OPEN ACCESS

Original research

Activation of innate-adaptive immune machinery by poly(I:C) exposes a therapeutic vulnerability to prevent relapse in stroma-rich colon cancer

Shania M Corry,[1] Amy MB McCorry,[1] Tamsin RM Lannagan,[2] Niamh A Leonard,[3,4] Natalie C Fisher,[1] Ryan M Byrne ,[1] Petros Tsantoulis,[5] Xabier Cortes-Lavaud,[2] Raheleh Amirkhah,[1] Keara L Redmond,[1] Aoife J McCooey,[1] Sudhir B Malla,[1] Emily Rogan,[1] Svetlana Sakhnevych,[1] Michael A Gillespie,[2,6] Mark White,[2,6] Susan D Richman,[7] Rene-Filip Jackstadt,[2,8] Andrew D Campbell,[2] Sarah Maguire,[1] S:CORT and ACRCelerate consortia, Simon S McDade,[1] Daniel B Longley,[1] Maurice B Loughrey ,[1,9,10] Helen G Coleman,[10] Emma M Kerr,[1] Sabine Tejpar,[11] Timothy Maughan,[12] Simon J Leedham,[13] Donna M Small,[1] Aideen E Ryan,[3,4] Owen J Sansom,[2,6] Mark Lawler ,[1] Philip D Dunne [1]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2021-326183).

For numbered affiliations see end of article.

Correspondence to
Dr Philip D Dunne, Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Belfast BT7 1NN, UK; p.dunne@qub.ac.uk

SMC, AMM and TRL are joint first authors.

Received 23 September 2021
Accepted 12 March 2022
Published Online First 27 April 2022

To cite: Corry SM, McCorry AMB, Lannagan TRM, *et al*. *Gut* 2022;**71**:2502–2517.

ABSTRACT

Objective Stroma-rich tumours represent a poor prognostic subtype in stage II/III colon cancer (CC), with high relapse rates and limited response to standard adjuvant chemotherapy.

Design To address the lack of efficacious therapeutic options for patients with stroma-rich CC, we stratified our human tumour cohorts according to stromal content, enabling identification of the biology underpinning relapse and potential therapeutic vulnerabilities specifically within stroma-rich tumours that could be exploited clinically. Following human tumour-based discovery and independent clinical validation, we use a series of *in vitro* and stroma-rich *in vivo* models to test and validate the therapeutic potential of elevating the biology associated with reduced relapse in human tumours.

Results By performing our analyses specifically within the stroma-rich/high-fibroblast (HiFi) subtype of CC, we identify and validate the clinical value of a HiFi-specific prognostic signature (HPS), which stratifies tumours based on STAT1-related signalling (High-HPS v Low-HPS=HR 0.093, CI 0.019 to 0.466). Using *in silico*, *in vitro* and *in vivo* models, we demonstrate that the HPS is associated with antigen processing and presentation within discrete immune lineages in stroma-rich CC, downstream of double-stranded RNA and viral response signalling. Treatment with the TLR3 agonist poly(I:C) elevated the HPS signalling and antigen processing phenotype across *in vitro* and *in vivo* models. In an *in vivo* model of stroma-rich CC, poly(I:C) treatment significantly increased systemic cytotoxic T cell activity ($p<0.05$) and reduced liver metastases ($p<0.0002$).

Conclusion This study reveals new biological insight that offers a novel therapeutic option to reduce relapse rates in patients with the worst prognosis CC.

Significance of this study

What is already known on this subject?

⇒ Stroma-rich tumour composition is associated with poor prognosis in patients with stage II/III colon cancer (CC), with a relapse rate of approximately 50% in this setting, even when patients are treated with standard adjuvant chemotherapy. Elevation of transforming growth factor-β (TGF-β) signalling is observed in stroma-rich CCs, which has been used as the basis for trials based on TGF-β blockade.

What are the new findings?

⇒ In this study, we push beyond the established association between stromal-derived TGF-β and poor prognosis, to identify, characterise and therapeutically exploit the biology that underpins relapse specifically within this TGF-β-high poor prognostic group. This stroma-rich subtype-specific approach reveals that STAT1-mediated antigen processing and viral response signalling is a targetable therapeutic vulnerability, via toll-like receptor 3 (TLR3) agonist poly(I:C), specifically in stroma-rich CC.

How might it impact on clinical practice in the foreseeable future?

⇒ This study reveals a new insight into the biology underpinning relapse in stroma-rich tumours, and offers a novel therapeutic option to reduce relapse rates in patients with stroma-rich tumours, which represents the worst prognostic subgroup in early stage CC.

INTRODUCTION

Colorectal cancer (CRC) is the third most commonly diagnosed cancer worldwide, with around 1.3 million cases diagnosed each year.[1]

Despite improvements in both surgical management and adjuvant treatment options, many stage II and III colon cancer (CC) patients still experience relapse following surgery; ~20% and 36% of patients within each stage respectively.[2] Classification of CRC patients into molecular subtypes, based on their underlying transcriptional signalling, revealed four consensus molecular subtypes (CMS1-4), where the stromal subtype (CMS4)[3] has the most dismal prognosis. Alongside molecular subtyping, these poor prognostic stroma-rich tumours can be also be identified using histology.[4–7] Based on this evidence, the stroma-rich or high-fibroblast subtype (HiFi) represents a poor prognostic subgroup in stage II/III CC, with relapse rates of ~50%–60%.[3 8] Importantly, this poor prognosis remains an issue even when stroma-rich patients receive adjuvant treatment following surgery; limited benefits from FOLFOX (bolus and infused fluorouracil with oxaliplatin) and capecitabine with oxaliplatin regimes were observed in patients with stroma-rich tumours in the short course oncology therapy (SCOT) clinical trial[9] and in a recent meta-analysis where adjuvant chemotherapy was found not to be effective in CMS4 tumours.[10]

Numerous studies, including our own, have defined and characterised the biology underpinning stromal-rich tumours compared with epithelium-rich (stromal-low) tumours, which is dominated by elevated transforming growth factor-β (TGF-β) signalling or other markers of mesenchymal/CMS4 biology.[11–14] Elevation of TGF-β and stromal signalling cascades have been proposed as targets themselves, however no evidence has been shown that such biology is driving the differential outcomes in the ~50%–60% of stroma-rich tumours that eventually relapse, compared with those that do not. Identification and understanding of the biology underpinning disease relapse specifically within stroma-rich tumours, rather than simply the characteristics of stroma-rich vs stroma-low tumours, could be used to develop novel therapeutic interventions specifically for patients with stroma-rich/CMS4 tumours that relapse following surgery.

To elucidate biology associated with patient outcome in the stroma-rich histological subtype, we combined fibroblast stratification with supervised transcriptomic analysis, based on risk of relapse, to uncover biology of specific relevance in stroma-rich localised (stage II/III) CC. To exploit this new understanding, we performed a series of in silico analyses to identify potential molecular vulnerabilities and a therapeutic candidate. Using a number of in vitro and in vivo models, we tested and validated the functional significance and potential clinical utility of poly(I:C) as a subtype-specific treatment option aimed at preventing metastatic relapse specifically within stroma-rich CC.

RESULTS

Prognostic value of morpho-molecular fibroblast measurement in tumour samples

To test the overlap between stromal gene signatures and histology, we used patient-matched transcriptional data and a QuPath[15]-derived H&E stromal classifier from colon resections (FOCUS cohort, n=361) and rectal pretreatment biopsies (Grampian cohort; n=225), previously characterised within the S:CORT stratified CRC programme[16] (figure 1A). Strong correlations were observed between H&E digital stroma scores and a number of previously established transcriptional signatures, including StromalScore using ESTIMATE,[17] cancer-associated fibroblast (CAF) score from Isella *et al*[18] and fibroblast score from MCPcounter,[19] alongside individual CAF markers *ACTA2* (alpha-smooth muscle actin; αSMA) and FAP (figure 1B,C). Combining *ACTA2* and *FAP* gene expression with the existing MCP fibroblast signature generated a single-sample gene set enrichment analysis (ssGSEA) 'fibroblast score' transcriptional classifier, with a correlation higher than other methods (figure 1C; Pearson correlation=0.856, online supplemental table 1). This approach enabled us to employ transcriptional data from cohorts where no H&E images are available, with the understanding that our findings can be translatable to the stroma-rich histological subtype, traditionally identifiable from patient-matched H&E slides.

Using transcriptional data from a discovery cohort of n=215 untreated stage II CC tumours[20] (online supplemental table 2), our ssGSEA fibroblast score was significantly higher in CMS4 tumours, compared with the other subtypes (figure 1D; t-test $p<0.0001$ for all). As fibroblast content is an already well-established prognostic biomarker, we defined an optimum prognostic cut-off level for our ssGSEA fibroblast scores, using relapse-free survival (RFS) data and Cox modelling (online supplemental figure 1). This resulted in stratification of the n=215 patients into high-fibroblast (HiFi; n=75; 35% of cohort) and low-fibroblast (LoFi; n=140; 65% of cohort) subgroups, with a larger proportion of HiFi tumours classified as CMS4 compared with LoFi tumours (69.1% and 7.1% respectively; figure 1E; Fisher's exact test $p<2.2\times10^{-16}$, (online supplemental figure 1). The epithelium-rich CMS2 subtype is more prevalent in the LoFi group compared with HiFi (32.1% and 9.3% respectively; figure 1E; Fisher's exact test $p=2.28\times10^{-06}$, (online supplemental figure 1). In line with previous studies, we observed significantly worse outcome in HiFi tumours compared with LoFi tumours, with patient relapse rates of 45.3% and 27.9%, respectively (figure 1F; log-rank p=0.00779, HR= 1.851, 95% CI (1.168 to 2.932)). In agreement with our initial correlative analyses (figure 1C), HiFi tumours also had significantly higher StromalScore using the ESTIMATE geneset, and fibroblast scores compared with LoFi tumours (figure 1G; adjusted $p<0.15$), alongside gene sets that we have previously directly associated with CAF infiltration,[21] including the epithelial to mesenchymal transition (figure 1G).

A number of previously identified prognostic factors are not prognostic in HiFi tumours

Although HiFi tumours in our discovery cohort have a significantly worse prognosis compared with LoFi, the relapse rates in the HiFi subgroup remain ~40%–50% (figure 1F), meaning that approximately half of patients with stage II stroma-rich tumours are cured by surgery alone. In line with previous studies, we demonstrate the ability of TGF-β signalling, as assessed using a number of transcriptional signatures, to identify the stroma-rich subtype (online supplemental figure 2A). Importantly, however, when the prognostic value of these signatures are assessed specifically within the HiFi subtype, they do not stratify patients based on relapse status online supplemental figure 2B). Assessment of previously defined CAF subtypes developed in CC and pancreatic cancer[22–24] failed to discriminate HiFi tumours based on relapse (online supplemental figure 2C); CRC CAF-A and CAF-B (upper left), pancreatic myCAF and iCAF (upper right), CRC differential contractility (lower left) and inflammatory-related fibroblasts $CD34^{-}THY1^{+}$, $CD34^{-}THY1^{-}$ and $CD34^{+}$ CAF (lower right)). Similarly, stratification based on fibroblast stiffness-related matrix index,[25] p53 activity (Hallmark gene set ssGSEA), stem-like markers, or overall fibroblast levels according to our ssGSEA score (online supplemental figure 2D) all failed to segregate the HiFi relapse and non-relapse tumours. Moreover, while unsupervised clustering of HiFi tumours identified two clusters, these subgroups did not have different prognostic outcomes

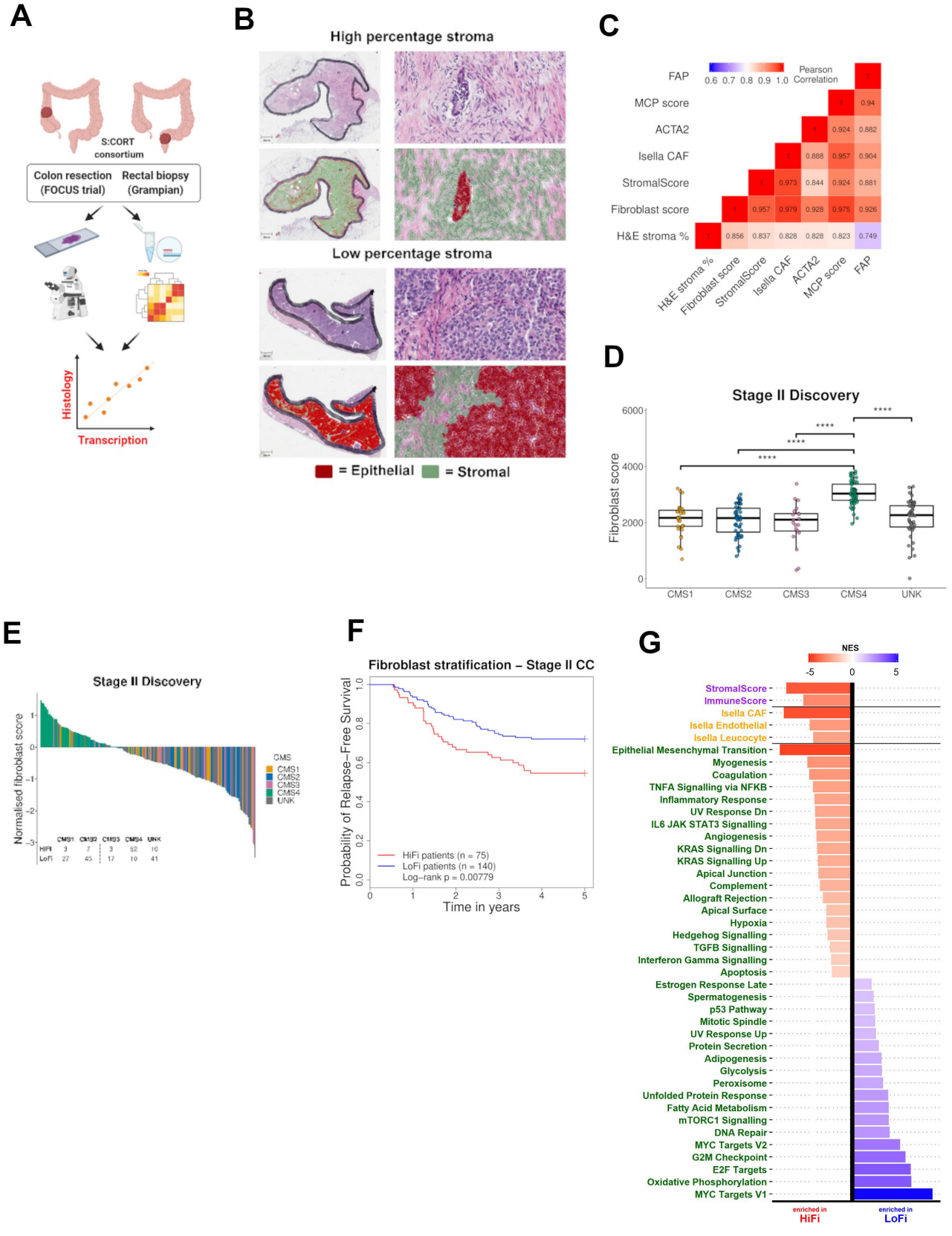


Figure 1 Development and validation of our transcriptional fibroblast score. (A) Schematic of correlation between stromal/fibroblasts scores via histology and transcriptomics. (B) H&E slide with the digital pathology stromal classifier applied to a sample with a high/low percentage stroma from the focus cohort. (C) Correlation matrix with histological stroma and transcriptional classifiers (Pearson's correlation). (D) CMS classification according to our fibroblast score. (UNK=unknown/mixed CMS classification) (t-test). (E) Waterfall plot of fibroblast scores indicating CMS classification. High-fibroblast (HiFi) n=75 and low-fibroblast (LoFi) n=140. (F) HiFi tumours have a worse prognosis than LoFi in discovery cohort (log-rank p=0.00779) (G). Comparison of HiFi and LoFi samples revealed that previously published stromal signatures and gene sets have significantly higher expression in the HiFi samples than the LoFi (adjusted p<0.15). CC, colon cancer; CMS, consensus molecular subtypes. **** denotes p<0.0001.

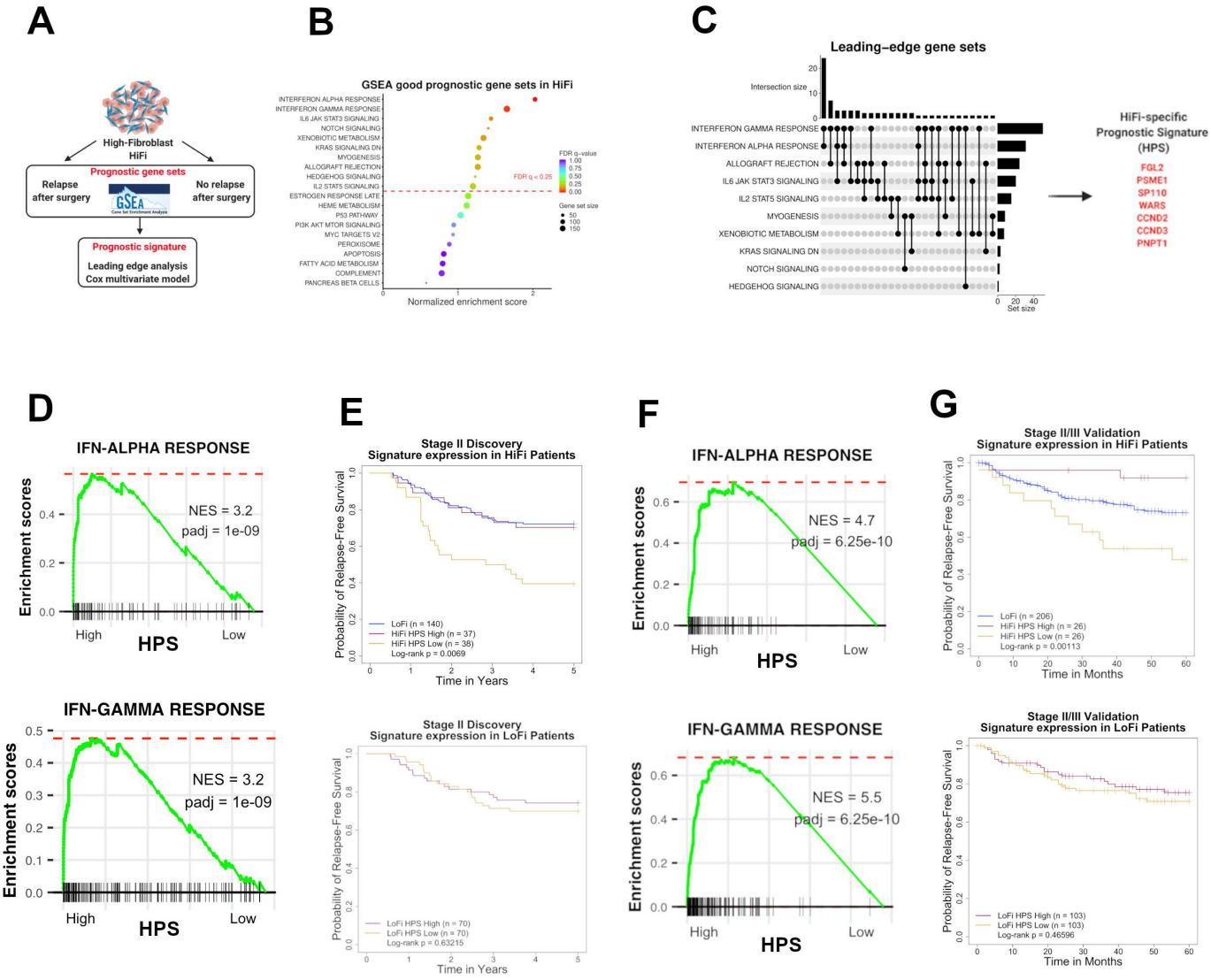


Figure 2 Identification of HiFi-specific prognostic biology (A) workflow summary of our supervised analyses. (B) Significant gene sets associated with good prognosis specifically within HiFi tumours from supervised GSEA analysis. (C) Leading-edge analysis (LEA) of the 10 gene sets demonstrating that, of the 71 genes, many of them overlap between the interferon response gene sets leading to identification of a seven gene HPS. (D) High expression of the HPS in HiFi tumours is associated with enriched IFN alpha and gamma response signalling in discovery cohort. (E) HPS has a strong prognostic value in HiFi tumours based on a median split in discovery cohort (log-rank p=0.0069; top). HPS has no prognostic value in the LoFi samples in discovery cohort (log-rank p=0.63215; bottom). (F) High expression of the HPS (n=26) in HIFI tumours is associated with enriched IFN alpha and gamma response signalling. (G) HPS can stratify HiFi samples into two groups in the validation cohort, one with significantly poorer RFS and another with RFS even better than the LoFi patients (log-rank p=0.00113; top). HPS has no prognostic value in the LoFi samples (log-rank p=0.46596; bottom). HiFi, high-fibroblast; LOFI, low-fibroblast; RFS, relapse-free survival.

