on 68 samples. The study was performed in accordance with the Declaration of Helsinki and protocols had been approved by an ethics committee (approval numbers: 18/LO/0165, 2015/405, 14/EE0024). All patients had signed written informed consent for study participation. Details for sample collection and preparation, targeted DNA sequencing, mutation calling, HLA mutation analysis, DNA copy number analysis, identification of probable driver aberrations, linear regression model for extrapolation from targeted sequencing to exome mutation number in MMRd CRC, phylogenetic tree reconstruction, IHC, 3'-RNA sequencing, immune cell abundance analysis using single sample gene set enrichment analysis (ssGSEA), expression analysis of CRC cells lines, and statistical analyses are presented in Supplementary materials and methods.

Results

Targeted sequencing of 71 samples from 20 CRCs (Table 1) where IHC reported MMRd enabled high sequencing depths (median: 2659×). This was critical

to avoid ITH overestimation in samples with low cancer cell content, such as lymph node metastases. The median non-silent mutation number of individual tumour regions was 58 (Figure 1A, and supplementary material, Table S2). Nineteen tumours had high numbers of mutations (28–90 mutations/region). Only 10 mutations/region were identified in T20, suggesting it was MMRp. T20 was excluded after histology review showed patchy loss of MLH1/PMS2, which is a recognised feature that can cause misclassification as MMRd [25]. A median of 44 mutations were ubiquitously present in all regions of individual tumours. Thus, an average of 76% of mutations/tumour were truncal.

ITH by subtype and stage

ITH was identified in all cases (median: 16.1% heterogeneous mutations/region, Figure 1A). It was significantly higher in tumours with MLH1/PMS2 than with MSH2/MSH6 loss (Figure 1B). ITH was numerically but not significantly higher in *BRAF* V600E mutant tumours and did not differ by stage. Ubiquitous mutation numbers did not significantly differ by *BRAF* status, pattern of MMR protein expression loss, or stage (supplementary material, Figure S1). A high proportion of indels is characteristic of MMRd and a median of 60.8% of non-silent mutations were indels (Figure 1C). The higher indel fraction compared with whole-exome data from MMRd CRCs [10] is probably a consequence of over-representation of microsatellite repeats in driver genes of our panel, high sequencing depths, and the use of a contemporaneous mutation caller with increased sensitivity for indels [26]. Two tumours had low indel fractions (T6, T10). Both showed only isolated MSH6 loss. This can be explained by the predominant recognition of base–base mismatches by the MSH2/MSH6 heterodimer, whereas the MSH2/MSH3 and MLH1/PMS2 heterodimers are also important for recognition and repair of indels [27,28]. A third tumour with isolated MSH6 loss had a high indel fraction (T8). This tumour harboured an early somatic stop codon mutation in *MSH2* (E28X) in all regions. Lynch syndrome patients with *MHS2* start codon mutations and preserved MSH2 expression have been described, suggestive of a hypomorphic MSH2 isoform expressed from the M67 downstream start codon [29]. Expression of this isoform probably explains the high indel fraction despite preserved MSH2 expression.

Table 1. Patients and pathological characteristics.

	MMRd CRC cohort ($n = 20$)
Median age at resection (range)	61.2 (33.5–79.2)
Sex	
Male	30% (6)
Female	70% (14)*
Stage (AJCC/UICC 8th edition)	
1	15% (3)
2	15% (3)
3	65% (13)*
4	5% (1)
Tumour grade (WHO 2019)	
Low	75% (15)
High	25% (5)*
Predominant differentiation (WHO 2019)	
Adenocarcinoma (NOS)	70% (14)*
Mucinous	30% (6)

*T20 reclassified as MMRp following targeted gene sequencing and MMR IHC slide review.
NOS, not otherwise specified.

Comparison of mutation numbers to whole-exome sequencing studies

To compare the mutation burdens in our samples to those from published MMRd CRC whole-exome sequencing [23], we trained a linear regression model. The number of point mutations within target genes in our panel correlated significantly with mutation numbers from whole-exome sequencing in the training dataset ($r = 0.851$, supplementary material, Figure S2). Applying the derived linear model equation to our series extrapolated a median of 1,280 mutations in the exome per sample. This is similar to the whole-exome mutation numbers in MMRd CRCs from the training cohort (median: 994) and TCGA (median: 1166). The median extrapolated number of truncal mutations per case was 909 in our series. Long phylogenetic trunks are, hence, characteristic for MMRd CRCs.

Phylogenies reveal branched evolution

All tumours showed branched evolution in the phylogenetic analysis (Figure 1D and supplementary material, Figure S3). Metastases had diverged before detectable genetic diversification of the primary tumour in six of eight cases where multiple primary tumour regions and at least one metastasis were available to assess dissemination timing. Confidence for early dissemination was low in two of six cases (T13, T17) based on bootstrap values <50%. Thus, the ability to metastasise may already have been acquired on the phylogenetic trunk in at least four of eight tumours.

Driver aberration evolution is characterised by a clear hierarchy

The acquisition of new genetic drivers is arguably the most important consequence of cancer evolution, but has not been systematically assessed in MMRd CRC. We identified drivers by cataloguing hotspot mutations in known oncogenes, as well as disrupting aberrations [frameshift, splice-site, nonsense mutations, disruption of the signal peptide, and loss of heterozygosity (LOH)] that were likely to confer biallelic inactivation of tumour suppressor (TS) genes or IE driver genes (Figure 2A, the driver genes assessed are shown in supplementary material, Table S3). A hotspot mutation in an oncogene or LOH of a TS or IE gene combined with a disrupting mutation that is clonal in a sample reliably defines driver aberrations. Yet, many TS and IE genes harboured two disrupting mutations. Even when clonal, two distinct disrupting mutations can only inactivate a gene if they affect

