Natural killer-like signature observed post therapy in locally advanced rectal cancer is a determinant of pathological response and improved survival

Short title:

Natural killer cell response improves survival

Authors:

Matthew Alderdice[1], Philip D. Dunne[2], Aidan J Cole[2,3], Paul G. O'Reilly[2], Darragh G. McArt[1], Vicky Bingham[1], Marc-Aurel Fuchs[1], Stephen McQuaid[1], Maurice B. Loughrey[2,4], Graeme I Murray[5], Leslie M. Samuel [6], Mark Lawler [2], Richard H. Wilson[2,3], Manuel Salto-Tellez[1,2,4][*], Vicky M. Coyle[2,3][*]

Affiliations:

1. Northern Ireland Molecular Pathology Laboratory, Centre for Cancer Research and Cell Biology, Queens University Belfast, 97 Lisburn Road, BT9 7BL, Northern Ireland.

2. Centre for Cancer Research and Cell Biology, Queens University Belfast, 97 Lisburn Road, BT9 7BL, Northern Ireland.

3. Cancer Centre, Belfast Health and Social Care Trust, Lisburn Road, Belfast BT9 7BL, Northern Ireland.

4. Department of Tissue Pathology, Royal Victoria Hospital, Belfast Health and Social Care Trust, Grosvenor Road, Belfast, BT12 6BA.

5. Pathology, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Cornhill Road, Foresterhill, Aberdeen, AB25, 2ZD

6. Department of Clinical Oncology, Aberdeen Royal Infirmary, NHS Grampian, Aberdeen UK.

*These authors contributed equally to this manuscript

Correspondence:

Professor Manuel Salto-Tellez

97 Lisburn Road, BT9 7BL, Northern Ireland.

Phone: +44 (0) 28 9097 2760

m.salto-tellez@qub.ac.uk

Word count: 4408 (Introduction-Discussion)

Abbreviations:

 CT - Cancer testis, CTA - Cancer testis antigens, CD56 - Cluster of differentiation 56, CD3 - Cluster of differentiation 3, CD4 - Cluster of differentiation 4, CD8 - Cluster of differentiation 8, DASL - cDNA mediated, annealing, selection and ligation), DNA - Deoxyribonucleic acid, FDA - Food and Drug Agency, FOLFOX - Folinic Acid, 5 Fluorouracil and Oxaliplatin, Gy – Grays, H&E - Hematoxylin and eosin, hg19 - Human genome 19, ILC3 - Innate lympoid cell group 3, iNKT- invariant Natural killer T, LINCs - Library of integrated cellular signatures, MAIT - Mucosal associated Invariant T cell, MSI - Microsatellite instability, PCR- Polymerase chain reaction, RNA - Ribonucleic Acid, TILS - Tumour Infiltrating lymphocytes,

Abstract

Around 12-15% of patients with locally advanced rectal cancer undergo a pathologically complete response (Tumour Regression Grade 4) to long course preoperative chemoradiotherapy; the remainder exhibit a spectrum of tumour regression (tumour regression grade 1-3). Understanding therapy-related transcriptional alterations may enable better prediction of response as measured by progression-free and overall survival, in addition to aiding the development of improved strategies based on the underlying biology of the disease. To this end we performed high-throughput gene expression profiling in 40 pairs of formalin fixed paraffin embedded rectal cancer biopsies and matched resections following long course preoperative chemoradiotherapy (discovery cohort). Differential gene expression analysis was performed contrasting tumour regression grades in resections. Enumeration of the tumour microenvironment cell population was undertaken using *in silico* analysis of the transcriptional data, and real-time PCR validation of *NCR1* undertaken. Immunohistochemistry and survival analysis was employed to measure CD56+ cell populations in an independent cohort (n=150). Gene expression traits observed following long course preoperative chemoradiotherapy in the discovery cohort suggested an increased abundance of natural killer cells in tumours that displayed a clinical response to CRT in a tumour regression grade dependent manner. CD56+ natural killer cell populations were measured by immunohistochemistry and found to be significantly higher in tumour regression grade 3 patients compared to tumour regression grade 1-2 in the validation cohort. Furthermore, it was observed that patients positive for CD56 cells after therapy had a better overall survival (HR=0.282, 95%CI=0.109-0.729, χ^2=7.854, p=.005). In conclusion, we have identified a novel post-therapeutic natural killer-like transcription signature in patients responding to long course preoperative chemoradiotherapy. Furthermore, patients with a higher abundance of CD56 positive natural killer cells post- long course preoperative chemoradiotherapy had better overall survival. Therefore, harnessing a natural killer-like response after therapy may improve outcomes for locally advanced rectal cancer patients.