(online supplemental figure 2E). As previously identified prognostic factors and unsupervised clustering provided no additional clinical value for identifying HiFi patients that relapse, we next performed a supervised analysis of tumours in the stage II untreated discovery cohort, using GSEA followed by leading-edge analysis (LEA) and Cox survival modelling, contrasting HiFi patients that relapsed within 5 years of surgery (n=34) and HiFi patients who never experienced disease relapse (n=41) (figure 2A).

An interferon-related seven-gene signature identifies HiFi patients with significantly better prognosis

GSEA revealed 10 significant gene sets associated with good prognosis in the HiFi group, including elevated interferon alpha and interferon gamma response (figure 2B, (online supplemental figure 2F). Using a LEA, which reveals specific genes that contribute most to the gene sets associated with prognosis in HiFi tumours, we identified 71 genes shared by more than one of the LEA subsets (figure 2C; left). Cox survival analysis, followed by a multivariate model for each individual gene (to adjust for age, gender, pT stage, tumour location, tumour differentiation grade, lymphovascular invasion status and mucinous/non-mucinous subtype) filtered this list to seven LEA genes ($p<0.05$; table 1); namely *FGL2*, *PSME1*, *SP110*, *WARS*, *CCND2*, *CCND3*, *PNPT1*, which we term hereafter as a HiFi-specific prognostic signature (HPS) capable of distinguishing relapse from non-relapse (figure 2C; right). We next confirmed that stratification of HiFi patients using a median split of the HPS was sufficient to represent the same elevated interferon (IFN) alpha and IFN gamma response GSEA signatures (figure 2D), however HPS was not associated with the levels of TGF-β signalling in HiFi

Table 1 HiFi-specific prognostic signature and relapse free survival

Gene (median exp)	Univariate HR High versus Low (95% CI)	Univariate p value	Multivariate HR High versus Low (95% CI)	Multivariate p value	Total cases (relapse)	High exp (relapse)	Low exp (relapse)
FGL2 (5.497)	0.568 (0.282 to 1.142)	0.113	0.374 (0.166 to 0.840)	0.017	74 (33)	37 (13)	37 (20)
PSME1 (7.384)	0.541 (0.269 to 1.089)	0.085	0.364 (0.153 to 0.868)	0.023	74 (33)	37 (13)	37 (20)
SP110 (6.082)	0.498 (0.245 to 1.014)	0.055	0.339 (0.148 to 0.775)	0.010	74 (33)	37 (12)	37 (21)
WARS (6.510)	0.387 (0.187 to 0.801)	0.010	0.319 (0.145 to 0.705)	0.005	74 (33)	37 (33)	37 (20)
CCND2 (9.464)	0.574 (0.285 to 1.155)	0.120	0.438 (0.204 to 0.939)	0.034	74 (33)	37 (13)	37 (20)
CCND3 (7.715)	0.513 (0.252 to 1.043)	0.065	0.414 (0.192 to 0.893)	0.025	74 (33)	37 (12)	37 (21)
PNPT1 (2.560)	0.312 (0.148 to 0.657)	0.002	0.265 (0.118 to 0.593)	0.001	74 (33)	37 (10)	37 (23)
Dataset (median sig)	**Univariate HR High versus CI)**	**Univariate p value**	**Multivariate HR High versus Low (95% CI)**	**Multivariate p value**	**Total cases (relapse)**	**High sig (relapse)**	**Low sig (relapse)**
Discovery (6.482)	0.395 (0.191 to 0.816)	0.012	0.218 (0.087 to 0.544)	0.001	74 (33)	36 (10)	38 (23)
Validation (9.137)	0.123 (0.027 to 0.550)	0.006	0.093 (0.019 to 0.466)	0.004	52 (14)	26 (2)	26 (12)

Median gene expression values were used to dichotomise the HiFi patients into high and low expression groups in Discovery cohort. Multivariate Cox regression analysis adjusted for age, sex, pT stage, tumour location, tumour differentiation grade, tumour subtype (mucinous/non-mucinous), lymphovascular invasion and the number of lymph nodes with relapse free survival as the outcome variable.
Median signature expression was used to dichotomise the HiFi patients in each cohort into high and low expression groups in both Discovery and Validation cohorts. For the discovery cohort, the multivariate Cox regression analysis adjusted for age, sex, pT stage, tumour location, tumour differentiation grade, tumour subtype (mucinous/non-mucinous), lymphovascular invasion and the number of lymph nodes with relapse free survival as the outcome variable. For the validation cohort, the multivariate Cox regression analysis adjusted for age, sex, TNM stage and tumour location with relapse free survival as the outcome variable.
HiFi, high fibroblast.

tumours (online supplemental figure 2G). This HPS median split was closely aligned to an area under the receiver operating characteristic (AUROC) optimal cut-off (online supplemental figure 3), which could significantly stratify patients with HiFi tumours based on relapse, where lower expression was associated with reduced RFS (figure 2E; top; log-rank p=0.0069) and those with high expression of HPS genes displayed RFS outcomes similar to those of LoFi patients.

The HPS was prognostic in the discovery cohort using either univariate (HR 0.395, 95% CI (0.191 to 0.816), Wald test p=0.012; table 1) or multivariate analysis adjusting for age, gender, tumour location, tumour differentiation, lymphovascular invasion status, tumour subtype and the number of lymph nodes (HR 0.218, 95% CI (0.087 to 0.544), Wald test p=0.001; table 1). Additionally, the prognostic value of the HPS was subtype-specific for patients with HiFi tumours, as it had no significant prognostic value when it was used to stratify patients with LoFi tumours (figure 2E; bottom; log-rank p=0.63215).

To independently validate these findings, we applied our ssGSEA fibroblast scoring method to transcriptional profiles from an independent validation cohort of untreated stage II/III CC tumours[26] (GSE39582; n=258 (online supplemental table 3). Similar to the discovery cohort, ssGSEA fibroblast scores were significantly higher in the CMS4 tumours compared with all other subtypes (online supplemental figure 4A); t-test p<0.0001). In line with stroma-rich populations identified in publicly-available cohorts (online supplemental figure 4B), patients within the top 20% ssGSEA fibroblast score were classed as HiFi (n=52) and the remaining 80% classed as LoFi (n=206), where HiFi samples were largely, but not exclusively, CMS4 (online supplemental figure 4C). Conversely, LoFi samples predominantly consisted of epithelium-rich subtypes; CMS2 and CMS3 (online supplemental figure 4C,D). HiFi tumours displayed higher StromalScore, and higher fibroblast score, alongside an analogous pattern of enrichment to that of the discovery cohort (online supplemental figure 4E). Importantly, and in line with the discovery findings, stratification of the HiFi tumours in this independent validation cohort using the median of HPS (again closely aligned to AUROC optimal cut-off; online supplemental figure 3), revealed that those with a low expression (n=26) had significantly lower IFN alpha and IFN gamma response signalling (figure 2F) and poorer RFS compared with those with a high expression (n=26) (relapse rates of 46.2% and 7.7%, respectively; figure 2G; top; log-rank p=0.00113). The HPS was also significantly prognostic in the validation cohort using both univariate (HR 0.123, 95% CI (0.027 to 0.550), p=0.006; table 1) and multivariate analyses (HR 0.093, 95% CI (0.019 to 0.466), p=0.004; table 1), which equates to a >10-fold higher risk of relapse in the HPS-low group compared with the HPS-high. We confirm the subtype-specific nature of the HPS, as it again provides no clinical value in stratifying the LoFi population based on outcome (figure 2G; bottom; log-rank p=0.46596).

This validation cohort contained additional molecular features that were not available in our discovery cohort; however, we found no significant associations between the HPS and mismatch repair, CIMP or CIN status, nor mutations in *TP53*, *KRAS* and *BRAF* (online supplemental figure 4F). While the vast majority of HiFi tumours were CMS4, we found that there was also no

significant difference in the proportions of the various CMS[3] and colorectal intrinsic subtypes[27] groups between HPS groups (online supplemental figure 4F), suggesting that our approach has identified HiFi-specific biology not identifiable using established genetic and transcriptional subtype analysis.

STAT1-mediated biology defines relapse in HiFi tumours

While our HPS was sufficient to stratify tumours into two groups based on RFS, we next investigated the overall differential biology underpinning HPS-high vs HPS-low tumours. Independent differential gene expression analyses were performed, revealing 41 genes significantly (BH adjusted p<0.05) differentially expressed in both discovery and validation cohorts; 30 upregulated and 11 downregulated genes in tumours with high signature expression (figure 3A, table 2).

Using the STRING database (string-db.org) to identify and visualise interactions, upregulated genes formed a network around STAT1 (figure 3B). As these analyses are based on total gene expression levels for *STAT1* itself; to assess downstream activation, we next examined *STAT1* in combination with several of its target genes (*STAT1*, *PSMB9* (*LMP2*), *IRF1* and *TAP1)*, where we observed strong direct

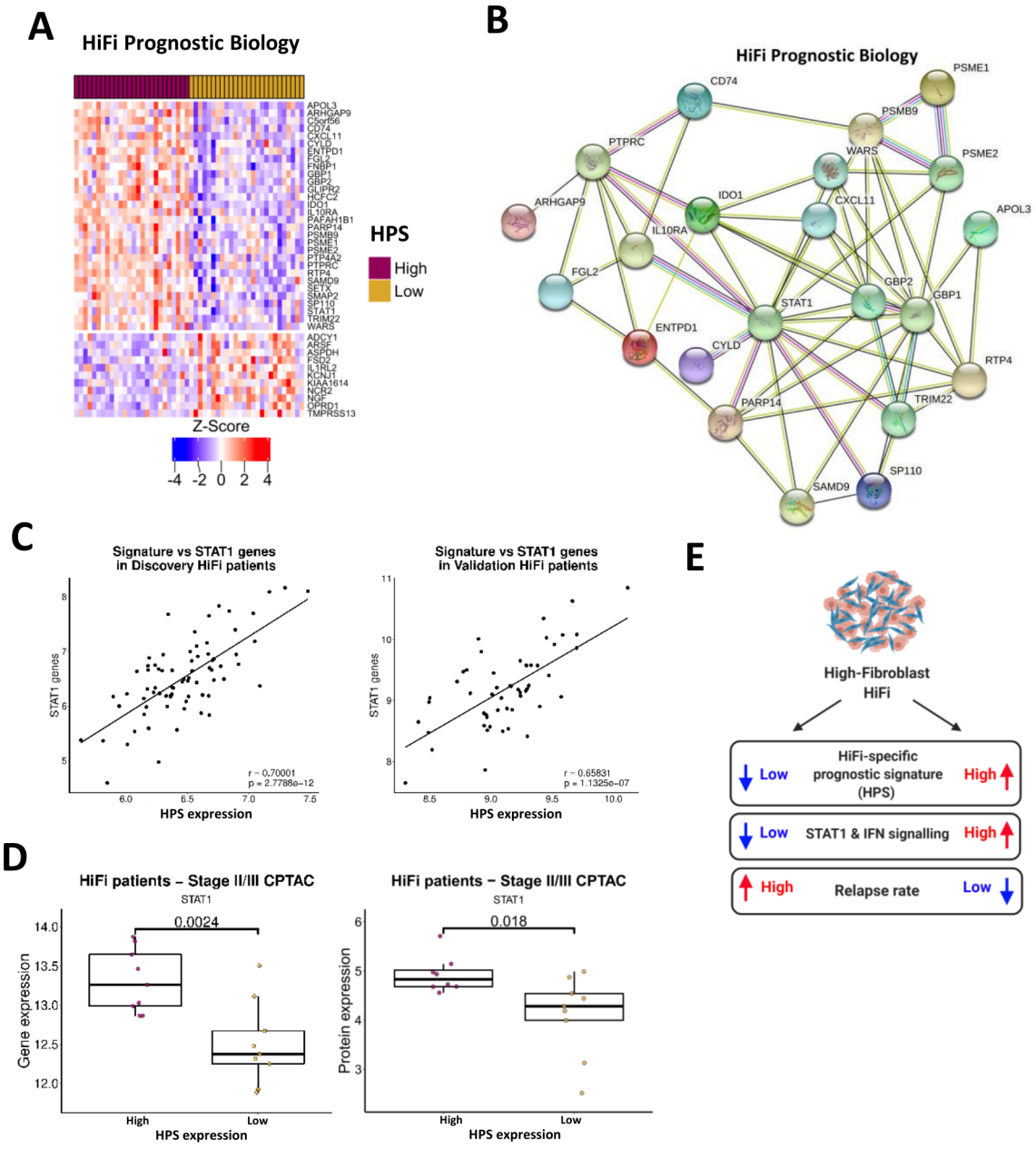


Figure 3 Validation of HiFi-specific prognostic biology and association with STAT1 (A) Heatmap displaying upregulated and downregulated genes shared by the differential comparisons between HPS expression groups in the discovery and validation cohorts (n=26 in each subgroup; 30 upregulated and 11 downregulated genes (adjusted p<0.05; table 2)). (B) String network formed by the upregulated genes form a cluster around STAT1. (C) Cumulative gene expression of *STAT1* and three of its target genes (*PSMB9, IRF1* and *TAP1*) correlated with expression of the HPS in the discovery (left) and validation cohort (right; t-test both p<0.00001). (D) Boxplots of *STAT1* gene expression (left) and protein levels (right) in HiFi patients in the CPTAC cohort according to HPS groups (high n=9 and low n=9; t-tests both p<0.05). (E) Schematic depicting the STAT1, IFN and relapse characteristics associated with the HiFi-specific prognostic signature within HiFi tumours. CPTAC, Clinical Proteomic Tumour Analysis Consortium; HiFi, high-fibroblast; HPS, HiFi-specific prognostic signature.

Table 2 Differential biology identified by HiFi-specific prognostic signature

Upregulated genes	Discovery Fold change	Discovery Adjusted p value	Validation Fold change	Validation Adjusted p value
APOL3	1.479163484	0.038046826	1.530912747	0.019648112
ARHGAP9	1.967295618	0.028903696	1.592821689	0.023805238
C5orf56	1.32317962	0.042212342	1.400667094	0.005369033
CD74	1.32309124	0.046116242	1.576548039	0.027523104
CXCL11	1.92974677	0.034123781	5.00897748	0.008189299
CYLD	1.438201775	0.028903696	1.469132166	0.010754914
ENTPD1	1.794582609	0.024260819	1.376059763	0.020745822
FGL2	1.345513	0.021933613	2.600331326	0.003587353
FNBP1	1.469824152	0.027721291	1.71533174	0.009536017
GBP1	1.65880111	0.024260819	1.618328245	0.046475584
GBP2	1.647099662	0.024260819	1.781673902	0.002288833
GLIPR2	1.476671309	0.024260819	1.611068902	0.007753646
HCFC2	1.525739539	0.017725045	1.421761963	0.012742342
IDO1	1.73588083	0.048494417	2.994630603	0.000754221
IL10RA	2.055572033	0.017725045	1.654583304	0.019098469
PAFAH1B1	1.309143492	0.038004231	1.448782114	0.015670524
PARP14	2.105452644	0.03660005	1.524254089	0.036762555
PSMB9	1.270092042	0.046116242	1.666844663	0.036265881
PSME1	1.642023214	0.000168244	1.250842547	0.03164453
PSME2	1.603403275	0.024059724	1.314548729	0.039143267
PTP4A2	1.346153934	0.043074478	1.213154499	0.02914658
PTPRC	1.476678757	0.036567059	2.33082532	0.008189299
RTP4	1.485506409	0.024260819	2.15318341	0.000552526
SAMD9	1.688418114	0.043449318	1.977749961	0.036984171
SETX	1.750386764	0.022838328	1.487177653	0.006733321
SMAP2	2.686823125	0.03156768	1.403024067	0.036891535
SP110	2.000918379	0.000168244	1.455018839	0.012537887
STAT1	1.859897307	0.013017895	1.518148697	0.020017274
TRIM22	1.539257527	0.036567059	1.730531226	0.023710413
WARS	1.460735784	0.024260819	1.696493454	0.013432331
Downregulated genes	**Discovery Fold change**	**Discovery Adjusted p value**	**Validation Fold change**	**Validation Adjusted p value**
ADCY1	−1.21661236	0.03714883	−1.242515217	0.040619631
ARSF	−1.259331053	0.030430316	−1.205077753	0.044850161
ASPDH	−1.209403567	0.040310191	−1.224072938	0.047792544
FSD2	−1.205195523	0.038100021	−1.128329641	0.04727427
IL1RL2	−1.181897912	0.047573384	−1.295819216	0.031084084
KCNJ1	−1.171753534	0.038739777	−1.121058095	0.029132685
KIAA1614	−1.239454355	0.031957568	−1.198516417	0.021006687
NCR2	−1.223150807	0.040766339	−1.16135162	0.046766842
NGF	−1.234256437	0.033460361	−1.283709027	0.006961439
OPRD1	−1.127349469	0.043716659	−1.188396135	0.044656705
TMPRSS13	−1.268170072	0.025078493	−1.34200588	0.0429152

HiFI, high-fibroblast.

positive correlation between their expression and the HPS in both discovery and validation cohorts (figure 3C; Pearson's Correlation r=0.70001 and r=0.65831, (online supplemental figure 4G). Furthermore, stratification of an additional independent cohort of stage II/III colon patients (Clinical Proteomic Tumour Analysis Consortium, CPTAC)[28] into HiFi and LoFi using our fibroblast score, followed by sub-stratification using the HPS, validated a significant enrichment for total STAT1 gene and protein expression in HiFi patients with high HPS expression (figure 3D; t-test p=0.0024 and p=0.018). Although these signalling pathways can be an indication of general tumour infiltration levels, we demonstrated that patient stratification based on the ESTIMATE ImmuneScore[17] is insufficient for prognostic stratification when applied specifically to HiFi patients, in either the discovery or validation cohorts (online supplemental figure 5A); left), and does not consistently align to HPS (online supplemental figure 5A; right). Furthermore, comparisons of the relative abundance of immune cells in the discovery and validation cohorts, using the CIBERSORT tool,[29] revealed a significantly larger proportion of dendritic cells (DCs) that was only apparent in the discovery cohort

HPS-high patients versus HPS-low and not recapitulated in the validation cohort (online supplemental figure 5B); t-test p=0.00076 and p=0.51333).

In summary, our HiFi-specific analyses identified that elevated expression of HPS, which distinguished primary tumours based on IFN-alpha, IFN-gamma and STAT1-related biological signalling, was significantly associated with disease relapse specifically within stroma-rich CC (figure 3E).

HiFi specific STAT1-related prognostic biology is associated with higher levels of immune lineage-specific antigen processing and presentation

Using transcriptional data derived from leucocyte, epithelial, fibroblast and endothelial lineages isolated from colorectal tumour tissue (GSE39396),[30] we determined that six of the seven HPS genes were highly associated with tumour infiltrating immune lineages compared with the other cell types (figure 4A). In line with functionally active STAT1 signalling (figure 3C), we observed increased expression of major histocompatibility (MHC) class I receptors, *HLA-A*, *HLA-B* and *HLA-C*, associated with HPS in both cohorts (online supplemental figure 5C); t-test p<0.05, above and below median HPS; *HLA-B* was not present on array used in the discovery cohort). In addition, elevated adaptive and innate immune signalling, alongside ssGSEA gene ontology scores for the antigen processing and presentation (APP) machinery (figure 4B–D, (online supplemental figure 5D) were all associated with high HPS. We next examined the association between HPS expression and APP specifically within purified immune lineages (GSE24759),[31] which revealed a significant and strong positive correlation between HPS expression (originally identified from bulk tumours; figure 4B–C) and APP signalling in mature antigen presenting cells (APC) (online supplemental figure 5E; Pearson's correlation r=0.89974, $p=1.36\times10^{-11}$). Interrogation of single-cell RNA-Seq data from tumour-infiltrating immune populations isolated from a further independent cohort of CRC tumours,[32] which again confirmed a significant elevation of HPS expression (figure 4E; t-test p<0.0001) and APP signalling (figure 4F; t-test all p<0.0001) in tumour infiltrating monocytes, macrophages and to a greater extent in DCs compared with epithelial and CAF populations.

Using transcriptional data derived from bone marrow-derived macrophages (BMDM) isolated from WT, *Stat1*Y701F (dominantnegative) or *Stat1*-/- mice (E-MTAB-3598),[33] we confirmed the essentiality of functional STAT1 in regulating gene expression of HPS and STAT1 targets (figure 4G), APP signalling using ssGSEA (figure 4H; t-test) and APP, IFN-alpha and IFN-gamma response signalling using pair-wise GSEA (figure 4I).