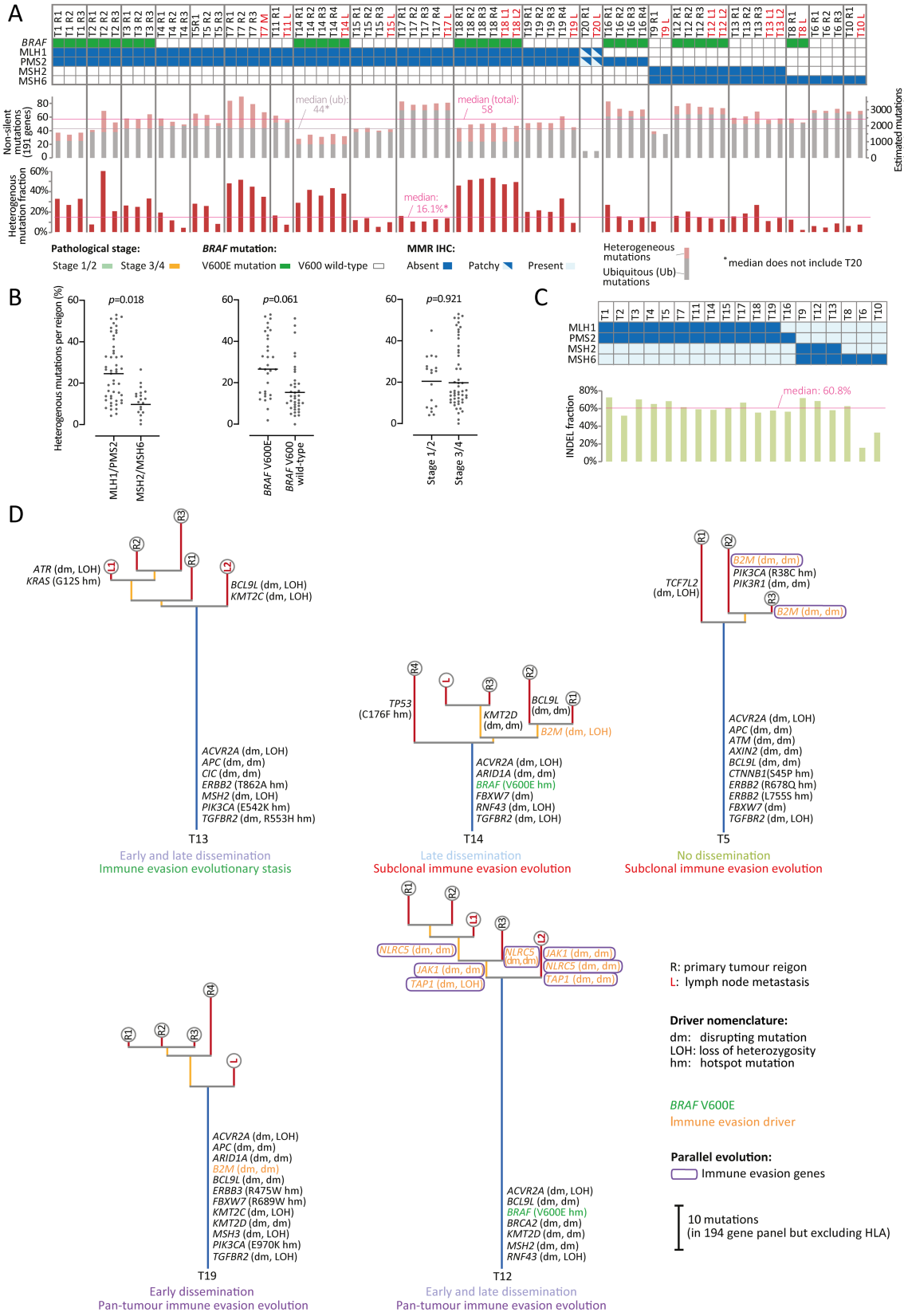

Figure 1. Multiregion sequencing analysis of MMRd CRCs. (A) Ubiquitous and heterogeneous mutations detected in 191 of 194 genes (excluding HLA-A/B/C) by tumour region/metastasis. Stage, *BRAF* V600E mutations, and MMR protein expression are shown. (B) Fraction of heterogeneous mutations per region by MMR protein loss, BRAF mutation status, and stage. The horizontal lines represent the estimated marginal mean. Significance was assessed with the nested *t*-test. (C) Fraction of indels among all called mutations by MMR protein loss. (D) Representative phylogenetic trees generated from mutation calls with the PHYLIP parsimony algorithm Pars, annotated with metastatic dissemination timing and IE evolutionary group. Driver aberrations in TS or IE genes were mapped on the branch where the second disrupting aberration had been acquired and in oncogenes on the branch where a hotspot mutation had been acquired. T, tumour; R, primary tumour region; L, lymph node metastasis. Bootstrap values (%) were calculated for phylogenetic trees where three or more tumour regions were available for genetic analysis.

both alleles. Most detected disrupting mutations were not in close enough proximity to assess, in the short-read sequencing data, whether they were present on different alleles or on the same allele. However, of five cases with multiple *B2M* mutations in the same sample, three harboured disrupting mutation pairs within the span of paired-end sequencing reads. In all three, mutations were mutually exclusive, i.e. never detected together in a single read (Figure 2B). We therefore accepted two independent clonal disrupting mutations in a TS or IE gene as a marker of biallelic inactivation. All identified driver aberrations (Figure 2A and supplementary material, Table S4) were mapped onto the phylogenetic trees (Figure 1D and supplementary material, Figure S3).

This revealed a clear hierarchy of driver evolution (Figure 2C): WNT/β-catenin (WNT) pathway aberrations, activating mutations of receptor tyrosine kinases/mitogen-activated protein kinase pathway (RTK-MAPK), and inactivation of TGF-β receptor family members were almost always truncal (87.0%, 86.4%, and 83.7%, respectively). These pathways were altered on the trunk of 89.5%, 94.7%, and 84.2% of tumours, respectively, demonstrating that they are critical for tumour initiation. Mutual exclusivity was observed for *APC* and *RNF43* disrupting mutations in the WNT pathway. This allowed further validation of our algorithm for the identification of drivers in TS genes, as *AXIN2* RNA overexpression is an established marker of ligand-independent WNT activation

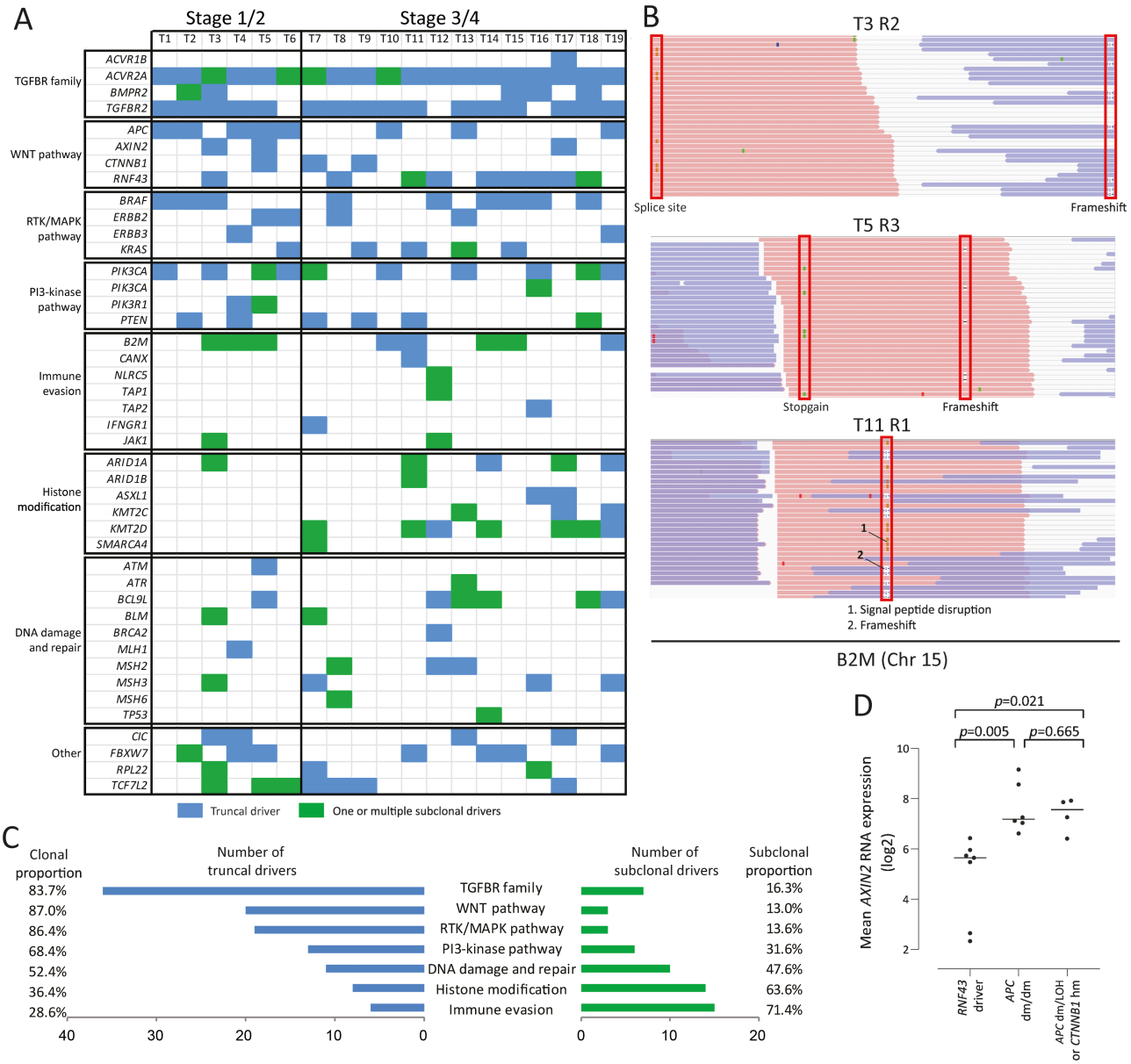

Figure 2. Driver aberrations and heterogeneity. (A) Heatmap of identified truncal or subclonal driver aberrations for each tumour. (B) Confirmation that distinct *B2M* mutations were located on different paired-end reads but never together. Screenshots were obtained from the Integrated Genomics Viewer. (C) Proportion of truncal and subclonal driver aberrations by functional group. (D) Mean *AXIN2* expression in primary tumour regions by truncal driver aberrations in *APC*, *RNF43*, and *CTNNB1*. Horizontal lines represent the mean and significance was assessed using a *t*-test.

(described to occur through *APC* inactivation or *CTNNB1* hotspot mutations) versus ligand-dependent activation (through *RNF43* inactivation) [30]. Six tumours with two truncal disrupting mutations in *APC* showed significantly higher *AXIN2* expression than tumours with truncal *RNF43* drivers (Figure 2D), and similar expression to those with truncal *APC* disrupting mutations/LOH or *CTNNB1* hotspot mutations. Thus, the detection of two independent disrupting mutations, even in a large TS gene such as *APC*, strongly indicates biallelic inactivation.