Abstract word count: 298

Key words:

Natural killer cells; rectal cancer; tumour regression; gene expression; chemoradiotherapy.

Long course preoperative chemoradiotherapy for treatment of locally advanced rectal cancer typically consists of a fluoropyrimidine (5-fluorouracil or capecitabine) in

combination with 45-50.4 Grays of pelvic radiation over 5-5.5 weeks in 1.8Gy fractions with an interval of 8-12 weeks prior to total mesorectal excision. Following long course preoperative chemoradiotherapy, 12-15% of patients experience a pathologically complete response [1]. The remaining locally advanced rectal cancer patients exhibit a full spectrum of response from minimal residual malignancy to progressive disease. There are at least six published classification systems for locally advanced rectal cancer tumour regression [2]. However, many institutions have adopted the 4-tier tumour regression grade system similar to that published by Rodel *et al* 2005, due to its improved ability to predict prognosis and scoring concordance between pathologists. The extent of tumour regression after therapy has previously been associated with disease-free survival [3], therefore, an increased understanding of the biology that mediates tumour regression is crucial. Additionally identifying therapy-related alterations to the tumour microenvironment architecture may provide guidance for alternative therapies for patients with poor outcomes. Modulation of key components of the immune system have observed in a variety of cancers treated preoperatively with radiotherapy and chemoradiotherapy, including sarcoma and rectal cancer. With regards to rectal cancer increased prevalence of CD3, CD4 and CD8 cell populations have been observed after long course preoperative chemoradiotherapy [4-7]. Based on these findings, we hypothesise that the transcriptome of post therapeutic tumour specimens and the tumour microenvironment will be representative of specific biology and thus may direct us to predictors of response to therapy.

Materials and Methods

Clinical Specimens, Specimen annotation, Nucleic acid extraction

Ethical permission for the study was authorised by the Northern Ireland Biobank (Research Ethics Committee number – 16/NI/0030. Project code: NIB13-0103). 60 consecutive patients were identified who received long course preoperative chemoradiotherapy for locally advanced rectal cancer. Where available, formalin fixed paraffin embedded diagnostic biopsies (n=40) and the corresponding rectal resections (n=40) specimens were obtained from these patients. All patients received long course CRT comprising 45Gy/25 fractions over 5 weeks along with 5-Fluorouracil or capecitabine. Patients underwent total mesorectal excision at an interval of 1-14 weeks (mean=9) following treatment. The clinical and pathological details of the patients are summarised in **Supplementary Table 1**. Tumour regression grade was scored using a four-tier system. Tumour regression grade 1 patients experience minor regression, tumour regression grade 2 patients experience significant regression, tumour regression grade 3 patients will have only small groups or single remaining malignant glands and tumour regression grade 4 will have no residual tumour after therapy. All specimens have been assessed by a standardised protocol, whereby, if no tumour is evident macroscopically, then the site of tumour, and any scars of areas of ulceration, are processed in entirety for histological examination. Tumour regression grading was carried out by a colorectal specialist pathologist (ML). H&E stains of each specimen were annotated by a pathologist (MST) enriching for the epithelial compartment of both biopsy and resection, while including the fibro-inflammatory component immediately adjacent to the epithelial component.

Total RNA was extracted using the RNeasy formalin fixed paraffin embedded kit (Qiagen) and genomic DNA was extracted using the DNeasy formalin fixed paraffin embedded kit (Qiagen). Integrity and purity of nucleic acids was assessed using the

4

Nanodrop 2000 (Thermofisher Scientific). Nucleic acids were quantified using the Qubit RNA HS Assay kit (Thermofisher scientific) and TaqMan RNase P Detection kit (ThermoFisher).

Whole Genome DASL array and Real-Time (RT)-PCR
Gene expression profiling was conducted on the Whole-Genome Gene expression DASL (cDNA-mediated, annealing, selection and ligation) high-throughput assay (Illumina) as reported elsewhere [8]. The RNA concentrations are detailed in **Supplementary table 2** and the total input of RNA for each array was approximately 50ng as quantified by Qubit (Thermofisher Scientific). This data is available via gene expression omnibus – accession GSE94104. Transcripts of interest were validated in samples were there was sufficient RNA remaining using Roche RealTime ready catalogue and designer assays; NCR1 (1000106257), SPANXA1 (100106293), GAPDH (100107995), and B2M (100108004). Reverse transcribed cDNA was generated using the First strand transcriptor kit (Roche) and pre-amplifed using the cDNA pre-amp kit (Roche) and assayed on the Roche Lightcycler 480 II. Suitable housekeeping genes were selected after assaying gene expression stability on a biopsy and tumour regression grade 1-3 resection specimens using the Human Reference Gene Panel (Roche). Gene expression fold changes ($\Delta\Delta$Cp Values) were calculated using the Second derivative method max instead of the fit point method for accurate quantification of low gene expression levels [9].