HPS signalling is associated with double stranded RNA and viral response cascades

Transcription factor (TF) activity prediction, using the DoRothEA resource, to identify potential regulons responsible for the signalling and phenotypes associated with HPS in HiFi tumours (online supplemental figure 6A) revealed a strong association with *STAT1, STAT2*, IFN (*IRF1, IRF9*) and NFκB (*NFKB1, REL, RELA, RELB*) TFs (figure 5A). In parallel, we used ingenuity pathway analysis in conjunction with the HPS differential genes identified earlier (figure 3A) to predict upstream regulators of the HiFi-specific prognostic biology, and in line with our findings thus far, interferon gamma (*IFNG*), *IRF7* and *STAT1* were all identified (figure 5B). In addition, the synthetic double stranded RNA (dsRNA) viral mimetic and TLR3 agonist, Poly(I:C), was also identified as an upstream regulator of, and potential therapeutic agent to activate, the STAT1-mediated signalling and APP phenotypes associated with prognosis in HiFi tumours (figure 5B). Poly(I:C) is a potent immune adjuvant via viral-mimicry that can be safely used for inducing both a transient innate immune response and maintained adaptive response, which notably is the same signalling we found was associated with the HPS (figures 4–5). We next investigated upstream events that could trigger the differential STAT1-mediated innate/adaptive immune activity and APP, and in line with poly(I:C) findings, these analyses revealed an enrichment of signalling associated with a viral response and the presence of dsRNA in non-relapsing HiFi tumours, with high HPS expression (figure 5C–E). Furthermore, this viral response relies on the presence of functional STAT1, emphasising the importance of this signalling cascade (figure 5F; t-test p<0.05).

Taken together, these data confirm the biology underpinning the bulk tumour-derived HPS is significantly associated with functional STAT1 activity and APP in tumour-infiltrating professional APC in CC, which may be downstream of a dsRNA and/or viral response in a subset of HiFi tumours (figure 5G).

The TLR3 agonist poly(I:C) elevates mechanistic phenotypes associated with improved outcome in HiFi CRC

Testing of IFN-alpha (IFNA), IFN-gamma (IFNG) or poly(I:C) in primary human macrophage immune lineages (GSE46599, GSE1925 and GSE41295) confirmed their ability to induce expression of the HPS genes, alongside increased expression of *STAT1* and its target genes (figure 6A). A therapeutic form of poly(I:C) has recently demonstrated favourable safety characteristics in a number of phase I clinical trials[34 35]; therefore, we selected poly(I:C) for further testing. Using a mouse DC model (GSE46478), we observed increased expression of the HPS genes and STAT1-related genes on treatment with poly(I:C), alongside significant induction of the same STAT, IFN and NFκB regulons (figure 6B) and IFN-alpha response, IFN-gamma response and APP associated with prognosis in HiFi tumours (figure 6C). These results were further confirmed using the RAW264.7 macrophage model (GSE15066; figure 6D and E).

Furthermore, to complement this transcriptional signalling, and to validate the utility of the *in silico* measure of APP, we next performed *in vitro* phenotypic measurements of antigen processing, using a fluorescent-labelled ova protein (DQ-ova) in the RAW264.7 macrophage model, cocultured with tumour-conditioned primary mesenchymal stromal cells to represent the stromal environment of a HiFi tumour microenvironment (TME) (figure 6F). In support of the potential therapeutic relevance of poly(I:C) in this setting, macrophages from the poly(I:C) treated cocultures had significantly higher DQ-ova fluorescence, and therefore induced antigen processing, in this model (figure 6F; t-test p=0.036, (online supplemental figure 6BC). To assess if the key characteristics associated with HPS in bulk tumour samples can be induced following a dsRNA/Poly(I:C) response in immune lineages, we created a 'Poly(I:C) Signature' of n=75 (human) differentially expressed genes from the Poly(I:C) treated DCs (figure 6B) (logFC >2 and adjusted p<0.001) (figure 6G, (online supplemental table 4). Using GSEA according to HPS subgroups in HiFi samples, the Poly(I:C) signature was significantly enriched in both the discovery and validation cohorts in HPS-high compared with HPS-low (figure 6H, (online supplemental figure 6D), further confirming that the biology underpinning HPS can be therapeutically induced via a viral-like dsRNA-response in immune cells.

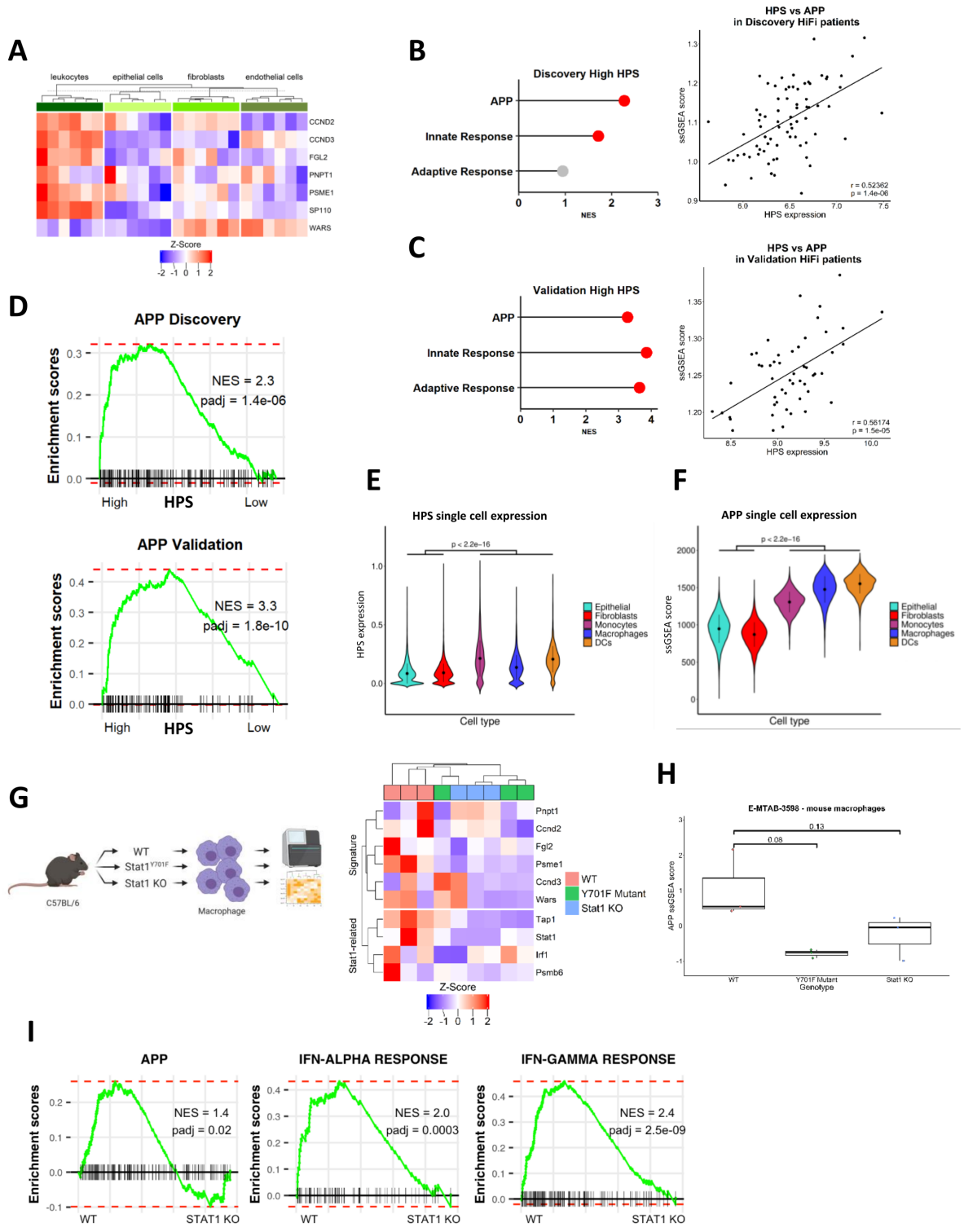


Figure 4 CRC tumour single-cell data confirms immune-specific nature of signature. (A) Gene expression of individual genes within the HPS according to a public dataset of CRC cell lineages purified by fluorescence-activated cell sorting (FACS) (n=4 populations from n=6 patients; total n=24). (B) Enrichment for APP, adaptive and innate signalling in HPS high group compared with low in HiFi tumours from the discovery cohort (left) (red=adjusted p<0.05). Correlation between ssGSEA scores for APP and HPS gene expression in the discovery cohort (Pearson's correlation r=0.5, p=1.4e-06; right). (C) Enrichment for APP, adaptive and innate signalling in HPS high group compared with low in HiFi tumours from the validation cohort (left). Correlation between ssGSEA scores for APP and HPS gene expression in the validation cohort (Pearson's correlation r=0.6, p=1.5e-05; right). (D) Enrichment for APP using pairwise GSEA in HPS high group compared with low in HIFI tumours from both the discovery and validation cohorts. (E, F) Immune cell populations have significantly higher expression of the HPS (E) and GO APP ssGSEA scores (F) than epithelial cells and fibroblasts (t-test both p<2.2e-16). (G) Expression levels of HPS genes and STAT1-related targets and (H) APP ssGSEA scores in bone-marrow derived macrophages with either wild-type (WT), mutant (Y701F mut) or knockout (KO) *STAT1* (n=3 for each genotype) (t-test). (I) Pairwise GSEA for GO APP, interferon alpha and gamma response in WT V STAT1 KO mouse macrophages. (n=3 per genotype). APP, antigen processing and presentation; CRC, colorectal cancer; GSEA, gene set enrichment analysis; HiFi, high-fibroblast; HPS, HiFi-specific prognostic signature; ssGSEA, single-sample GSEA.

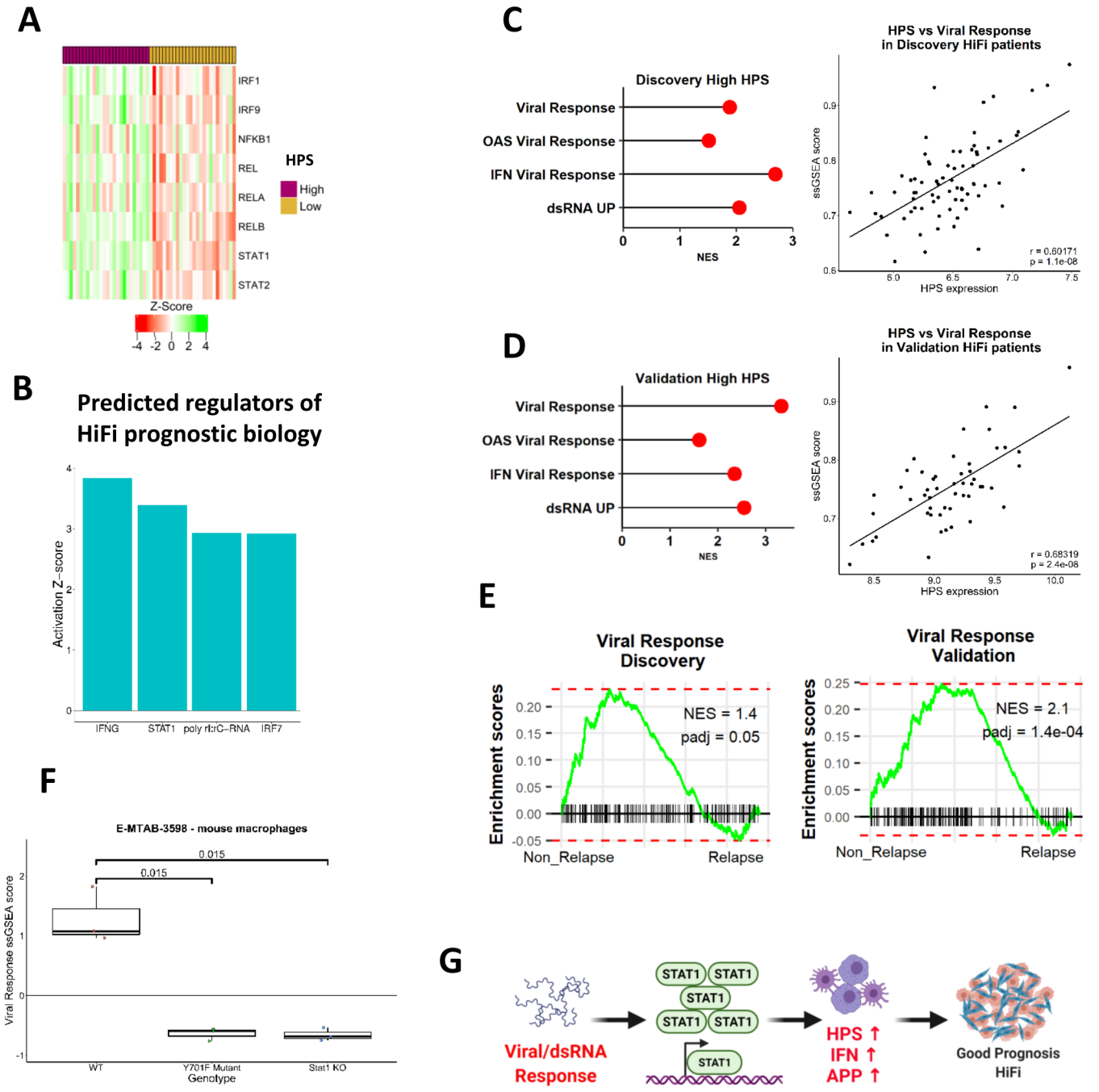


Figure 5 IFN and APP signalling cascades are associated with a STAT1-mediated viral/dsRNA response. (A) Activity status of key TF regulons according to HPS groups in the validation cohort (n=26 in each subgroup). (B) Top upstream regulators from an ingenuity pathway analysis (IPA) of the HPS differentially expressed genes in both the discovery and validation cohorts (table 2). (C) Enrichment for multiple viral response gene sets and dsRNA response in HPS high group compared with low in HiFi tumours in the discovery cohort (red=adjusted p<0.05; left). Correlation between ssGSEA scores for viral response and HPS gene expression in the discovery cohort (Pearson's correlation r=0.6, p=1.1e-08; right). (D) Enrichment for multiple viral response gene sets and dsRNA response in HPS high group compared with low in HiFi tumours in the validation cohort (left) (red=adjusted p<0.05). Correlation between ssGSEA scores for viral response and HPS gene expression in the validation cohort (Pearson's correlation r=0.7, p=2.4e-08; right). (E) Enrichment for viral response using pair-wise GSEA in non-relapse versus relapse HiFi tumours from both the discovery and validation cohorts. (F) viral response ssGSEA scores in bone-marrow derived macrophages with either wild-type (WT), mutant (Y701F mut) or knockout (KO) *STAT1*. (n=3 for each genotype) (t-test p<0.05). (G) Schematic detailing role for viral response/dsRNA signalling in regulating STAT1-mediated signalling cascades, HPS, APP and IFN signalling in immune lineages results in a good prognosis HiFi tumour. APP, antigen processing and presentation; CRC, colorectal cancer; ds RNA, double stranded RNA; HiFi, high-fibroblast; HPS, HiFi-specific prognostic signature; ssGSEA, single-sample GSEA; TF, transcription factor.

While previous studies have described the efficacy of poly(I:C) in tumour models, predominantly melanoma, its ability to reduce metastases in a CMS4-related genetically engineered mouse model (GEMM) has not been tested. To this end, we assessed a range of previously characterised GEMMs to identify genotypes associated with HiFi transcriptional signalling

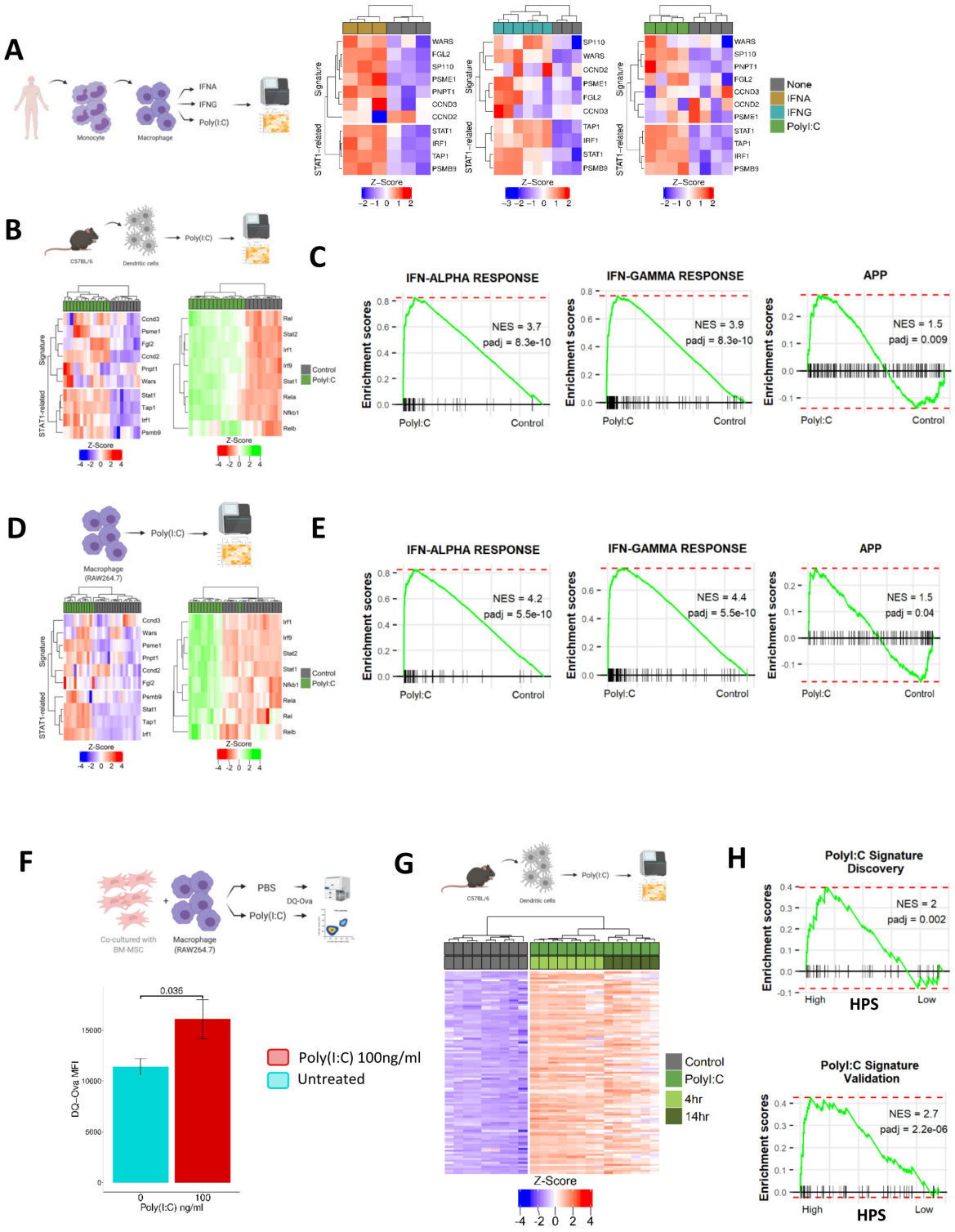


Figure 6 The TLR3 agonist poly(I:C) could be a potential treatment for HiFi (A) gene expression of HPS and STAT1 targets in human macrophages from different datasets treated with interferon (IFN) alpha (left) (n=3), IFN gamma (middle) (n=6) and poly(I:C) (right) (n=4) compared with untreated control samples (n=>3). (B) gene expression of HPS and STAT1 targets (left) and TF activity (right) in dendritic cells from mice treated with poly(I:C) (n=14) or untreated. (C) pair-wise GSEA of IFN alpha and gamma response, alongside APP gene sets in dendritic cells from mice treated with poly(I:C) or untreated. (D) Gene expression of HPS and STAT1 targets (left) and TF activity (right) in raw macrophage cells treated with poly(I:C) (n=12) or untreated. (E) Pair-wise GSEA of IFN alpha and gamma response, alongside APP gene sets in RAW macrophage cells treated with poly(I:C) or untreated. (F) Flow cytometry analysis of antigen processing in a co-culture comprised of primary mouse mesenchymal stromal cells (MSCs) and the mouse macrophage cell line RAW264.7, incubated with fluorescently labelled ovalbumin protein (DQ-Ova) and treated with either poly(I:C) or control (n=3) (t-test p<0.05). (G). differentially expressed genes (logFC >2 and adjusted p<0.001) in Poly(I:C) treated vs non-treated dendritic cells creating the 'Poly(I:C) Signature'. (H) Enrichment for Poly(I:C) Signature using pair-wise GSEA in HPS high group compared with low in HiFi tumours from both the discovery and validation cohorts. APP, antigen processing and presentation; HiFi, high-fibroblast; HPS, HiFi-specific prognostic signature; TF, transcription factor.

and histology. These analyses revealed that the recently developed stroma-rich CMS4 models; *Kras*$^{G12D/+}$, *Trp53*$^{fl/fl}$ (KP) and KP with constitutively activated NOTCH1 intracellular domain (KPN)[36] display significantly higher fibroblast scores (figure 7A) and stromal histology (figure 7B) compared with a number of *Apc*-based models.