Aberrations activating the PI3K pathway, inactivation of histone modifiers or of DNA damage response and repair genes were truncal in 36.4–68.4% of occurrences. Inactivation of HLA class I antigen presentation or IFN-γ signalling genes, previously shown to enable IE [12–15], were predominantly subclonal (71.4%). ITH hence impedes accurate assessment of these critical immunotherapy biomarkers and must be taken into account for biomarker analyses in primary tumour tissue. Disrupting mutations of HLA class I genes were infrequent (five mutations across the six HLA class I gene alleles per genome, see supplementary material, Figure S3) in comparison with drivers in other antigen processing/presentation genes (31 mutations in 15 antigen presentation genes according to supplementary material, Table S3). This perhaps suggests that the disruption of genes with broad effects on antigen presentation is a more effective evolutionary path to IE in tumours with such high neoantigen numbers.

Parallel evolution predominates among IE drivers

We next assessed parallel evolution, which is a strong indicator of Darwinian selection [20,31]. This was defined by the presence of at least two distinct driver aberrations affecting genes with similar function on different phylogenetic branches. Eight of 19 tumours showed 12 instances of parallel evolution in nine different driver genes (see supplementary material, Figure S3). Parallel evolution was most common for IE genes (six of nine genes), most frequently involving *B2M*. The most striking example was observed in T12: *NLRC5* (the master regulator of HLA class I expression) was inactivated in three and the antigen transporter *TAP1* and the IFN-γ pathway gene *JAK1* each in two different ways, leading to pan-tumour inactivation of these three genes in all regions. Pan-tumour IE was observed in one additional case. Importantly, parallel evolution of IE occurred in four of six tumours that harboured any subclonal IE drivers, indicating that these tumours were under intense immune selection pressure. Subclonal IE drivers were absent in all five tumours with truncal IE drivers. Parallel evolution, and mutual exclusivity of truncal and subclonal IE drivers in a total of 11 of 19 tumours, substantiates that these evolved through Darwinian selection.

Driver heterogeneity between primary tumours and metastases

We questioned whether specific drivers evolved in metastases (supplementary material, Figure S3). Only *KMT2D* acquired driver aberrations more than once (three of 15 metastases).

IE drivers were acquired by two metastases; *B2M* in T15 and *JAK1*, *NLRC5* as well as *TAP1* in T12. *PTEN* was inactivated on the common branch of two metastases in T18. No new drivers had evolved on the branches, leading to seven of 15 metastases. Thus, drivers can be unique to metastatic sites, but we found no strong evidence for recurrent selection of specific drivers, consistent with our above observation that metastatic potential may already be encoded on the trunk in many cases. These conclusions are limited to lymph node metastases, as only one distant metastasis was assessed.

Heterogeneity of immune infiltrates

Our next aim was to interrogate the ITH of tumour infiltrating T-cells, which are surrogate markers of cancer immunogenicity [11]. We quantified CD8 T-cells using IHC and computational image analysis. Mean CD8 T-cell densities in the primary tumour varied between cases (1.9–21.5% of nucleated cells), questioning how T-cell density is regulated. It did not significantly differ by *BRAF* status, pattern of MMR protein loss, or stage (Figure 3A). T-cell densities did not correlate with ubiquitous mutation or indel numbers, nor with heterogeneous mutation numbers (supplementary material, Table S5). This is perhaps unsurprising, as all tumours had sufficient truncal mutations (909 extrapolated exonic mutations/tumour) to encode for multiple neoantigens, even when taking conservative estimates that only 0.5–1% of somatic mutations give rise to HLA-presented neoantigens [32,33]. T-cell densities also varied within primary tumours (Figure 3B). Dichotomisation using the mean CD8 T-cell fraction in the primary tumour distinguished tumours with low (mean: 4.0%) from tumours with high (mean: 11.7%), albeit frequently variable, T-cell infiltrates. This suggested a tumour-intrinsic setpoint, which is, however, accompanied by marked variability within tumours with dense infiltrates.

To obtain more granular insights, we computationally inferred 15 immune cell subtypes (supplementary material, Table S2) from RNA expression data of each tumour region. IHC-measured CD8 T-cells correlated most strongly with activated CD8 T-cells ($r = 0.525$, supplementary material, Table S6), indicating that inference from expression data worked reliably. Hierarchical clustering showed that most lymph node metastases clustered together (Figure 3D). This cluster was enriched for immune cells expected in the lymph node environment, such as B-cells and activated dendritic cells, but also other immune cell types, including activated CD8 and CD4 cells. Although samples had been macrodissected we cannot completely rule out contamination with immune cells from surrounding lymph node tissue. Yet, this can be reliably excluded in CD8 T-cell IHC, which also supported higher T-cell densities within cancer cell areas in lymph nodes. Primary tumour regions from individual cancers cosegregated into one of two major clusters. Higher

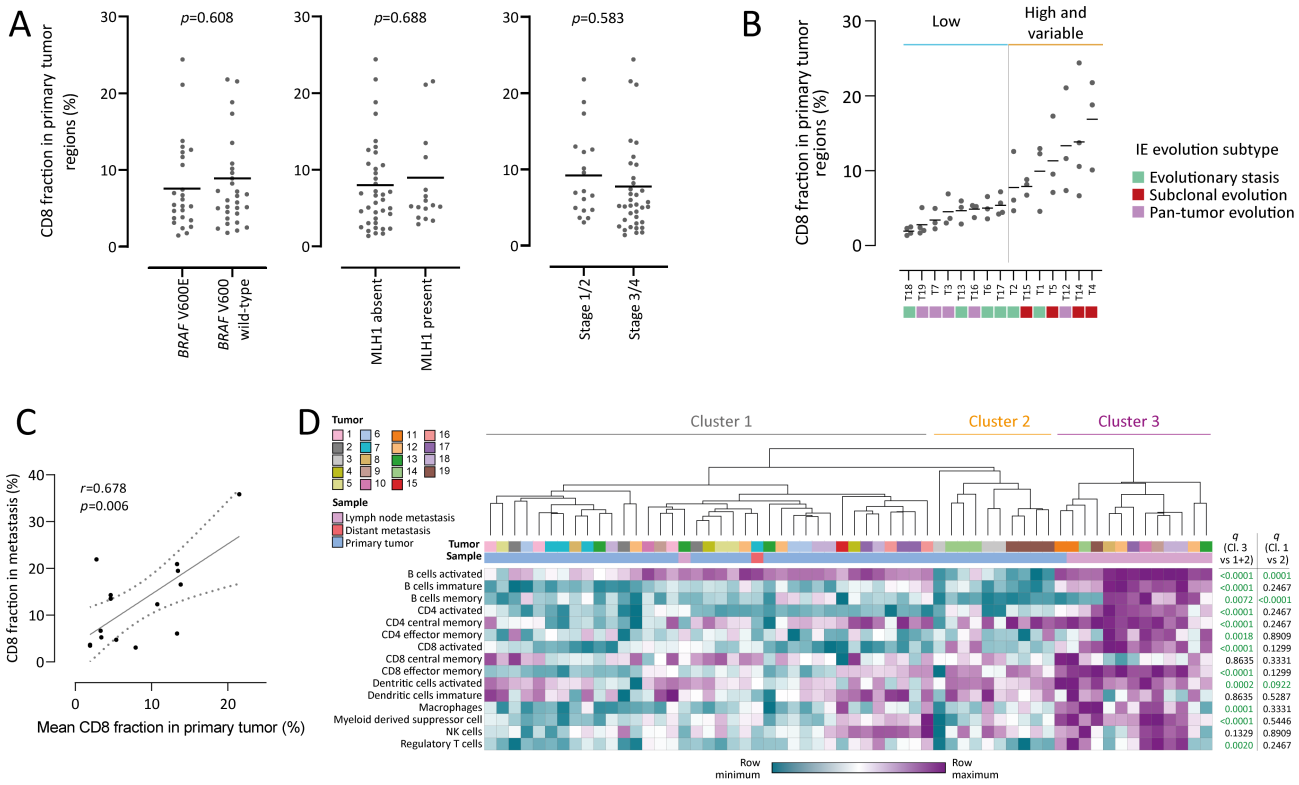

Figure 3. Immune landscapes of MMRd CRCs. (A) Fraction of CD8 T-cells (IHC) among all nucleated cells per primary tumour region by *BRAF* mutation status, MLH1 loss, and stage. The horizontal lines represent the estimated marginal mean. (B) CD8 T-cell fraction (IHC) in each primary tumour region and the mean for each tumour where at least three primary tumour regions were available. (C) Linear regression analysis of average CD8 T-cell fraction in the primary tumour and in matched lymph node metastases. The regression line (solid line) and 95% confidence intervals (dotted lines) are shown. (D) Euclidian clustering of 15 immune cell subtype abundances inferred from RNA expression data with ssGSEA; *q* values were generated with Morpheus with the *t*-test and multiple testing correction. Significance was assessed using the nested *t*-test (A) and a *t*-test (D). False discovery rate multiple testing correction was applied where *q* values are shown. $p < 0.05$ and $q < 0.1$ were considered significant. Correlation coefficients (*r*) were calculated using the Pearson test (C).

activated and memory B-cells in cluster 1 and higher activated dendritic cells in cluster 2 were the only significant differences between these (Figure 3D). Hence, these clusters did not explain differences in CD8 T-cell densities.