Immunohistochemistry

All H&E stains were performed using the Sakura Autostainer. Immunohistochemistry was employed using the Bond Max (Leica microsystems) on formalin fixed paraffin embedded sections (3µm-thick) and stained with CD56 antibody NCL-L-CD56-1B6

(Novocastra) at a 1:100 dilution. Digital images were taken using the Aperio ScanScope (X40 magnification). Patients were considered to be CD56 positive if a single core had ≥4 positive cells; adopted from a CD56 immunohistochemistry tissue microarray scoring system published by Sconnochia *et al* 2014 [10].

Tissue Microarray
Ethical permission was granted to investigate a post therapeutic rectal cancer tissue microarray provided by the Grampian Biorepository as a validation cohort (North of Scotland research ethics committee number - 11/NS/0015, Tissue request number TR71). The 150 patient resections were collected between 2005-2011 and all patients had long course chemoradiotherapy (45Gy) and concurrent capecitabine (825mg/m2 twice daily) as pre-operative treatment [11]. The tissue microarray was constructed using two 1mm ($0.79mm^2$) cores from each patient resection. The tumour regression grading of these samples was conducted using a similar classification system (Complete response, good partial, partial and minimal) by an expert gastrointestinal pathologist (G.I.M) allowing an easy conversion into the tumour regression grade 1-4 classification system adopted in our own study. The clinical and pathological details of the patients in this cohort are summarised in **Supplementary Table 3.** Patient all-cause mortality was identified from interrogation of the electronic hospital patient information system where status (alive/dead) was recorded. At the time of censoring of patient outcome data, there had been 35 (23.5%) deaths. The mean patient survival was 72 months [95% confidence interval (CI) 68–75 months]. Clinical follow-up was a minimum of 19 months up to maximum of 99 months.

Exclusion criteria and comparative assessment

The inclusion and exclusion process of patients in the discovery and validation cohorts is outlined in the consort diagram in figure 1. All tumour regression grade 4 patients

were excluded from the discovery cohort (n=5), as were patients that tumour had been exhausted by serial sectioning (n=12). A further three patients failed to produce transcriptional profiles. Similarly for the validation cohort tumour regression grade 4 patients were excluded from subsequent analysis (n=73) as were cores that lacked tumour with excessive fibrosis, muscle and mucin (n=98) which was in-line with the approach for transcriptional profiling of NIB13-0103 cohort.

Comparative assessment of clinical and pathological parameters of both patient cohorts demonstrated their suitability as clinically representative discovery and validation cohorts for investigating tumour regression in locally advanced rectal cancer. Patient specimens were obtained for the NIB13-0103 discovery cohort between 2004-2013 from the Belfast Health and Social Care Trust while the patient specimens from the validation cohort were collected between 2005-2011 by NHS Grampian. Only patients who received long course preoperative chemoradiotherapy consisting of fluoropyrimidines and 45Gy pelvic radiation and for whom there was sufficient tumour available for molecular and tissue analysis were included in the study.

Statistical analysis
Assessing statistically significant differences between variables was determined using the Students t test, Log-rank test, Chi-squared and Cox multivariate analysis. Survival was assessed using the Kaplan-Meier method assessed in SPSSv23 for windows. P-values <.05 were considered significant.

High-throughput gene expression analysis
The high-throughput gene expression dataset was exported using Illumina Genome studio. Data was background corrected and normalised using positive, negative and housekeeping controls from the WG-DASL assay automatically provided by Partek's

Report Plug-in. Principal components analysis, differential gene expression analysis (ANOVA between tumour regression grade 1 vs tumour regression grade 3) and hierarchical clustering analysis were conducted using Partek® Genomics Suite® software, version 6.6 Copyright ©; 2016. Differentially expressed genes were determined using fold change thresholds ranging between 2 and 10-fold expression change and only genes with an adjusted p-value <0.05 (Benjamini-Hochberg) were reported. Heat maps with hierarchical clustering were constructed using Ward's linkage and Euclidean distance. Row expression values were standardised to a mean of zero and scaled to a standard deviation of one. Ingenuity Pathway Analysis and Gene Set Enrichment Analysis (http://software.broadinstitute.org/gsea/index.jsp) were used for biological interpretation of differentially expressed genes.

Next Generation Sequencing
Mutational status was acquired from patient biopsies from approximately 10ng DNA. Sequencing was performed using the Ion Torrent Personal Genome machine, as previously published [12]. Only 38/40 patients had successful mutational profiles generated.