In line with our discovery and validation human cohorts (figure 1E and online supplemental figure 4C), we saw a strong association between CMS4 classification and fibroblast scores (figure 7C). In addition, assessment of the same signals observed in HiFi versus LoFi human tumours (figure 1G, (online supplemental figure 4E) revealed an analogous pattern of enrichment in HiFi-related signalling cascades such as EMT, myogenesis and TGF-β signalling in HiFi/CMS4 GEMMs when compared with LoFi GEMMs (figure 7D). While both KP and KPN models were associated with HiFi/CMS4 classification, the KPN model was most representative of a poor prognostic HiFi model given its previously-reported highly metastatic nature.[36]

In line with this poor prognosis, we observe a significantly reduced APP signalling in CMS4 KPNs compared with the CMS4 KP models (figure 7E; NES=1.8). We, therefore, selected the KPN model to test the in vivo efficacy of poly(I:C) in reducing metastatic tumour burden using an intra-splenic injection metastatic assay (figure 7F). Following splenic KPN implantation, treatment with poly(I:C) (4 mg/kg administered biweekly by intraperitoneal injection from 9 to 42 days post-surgery) significantly reduced liver metastases burden in vivo, as assessed using a digital histology assessment (figure 7G; pooled in vivo results Mann-Whitney U $p<0.0002$). (online supplemental figure 7A); individual in vivo experiments), validating our in silico and in vitro analyses, alongside supporting its clinical translation in this setting. At endpoint, FLOW analyses of liver metastases (online supplemental figure 7B) revealed a significant elevation of CD3 +CD8+cytotoxic T cells and a complementary significant reduction in CD3 +CD4+T cells in poly(I:C)-treated mice compared with saline control (figure 7H; Mann-Whitney U both $p<0.05$).

DISCUSSION

We and others have previously demonstrated how CRC subtypes are heavily influenced by the composition of the TME, and the significant association with poor prognosis in the fibroblast-rich, CMS4 and stem-like subtype.[37–42] When compared with epithelial-rich subtypes, stroma-rich tumours display elevated signalling related to TGF-β and other stromal biologies, and this elevated signalling in general has been used as the rationale for targeting these pathways as potential therapeutic options. While substantial preclinical data supports TGF-β blockade as a promising target, the positive results obtained in *in vitro* and *in vivo* studies have not translated into clinical efficacy for stroma-rich tumours, even after numerous clinical trials in the past decade.[43 44] Currently there are significant efforts aimed at combining TGF-β blockade with immunotherapy, which may yet yield clinical benefit. However, in this study we reveal that the biology associated with disease relapse *within* stroma-rich tumours (and which are uniformly elevated for TGF-β signalling) is *not* associated with the biology that distinguishes between stroma-rich and epithelial-rich subtypes, nor is it associated with factors that are prognostic in general across unstratified stage II/III CC cohorts. Therefore, we set out to identify prognostic biology underpinning relapse specifically *within* stroma-rich tumours, and to use this new understanding to identify therapeutic vulnerabilities that could be exploited to reduce relapse rates in this poor prognostic patient group. This approach revealed that elevated levels of STAT1-mediated APP signalling downstream of a viral/dsRNA response in immune lineages correlated with improved RFS only in HiFi tumours; signalling that provides no prognostic value in the relatively good prognostic LoFi group. Furthermore, we demonstrate the therapeutic potential of this biology, as treatment with poly(I:C) resulted in elevation of the signalling and phenotype associated with good prognosis in HiFi tumours and, most importantly, a significant reduction of liver metastases in a mouse model of stroma-rich CC. Data presented here reveals a subtype-specific therapeutic approach, mediated via poly(I:C), that could potentially improve outcome for patients in the poor prognostic, high-fibroblast subtype of early stage CC (figure 8).

Ontology/pathway-led approaches for transcriptional analyses have the advantage of identifying biologically meaningful information associated with a particular subgroup, in our case relapse within HiFi tumours, rather than individual genes which can be confounded by issues such as intratumoural heterogeneity[39 45] or technical variations between profiling platforms/methods.[46] In addition, our study was designed to identify elevated phenotypes, and their regulators, associated with improved outcome to inform new treatment options that boost this biology; an approach that is known to deliver increased biological insights and more successful therapeutic outcomes when compared with those that rely on trying to downregulate or repress biology, which can be confounded by off-target effects.[47] The holistic discovery approach used here, which incorporates biological knowledge using experimentally validated signatures from the Hallmarks collection,[48] indicated that STAT1-mediated APP downstream of viral/dsRNA response was associated with reduced relapse rates in HiFi tumours, which in turn could be induced through treatment with IFNA, IFNG and the dsRNA TLR3 agonist, poly(I:C).

Interferon therapies have been trialled in multiple cancer types, but efficacy has been hindered by dose-limiting toxicities.[49] Trials using TLR agonists can induce IFN production, while causing fewer side effects compared with exogenous IFN treatment.[49] Following treatment with poly(I:C), a number of human or mouse macrophages and DCs display the same transcriptional signalling and APP activation that distinguished good from poor prognostic HiFi tumours and was sufficient to reduce metastatic lesions in a HiFi-specific mouse model. A recent breast cancer study also demonstrated that the prognostic value of DC subsets was dependent on the subtype of the tumours themselves, as signatures specific to plasmacytoid DCs and cDC2 cells were prognostic in triple-negative breast cancers, but not in the luminal subtype.[50] In line with this, the biology we identify as associated with prognosis in HiFi tumours provides no prognostic value in LoFi patients.

Elevation of dsRNA and viral response signalling was observed in HiFi tumours with reduced relapse rates and may provide a biological explanation for the differential activation status of this STAT1 and IFN-related biology in the different subsets of HiFi tumours. Activation of these same cascades were noted in a recent pancreatic cancer study as a downstream consequence of increased expression of endogenous retroviral transcripts.[51] In our study, we highlighted the essential nature of functional STAT1 in regulating this potential viral mimicry signalling in specific immune lineages and in the APP phenotype associated with improved outcome in HiFi tumours. Increased abundance and functionality of tumour infiltrating DCs correlate with prognosis in a variety of cancers,[52] and while dsRNA and viral signalling can drive their activation, we cannot rule out that these DCs are also regulated by other undiscovered TME-related factors

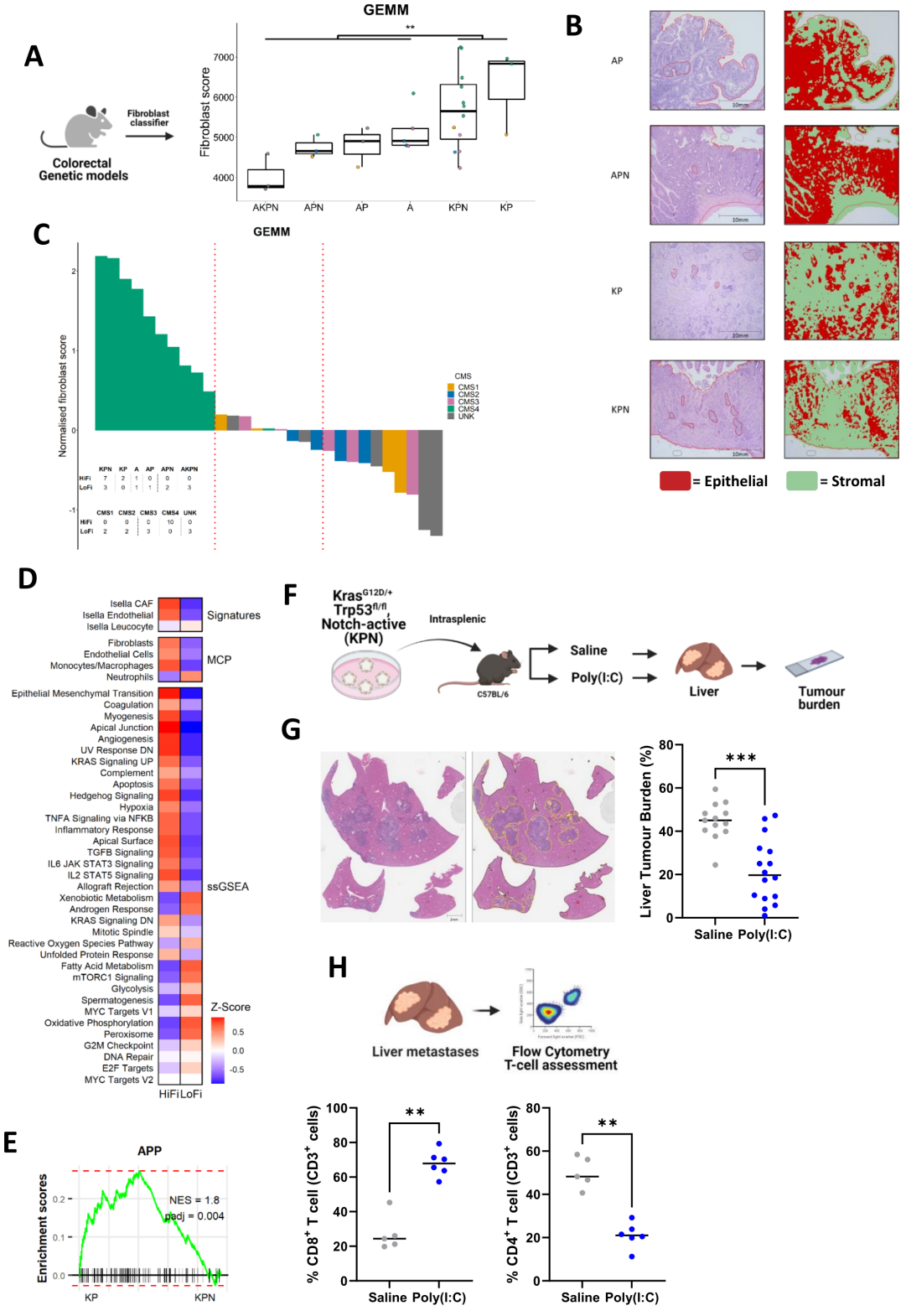


Figure 7 In vivo validation of poly(I:C) in HiFi model (A) CMS classification according to fibroblast score of GEMM genotypes. (A=*Apc*$^{fl/fl}$, K=*Kras*$^{G12D/+}$, p=*p53*$^{fl/fl}$ and n=*Notch1Tgl*$^{/+}$) (t-test p<0.01). (B) Stromal scores from H&E slides using digital pathology applied to GEMM tissue. (C) Waterfall plot of fibroblast scores indicating CMS classification in GEMMs. (D) comparison of HiFi (n=10) and LoFi (n=10) GEMMs using previously published stromal signatures and gene sets as assessed in figure 1G. (E) Pairwise GSEA of the APP gene set in CMS4 KP compared with CMS4 KPN. (F) Intrasplenic metastasis assay with KPN models in vivo treated with poly(I:C) compared with saline control. (G) Digital pathology assessment of H&E from in vivo studies demonstrates reduced liver metastasis in mice treated with poly(I:C) (n=16) compared with saline control (n=13) (Mann-Whitney U test). (H) Flow cytometry assessment of CD3 +cell populations from liver metastases in treatment groups highlight significant elevation of CD8 +T cells alongside significant reduction in CD4 +T cells in poly(I:C) arm (n=6) compared with saline (n=5) (Mann-Whitney U test; both p<0.05). APP, antigen processing and presentation; CMS, consensus molecular subtypes; GEMM, genetically engineered mouse model; HiFi, high fibroblast; LoFi, low fibroblast. *** denotes p<0.0002, ** denotes p<0.05.

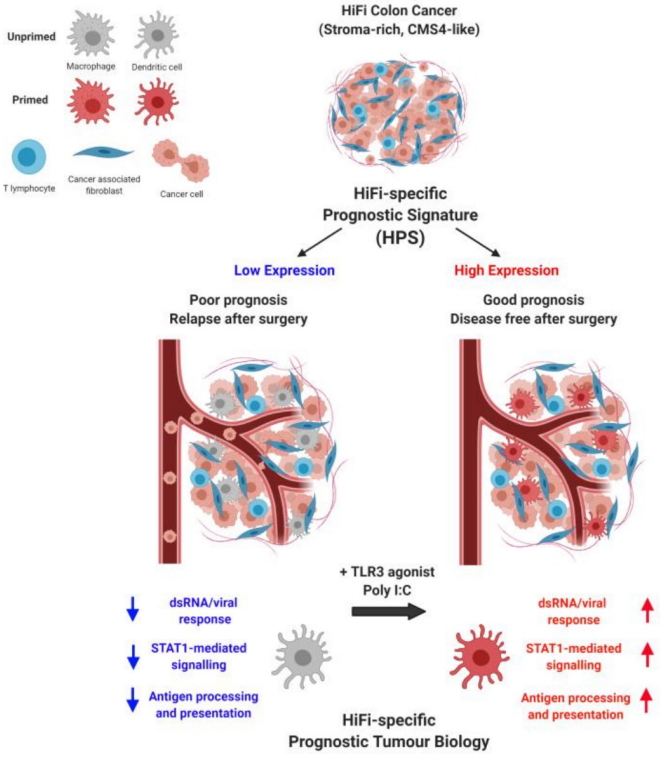


Figure 8 Graphical summary

within our study, including cytokines and other soluble factors. Data presented here have revealed a number of critical biological cascades underpinning poor prognosis in HiFi tumours, however additional secretome, epigenetic and microbiome profiling would be of interest to identify factors that regulate activation and survival in specific immune lineages, or in the case of DCs, by quantification of factors regulating their commitment, turnover and dendropoiesis.

Expansion and activation of DC lineages, via pretreatment with the growth factor FLT3L, which promotes DC commitment in haematopoietic progenitor cells and subsequent DC activation and growth, followed by poly(I:C) treatment, has demonstrated utility as an approach to improve response to checkpoint blockade in melanoma models.[53] Very recent data on the safety and efficacy of neoadjuvant immuno-oncology treatment (combined CTLA4 and PD1 blockade) in CC confirm the shifting clinical landscape for neoadjuvant treatment scheduling for localised colon tumours.[54] Results from that clinical trial indicate that while microsatellite instability-high tumours universally displayed pathological response to treatment, only 27% of microsatellite stable tumours were responsive; a group that urgently requires therapeutic interventions that can reprime the suppressed innate immune system and ultimately reinitiate tumour immune surveillance. Taken together, our subtype-specific *in silico* data, alongside the *in vitro* and *in vivo* data presented here, strongly support the clinical testing of poly(I:C) as a novel treatment option in the neoadjuvant setting to complement the current adjuvant standard of care, to reduce relapse rates for patients with stroma-rich CC.

In conclusion, our tumour-based discovery and validations, alongside *in vitro* and *in vivo* models, have identified a key role for viral/dsRNA response and IFN signalling, upstream of a STAT1-mediated cascade, which in turn drives an innate-adaptive immune activation, as a critical mediator of relapse in stroma-rich CRC. Data presented here provide a strong biological rationale for clinical testing of poly(I:C) as a novel therapeutic option to reduce metastatic relapse rates in the worst prognostic group of early stage CC.

METHODS

Additional Methods details are available within online supplemental material 1. A study overview including methods and criteria used is presented in online supplemental figure 8. Schematics designed using BioRender.com.

Patient datasets and data processing

The discovery transcriptional dataset was previously assembled for the development of the FDA-approved stage II ColDx/GeneFx assay,[20] consisting of 215 stage II CC patients (ArrayExpress accession number E-MTAB-863). As this was an existing cohort, we did not perform sample size power calculations. Clinicopathological information is in online supplemental table 2). The stage II/III untreated CC validation dataset (GSE39582) was downloaded as CEL files from GEO, processed then collapsed in the same way as the discovery dataset (detailed further in online supplemental methods). Clinicopathological information is in online supplemental table 3. RNA-seq and label-free proteomic data from the CPTAC[28] colon adenocarcinoma cohort (n=100) were downloaded from http://linkedomicsorg/cptac-colon/. The use of patient material from the S:CORT programme was approved by the ethics commission (REC 15/EE/0241). GSE39396: Fluorescence-activated cell sorted (FACS) purified cells (CD45 +leukocytes, FAP +fibroblasts, CD31 +endothelial cells and Epcam +epithelial cells) from 6 CRC patients. Data were retrieved from GEO in its log2 RMA normalised form.

Survival analysis

All survival analyses were performed in R using the survival package (V.3.2–13). For Kaplan-Meier curves and Cox proportional hazards models, median value of gene expression was used to dichotomise patients into high/low groups. In the discovery cohort, univariate Cox proportional hazards regression analysis was performed to identify genes that correlated with RFS. Genes with a likelihood $p<0.20$ were subjected to multivariate Cox analysis adjusted for age, sex, pT stage, tumour location, tumour differentiation grade, tumour subtype (mucinous/non-mucinous), lymphovascular invasion and the number of lymph nodes. A total of 214 samples were included (one sample was removed due to lack of clinicpathology information). The multivariate analysis in the validation dataset was adjusted for age, sex, TNM stage and tumour location. All the variables in the model were non-significant ($p>0.05$) and proportional hazards assumptions were met for both the discovery and validation cohorts using cox.zph function.

Immune lineage datasets

GSE24759: 38 purified populations of human haematopoietic cells.[31] GSE46599: primary human monocytes differentiated into macrophages and treated with interferon alpha.[55] GSE1925: primary human monocytes which were differentiated into macrophages and treated with interferon gamma.[56] GSE41295; primary human monocytes differentiated into macrophages and treated with poly(I:C).[57] GSE46478: primary DCs from C57BL/6 mice.[58] GSE15066: mouse macrophage cell line RAW264.7 stimulated with poly(I:C).[59] E-MTAB-3598: BMDM isolated from WT, $Stat1^{Y701F}$ or $Stat1^{-/-}$ mice.[33] All of these datasets were collapsed to gene-level using the same method for patient data outlined in online supplemental methods section.

The generation of the single-cell RNA-Seq data has been previously published[32] and is detailed in online supplemental methods.

GEMM dataset descriptor and histology

Detailed information is available in online supplemental methods. Briefly, all animal experiments were performed in accordance with a UK Home Office Project Licence (70/9112), observed ARRIVE guidelines and were reviewed by local animal welfare and the ethical review committee at the University of Glasgow. Histology assessments were performed on a previously established and described cohort as part of the ACRCelerator programme (https://www.beatson.gla.ac.uk/ACRCelerate/acrcelerate.html).

KPN intrasplenic model

Intrasplenic injection was performed as previously described (Jackstadt *et al*)[36] using a cell suspension of liver metastasis organoids derived from a single C57BL/6 *Kras*$^{G12D/+}$, *Trp53*$^{fl/fl}$, constitutively activated NOTCH1 (KPN) mouse. Organoid donor and recipient mice were sex and strain matched. Nine days postimplantation mice were administered poly(I:C) at 4 mg/kg in saline by biweekly intraperitoneal injection (n=6) or saline vehicle control (n=6; n=5 used for tissue processing) until sampling on day 42. Blind to treatment, body weight and liver weight was recorded and gross liver metastasis quantified. The study was repeated with n=12 poly(I:C) and n=12 saline with similar results.

Mouse model tissue processing

A biopsy of liver metastasis was taken for flow cytometry with the remainder fixed in 10% neutral buffered formalin and processed by standard histological processing into paraffin. 5 µm sections were cut and stained for H&E.

SAMPLE PROCESSING AND STAINING FOR FLOW CYTOMETRY

A detailed protocol and gating strategy is included in the online supplemental methods, online supplemental figures section. Analysis was conducted on FlowJo V.10.7.2. Cells were gated based on live cell status from live/dead stain, single cells and CD45 +cells selected. From here, data was down-sampled to 12 000 events per sample and CD3 +CD4+ and CD3+CD8+ cells were identified.

METASTASIS SCORING

H&E sections of liver were scanned using an Aperio AT2 slide scanner at 20x and svs files imported) into QuPath v0.2.3. A pixel thresholder (Resolution: 4.02 µm/px; Channel: Average channels; Prefilter: Gaussian; Smoothing sigma: 3.0; Threshold: 200.0) was applied to quantify the total tissue area. Liver metastasis were manually annotated on each whole slide image. The mean tissue and liver metastasis area was utilised to calculate a tumour burden percentage per mouse.