Taken together, we found no correlation of CD8 T-cell densities with mutation numbers or MMRd CRC subtypes. We therefore investigated whether different T-cell densities are a consequence of IE driver evolution.

Coevolution of genetic and immune landscapes

Three distinct patterns of genetic IE evolution were apparent in cases where at least three primary tumour regions had been analysed (Table 2 and supplementary material, Figure S3). These patterns were: (1) tumours without evolution of any genetic IE drivers, which we refer to as IE evolution stasis (T1, T2, T6, T13, T17, T18); (2) tumours with subclonal IE driver evolution in some but not all regions (T4, T5, T14, T15), and (3) tumours with pan-tumour IE through truncal aberrations or parallel evolution in all tumour regions (T3, T7, T12, T16, T19).

CD8 T-cell densities were significantly higher in group 2 (subclonal evolution) compared with groups 1 (stasis) and 3 (pan-tumour evolution) (Figure 4A). Deconvolution from RNA expression data showed that activated CD8 T-cells, effector memory CD8 T-cells, and activated CD4 T-cells were particularly enriched in group 2 (Figure 4B, sample-level data: Figure 4C). This indicates that active selection pressure from abundant CD8 T-cells is the proximate cause for subclonal IE evolution in group 2. Conversely, the absence of this selection pressure is one likely explanation for evolutionary stasis with respect to known IE mechanisms in group 1, whereas pan-tumour IE is probably responsible for the low infiltrates of CD8 T-cells in group 3. These subtypes may also explain the ITH of CD8 T-cell infiltrates (Figure 3B), as group 2 was only found in tumours with high variability, whereas groups 1 and 3 predominated among tumours with low mean T-cell densities and limited variability between regions.

We hypothesised that low CD8 T-cell infiltrates in tumours with IE evolution stasis (group 1) may indicate a previously unrecognised IE mechanism. This was not the consequence of lower numbers of mutations (median group 1: 56, group 2: 51). We assessed CD8 T-cell densities at the tumour margin using IHC (supplementary material, Table S2) and found that this was significantly ($p = 0.045$) lowest in the evolutionary

Table 2. Identified aberrant IE genes incorporated into MMRd CRC evolutionary pattern classification.

Gene symbol	Protein function
B2M	Light chain of MHC class 1
TAP1	Complex with TAP2 mediating transport of degraded cytosolic peptides across endoplasmic reticulum membrane for assemble with MHC class 1
TAP2	Complex with TAP1 mediating transport of degraded cytosolic peptides across endoplasmic reticulum membrane for assemble with MHC class 1
NLRC5	Master transcription activator of MHC class 1 genes
IFNGR1	Antigen presentation enhancer and anti-tumour effector cell activator
JAK1	Antigen presentation enhancer and anti-tumour effector cell activator
CANX	MHC class 1 chaperone

stasis group (Figure 4D), suggesting that T-cell recruitment may be impaired. This was supported by the numerically lower expression of the CD8 T-cell chemoattractants CXCL9-11 compared with group 2 (Figure 4E). Regulatory T-cells were also more abundant in group 1, but not other immunosuppressive cells, such as myeloid-derived suppressor cells or macrophages (Figure 4B). A recent mouse study showed that cGAS-STING pathway activation promotes CD8 T-cell infiltration into MMRd CRCs [34]. Loss of STING expression occurs in CRCs [35] and may impair immunogenicity, yet we detected STING in all group 1 tumours (Figure 4F). Overall, our data are most consistent with impaired recruitment rather than inactivation of CD8 T-cells as an explanation for the absence of IE evolution in group 1.

Distinct effects of impaired antigen presentation and IFN-γ signalling on immune infiltrates

We next assessed whether the disruption of antigen presentation versus the IFN-γ pathway had different impacts on immune infiltrates. We compared immune profiles of primary tumour regions that lost genes in the antigen presentation pathway ($n = 7$) with proficient regions ($n = 4$) in the same tumours. *B2M* was the affected gene in all cases. RNA-inferred activated CD4 and activated CD8 T-cells were significantly less abundant, whereas macrophages, immature B-cells, NK cells, and CD8 effector memory cells were significantly higher in regions with defective *B2M* (Figure 4G,H). The impact of IFN-γ pathway inactivation could not be assessed within tumours, as only one case contained both subclonal driver and wild-type regions. We therefore compared regions with IFN-γ pathway inactivation ($n = 7$) to those with defective *B2M* ($n = 13$). This revealed no significant difference in activated CD4 or CD8 T-cells, but lower infiltrates for most other immune cell subtypes, including NK cells, demonstrating broad depression of RNA-inferred immune infiltrates (Figure 4I). The enrichment of NK cells is an expected consequence of *B2M* loss and suggests opportunities for NK cell therapies [13]. Sparse infiltrates of most immune cell subtypes in tumours with IFN-γ pathway defects indicate that distinct treatment strategies are required for these.

Mechanisms influencing PD-L1 expression

Expression of PD-L1 by cancer cells, cells in the tumour stroma (predominantly by immune cells, mainly macrophages, in the stroma [36]), or both, is a major non-genetic IE mechanism [37–39]. To assess the heterogeneity of PD-L1 expression we categorised samples with accepted cut-offs (0: 0%; 1+: 1–4%; 2+: 5–49%; 3+: 50–100% [40]) based on the percentage of IHC PD-L1-positive cancer or stromal cells (supplementary material, Table S2). Stromal expression was common (1+ to 3+ staining in 97.1% of samples) and ITH was modest (Figure 5A,B). Only four of 19 tumours harboured PD-L1-expressing cancer cells and expression was detected in all regions of these. Cancer cell PD-L1 expression has been associated with more aggressive tumour behaviour than stromal PD-L1 expression in CRCs [39,41]. We show that this can be a stable characteristic, similar to truncal IE through genetic mechanisms. Absent cancer cell PD-L1 expression in 15 cases was surprising in view of the abundant stromal PD-L1 expression, but is consistent with previous reports in MMRd CRCs [36].

We correlated PD-L1 expression with immune cell infiltrates to investigate the determinants of PD-L1 expression. Because of the bias towards higher immune infiltrates in lymph nodes, we only analysed primary tumour regions. Stromal PD-L1 expression correlated significantly with IHC-quantified CD8 T-cells ($r = 0.355$, Figure 5C). Comparison with 15 immune cell subtypes from RNA expression showed an even stronger correlation of stromal PD-L1 with activated CD8 T-cells ($r = 0.507$) and with activated CD4 and dendritic cells (Figure 5D,E). Thus, an adaptive immune response predominated by T-cells strongly associated with stromal PD-L1 expression. In contrast, cancer cell PD-L1 expression did not correlate with IHC-quantified CD8 T-cells and only weakly with activated CD8 T-cells inferred from RNA expression (Figure 5F). Thus, we hypothesised that PD-L1 expression on cancer cells requires additional permissive factors.

We investigated whether genetic drivers were specific to tumours with PD-L1-expressing cancer cells. *CD274* (encoding PD-L1) was not amplified (supplementary material, Table S2). We found no enrichment of specific truncal drivers in tumours with PD-L1-positive cancer cells (supplementary material, Figure S3). Published data showed higher PD-L1 expression in CRCs with low expression of the intestinal homeobox transcription factor CDX2 but this did not distinguish PD-L1 expression by stromal or cancer cells nor between MMRd and MMRp tumours [42]. In our series, *CDX2* RNA expression was significantly negatively correlated with PD-L1-positive cancer cells ($r = -0.683$) but not with stromal PD-L1 ($r = -0.269$, Figure 5G). Remarkably,

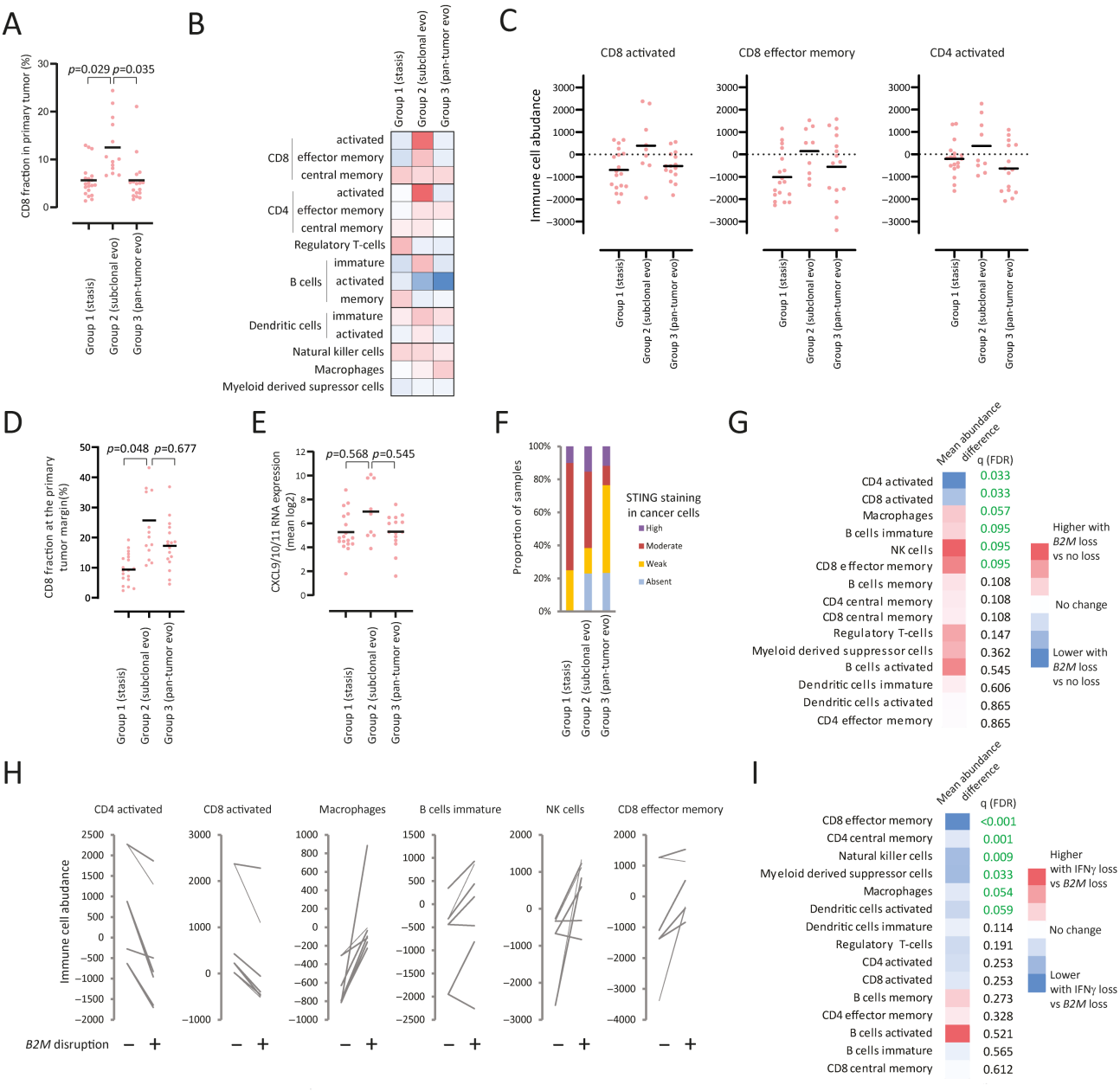

Figure 4. Coevolution of genetic and immune landscapes. (A) CD8 T-cell fraction (IHC) measured in 50 primary tumour regions from 15 MMRd CRCs that were classified into three groups based on IE driver evolution patterns. The horizontal lines represent the estimated marginal mean. (B) Abundance of 15 immune cell subtypes inferred from RNA expression data in the three evolution groups. Statistical results were not significant after multiple testing correction and values are not shown. (C) Sample-level data and means (horizontal bars) for immune cell subtypes in (B). (D) CD8 T-cell density at the tumour margin by evolution group. The horizontal lines represent the estimated marginal mean. (E) Expression of the CD8 T-cell chemoattractants *CXCL9*, *CXCL10*, and *CXCL11*. The horizontal lines represent the estimated marginal means. (F) Staining intensity for STING (using IHC) in cancer cells. (G) Comparison of primary tumour regions with *B2M* defects ($n = 7$) to all remaining primary tumour regions from the same cases ($n = 4$). The heatmap shows the mean abundance in regions with *B2M* defects minus the mean abundance in regions without defects. (H) Abundance difference of immune cell subtypes that were significant in (G), comparing tumour regions with *B2M* disruption (+) to regions without disruption (−). (I) Comparison of immune cell abundancies in all primary tumour regions with defects of IFN-γ pathway genes ($n = 7$) to all remaining primary tumour regions with *B2M* disruption ($n = 13$). The heatmap shows the mean abundance in regions with IFN-γ defects minus the mean abundance in regions with *B2M* disruption. Significance analyses were performed using the nested one-way ANOVA and Tukey's comparison of means test (A, D, and E), a paired *t*-test (G), and a nested *t*-test (I).

CDX2 expression was 24.7-fold lower in tumours with PD-L1-positive cancer cells. We sought to validate this by IHC [8] in an independent cohort of 23 MMRd CRCs (supplementary material, Table S7). CDX2 staining was low or absent in tumours with PD-L1-positive cancer cells and those without PD-L1-positive cancer cells showed significantly higher CDX2 (Figure 5H). Moreover, analysis of 57 CRC cell lines [43] confirmed a significant negative correlation of *CDX2* and *CD274* RNA expression (Figure 5I).