Author affiliations
[1]Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Belfast, UK
[2]Cancer Research UK, Beatson Institute for Cancer Research, Glasgow, UK
[3]Lambe Institute for Translational Research, College of Medicine Nursing and Health Sciences, National University of Ireland, Galway, Ireland
[4]Discipline of Pharmacology & Therapeutics, School of Medicine, National University of Ireland, Galway, Ireland
[5]Université de Genève, Geneva, Switzerland
[6]Institute of Cancer Sciences, University of Glasgow, Glasgow, UK
[7]Leeds Institute of Medical Research, University of Leeds, Leeds, UK
[8]Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH) and Cancer Progression and Metastasis Group, German Cancer Research Center (DKFZ), Heidelberg, Germany
[9]Cellular Pathology, Belfast Health and Social Care Trust, Belfast, UK
[10]Centre for Public Health, Queens University Belfast, Belfast, UK
[11]Digestive Oncology Unit, University Ospital Gasthuisberg, Leuven, Belgium
[12]Department of Oncology, University of Oxford, Oxford, UK
[13]Wellcome Trust Centre Human Genetics, University of Oxford, Oxford, UK

Twitter Shania M Corry @ShaniaMCorry, Niamh A Leonard @NiamhLeonard5 and Philip D Dunne @DrPipDunne

Contributors Study concept and design: SC, AmM, ML, PDD. Performing experiments and interpretation of results: SC, AmM, TL, NL, NF, RMB, PT, XL, RA, KLR, AoM, SaM, MG, MW, SDR, RJ, ADC, SaM, SiM, DL, MBL, HGC, EK, ST, TM, SJL, DS, AR, OJS, ML, PDD. Data generation: SC, AmM, TL, NL, NF, PT, XL, RA, KLR, ER, SS, SDR, DS, AR, PDD. Manuscript draft: SC, AmM, ML, PDD. Manuscript review: all authors. Guarantor: PDD.

Funding This research was supported by a CRUK early detection project grant (A29834; PDD), an International Accelerator Award, ACRCelerate, jointly funded by Cancer Research UK (A26825 and A28223), FC AECC (GEACC18004TAB) and AIRC (22795), CRUK core funding (A21139; OJS. A17196 and A31287; CRUK Beatson Institute), a Queen's University Belfast Foundation grant, the CRUK Belfast Experimental Cancer Medicine Centre, a joint MRC-CRUK Stratified Medicine Programme Grant (S:CORT; MR/M016587/1) and a Health Data Research UK Substantive Site grant. We thank the Patrick G Johnston Centre for Cancer Research Applied Genomics Hub for Bioinformatics Support.

Disclaimer This manuscript reflects the views of the authors and may not reflect the opinions or views of the funders.

Competing interests ML has received honoraria from Pfizer, EMF Serono and Roche for presentations unrelated to this work; ML has received an unrestricted educational grant from Pfizer for research unrelated to this work. PT has received honoraria and travel expenses from BMS, Merck, Roche, Lilly and Sanofi-Aventis for contributions that are not related to the present work. The authors declare no other potential conflicts of interest.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Scripts and source files to reproduce all results will be deposited on our lab website (www.dunne-lab.com)

ORCID iDs
Ryan M Byrne http://orcid.org/0000-0001-6505-6345
Maurice B Loughrey http://orcid.org/0000-0001-8424-1765
Mark Lawler http://orcid.org/0000-0002-3930-8470
Philip D Dunne http://orcid.org/0000-0001-9160-283X

REFERENCES

1 Ferlay J, Soerjomataram I, Dikshit R, *et al*. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.
2 Böckelman C, Engelmann BE, Kaprio T, *et al*. Risk of recurrence in patients with colon cancer stage II and III: a systematic review and meta-analysis of recent literature. *Acta Oncol* 2015;54:5–16.

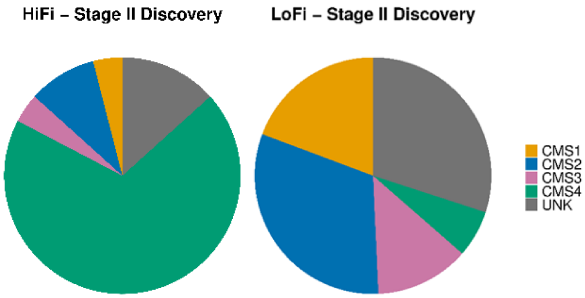
3 Guinney J, Dienstmann R, Wang X, *et al*. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015;21:1350–6.
4 Jass JR, Love SB, Northover JM. A new prognostic classification of rectal cancer. *Lancet* 1987;1:1303–6.
5 Jass JR, Morson BC. Reporting colorectal cancer. *J Clin Pathol* 1987;40:1016–23.
6 Hynes SO, Coleman HG, Kelly PJ, *et al*. Back to the future: routine morphological assessment of the tumour microenvironment is prognostic in stage II/III colon cancer in a large population-based study. *Histopathology* 2017;71:12–26.
7 Park JH, van Wyk H, McMillan DC, *et al*. Preoperative, biopsy-based assessment of the tumour microenvironment in patients with primary operable colorectal cancer. *J Pathol Clin Res* 2020;6:30–9.
8 Calon A, Lonardo E, Berenguer-Llergo A, *et al*. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet* 2015;47:320–9.
9 Roseweir AK, Park JH, Hoorn ST, *et al*. Histological phenotypic subtypes predict recurrence risk and response to adjuvant chemotherapy in patients with stage III colorectal cancer. *J Pathol Clin Res* 2020;6:283–96.
10 Ten Hoorn S, de Back TR, Sommeijer DW, *et al*. Clinical value of consensus molecular subtypes in colorectal cancer: a systematic review and meta-analysis. *J Natl Cancer Inst* 2021. doi:10.1093/jnci/djab106. [Epub ahead of print: 02 Jun 2021].
11 Alazzouzi H, Alhopuro P, Salovaara R, *et al*. Smad4 as a prognostic marker in colorectal cancer. *Clin Cancer Res* 2005;11:2606–11.
12 Calon A, Espinet E, Palomo-Ponce S, *et al*. Dependency of colorectal cancer on a TGF-β-Driven program in stromal cells for metastasis initiation. *Cancer Cell* 2012;22:571–84.
13 Herzig DO, Tsikitis VL. Molecular markers for colon diagnosis, prognosis and targeted therapy. *J Surg Oncol* 2015;111:96–102.
14 Tsushima H, Kawata S, Tamura S, *et al*. High levels of transforming growth factor beta 1 in patients with colorectal cancer: association with disease progression. *Gastroenterology* 1996;110:375–82.
15 Bankhead P, Loughrey MB, Fernández JA, *et al*. QuPath: open source software for digital pathology image analysis. *Sci Rep* 2017;7:16878.
16 Lawler M, Kaplan R, Wilson RH, *et al*. Changing the Paradigm—Multistage Multiarm randomized trials and stratified cancer medicine. *Oncologist* 2015;20:849–51.
17 Yoshihara K, Shahmoradgoli M, Martínez E, *et al*. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612 https://www.nature.com/articles/ncomms3612#supplementary-information
18 Isella C, Terrasi A, Bellomo SE, *et al*. Stromal contribution to the colorectal cancer transcriptome. *Nat Genet* 2015;47:312–9.
19 Becht E, Giraldo NA, Lacroix L, *et al*. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* 2016;17:218.
20 Kennedy RD, Bylesjo M, Kerr P, *et al*. Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue. *J Clin Oncol* 2011;29:4620–6.
21 McCorry AM, Loughrey MB, Longley DB, *et al*. Epithelial-to-mesenchymal transition signature assessment in colorectal cancer quantifies tumour stromal content rather than true transition. *J Pathol* 2018;246:422–6.
22 Glentis A, Oertle P, Mariani P, *et al*. Cancer-associated fibroblasts induce metalloprotease-independent cancer cell invasion of the basement membrane. *Nat Commun* 2017;8:924.
23 Li H, Courtois ET, Sengupta D, *et al*. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* 2017;49:708–18.
24 Mizoguchi F, Slowikowski K, Wei K, *et al*. Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nat Commun* 2018;9:789.
25 Pearce OMT, Delaine-Smith RM, Maniati E, *et al*. Deconstruction of a metastatic tumor microenvironment reveals a common matrix response in human cancers. *Cancer Discov* 2018;8:304–19.
26 Marisa L, de Reyniès A, Duval A, *et al*. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013;10:e1001453.
27 Isella C, Brundu F, Bellomo SE, *et al*. Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nat Commun* 2017;8:15107.
28 Edwards NJ, Oberti M, Thangudu RR, *et al*. The CPTAC data portal: a resource for cancer proteomics research. *J Proteome Res* 2015;14:2707–13.
29 Newman AM, Liu CL, Green MR, *et al*. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12:453–7.
30 Calon A, Espinet E, Palomo-Ponce S, *et al*. Dependency of colorectal cancer on a TGF-β-driven program in stromal cells for metastasis initiation. *Cancer Cell* 2012;22:571–84.
31 Novershtern N, Subramanian A, Lawton LN, *et al*. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 2011;144:296–309.
32 Qian J, Olbrecht S, Boeckx B, *et al*. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. *Cell Res* 2020;30:745–62.
33 Majoros A, Platanitis E, Szappanos D, *et al*. Response to interferons and antibacterial innate immunity in the absence of tyrosine-phosphorylated STAT1. *EMBO Rep* 2016;17:367–82.
34 Mehrotra S, Britten CD, Chin S, *et al*. Vaccination with poly(IC:LC) and peptide-pulsed autologous dendritic cells in patients with pancreatic cancer. *J Hematol Oncol* 2017;10:82.
35 Rodríguez-Ruiz ME, Perez-Gracia JL, Rodríguez I, *et al*. Combined immunotherapy encompassing intratumoral poly-ICLC, dendritic-cell vaccination and radiotherapy in advanced cancer patients. *Ann Oncol* 2018;29:1312–9.
36 Jackstadt R, van Hooff SR, Leach JD, *et al*. Epithelial NOTCH signaling rewires the tumor microenvironment of colorectal cancer to drive poor-prognosis subtypes and metastasis. *Cancer Cell* 2019;36:319–36.
37 Budinska E, Popovici V, Tejpar S, *et al*. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol* 2013;231:63–76.
38 De Sousa E Melo F, Wang X, Jansen M, *et al*. Poor-Prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med* 2013;19:614–8.
39 Dunne PD, Alderdice M, O'Reilly PG, *et al*. Cancer-cell intrinsic gene expression signatures overcome intratumoural heterogeneity bias in colorectal cancer patient classification. *Nat Commun* 2017;8:15657.
40 Marisa L, de Reyniès A, Duval A, *et al*. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013;10:e1001453.
41 Sadanandam A, Lyssiotis CA, Homicsko K, *et al*. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* 2013;19:619–25 https://www.nature.com/articles/nm.3175#supplementary-information
42 Schlicker A, Beran G, Chresta CM, *et al*. Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med Genomics* 2012;5:66.
43 Ciardiello D, Elez E, Tabernero J, *et al*. Clinical development of therapies targeting TGFβ: current knowledge and future perspectives. *Ann Oncol* 2020;31:1336–49.
44 Tauriello DVF, Palomo-Ponce S, Stork D, *et al*. TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* 2018;554:538–43.
45 Dunne PD, McArt DG, Bradley CA, *et al*. Challenging the cancer molecular stratification dogma: intratumoral heterogeneity undermines consensus molecular subtypes and potential diagnostic value in colorectal cancer. *Clin Cancer Res* 2016;22:4095–104.
46 Leek JT, Scharpf RB, Bravo HC, *et al*. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010;11:733–9.
47 Kaelin WG. Common pitfalls in preclinical cancer target validation. *Nat Rev Cancer* 2017;17:441–50.
48 Liberzon A, Birger C, Thorvaldsdóttir H, *et al*. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–25.
49 Parker BS, Rautela J, Hertzog PJ. Antitumour actions of interferons: implications for cancer therapy. *Nat Rev Cancer* 2016;16:131–44.
50 Michea P, Noël F, Zakine E, *et al*. Adjustment of dendritic cells to the breast-cancer microenvironment is subset specific. *Nat Immunol* 2018;19:885–97.
51 Espinet E, Gu Z, Imbusch CD, *et al*. Aggressive PDACs show hypomethylation of repetitive elements and the execution of an intrinsic IFN program linked to a ductal cell of origin. *Cancer Discov* 2021;11:638–59.
52 Karthaus N, Torensma R, Tel J. Deciphering the message broadcast by tumor-infiltrating dendritic cells. *Am J Pathol* 2012;181:733–42.
53 Salmon H, Idoyaga J, Rahman A, *et al*. Expansion and Activation of CD103(+) Dendritic Cell Progenitors at the Tumor Site Enhances Tumor Responses to Therapeutic PD-L1 and BRAF Inhibition. *Immunity* 2016;44:924–38.
54 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;26:566–76.
55 Goujon C, Moncorgé O, Bauby H, *et al*. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* 2013;502:559–62.
56 Hu X, Park-Min K-H, Ho HH, *et al*. IFN-gamma-primed macrophages exhibit increased CCR2-dependent migration and altered IFN-gamma responses mediated by STAT1. *J Immunol* 2005;175:3637–47.
57 Short KR, Vissers M, de Kleijn S, *et al*. Bacterial lipopolysaccharide inhibits influenza virus infection of human macrophages and the consequent induction of CD8+ T cell immunity. *J Innate Immun* 2014;6:129–39.
58 Pantel A, Teixeira A, Haddad E, *et al*. Direct type I IFN but not MDA5/TLR3 activation of dendritic cells is required for maturation and metabolic shift to glycolysis after poly IC stimulation. *PLoS Biol* 2014;12:e1001759.
59 Yang IV, Jiang W, Rutledge HR, *et al*. Identification of novel innate immune genes by transcriptional profiling of macrophages stimulated with TLR ligands. *Mol Immunol* 2011;48:1886–95.

Supplementary Figure 1

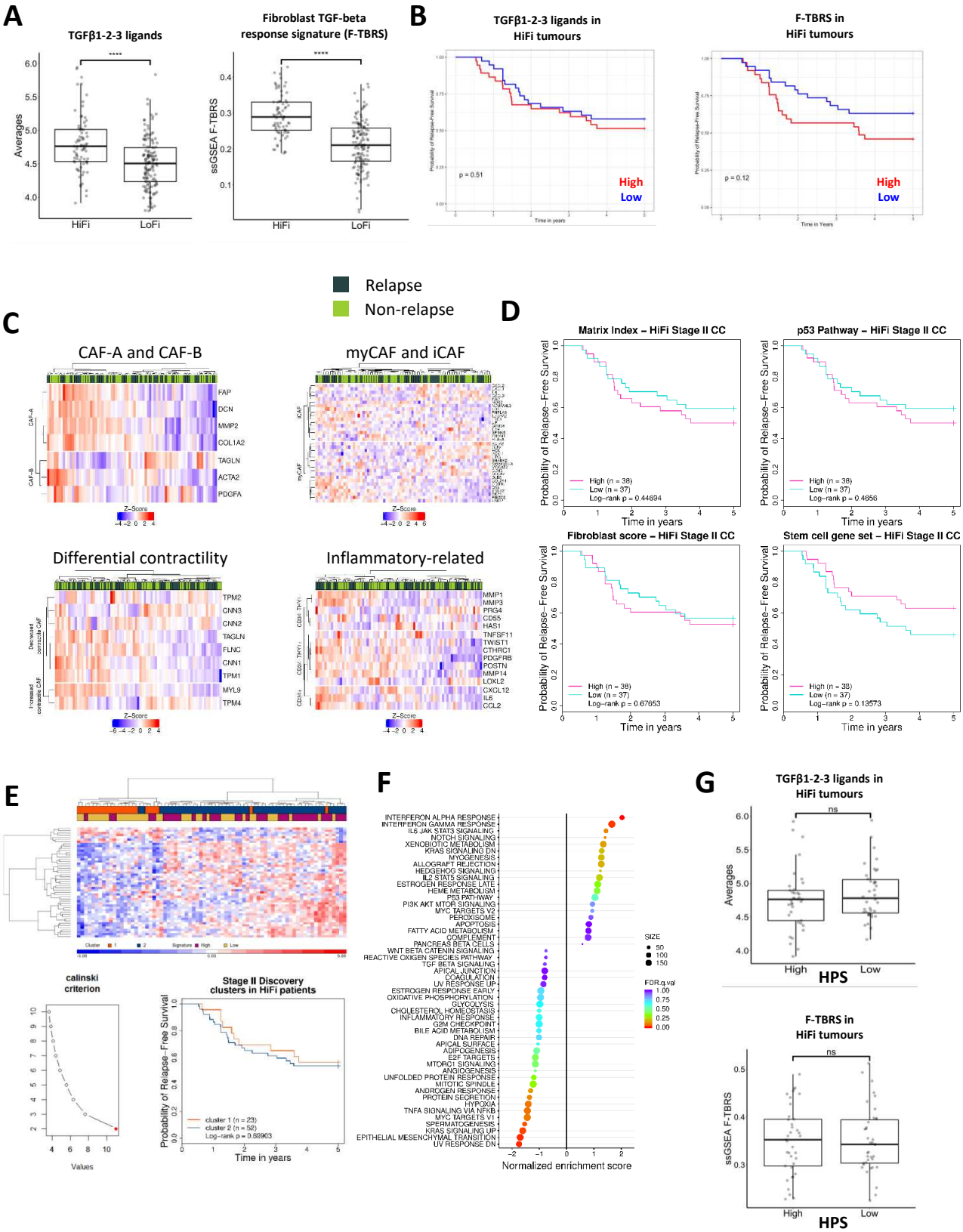
A

Cut1	HR1	P1
1473.97	0.909811877	0.812063248
1512.056241	0.669377513	0.219964264
1550.142482	0.750410578	0.380201791
1588.228723	0.791699836	0.475315669
1626.314964	0.833520265	0.577835972
1664.401205	0.875883881	0.685450566
1702.487446	0.875883881	0.685450566
1740.573687	0.822405819	0.522782527
1778.659928	0.871546038	0.643757589
1816.746169	0.83781441	0.541295664
1854.83241	0.906726625	0.729292178
1892.918651	1.037528078	0.896398444
1931.004892	1.09257795	0.749181528
1969.091133	1.103278018	0.712536404
2007.177374	1.166347115	0.564046108
2045.263615	1.13913282	0.619640478
2083.349856	1.056373612	0.829774604
2121.436097	1.121894773	0.648049994
2159.522338	1.136163904	0.608429835
2197.608579	1.282838944	0.317540457
2235.69482	1.326277252	0.252390041
2273.781061	1.422476785	0.153210865
2311.867302	1.594979363	0.056234447
2349.953543	1.567089383	0.062123833
2388.039784	1.521138426	0.077987899
2426.126025	1.666354256	0.03193368
2464.212266	1.785113236	0.014929252
2502.298507	1.702188545	0.024214375
2540.384748	1.584347371	0.04942289
2578.470989	1.702595408	0.02308708
2616.55723	1.601057844	0.044604403
2654.643471	1.640883615	0.034589522
2692.729712	**1.850538756**	**0.008755126**
2730.815953	1.853195883	0.008963489
2768.902194	1.863391509	0.008939134
2806.988435	1.640038686	0.044984899
2845.074676	1.692204722	0.03304536
2883.160917	1.729891398	0.029714302
2921.247158	1.729891398	0.029714302
2959.333399	1.679502085	0.045077266
2997.41964	1.725587137	0.040985951
3035.505881	1.75053168	0.039389894
3073.592122	1.939696431	0.019326126
3111.678363	1.876330989	0.030005828
3149.764604	1.598088372	0.125699913
3187.850845	0.9750217	0.94926715

B

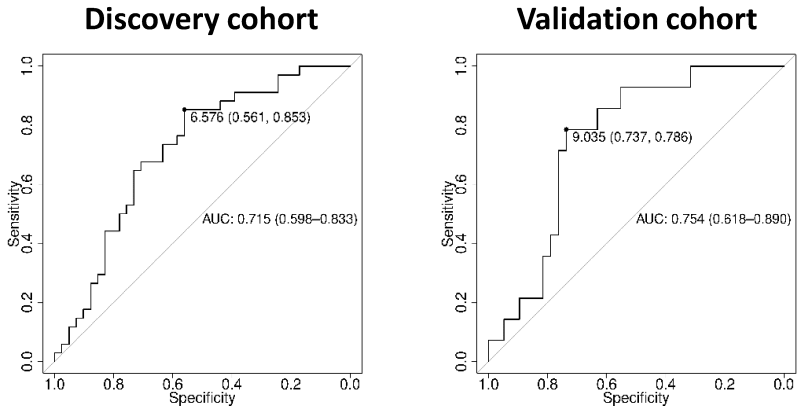


Supplementary Figure 2



Supplementary Figure 3

A



B

Discovery cohort

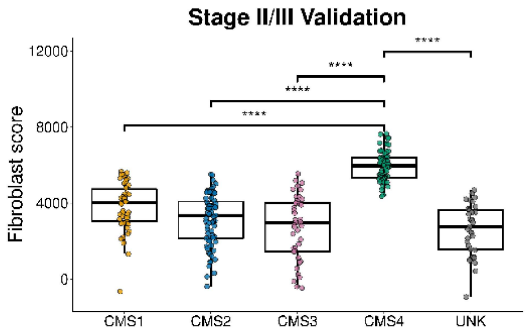
Method of stratification	Expression threshold	Sensitivity	Specificity
ROC	6.5755150	0.5609756	0.8529412
Median	6.4744600	0.6341463	0.6764706

Validation cohort

Method of stratification	Expression threshold	Sensitivity	Specificity
ROC	9.0351230	0.7368421	0.7857143
Median	9.1369379	0.6315789	0.8571429