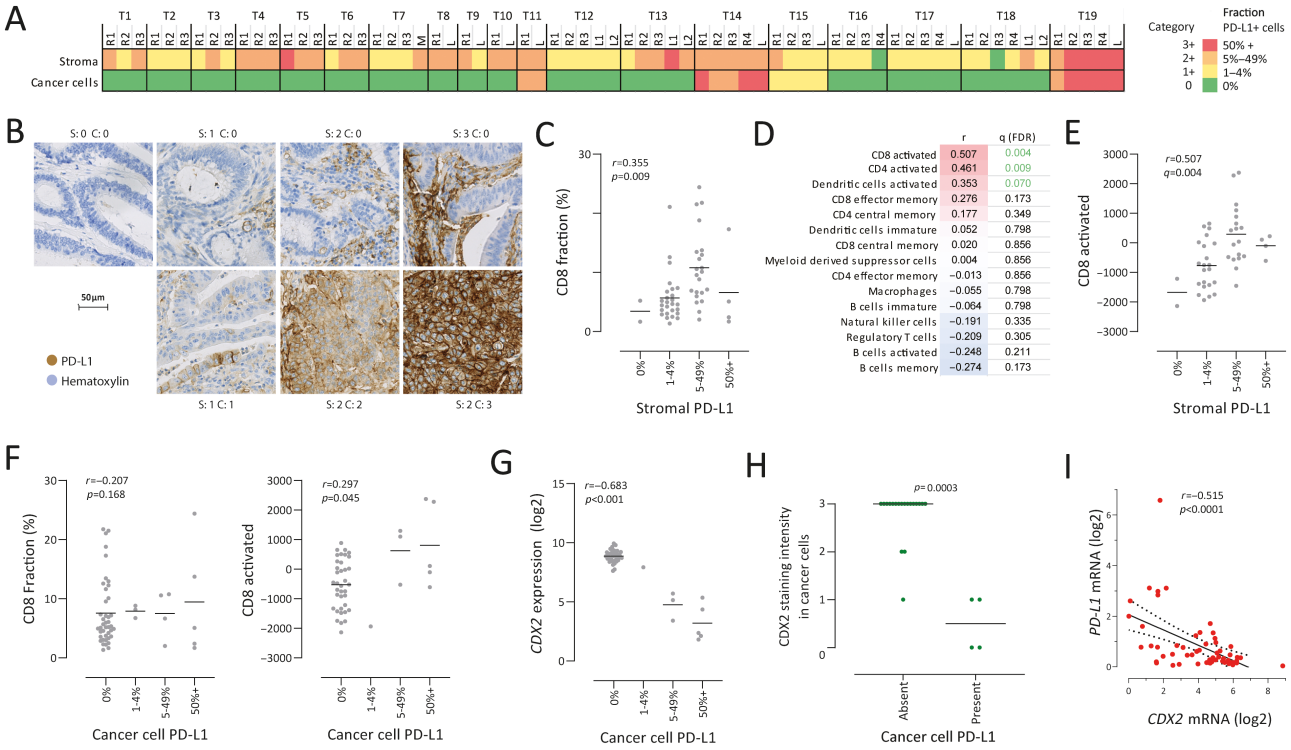

Figure 5. Distinct mechanisms influence PD-L1 expression by stromal and cancer cells. (A) Proportion of PD-L1-positive stromal cells and cancer cells by IHC in each analysed sample. (B) Examples of PD-L1 staining in stroma and cancer cells (S: stromal staining, C: cancer cell staining). (C) CD8 T-cell fraction (using IHC) in primary tumour regions by PD-L1 positivity in stromal cells. (D) Correlation of immune cell subtype abundance assessed by ssGSEA with stromal PD-L1 expression. (E) Individual sample data of activated CD8 T-cell abundance (inferred from RNA expression) in primary tumour regions by PD-L1 positivity. (F) CD8 T-cell fraction IHC (left) and activated CD8 T-cells from RNA expression (right) in primary tumour regions by PD-L1 positivity in cancer cells. (G) *CDX2* RNA expression in primary tumour regions by PD-L1 positivity in cancer cells. (H) CDX2 staining intensity (0: absent, 1: weak, 2: moderate, 3: strong) by cancer cell PD-L1 expression in an independent validation cohort of 23 MMRd CRCs. (I) Linear regression analysis of PD-L1/*CD274* and *CDX2* mRNA expression in 57 CRC cell lines from the Cancer Cell Line Encyclopedia. Horizontal bars show the medians in (H) and means elsewhere. Significance was assessed using a Mann–Whitney test (H). False discovery rate multiple testing correction was applied where q values are shown. $p < 0.05$ and $q < 0.1$ were considered significant. Correlation coefficients (r) were calculated with Pearson (I) and Spearman rank (C–G) tests.

Taken together, stromal PD-L1 was highest in tumour regions with abundant CD8 T-cells, where it probably protects cancer cells from immune-mediated destruction, whereas PD-L1 expression by cancer cells was restricted to cases with low or absent CDX2.

IE and metastatic dissemination

We finally explored whether IE mechanisms in cancer cells (including drivers in IE genes or cancer cell PD-L1 expression) associated with the presence of metastases. We found no significant difference between the three evolutionary subtypes (Figure 6A) and the proportion of tumours harbouring any IE mechanism was similar between stage 1/2 tumours and Stage 3/4 tumours (Figure 6B). However, truncal IE mechanisms were exclusively identified in Stage 3/4 tumours (Figure 6C), suggesting that early acquisition of these may foster metastasis development. This was significant ($p = 0.044$), but larger studies are required to substantiate this.

Discussion

Novel findings of this systematic analysis of MMRd CRC evolution include the identification of a clear hierarchy of clonal and subclonal driver pathways, the correlation of CD8 T-cell densities with phylogenetic patterns of IE driver evolution, and an association of PD-L1 expression in cancer cells with CDX2 loss. That T-cell densities did not correlate with mutation numbers is perhaps unsurprising, as all tumours had sufficient truncal mutations (909 extrapolated exonic mutations/tumour) to encode for multiple neoantigens, even when considering conservative estimates that only 0.5–1% of somatic mutations give rise to HLA-presented neoantigens [32,33]. The pervasive ITH in these tumours with high CPI response rates reinforces that abundant subclonal mutations do not seem to hinder effective immune responses during these treatments [19]. The temporal hierarchy of driver evolution shows that the WNT pathway, RTK/MAPK pathway, and TGF-β receptor family members are critical for cancer initiation. The increasing prevalence of subclonal drivers in PI3K, DNA damage response and repair genes, histone

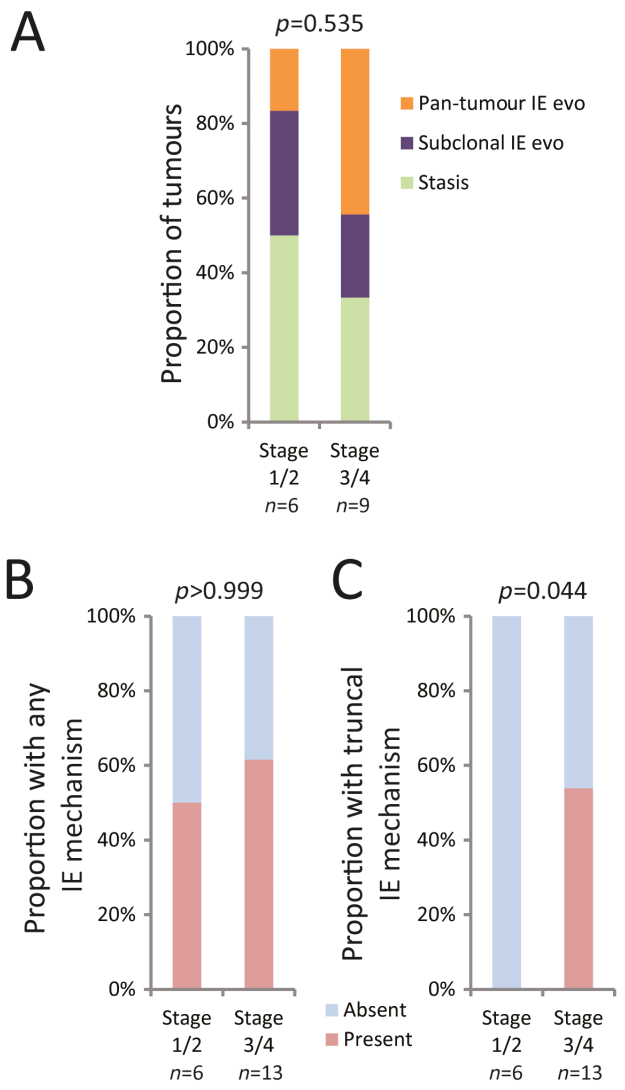

Figure 6. IE mechanisms in tumours with and without metastases. (A) Three IE evolution subtypes by stage. (B) Presence of any IE mechanisms in cancer cells (clonal or subclonal IE driver aberrations from Figure 2A or cancer cell PD-L1 expression) by stage. (C) Presence of truncal IE mechanisms in cancer cells by stage. Cancer cell PD-L1 expression was considered equivalent to a truncal genetic IE driver as it was detected in all tumour regions where present. Statistical analyses were performed using a χ^2 test for (A) and Fisher's exact test for (B) and (C).

modifier, and IE genes support their relevance in promoting cancer progression. IFN-γ pathway inactivation is associated with resistance to PD1 inhibitors [12,14] and CTLA4 inhibitors [15] in lung cancers and melanoma. B2M inactivation correlated with better responses in MMRd CRCs. IE through mutations in these pathways is known to occur frequently in MMRd CRCs [23,44–46], but neither these studies nor biomarker analyses in CPI-treated MMRd CRCs interrogated ITH. Our results show that clonality assessment of IE mutations must be taken into account for predictive CPI biomarker development from primary tumour tissue. Multiregion sequencing is unlikely to be clinically feasible. However, circulating tumour DNA sequencing is effective in CRC and methods to determine driver aberration clonality have been developed [47,48].

Tumours with subclonal IE evolution showed the highest CD8 T-cell densities. This may appear counter-intuitive, but it is probably testament to ongoing evolutionary adaptation to the immune selection pressure that is active in these tumours at the time of sampling. Low immune cell infiltrates in cases with pan-tumour IE patterns are plausibly explained by the lack of antigen presentation or IFN-γ pathway inactivation. In contrast, the low immune infiltrates in tumours without any IE mutations suggest a novel mode of IE, as no known IE mechanisms were apparent. The specific mechanism could not be identified in this small cohort, but our data indicate T-cell recruitment may be impaired. The low CD8 T-cell densities and expression of T-cell chemoattractants further associated with low myeloid and dendritic cell abundancies. This may also represent failure to trigger any T-cell immune responses, for example through insufficient danger signals. IFN-γ pathway inactivation was associated with broad depression of RNA-inferred immune cell abundances, even in comparison with *B2M* inactivation. Whether IFN-γ pathway aberrations lead to CPI resistance is therefore an important question.

We finally found that PD-L1 expression by cancer cells strongly associated with CDX2 loss, suggesting a permissive effect for PD-L1 upregulation. Impaired CDX2 expression in bowel cancer confers an increased recurrence risk [8,49,50]. This is thought to be a consequence of differentiation loss and increased invasiveness [51]. Our data indicate that increased T-cell suppression through PD-L1 upregulation may contribute. However, the role and effect of other coinhibitors of immune responses, such as IDO or LAG3, was not assessed in our study. In addition, cancer cell PD-L1 expression conferred higher susceptibility to CPIs in mouse models [37,39]. Whether CDX2 loss identifies tumours that particularly benefit from adjuvant CPIs should be investigated in ongoing trials (e.g. NCT02912559).

The small number of 19 MMRd CRCs is the main limitation of this study. Further validation of our findings in larger cohorts, preferentially from clinical trials that used immunotherapy, should be performed. As multiregion tissue samples are often unavailable, circulating tumour DNA-based assessment of ITH may be a suitable approach. Furthermore, measuring immune cells from RNA expression data can be error prone. These results need to be interpreted cautiously in the absence of an independent validation method.

Acknowledgements

We thank the patients who donated samples for research and the clinical research teams for trial support and sample collection. The study was supported by funding to MG from the European Research Council under the European Union's Horizon 2020 research and

innovation programme (grant agreement no. 820137), from the NIHR Barts Biomedical Research Centre Cancer Theme and from the Royal Marsden Hospital/ICR NIHR Biomedical Research Centre for Cancer. JR-B receives funding from a Río Hortega fellowship from the Institute of Health Carlos III (CM19/00087) and a 2020 TTD Research Grant from the Spanish Cooperative Group for the Treatment of Digestive Tumors (TTD). BRC receives funding from Cancer Research UK and The Wellcome Trust. The POLEM clinical trial was supported by a research grant from Merck KGaA. JM, JLA and JK received funding from the NIHR Imperial College Biomedical Research Centre. We thank Dr Gideon Coster at The Institute of Cancer Research, London for critical discussions.

Author contributions statement

BRC and MG designed the study. BRC, MG, JR-B and JM obtained funding. BE, RE, SF, MH, DK, NM, LN, VP, MS, KS, TD, DCu, IC, NSt and MG recruited patients to the POLEM trial. TD, HP, DCh, DCu, IC and NSt developed and coordinated the POLEM clinical trial. RB coordinated the clinical sample collection. RC and AT coordinated clinical data collection. JR-B, JLA, JK, DL, GA and CF provided archival samples and data. BRC, HL-I, KvL, HBS and MD-A processed the tissue. BRC and LJB prepared sequencing libraries. IA and KF performed NGS sequencing. AW, BRC and MG performed bioinformatic analyses. BRC, TL, DB, NSi and HBS performed immunohistochemical staining. BRC, AW, MB, SG and MG interpreted the data. BRC, LJB and MG wrote the manuscript. All authors read and gave final approval of the manuscript.

Data availability statement

The accession number for the DNA sequencing data reported in this paper is EGA: EGAS00001005769. Datasets are password protected and will be shared with researchers subject to signing a data-sharing agreement.

References

1. Raynaud F, Mina M, Tavernari D, *et al*. Pan-cancer inference of intra-tumor heterogeneity reveals associations with different forms of genomic instability. *PLoS Genet* 2018; **14:** e1007669.
2. Domingo E, Freeman-Mills L, Rayner E, *et al*. Somatic POLE proof-reading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *Lancet Gastroenterol Hepatol* 2016; **1:** 207–216.
3. Kloor M, Michel S, von Knebel Doeberitz M. Immune evasion of microsatellite unstable colorectal cancers. *Int J Cancer* 2010; **127:** 1001–1010.
4. Le DT, Uram JN, Wang H, *et al*. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* 2015; **372:** 2509–2520.
5. André T, Shiu KK, Kim TW, *et al*. Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. *N Engl J Med* 2020; **383:** 2207–2218.
6. Lenz H-J, Lonardi S, Zagonel V, *et al*. Nivolumab (NIVO)+ low-dose ipilimumab (IPI) as first-line (1L) therapy in microsatellite instability-high/mismatch repair-deficient (MSI-H/dMMR) metastatic colorectal cancer (mCRC): two-year clinical update. *J Clin Oncol* 2020; **38:** 4040.
7. Chalabi M, Fanchi LF, Dijkstra KK, *et al*. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat Med* 2020; **26:** 566–576.
8. Dalerba P, Sahoo D, Paik S, *et al*. CDX2 as a prognostic biomarker in stage II and stage III colon cancer. *N Engl J Med* 2016; **374:** 211–222.
9. Hansen TF, Kjær-Frifeldt S, Eriksen AC, *et al*. Prognostic impact of CDX2 in stage II colon cancer: results from two nationwide cohorts. *Br J Cancer* 2018; **119:** 1367–1373.
10. Turajlic S, Litchfield K, Xu H, *et al*. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol* 2017; **18:** 1009–1021.
11. Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer* 2019; **19:** 133–150.
12. Zaretsky JM, Garcia-Diaz A, Shin DS, *et al*. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N Engl J Med* 2016; **375:** 819–829.
13. Sade-Feldman M, Jiao YJ, Chen JH, *et al*. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat Commun* 2017; **8:** 1136.
14. Shin DS, Zaretsky JM, Escuin-Ordinas H, *et al*. Primary resistance to PD-1 blockade mediated by JAK1/2 mutations. *Cancer Discov* 2017; **7:** 188–201.
15. Gao J, Shi LZ, Zhao H, *et al*. Loss of IFN-γ pathway genes in tumor cells as a mechanism of resistance to anti-CTLA-4 therapy. *Cell* 2016; **167:** 397–404.e9.
16. Overman MJ, McDermott R, Leach JL, *et al*. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol* 2017; **18:** 1182–1191.
17. Middha S, Yaeger R, Shia J, *et al*. Majority of B2M-mutant and -deficient colorectal carcinomas achieve clinical benefit from immune checkpoint inhibitor therapy and are microsatellite instability-high. *JCO Precis Oncol* 2019; **3:** 1–14.
18. de Vries NL, van de Haar J, Veninga V, *et al*. γδ T cells are effectors of immunotherapy in cancers with HLA class I defects. *Nature* 2023; **613:** 743–750.
19. von Loga K, Woolston A, Punta M, *et al*. Extreme intratumour heterogeneity and driver evolution in mismatch repair deficient gastro-oesophageal cancer. *Nat Commun* 2020; **11:** 139.
20. Gerlinger M, Horswell S, Larkin J, *et al*. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 2014; **46:** 225–233.
21. Gulati S, Martinez P, Joshi T, *et al*. Systematic evaluation of the prognostic impact and intratumour heterogeneity of clear cell renal cell carcinoma biomarkers. *Eur Urol* 2014; **66:** 936–948.
22. Cortes-Ciriano I, Lee S, Park WY, *et al*. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* 2017; **8:** 15180.
23. Giannakis M, Mu XJ, Shukla SA, *et al*. Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Rep* 2016; **15:** 857–865.
24. Grasso CS, Giannakis M, Wells DK, *et al*. Genetic mechanisms of immune evasion in colorectal cancer. *Cancer Discov* 2018; **8:** 730–749.
25. Markow M, Chen W, Frankel WL. Immunohistochemical pitfalls: common mistakes in the evaluation of lynch syndrome. *Surg Pathol Clin* 2017; **10:** 977–1007.