Supplementary Figure 4

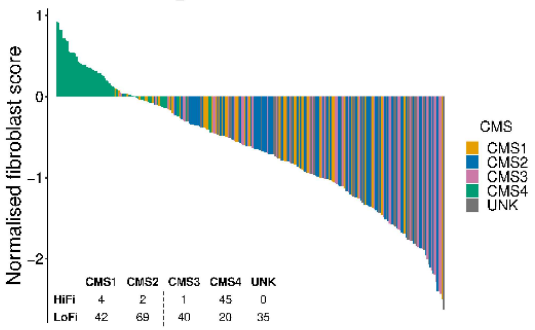
A



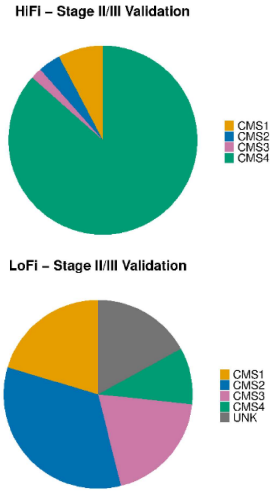
B

First author of study	Subgroup of poor prognostic mesenchymal/stem biology	Source of percentage data	Percentage of subgroup in cohort
Guinney	CMS4	Main paper text	23
Roepman	C-type	Main paper text	16
Budinska	D	Supp. Appendix S1	22.7
Sadanandam	Stem-like	Supp. Table 3	20
De Sousa E Melo	CCS3	Main paper text	27
Marisa	C4	Supp. Table S1	10.4
		Average percentage	**19.85**

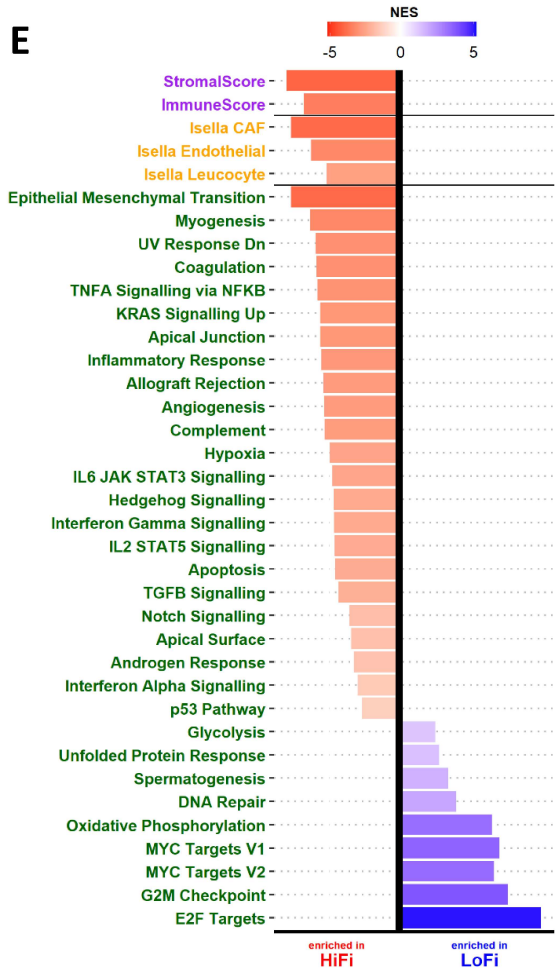
C



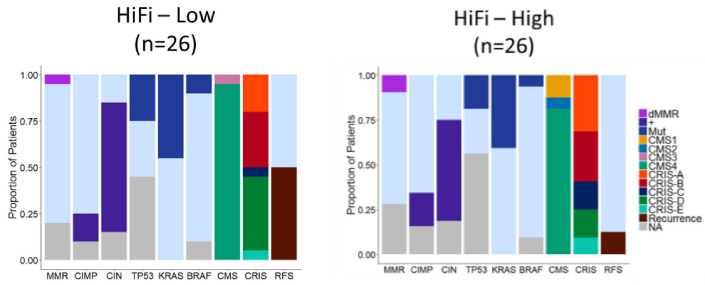
D



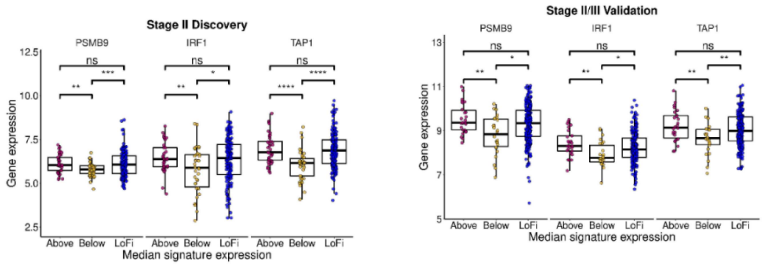
E



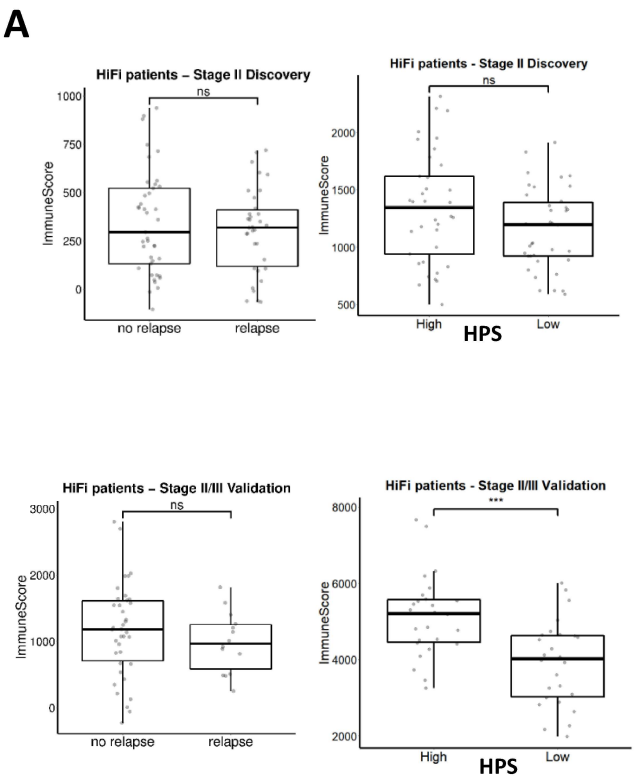
F

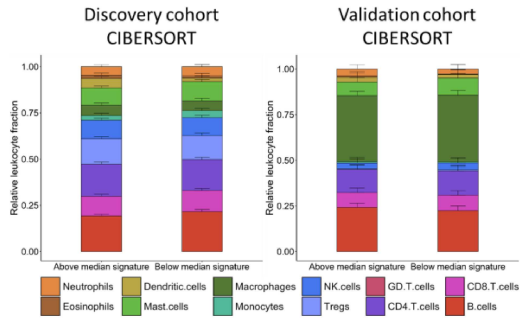


G

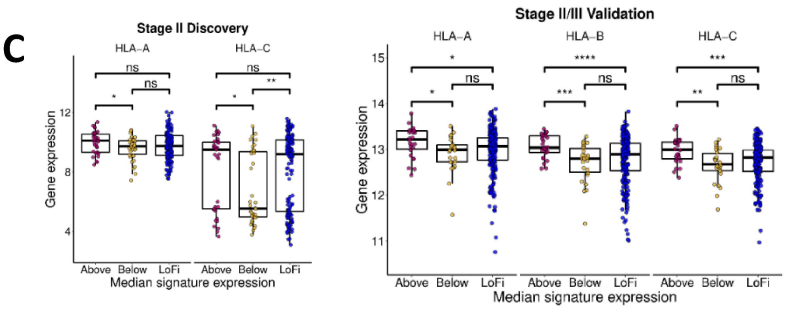


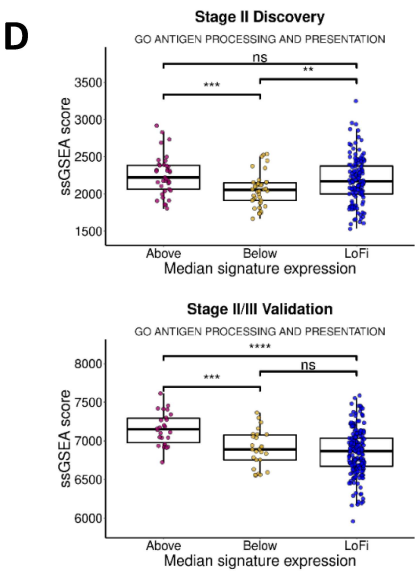
Supplementary Figure 5

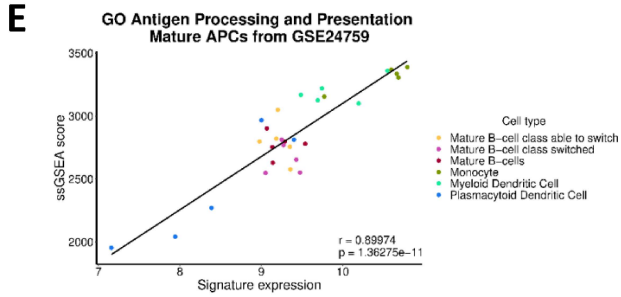




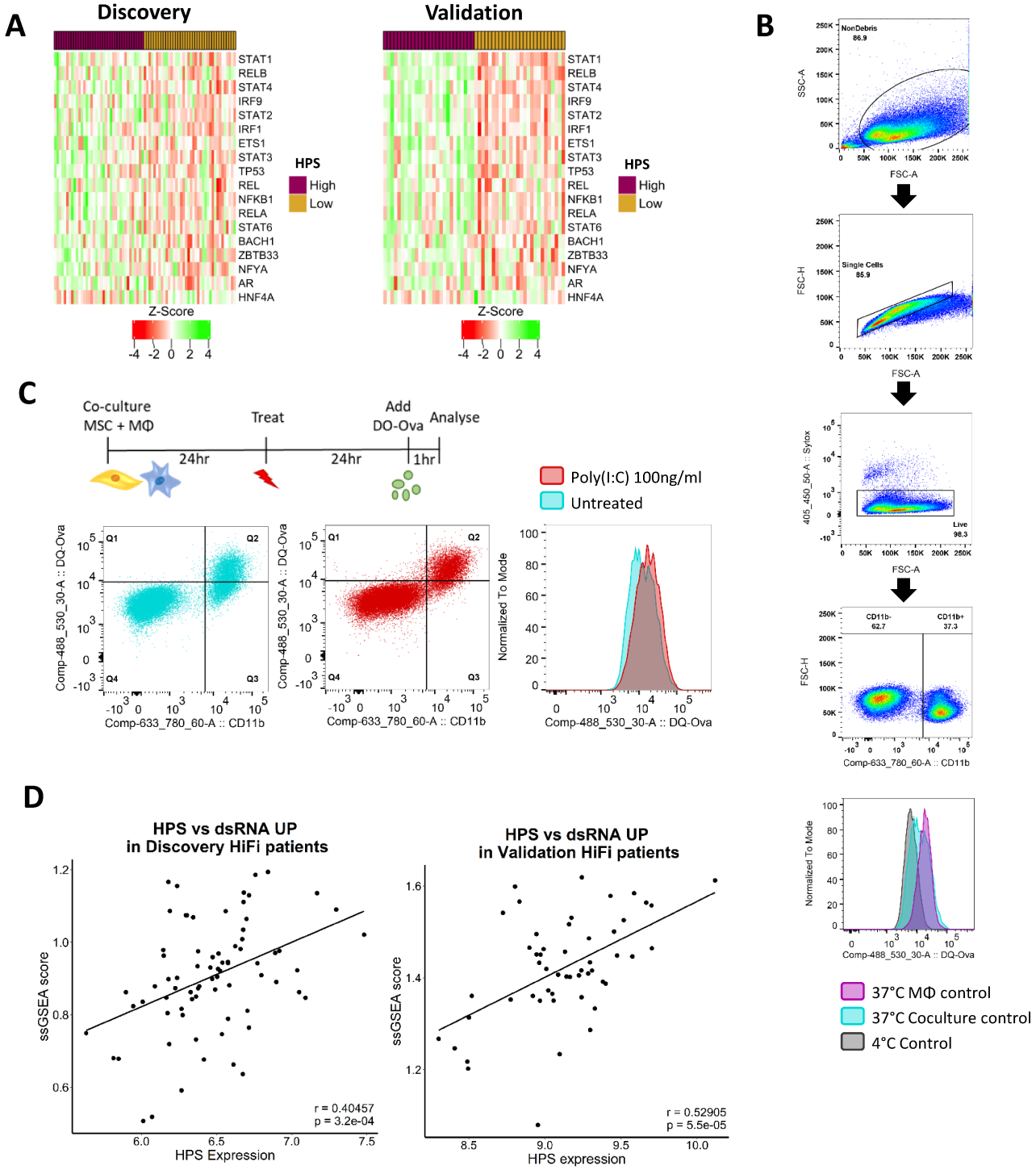
Cell population	Discovery	Validation
Neutrophils	0.74663	0.33569
Eosinophils	0.26928	0.03057
Dendritic cells	**0.00076**	0.51333
Mast cells	0.61458	0.27824
Macrophages	0.84487	0.85677
Monocytes	0.06597	0.22903
NK cells	0.73958	0.30186
Tregs	0.53897	0.35915
GD T cells	NA	0.18378
CD4 T cells	0.83181	0.58254
CD8 T cells	0.63046	0.90787
B cells	0.15901	0.62866





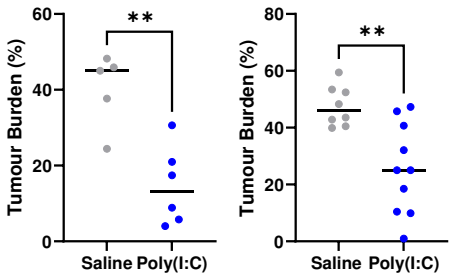


Supplementary Figure 6

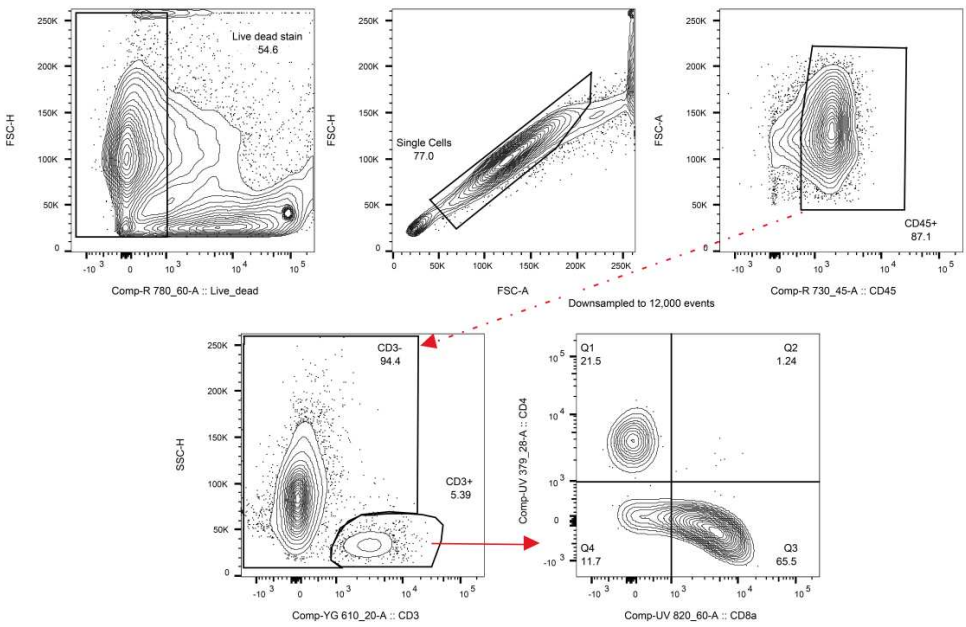


Supplementary Figure 7

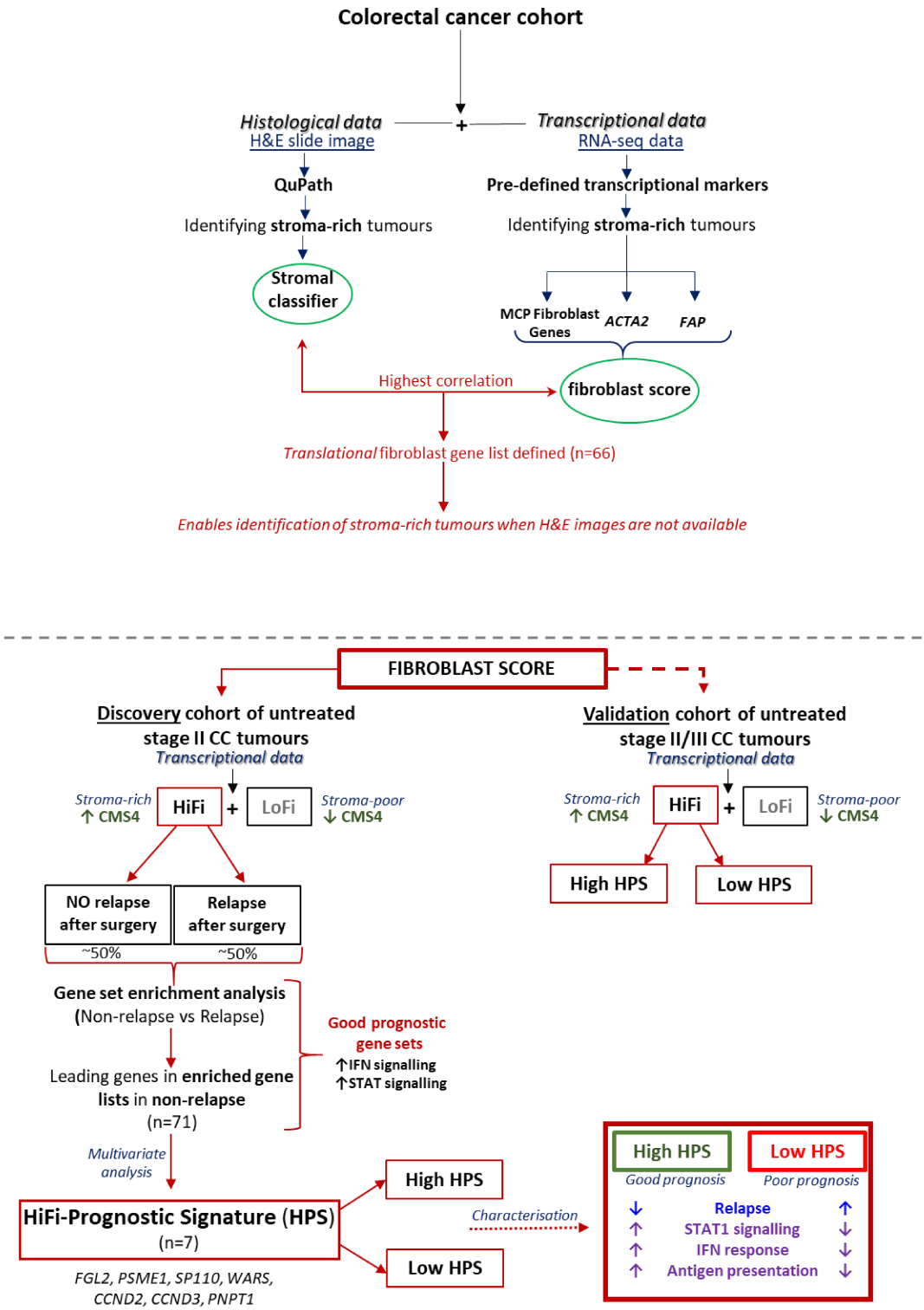
A



B



Supplementary Figure 8



Supplementary Table 1

Custom fibroblast gene set

ACTA2	COL12A1	FAP	IGFBP5	MFGE8	PLXNA3	THY1
ADAMTS2	COL1A1	FBLN1	ITGA11	MTSS1L	PODN	TMEM119
ANGPTL2	COL3A1	FBLN5	KRTAP1-5	MXRA8	PPP1R3C	TNFRSF11B
C1R	COL6A1	FGF7	LINC01279	MYL9	PRR16	VASN
C1S	COL6A2	FIBIN	LOC100287387	PAMR1	PRRX2	VGLL3
CCDC80	COPZ2	GLT8D2	LOC100507165	PARVA	RGMB	WISP1
CD248	CREB3L1	GREM1	LOX	PCDH18	SCARF2	
CEMIP	DCN	GREM2	LPAR1	PDGFRA	STC2	
CNN1	EFEMP2	HSPB6	LRRN4CL	PDGFRB	SVEP1	
CNTN3	ELN	HSPB7	MASP1	PLAC9	TAGLN	

Supplementary Table 2

Clinical information for discovery data

Clinical Characteristic		**Total cohort (n = 215) (%)**	**HiFi (n = 75) (%)**	**LoFi (n = 140) (%)**
Age	Median (range)	72 (45 - 95)	72 (45 - 95)	72 (45 - 95)
Sex	Male	106 (49.3)	41 (54.7)	65 (46.3)
	Female	109 (50.7)	34 (45.3)	75 (53.6)
Recurrence (within 5 years)	Yes	73 (34.0)	34 (45.3)	39 (27.9)
	No	142 (67.4)	41 (54.7)	101 (72.1)
pT stage	3	188 (87.4)	65 (86.7)	123 (87.9)
	4	27 (12.6)	10 (13.3)	17 (12.1)
Tumor location	Proximal	136 (63.3)	44 (58.7)	92 (65.7)
	Distal	79 (36.7)	31 (41.3)	48 (34.3)
Tumor differentiation grade	Well	9 (4.2)	5 (6.7)	4 (2.9)
	Moderate	174 (80.9)	61 (81.3)	113 (80.7)
	Poor	33 (15.3)	9 (12.0)	23 (16.4)
Tumor subtype	Mucinous	37 (17.2)	13 (17.3)	24 (17.1)
	Non-mucinous	177 (82.3)	61 (81.3)	116 (82.9)
	No information	1 (0.5)	1 (1.3)	0 (0.0)
Lymphovascular invasion	Yes	36 (16.7)	10 (13.3)	26 (18.6)
	No	127 (59.1)	43 (57.3)	84 (60.0)
	No information	52 (24.2)	22 (29.3)	30 (21.4)
Number of lymph nodes assessed	Median (range)	13 (6 - 40)	11 (6 - 36)	14 (6 - 40)

HiFi patients were defined as those with ssGSEA fibroblast scores greater than the cutoff for the cohort which was generated using the findcut function.

Supplementary Table 3

Clinical information for validation data

Clinical Characteristic		**Total cohort (n = 258) (%)**	**HiFi (n = 52) (%)**	**LoFi (n = 206) (%)**
Age	Median (range)	73 (24 - 94)	70 (40 - 93)	73 (24 - 94)
Sex	Male	147 (57.0)	24 (46.2)	123 (59.7)
	Female	111 (43.0)	28 (53.8)	83 (40.3)
Recurrence (within 5 years)	Yes	62 (24.0)	14 (23.1)	47 (22.8)
	No	196 (76.0)	38 (73.1)	159 (77.2)
TNM stage	2	203 (78.7)	39 (75.0)	164 (79.6)
	3	55 (21.3)	13 (25.0)	42 (20.4)
Tumor location	Proximal	118 (45.7)	28 (53.8)	90 (43.7)
	Distal	140 (54.3)	24 (46.2)	116 (56.3)
MMR status	pMMR	180 (69.8)	35 (67.3)	145 (70.4)
	dMMR	47 (18.2)	4 (7.7)	43 (20.9)
	No information	31 (12.0)	13 (25.0)	18 (8.7)
CIMP status	+	52 (20.2)	9 (17.3)	43 (20.9)
	-	178 (69.0)	36 (69.2)	142 (68.9)
	No information	28 (10.9)	7 (13.5)	21 (10.2)
CIN status	+	158 (61.2)	32 (61.5)	126 (61.2)
	-	65 (25.2)	11 (21.2)	54 (26.2)
	No information	35 (13.6)	9 (17.3)	26 (12.6)
tp53 mutation	Mutant	79 (30.6)	11 (21.2)	68 (33.0)
	Wild-type	80 (31.0)	14 (26.9)	66 (32.0)
	No information	99 (38.4)	27 (51.9)	72 (35.0)
KRAS mutation	Mutant	86 (33.3)	22 (42.3)	64 (31.1)
	Wild-type	163 (63.2)	30 (57.7)	133 (64.6)
	No information	9 (3.5)	0 (0.0)	9 (4.4)
BRAF mutation	Mutant	33 (12.8)	4 (7.7)	29 (14.1)
	Wild-type	206 (79.8)	43 (82.7)	163 (79.1)
	No information	19 (7.4)	5 (9.6)	14 (6.8)

HiFi patients were defined as those within the top 20% of the ssGSEA fibroblast scores for the cohort, with the other 80% considered LoFi.

Supplementary Table 4

Poly(I:C) Signature

INSL6	IFNA10	PNP	LGALS16	PARP14	C15orf48	RAB3D
TSPO	IFNA4	IFIT3	LGALS14	NDRG1	GBP6	OASL
IFNA16	IFNA8	MLKL	XDH	CTSE	IFIH1	AGRN
IFNA2	MX1	SCT	IFNA5	GNB4	IRGM	MS4A6A
IFNA6	GLIPR2	CA13	IFNA21	SLFN12	GBP5	MS4A6E
IFNA13	SAMHD1	ZBP1	CXCL10	SLFN12L	BST2	LGALS13
IFNA7	CD40	IRF7	IFIT2	STAT2	LGALS3BP	CISH
IFNA17	CDK5R1	IFNB1	GJB2	TOR3A	CD86	AXL
CDKN2B	IL15	IL27	DDX58	ISG20	GCA	APOD
IFNA1	USP41	UBE2L6	CCND2	NT5C3A	OAS2	
IFNA14	USP18	FSCN1	C2	MNDA	DAXX	

Supplementary Figure Legends

Supplementary Figure 1.

A Tested cutoff points of fibroblast scores using "findcut" in the discovery cohort and their corresponding hazard ratio (HR1) and p-values (P1), with the chosen cutoff shown in red. **B** Pie charts of the CMS classification in the discovery cohort (n=215 patients) in the HiFi patients (n=75) (left) and LoFi patients (n=140) (right).

Supplementary Figure 2.

A Assessment of previously published TGF-β signalling signatures are significantly associated with HiFi and LoFi groups in our discovery cohort. **B** Kaplan-Meier (KM) plots showing no significant difference between HiFi patients with high or low scores (median split) for the TGF-β signalling signatures**. C** Heatmaps showing previous cancer-associated fibroblast (CAF) classification. Li *et al.* CAF-A and CAF-B (top left), Öhlund *et al.* myCAF and iCAF (top right), Glentis *et al.* CAF with increased contractility and decreased contractility (bottom left) and Mizoguchi *et al.* $CD34^{-}THY1^{+}$, $CD34^{-}THY1^{-}$ and $CD34^{+}$ CAF (bottom right); none of the classifiers clustered the patients in the discovery cohort (n=215 patients) by relapse (dark green = relapse, light green = no relapse). **D** Kaplan-Meier (KM) plots showing no significant difference between HiFi patients with high or low scores (median split) for matrix index (p=0.44694), p53 signaling determined using the Hallmark gene set from MSigDB (p=0.4656), fibroblast scores determined using the custom fibroblast gene set (p=0.67653) or an expression signature of stem markers (p=0.13573). **E** Unsupervised clustering was performed in the HiFi patients from the discovery cohort. Calinski-Harabasz criterion was used to determine the optimal number of clusters (bottom left) which was 2. The heatmap (top) compares the unsupervised clusters (using k-means clustering; labelled with orange and navy) to the HiFi-specific prognostic signature (HPS) above and below median patients (labelled purple and yellow). KM of the HiFi patients (bottom right) split based on the two clusters identified using k-means clustering, showing no significant difference in RFS (log-rank p=0.69903). **F** Full GSEA results from the discovery cohort comparing HiFi patients who experienced relapse after surgery vs those who did not. There are 9 gene sets (UV RESPONSE DN, EPITHELIAL MESENCHYMAL TRANSITION, KRAS SIGNALING UP, SPERMATOGENESIS, MYC TAGRETS V1, TNFA SIGNALING VIA NFKB, HYPOXIA, PROTEIN SECRETION and ANDROGEN RESPONSE) significantly enriched (FDR<0.25) in the poor prognosis (relapse after surgery) patients and 10 gene sets significantly enriched in the good prognosis (no relapse after surgery) patients (INTERFERON ALPHA REPONSE,INTERFERON GAMMA REPONSE, IL6 JAK STAT3 SIGNALING, NOTCH SIGNALING, XENOBIOTIC METABOLISM, KRAS SIGNALING DN, MYOGENESIS, ALLOGRAFT REJECTION, HEDGEHOG SIGNALING and IL2 STAT5 SIGNALING). **G** Assessment of TGF-β signalling signatures reveals no association with our new HiFi-specific prognostic signature (HPS).

Supplementary Figure 3.

A. Receiver operating characteristic (ROC) curves for the HPS obtained when predicting relapse in HiFi patients from the discovery (n=75) (left) and validation cohorts (n=52) (right). The optimal cutoff which maximized the area under curve (AUC) for each cohort are shown, along with the maximum AUC value with 95% confidence intervals. **B** Comparison of the HPS cutoffs determined using either ROC or median HPS expression in the discovery (upper) and validation cohorts (lower), showing the sensitivity and specificity obtained using each cutoff.

Supplementary Figure 4.

A CMS classification according to fibroblast score. **B** Percentage of the stroma-rich, mesenchymal/stem subgroups in previously published colorectal cancer (CRC) subtyping studies. **C** Waterfall plot of fibroblast scores indicating CMS classification. High-fibroblast (HiFi) n=52 and low-fibroblast (LoFi) n=206 . **D** Pie charts of the CMS classifications in the validation cohort in the HiFi patients (n=52) (left) and LoFi patients (n=206) (right). **E** Comparison of HiFi and LoFi samples revealed that previously published stromal signatures and gene sets have significantly higher expression in the HiFi samples than the LoFi (adjusted p-value<0.15). **F** Comparison of the mismatch repair status, *CIMP, CIN*, *TP53*, *KRAS* and *BRAF* mutations, CMS and CRIS classifications and relapse status of HiFi patients from the validation cohort with low (below median) or high (above median) expression of the HPS. **G** The expression of the three *STAT1* target genes was significantly higher in the above median signature samples (n=37) compared to the below median samples (n=38) in the discovery cohort (upper; all p<0.01) and in the validation cohort (n=26 in each subgroup) (lower; all p<0.01). (Reference for all figures: ns = p>0.05, * = p≤0.05, ** = p≤0.01, *** = p≤0.001, **** = p≤0.0001).

Supplementary Figure 5.

A ImmuneScore derived from ESTIMATE in discovery (relapse=34, non-relapse=41) and validation cohorts (relapse=14, non-relapse=38) according to relapse status (left). ImmuneScore derived from ESTIMATE in discovery and validation cohorts according to HPS subgroups (right). **B** Relative CIBERSORT scores for immune cell populations in HiFi patients with above and below median HPS in the discovery (left) and validation cohort (right). The table shows each immune cell population and whether there was a significant difference in the scores between signature groups in each cohort. **C** The expression of the major histocompatibility class I genes *HLA-A* and *HLA-C* was significantly higher in the above median HPS samples (n=37) in the discovery cohort compared to the below median samples (n=38) (upper; both p< 0.05; note *HLA-B* was not on the microarray used for this cohort) and *HLA-A*, *HLA-B* and *HLA-C* were significantly higher in the above median HPS samples in the validation cohort (n=26) compared to the below median samples (n=26) (lower; all p<0.05). **D** HiFi patients with above median HPS expression had significantly higher GO APP ssGSEA scores compared to those with below median HPS expression in the discovery cohort (left; p<0.001) and in the validation cohort (right; p<0.001). **E** Correlation between ssGSEA scores for APP and HPS gene expression in purified mature antigen presenting cells. (Reference for all figures: ns = p>0.05, * = p≤0.05, ** = p≤0.01, *** = p≤0.001, **** = p≤0.0001).

Supplementary Figure 6.

A Heatmap showing the transcription factor activity scores according to HPS groups in the discovery (High=37, Low=38) and validation cohorts (n=26 for each subgroup). **B** Gating strategy and **C** experimental set up for flow cytometry analysis of macrophage mediated antigen uptake and processing of fluorescently labelled ovalbumin protein (DQ-Ova) with or without co-culture with tumor conditioned MSCs. **D** Correlation between ssGSEA scores for dsRNA and HPS gene expression in the discovery and validation cohorts (Pearson's correlation).

Supplementary Figure 7.

A Digital pathology assessment of H&Es from individual *in vivo* studies demonstrates reduced liver metastasis in mice treated with poly(I:C) compared to saline control. **B** Gating strategy; Cells were gated based on live cell status from live/dead stain, single cells and CD45+ cells selected. From here, data was downsampled to 12,000 events per sample and CD3+CD4+ and CD3+CD8+ cells were identified.

Supplementary Figure 8.

Schematic overview of initial fibroblast stratification, followed by identification and characterization of HiFi-Prognostic Signature (HPS).

Supplementary Methods

Datasets and data processing

The CEL files for the discovery cohort (E-MTAB-863) were imported into Partek Genomics Suite (PGS; v6.6) and RMA normalized then log2 transformed. As Gene Set Enrichment Analysis (GSEA) uses gene-level data, the probesets on the array were collapsed. This was done by importing the normalized data into R (v3.3.2 or later) and, using the 'collapseRows' function from WGCNA (Weighted Gene Coexpression Network Analysis, RRID:SCR_003302) package (v1.61) [1], selecting the probeset with the highest mean expression per gene [2]. For the S:CORT cohorts, hematoxylin and eosin (H&E) stained slide images were also obtained. Transcriptional data for the FOCUS and Grampian cohorts was generated using the Almac Diagnostics proprietary Xcel™ GeneChip array (Affymetrix, Santa Clara, CA).

GSE41295; primary human monocytes differentiated into macrophages and treated with poly(I:C), the raw CEL files were downloaded from GEO (due to issues matching the probe IDs in the series matrix to the array annotation file) then imported into PGS and RMA normalized and log2 transformed. GSE15066: mouse macrophage cell line RAW264.7 stimulated with poly(I:C) [3]. The RAW data was downloaded from GEO, as the series matrix contained null values, and imported into R and background corrected and normalized using the limma package (LIMMA, RRID:SCR_010943; v3.30.13). GSE1925 series matrix required a log2 transformation prior to its collapse to gene-level. In the single cell cohort, dendritic cells, monocytes and macrophages were extracted and further manually curated based on graph clustering with the igraph package (v1.2.5). The dendritic population was subclassified based on the expression of highly specific marker genes (pDC: *LILRA4, IRF7, CLEC4C*; cDC1: *CLEC9A, XCR1*; cDC2: *CD1C, CLEC10A*; migDC: *CCR7, CCL17).*

Digital pathology stroma classifier

The FOCUS cohort (n = 361; FFPE resections from a randomized clinical trial)[4] and the Grampian cohort (n = 225; FFPE pre-treatment rectal cancer biopsies), were obtained from the S:CORT data portal (https://www.s-cort.org/). For FOCUS and Grampian cohorts, H&E stained whole slide images with existing pathologist annotations were scanned at 40x magnification and imported into QuPath (v2.0 milestone 9)[5]. Batch analysis was applied across all slides to detect cells (Watershed cell detection with default setup, nucleus, cell and general parameters; Intensity threshold: 0.1; Maximum background intensity: 2) within the tissue annotation. Gaussian smoothed (Radius: 25µm) was applied. Annotated regions of tumour epithelium and stroma, on a selection of cases were used to build the Random Trees object classifier (Object filter: detections (all); Features: all measurements; Classes: all classes).

Single-sample Gene Set Enrichment Analysis (ssGSEA) fibroblast classifier

The 64 fibroblast genes from MCPcounter, smooth-muscle actin (*ACTA2)* and fibroblast activation protein (*FAP)* (Supplementary table 1)[6] were used with ssGSEA to generate a fibroblast score for each patient. Discovery cohort patients were split into high-fibroblast (HiFi) or low-fibroblast (LoFi) using the 'findcut' function[7]. The 'findcut' function was used to determine the optimal cutpoint for the ssGSEA fibroblast score by identifying the cutpoint value resulting in the most significant association between LoFi and HiFi groups and relapse-free survival. For all other human transcriptional datasets, samples with the highest 20% of fibroblast scores were classified as HiFi. In GEMMs, the median split was used for the waterfall plot and top/bottom tertile was used as a surrogate for HiFi/LoFi.

HiFi cut off

We would expect that the stroma-rich group accounts from ~20% in the CRC population, which has been quite a stable and robust proportion identified in multiple histological and molecular studies. However, to ensure additional power for the Discovery work in our current study, we have utilised the cohort assembled by Almac diagnostics (published study by Kennedy et al., https://pubmed.ncbi.nlm.nih.gov/22067406/). This Discovery cohort we use in this current study was previously used for the development of the now FDA-approved ColDx/GeneFx stage II prognostic assay (https://www.almacgroup.com/diagnostics/portfolio-overview/coldx/). When this cohort was assembled, it was compared with the Surveillance, Epidemiology, and End Results (SEER) database to ensure it represented a general population with stage II colon cancer. However, although it represents the general clinical population demographics, to ensure sufficient relapse "events" for prognostic biomarker development, the tumours included were pre-selected to ensure a 2:1 ratio of "low-risk" to "high-risk". Low-risk patients were defined as those with no cancer recurrence within 5 years of primary surgery. High-risk patients were defined as those with metastatic cancer recurrence within 5 years of primary surgery. Therefore, after balancing for clinical factors and applying quality control criteria, the Discovery cohort contains 215 patients (142 low-risk and 73 high-risk patients), with a relapse rate of 34% in this cohort. By having a higher relapse rate, given the relationship between stromal content and relapse in stage II, it also contains an increased proportion of stromal tumours when compared to more general stage II or stage II/III cohorts. Discovery cohort patients were split into high-fibroblast (HiFi) or low-fibroblast (LoFi) using the 'findcut' function[7].

For these reasons, we did not use the widely-used 20% cut-off, and our optimal fibroblast score (Supplementary Figure 1A) was defined as 35% in this cohort. This enrichment is precisely why it is valuable as a Discovery cohort in our study and for the generation of prognostic classifiers and subsequent characterisation of prognostic biology. For Validation, we have utilised the well-characterised "Marisa" cohort used in the development of the CMS classification system by the CRCSC consortium (https://www.synapse.org/#!Synapse:syn2623706/wiki/67246), a cohort that isn't pre-selected in the same way as our Discovery cohort. In this case, we have identified that in line with stroma-rich proportions across all CRCSC cohorts (Supplementary Figure 4B), the consensus proportion of this stroma-rich histological group is 20%.

Gene Set Enrichment Analysis (GSEA) and Leading Edge Analysis (LEA)

GSEA using Hallmark gene sets from the Molecular Signature Database (MSigDB) with default settings (except gene_set permutations, random seed = 367707) was performed. Gene sets with an FDR < 0.25 were considered significant. LEA was used to identify genes that were shared by the leading-edge subsets of two or more significant gene sets. Following LEA (and a multivariate model) identifying the HPS, fgsea (V 1.16.0) (permutation = 1000, set.seed =121) using Hallmarks and immune-related gene lists were assessed (detailed in GitHub scripts associated with this study). Gene lists that required conversion from human to mouse or vice versa, was converted using biomaRt (V 2.46.3).

Unsupervised clustering of HiFi tumors

Transcriptional data from the HiFi patients from the discovery cohort were filtered to keep the top 50% variant genes that also had above mean expression (n = 4833). The data were then scaled using the 'scale' function from the base package. To determine the optimal number of clusters, the 'cascadeKM' function from the vegan package (v2.5-1) was used to generate a Calinski criterion for 1-

10 clusters. The optimal number of clusters was determined to be two and the 'kmeans' function from the stats package was used to split the discovery cohort into two clusters.

CMS and CRIS classifications

To classify patients into the CMS groups, the R package CMScaller (v0.99.1) was used with default settings and seed set to 367707 [8]. CRIS classifications were performed with the Nearest Template Prediction module on GenePattern (https://cloud.genepattern.org/gp; RRID:SCR_003201) with a template containing the CRIS genes from the original study [9].

Estimating relative abundance of immune and stromal cell population

Four different methods were employed to estimate the relative abundance of immune and stromal cell populations in transcriptional data from bulk tumor samples. (i) MCPcounter R package (v1.1.0), which assigns scores for eight immune and two stromal cell populations; (ii) mean expression of transcriptional signatures for cancer-associated fibroblasts, endothelial cells and leukocytes [10]. (iii) ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) method implemented in the R package estimate which assigns scores of general immune and stromal infiltrate [11]. (iv) the online CIBERSORT tool (https://cibersort.stanford.edu/; RRID:SCR_016955), assigning scores for 22 immune cell populations [12]. Before analysis of the CIBERSORT scores, certain populations were combined as described previously by Zemek *et al.* [13].

Transcription Factor Activity Analysis

The dorothea (V 1.2.2) and viper (V 1.24.0) packages were used to investigate transcription factor activity, with human (dorothea_hs) or mouse (dorothea_mm) versions applied depending on the dataset species.