26. Benjamin D, Sato T, Cibulskis K, *et al*. Calling somatic SNVs and indels with Mutect2. *bioRxiv* 2019; 861054. [Not peer reviewed].
27. Houlleberghs H, Goverde A, Lusseveld J, *et al*. Suspected lynch syndrome associated MSH6 variants: a functional assay to determine their pathogenicity. *PLoS Genet* 2017; **13:** e1006765.
28. Helderman NC, Van Der Werf-'t Lam AS, MSH6 Tumor Group, *et al*. Molecular profile of MSH6-associated colorectal carcinomas shows distinct features from other lynch syndrome-associated colorectal carcinomas. *Gastroenterology* 2023; **165:** 271–274.e2.
29. Kets CM, Hoogerbrugge N, van Krieken JH, *et al*. Compound heterozygosity for two MSH2 mutations suggests mild consequences of the initiation codon variant c.1A>G of MSH2. *Eur J Hum Genet* 2009; **17:** 159–164.
30. Kleeman SO, Leedham SJ. Not all Wnt activation is equal: ligand-dependent versus ligand-independent Wnt activation in colorectal cancer. *Cancers (Basel)* 2020; **12:** 3355.
31. Gerlinger M, Rowan AJ, Horswell S, *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012; **366:** 883–892.
32. Newey A, Griffiths B, Michaux J, *et al*. Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. *J Immunother Cancer* 2019; **7:** 309.
33. Schumacher TN, Scheper W, Kvistborg P. Cancer neoantigens. *Annu Rev Immunol* 2019; **37:** 173–200.
34. Lu C, Guan J, Lu S, *et al*. DNA sensing in mismatch repair-deficient tumor cells is essential for anti-tumor immunity. *Cancer Cell* 2021; **39:** 96–108.e6.
35. Xia T, Konno H, Ahn J, *et al*. Deregulation of STING signaling in colorectal carcinoma constrains DNA damage responses and correlates with tumorigenesis. *Cell Rep* 2016; **14:** 282–297.
36. Llosa NJ, Cruise M, Tam A, *et al*. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov* 2015; **5:** 43–51.
37. Lau J, Cheung J, Navarro A, *et al*. Tumour and host cell PD-L1 is required to mediate suppression of anti-tumour immunity in mice. *Nat Commun* 2017; **8:** 14572.
38. Juneja VR, McGuire KA, Manguso RT, *et al*. PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity. *J Exp Med* 2017; **214:** 895–904.
39. Kleinovink JW, Marijt KA, Schoonderwoerd MJA, *et al*. PD-L1 expression on malignant cells is no prerequisite for checkpoint therapy. *Oncoimmunology* 2017; **6:** e1294299.
40. Möller K, Blessin NC, Höflmayer D, *et al*. High density of cytotoxic T-lymphocytes is linked to tumoral PD-L1 expression regardless of the mismatch repair status in colorectal cancer. *Acta Oncol* 2021; **60:** 1210–1217.
41. Wyss J, Dislich B, Koelzer VH, *et al*. Stromal PD-1/PD-L1 expression predicts outcome in colon cancer patients. *Clin Colorectal Cancer* 2019; **18:** e20–e38.
42. Inaguma S, Lasota J, Wang Z, *et al*. Clinicopathologic profile, immunophenotype, and genotype of CD274 (PD-L1)-positive colorectal carcinomas. *Mod Pathol* 2017; **30:** 278–285.
43. Ghandi M, Huang FW, Jané-Valbuena J, *et al*. Next-generation characterization of the cancer cell line encyclopedia. *Nature* 2019; **569:** 503–508.
44. Ijsselsteijn ME, Petitprez F, Lacroix L, *et al*. Revisiting immune escape in colorectal cancer in the era of immunotherapy. *Br J Cancer* 2019; **120:** 815–818.
45. Kloor M, Michel S, Buckowitz B, *et al*. Beta2-microglobulin mutations in microsatellite unstable colorectal tumors. *Int J Cancer* 2007; **121:** 454–458.
46. Ozcan M, Janikovits J, von Knebel DM, *et al*. Complex pattern of immune evasion in MSI colorectal cancer. *Oncoimmunology* 2018; **7:** e1445453.
47. Woolston A, Khan K, Spain G, *et al*. Genomic and transcriptomic determinants of therapy resistance and immune landscape evolution during anti-EGFR treatment in colorectal cancer. *Cancer Cell* 2019; **36:** 35–50.e9.
48. Knebel FH, Barber LJ, Newey A, *et al*. Circulating tumour DNA sequencing identifies a genetic resistance-gap in colorectal cancers with acquired resistance to EGFR-antibodies and chemotherapy. *Cancers (Basel)* 2020; **12:** 3736.
49. Tomasello G, Barni S, Turati L, *et al*. Association of CDX2 expression with survival in early colorectal cancer: a systematic review and meta-analysis. *Clin Colorectal Cancer* 2018; **17:** 97–103.
50. Jun SY, Eom DW, Park H, *et al*. Prognostic significance of CDX2 and mucin expression in small intestinal adenocarcinoma. *Mod Pathol* 2014; **27:** 1364–1374.
51. Hryniuk A, Grainger S, Savory JG, *et al*. Cdx1 and Cdx2 function as tumor suppressors. *J Biol Chem* 2014; **289:** 33343–33354.
52. Lau D, Kalaitzaki E, Church DN, *et al*. Rationale and design of the POLEM trial: avelumab plus fluoropyrimidine-based chemotherapy as adjuvant treatment for stage III mismatch repair deficient or POLE exonuclease domain mutant colon cancer: a phase III randomised study. *ESMO Open* 2020; **5:** e000638.
53. Shukla SA, Rooney MS, Rajasagi M, *et al*. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* 2015; **33:** 1152–1158.
54. Robinson JT, Thorvaldsdóttir H, Winckler W, *et al*. Integrative genomics viewer. *Nat Biotechnol* 2011; **29:** 24–26.
55. Yeh CH, Bellon M, Nicot C. FBXW7: a critical tumor suppressor of human cancers. *Mol Cancer* 2018; **17:** 115.
56. Smith IN, Briggs JM. Structural mutation analysis of PTEN and its genotype-phenotype correlations in endometriosis and cancer. *Proteins* 2016; **84:** 1625–1643.
57. Leslie NR, Longy M. Inherited PTEN mutations and the prediction of phenotype. *Semin Cell Dev Biol* 2016; **52:** 30–38.
58. Boc A, Diallo AB, Makarenkov V. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res* 2012; **40:** W573–W579.
59. Bankhead P, Loughrey MB, Fernández JA, *et al*. QuPath: open source software for digital pathology image analysis. *Sci Rep* 2017; **7:** 16878.
60. Charoentong P, Finotello F, Angelova M, *et al*. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* 2017; **18:** 248–262.
61. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 2006; **93:** 491–507.

References 52–61 are cited only in the supplementary material

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Figure S1. Ubiquitous mutation load by *BRAF*, MMR loss pattern, and stage

Figure S2. Linear regression model generated using GraphPad PRISM by analysing non-silent mutation calls in the exome versus those in 191 target genes (excluding class I HLA genes) in our targeted sequencing panel for 518 MSI and MSS CRCs (POLE cases were excluded)

Figure S3. Phylogenetic trees generated from mutation calls with the PHYLIP parsimony algorithm Pars, grouped according to stage and metastatic dissemination timing

Figure S4. Signal peptide location in the B2M protein sequence

Table S1. The 194 genes included in the targeted sequencing panel

Table S2. Patient characteristics, sequencing metrics, mutation and copy number data, IHC data, and ssGSEA immune cell abundance

Table S3. Known oncogenes, TS genes, and IE genes (involved in class I HLA antigen presentation or IFN-γ signalling) that were assessed for driver aberrations

Table S4. Data supporting the identified driver aberrations

Table S5. Spearman correlation coefficient (*r*) of mean CD8 fraction in primary tumour regions versus mutation metrics in 19 MMRd CRCs and significance analysis

Table S6. Pearson correlation analysis of CD8 T-cell fractions assessed by IHC versus immune cell subtype abundances assessed by ssGSEA in 60 tumour regions from which RNA-sequencing data were available

Table S7. Patient and pathological characteristics of the validation cohort of 23 MMRd CRCs

Table S8. Detailed IHC autostaining protocols and reagents

Table S9. Qupath algorithm settings for chromogenic CD8 IHC stain computational detection

Table S10. Normalised RNA expression data