Statistics

Pearon's Correlation: Figure 1C: Correlation matrix with histological stroma and transcriptional classifiers using Pearson's correlation. **Figure 3C:** Cumulative gene expression of *STAT1* and three of its target genes (*PSMB9, IRF1* and *TAP1*) correlated with expression of the HPS in the discovery (left) and validation cohort (right; both p<0.00001) using Pearson's correlation. **Figure 4B:** Correlation between ssGSEA scores for APP and HPS gene expression in the discovery cohort (r=0.5, p=1.4e-06) using Pearson's correlation. **Figure 4C:** Correlation between ssGSEA scores for APP and HPS gene expression in the validation cohort (r=0.6, p=1.5e-05) using Pearson's correlation. **Figure 5C:** Correlation between ssGSEA scores for viral response and HPS gene expression in the discovery cohort (r=0.6, p=1.1e-08) using Pearson's correlation **Figure 5D:** Correlation between ssGSEA scores for viral response and HPS gene expression in the validation cohort (r=0.7, p=2.4e-08) using Pearson's correlation.

T-test:

Figure 1D: CMS classification according to our fibroblast score in discovery cohort. (UNK = unknown/mixed CMS classification) using a t-test. **Figure 3D:** Boxplots of *STAT1* gene expression (left)

and protein levels (right) in HiFi patients in the CPTAC cohort according to HPS groups (High n=9 and Low n=9) using t.test**. Figure 4E, 4F:** Immune cell populations have significantly higher expression of the HPS **(4F)** and GO APP ssGSEA scores than epithelial cells and fibroblasts (both p<2.2e-16) using t-test. **Figure 4H:** APP ssGSEA scores in bone-marrow derived macrophages with either wild-type (WT), mutant (Y701F mut) or knockout (KO) *STAT1*. (n=3 for each genotype) using t.test **Figure 5A:** Activity status of key TF regulons according to HPS groups in the validation cohort (n=26 in each subgroup), p value (<0.05) determined using row t-test. **Figure 5F:** Viral response ssGSEA scores in bone-marrow derived macrophages with either wild-type (WT), mutant (Y701F mut) or knockout (KO) *STAT1*. (n=3 for each genotype) using t-test**. Figure 7A:** CMS classification according to fibroblast score of GEMM genotypes. (A=*Apc*$^{fl/fl}$, K=*Kras*$^{G12D/+}$, P=*p53*$^{fl/fl}$ and N=*Notch1Tg/*$^{+}$) using t-test.

Survival:

Figure 1F: Kaplan-Meier (KM) curve showing HiFi patients have significantly poorer relapse-free survival (RFS) than LoFi patients (log-rank p=0.00779) p-value extracted from survival package (V 3.2-13). **Figure 2E:** KM curve showed HPS has strong prognostic value in HiFi tumours based on a median split in Discovery cohort (log-rank p=0.0069) (top). KM curve showed no prognostic value in the LoFi samples in Discovery cohort (log-rank p=0.63215) (bottom). P-value was extracted from survival package (V 3.2-13). **Figure 2G:** HPS can stratify HiFi samples into two groups in the validation cohort, one with significantly poorer RFS and another with RFS even better than the LoFi patients (log-rank p=0.00113) (top). HPS has no prognostic value in the LoFi samples (log-rank p=0.46596) (bottom). P-value was extracted from survival package (V 3.2-13).

Other:

Figure 1G: Comparison of HiFi and LoFi samples revealed that previously published stromal signatures and gene sets have significantly higher expression in the HiFi samples than the LoFi (adjusted p value provided by GSEA <0.15). **Figure 6G:** Differentially expressed genes (logFC>2 and adjusted p-value <0.001) in Poly(I:C) treated vs non-treated dendritic cells creating the 'Poly(I:C) Signature'. **Figure 7G:** Digital pathology assessment of H&Es from in vivo studies demonstrates reduced liver metastasis in mice treated with poly(I:C) (n=16) compared to saline control (n=13) (Mann Whitney U test). **Figure 7H:** Flow cytometry assessment of CD3+ cell populations from liver metastases in treatment groups highlight significant elevation of CD8+ T cells alongside significant reduction in CD4+ T cells in poly(I:C) arm (n=6) compared to saline (n=5) (Mann Whitney U test).

Correlation analysis

All correlations were performed in R using the 'cor.test' function from the stats package, with the "pearson" method.

Single-sample Gene Set Enrichment Analysis (ssGSEA)

ssGSEA was also used with gene sets from the Molecular Signature Database (MSigDB; v6.0 or later) to assign TGF-β scores, using the "HALLMARK_TGF_BETA_SIGNALING" gene set from the Hallmark collection (56); p53 scores using the "HALLMARK_P53_PATHWAY" gene set from the Hallmark collection; and antigen processing and presentation scores, using the "GO_ANTIGEN_PROCESSING_AND_PRESENTATION" gene set from the C5 collection.

Analysis of previously identified prognostic factors in HiFi tumors

Signatures from four studies that have attempted to identify subgroups of fibroblasts were tested in the HiFi tumors from the discovery cohort. (A) Li *et al.* [14] who used a single-cell RNA sequencing approach and profiled cells from 11 primary colorectal tumors and normal tissue, and found within the fibroblasts from the tumors samples that two groups could be identified (CAF-A (with ECM remodeling gene expression) and CAF-B (with activated myofibroblasts signaling)). (B) Öhlund *et al.* [15] who used co-culture models of pancreatic ductal adenocarcinoma and found two reversible subtypes of CAFs existed, one with higher expression of αSMA and lower expression of IL-6 (myCAFs) and the other with lower levels of αSMA and higher expression of IL-6 (iCAFs). (C) Glenis *et al.* [16], who utilized proteomics from primary human normal fibroblasts and CAFs, and identified that CAFs had increased expression of highly contractile proteins and reduced expression of proteins associated with reduced contractility compared to normal fibroblasts. (D) Mizoguchi *et al.* [17] who looked at heterogeneity in fibroblasts using transcriptional data derived from fibroblasts from joints of patients with rheumatoid arthritis or osteoarthritis. They identified three types of fibroblasts, including a group of PDPN+CD34–THY1+ cells that were involved in immune cell recruitment and matrix invasion. Several other prognostic factors were tested in the HiFi tumors from the discovery cohort including matrix index, a measure of matrix stiffness utilizing 22 genes described by Pearce *et al.* [18]; p53 signaling, determined using ssSGEA scores as described above; our own fibroblast score, calculated using ssGSEA as described above; and the expression of *CD44*, a known marker of stemness in cancer.

ROC analysis

ROC curves were generated using the 'roc' function from the pROC R package (v1.10.0) [19]. Sensitivity and specificity values for the median cutoff were found using the 'coords' function.

Heatmaps of signature and STAT1 gene expression

Heatmaps were generated in R by first scaling the data using the 'scale' function from the base package to generate Z-scores, then clustering (using Ward's linkage and Euclidean distance) and plotting them using the 'Heatmap' function from the ComplexHeatmap package (RRID:SCR_017270; v2.0.0 or later) [20].

Differential gene expression analysis

The function 'rowttests' from the package genefilter R package (v1.56.0) was used to calculate t-tests per gene between above median and below median signature expression tumors in the discovery and validation cohorts. The p-values from these tests were adjusted using the Benjamini-Hochberg method utilized in the function 'mt.rawp2adjp' from the multtest package (v2.30.0) [21]. The final gene list consisted of genes that were significantly differentially expressed in the same direction in both cohorts.

Ingenuity Pathway Analysis (IPA)

The genes identified as differentially expressed between the HPS groups in the discovery and validation cohorts were entered into IPA (QIAGEN Inc,; RRID:SCR_008653 https://digitalinsights.qiagen.com/qiagen-ipa), a web-based application that enables the analysis of

gene expression patterns using a literature-based database, to identify upstream regulators of the differentially expressed genes and analyzed using the default settings for a core analysis [22].

Single-cell RNA-Seq data

10X sequencing of 15 distinct patients, including samples from cancer (core and tumour border) and normal tissue was performed. Cells were filtered to contain at least 300 unique genes and less than 6000 unique genes, and a mitochondrial content of less than 20%. Batch normalization with scran (v1.14) and initial cell classification with SingleR (v1.0.5) against the Human Primary Cell Atlas[23], which is a human immune cell atlas[24] and a previously published PBMC dataset[25] (https://support.10xgenomics.com/single-cell-gene-expression/datasets), were performed. After manual curation, the two main batches were integrated with MNN dimension [26] reduction for plotting.

Cell culture

CT26 cells (murine colon adenocarcinoma cells derived from Balb/c mice. RRID:CVCL_7256) were cultured in Dulbecco's Modified Eagles Medium (DMEM) High glucose with L-glutamine (Hyclone; Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% fetal bovine serum (FBS; Sigma) and 1% penicillin/streptomycin (p/s; Gibco; Thermo Fisher Scientific, Waltham, MA, USA). RAW264.7 cells (murine macrophages derived from Balb/c mice) were cultured in DMEM high glucose, GlutaMAX™ with sodium pyruvate (Gibco) supplemented with 10% FBS and 1% p/s. Primary mesenchymal stromal cells (MSCs) were isolated from the bone marrow of Balb/c mice as previously described [27]. MSCs were cultured in Minimum Essential Medium (MEM) alpha (Gibco) supplemented with 10% FBS and 1% p/s. All cells were incubated at 37°C in 5% CO_2 and experiments were performed with MSCs between passages 8-11. Cell lines were purchased from the ATCC, confirmed mycoplasma negative (MycoAlert; Lonza; Basel, Switzerland), expanded, frozen and used within 15 passages of testing for all subsequent experiments.

MSC conditioning, co-culture and poly(I:C) treatment

Tumor-conditioned medium (TCM) was generated by seeding 1 x 10^6 CT26 cells (RRID:CVCL_7256) in a T175 flask in 25ml of medium. After 72 hours of culture, the medium was collected and spun at 1,000 RCF to remove any debris.

MSCs were seeded at a density of 9.0 x 10^4 cells per well in a 6-well plate in 2ml of culture medium. After 8 hours, the medium was removed and replaced with 40% fresh MSC medium and 60% TCM. After 48 hours, the conditioned MSCs were co-cultured with RAW264.7 cells (RRID:CVCL_0493) by seeding both cells at a density of 2.0 x 10^4 cells per well in a flat-bottom 96-well plate. A monoculture of RAW264.7 cells was also plated at a density of 2.0 x 10^4 cells per well. After 24 hours, the cultures were treated overnight with either 100ng/ml poly(I:C) (InvivoGen; San Diego, CA, USA) or RAW264.7 culture medium to act as an untreated control.

DQ-Ova flow cytometry

To assess antigen processing, we used a BODIPY FL labelled DQ™ Ovalbumin (DQ-Ova; Thermo Fischer Scientific). RAW264.7 cells (RRID:CVCL_0493), either alone or in co-culture with conditioned MSCs,

were incubated with DQ-Ova at a final concentration of 1µg/ml for one hour, washed with PBS twice, trypsinized and transferred to a v-bottom 96-well plate before being spun at 600 RCF and resuspended in FACS buffer (PBS supplemented with 2% FBS and 0.05% sodium azide). Staining was performed by incubating the cells with the CD11b antibody (Cat# 101226, RRID:AB_830642; APC/Cy7, Clone M1/70; BioLegend; San Diego, CA, USA) in FACS buffer at 4°C for 15 minutes. After two washes in FACS buffer, the cells were resuspended in FACS buffer containing SYTOX™ Blue (Thermo Fischer Scientific) for viability assessment. Samples were analyzed on a FACSCanto™II (BD Biosciences; Becton, Dickinson and Company, Franklin Lakes, NJ, USA) and analyzed using FlowJo (FlowJo, RRID:SCR_008520). Gating strategy in Supplementary Figure 6 (v10.6.1; BD Biosciences).

GEMM dataset descriptor and histology

Both genders of AKPN, APN, AP, A, KPN, KP mice were induced with a single intraperitoneal injection of 2mg tamoxifen between 7 and 14 weeks of age (n of at least 3 per genotype). Mice were sampled at clinical endpoint according to defined criteria including weight loss and/or hunching and/or anemia and/or loss of body condition. Intestinal tumours were sampled for histology and RNAseq analysis.

In vivo tissue processing by flow cytometry

The tissues for flow cytometry analysis of in vivo samples were processed the following way: The blood samples were mixed with an erythrocyte lysis buffer (For 1L: 8.29g NH4Cl, Sigma-Aldrich, Cat#A9434; 1.0g KHCO3, Sigma-Aldrich, Cat#60339; 37.2mg Na2EDTA, Sigma-Aldrich, Cat#E5134; 1L dH2O, adjusting pH to 7.2-7.4). The blood cells were incubated for 10 minutes at 4°C. The liver metastasis samples were dissected and digested using the Mouse Tumour Dissociation Kit (Miltenyi Biotec, 130-096-730) and the GentleMACS Octo Dissociator (Miltenyi Biotec, 130-095-937), following the manufacturer's instructions for these tumours.

Cells from both tissue samples were passed through a 70µm cell strainer, washed with RPMI 1640, 10% fetal bovine serum, 2mM EDTA, and spun down at 400g, 4°C for 5 minutes. The cells were washed with PBS and spun down at 400g, 4°C for 5 minutes. After discarding the supernatant, the cells were stained with 50 µL of the Live/Dead NIR staining kit, at 1:1000 dilution in PBS, for 20min at 4°C in the dark. After that, the cells were washed with PBS 1% BSA and spun down at 400g, 4°C for 5 minutes. After discarding the supernatant, the cells were resuspended in 25µL of FcBlock (BD Pharmingen, 553141) 1:200 in PBS 1% BSA, and incubated for 10 minutes on ice in the dark. Then, 25µL of the antibody staining mix at 2x concentration were added (see table below for antibody details).

Marker	Fluorochrome	Company	Catalog no.
MHCI	PerCP/Cy5.5	Biolegend	114620
CD86	FITC	Biolegend	105110
CSF1R	BV786	BD Biosciences	750888
CD103	BV711	Biolegend	121435
CD11b	BV650	Biolegend	101259
CD11c	BV605	Biolegend	117334

F4/80	BV510	Biolegend	123135
MHCII	BV421	Biolegend	107632
CD8a	BUV805	BD Biosciences	612898
CD69	BUV737	BD Biosciences	612793
CD4	BUV395	BD Biosciences	563790
CD45	AF700	Biolegend	103128
Ly6G	AF647	Biolegend	127610
CD64	PE/Cy7	Biolegend	139314
CD3	PE/Dazzle594	Biolegend	100246
CD26	PE	Biolegend	137804

The cells were incubated for 30 minutes on ice in the dark. The cells were washed with PBS and re-suspended in 50µL PBS. The cells were fixed adding 50µL 4% PFA in PBS and incubated for 15 minutes at room temperature in the dark. The cells were washed with PBS two times spinning down at 400g for 5 minutes. They were resuspended in 400µL PBS for data acquisition in the flow cytometer. Countbright beads (ThermoFisher Scientific, C36950) were added for absolute cell quantifications at this step.

For initial QC, immune cells populations were identified in a BD LSRFortessa (BD Biosciences) using the FACSDiva software, with the following gating strategy: cells in a FSC-A and SSC-A plot; doublet discrimination by discrepancy between FSC-A and FSC-H signals; Live/Dead- live cells; CD45+ for all immune cells. The T-cells were identified as CD3+, then looked for CD69, CD4 and CD8 markers. Neutrophils where identified as CD3-CD11b+Ly6G+. Eosinophils as CD3-SSChiLy6G-. Monocytes (in blood only) as CD11b+Ly6G-CSF1R+. Macrophages (in liver metastatic tissue only) as CD11b+Ly6G-F4/80+CD64+. Dendritic cells were identified as CD3-Ly6G-MHCII+CD11c+F4/80-CD64-, then looked for MHCI, CD26, CD86, CD11b and CD103. The flow cytometry data were analysed using FlowJo v10.7.2.

Patient and Public Involvement: Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Data availability: RRIDs (research resource identifiers) are used to identify resource within the Methods section. Any data used in this study that are not already publically-available will be shared, please contact corresponding author (p.dunne@qub.ac.uk). In addition, code to reproduce all results will be deposited on our lab website (www.dunne-lab.com).

Supplementary Method References:

1. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9(1):559. doi: 10.1186/1471-2105-9-559
2. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012;481(7382):506-10. doi: 10.1038/nature10738; 10.1038/nature10738
3. Yang IV, Jiang W, Rutledge HR, et al. Identification of novel innate immune genes by transcriptional profiling of macrophages stimulated with TLR ligands. *Molecular Immunology* 2011;48(15):1886-95. doi: https://doi.org/10.1016/j.molimm.2011.05.015
4. Seymour MT, Maughan TS, Ledermann JA, et al. Different strategies of sequential and combination chemotherapy for patients with poor prognosis advanced colorectal cancer (MRC FOCUS): a randomised controlled trial. *The Lancet* 2007;370(9582):143-52. doi: 10.1016/S0140-6736(07)61087-3
5. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. *Scientific Reports* 2017;7(1):16878. doi: 10.1038/s41598-017-17204-5
6. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome biology* 2016;17(1):218. doi: 10.1186/s13059-016-1070-5
7. Chang C, Hsieh M-K, Chang W-Y, et al. Determining the optimal number and location of cutoff points with application to data of cervical cancer. *PLOS ONE* 2017;12(4):e0176231.
8. Eide PW, Bruun J, Lothe RA, et al. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Scientific Reports* 2017;7(1):16618. doi: 10.1038/s41598-017-16747-x
9. Alderdice M, Richman SD, Gollins S, et al. Prospective patient stratification into robust cancer-cell intrinsic subtypes from colorectal cancer biopsies. *The Journal of pathology* 2018;245(1):19-28. doi: 10.1002/path.5051 [doi]
10. Isella C, Terrasi A, Bellomo SE, et al. Stromal contribution to the colorectal cancer transcriptome. *Nature genetics* 2015;47(4):312–19. doi: 10.1038/ng.3224
11. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications* 2013;4:2612. doi: 10.1038/ncomms3612 [doi]
12. Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nature medicine* 2015;21(8):938-45. doi: 10.1038/nm.3909 [doi]
13. Zemek RM, De Jong E, Chin WL, et al. Sensitization to immune checkpoint blockade through activation of a STAT1/NK axis in the tumor microenvironment. *Science Translational Medicine* 2019;11(501):eaav7816. doi: 10.1126/scitranslmed.aav7816
14. QuPath: Open source software for digital pathology image analysis. *bioRxiv* 2017 doi: 10.1101/099796 "A1 Bankhead, Peter" "A1 Loughrey, Maurice B" "A1 Fernández, José A" "A1 Dombrowski, Yvonne" "A1 McArt, Darragh G" "A1 Dunne, Philip D" "A1 McQuaid, Stephen" "A1 Gray, Ronan T" "A1 Murray, Liam J" "A1 Coleman, Helen G" "A1 James(TRUNCATED)
15. Öhlund D, Handly-Santana A, Biffi G, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *The Journal of experimental medicine* 2017 doi: 10.1084/jem.20162024
16. Glentis A, Oertle P, Mariani P, et al. Cancer-associated fibroblasts induce metalloprotease-independent cancer cell invasion of the basement membrane. *Nature Communications* 2017;8(1):924. doi: 10.1038/s41467-017-00985-8
17. Mizoguchi F, Slowikowski K, Wei K, et al. Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nature Communications* 2018;9(1):789. doi: 10.1038/s41467-018-02892-y

18. Pearce OMT, Delaine-Smith RM, Maniati E, et al. Deconstruction of a Metastatic Tumor Microenvironment Reveals a Common Matrix Response in Human Cancers. *Cancer Discov* 2018;8(3):304-19. doi: 10.1158/2159-8290.CD-17-0284 [published Online First: 2017/12/01]
19. Jansen FH, van Rijswijk A, Teubel W, et al. Profiling of Antibody Production against Xenograft-released Proteins by Protein Microarrays Discovers Prostate Cancer Markers. *Journal of proteome research* 2012;11(2):728-35. doi: 10.1021/pr2006473
20. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32(18):2847-49. doi: 10.1093/bioinformatics/btw313
21. Pollard KS, Dudoit S, van der Laan MJ. Multiple Testing Procedures: the multtest Package and Applications to Genomics. In: Gentleman R, Carey VJ, Huber W, et al., eds. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. New York, NY: Springer New York 2005:249-71.
22. Krämer A, Green J, Pollard J, Jr, et al. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2013;30(4):523-30. doi: 10.1093/bioinformatics/btt703
23. Mabbott NA, Baillie JK, Brown H, et al. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics* 2013;14(1):632. doi: 10.1186/1471-2164-14-632
24. Monaco G, Lee B, Xu W, et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Reports* 2019;26(6):1627-40.e7. doi: 10.1016/j.celrep.2019.01.041
25. Zheng GXY, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nature Communications* 2017;8(1):14049. doi: 10.1038/ncomms14049
26. Haghverdi L, Lun ATL, Morgan MD, et al. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology* 2018;36(5):421-27. doi: 10.1038/nbt.4091
27. O'Malley G, Treacy O, Lynch K, et al. Stromal cell PD-L1 inhibits CD8+ T-cell antitumor immune responses and promotes colon cancer. *Cancer Immunology Research* 2018:canimm.0443.2017. doi: 10.1158/2326-6066.Cir-17-0443

Graphical summary