Defining microsatellite instability

Microsatellite instability was assessed using MSI Analysis system, Version 1.2 (Promega) on the Applied Biosystems 3500 analyser (Thermofisher Scientific) using Genemapper v4.1 software. Microsatellite instability was determined by the instability of two or more of the five markers (NR24, BAT26, NR21, BAT25 and MONO27) [8].

Microenvironment Cell Populations Counter

Microenvironment cell population counter is a peer reviewed software tool that robustly measures the abundance of microenvironment cell populations using transcriptional data generated from heterogeneous tissues. The microenvironment cell population

counter R package was downloaded from - . An estimate score for NK cells which equates to the absolute abundance of cell type was generated for each NIB13-0103 transcriptional profile [13].

QUADrATiC Analysis

The 27 gene signature generated from differential gene expression analysis in Partek® Genomics Suite® software, version 6.6 Copyright ©; (see Supplementary Table 4) was used to perform connectivity map analysis using the QUADrATiC software [14]. The 27 genes were mapped using annotation files to probe IDs for the Affymetrix HG U133A microarray, since QUADrATiC requires its input signatures in this form. Running the analysis produces a candidate list of positive connections, i.e. those FDA-approved compounds within the LINCS database which appear to induce similar expression patterns to that of the 27 genes, in particular cell lines. This candidate list was further investigated for any compounds with immune-modulating properties.

Results

Comparative analysis of the discovery and validation cohorts revealed no statistically significant differences observed between the proportions of regression grades (tumour regression grade 3 27.5% vs 37%, tumour regression grade 2 42.5% vs 46%, tumour regression grade 1 30% vs 17%), as determined by Chi-squared (p= 0.147). This ensured no regression-related sample enrichment or bias in our selected sample sets.

Unsupervised principal components analysis was employed to visualise gene expression data from the NIB13-013 post-therapeutic resection specimens (Figure 2A). The principal components analysis plot highlighted a distinct separation between

post-treatment profiles from tumour regression grade 1 and 3 patients with an overlap of tumour regression grade 2 patients with both groups which highlights major differences in gene expression across the patient cohort. Using these tumour transcriptomic profiles we conducted differential gene expression analysis to identify specific genes of interest for downstream investigation by comparing tumours with either tumour regression grade 1 or tumour regression grade 3 response. Using this supervised approach, we identified 27 genes which had over 10-fold differential expression changes between our contrasting tumour regression grade subgroups. In order to test the robustness of classifying samples based on tumour regression grade, we used this 27 gene signature generated from contrasting tumour regression grade 1 with tumour regression grade 3, to classify tumour regression grade 2 tumours. Importantly, we found that our signature clustered tumour regression grade 2 tumours as having intermediate signature expression compared to tumour regression grade 1 and tumour regression grade 3, indicating a dose response of expression according to tumour regression grade and confirming its potential clinical utility (Figure 2B). The 27 genes consisted of natural killer cell receptors, cancer/testis antigens, olfactory receptors, and solute carriers (supplementary table 4). Ingenuity pathway analysis of the 27 genes suggested that "Natural Killer Cell signalling" and "Crosstalk between Dendritic Cells and Natural Killer Cells" are the two defining pathways involved with this tumour regression grade associated gene list (Figure 2C).

To identify potential FDA-approved compounds that may induce the gene expression phenotype of responders after therapy and harness the good prognosis natural killer-like biology, we utilised the QUADrATiC software. QUADrATiC requires an input of Affymetrix probe IDs. Therefore, the analysis was performed using 21 Affymetrix probe IDs that could be matched from the 27 Illumina probe IDs identified. Statistically

significant positive connection scores were generated for the immunomodulatory drug (IMiD) Lenalidomide and the proteosomal inhibitor bortezomib (Supplementary figure 1a). Both of these agents are already established in routine clinical practice in the treatment of haematological malignancies [15,16]. The heatmap shows the relative contribution (or Contribution Fraction) of each individual probes input to the overall score and we can conclude from this that the cancer-testis antigen families SPANX and GAGE were the most significant contributors to the connection scores and hence are potential targets for the mechanism of action of the two drugs (Supplementary figure 1a). This result prompted us to cluster resection specimens based upon the expression of a list of cancer testis antigens genes. Robust clustering in a tumour regression grade dependent manner was observed with a higher expression in the responders (tumour regression grade 3) than in tumour regression grade 1/non-responders (Supplementary figure 1b). We selected the cancer testis antigen SPANXA1 to validate using RT-PCR. Results indicated that the average fold change between the tumour regression grade 1 group and tumour regression grade 3 patient group was 7.3 (Supplementary figure 1c).

Based upon our observations, microenvironment cell population counter was used to assess the abundance of natural killer cells in our resection specimens. Where we observed significantly increased natural killer cell populations in a tumour regression grade dependent fashion (figure 3A, $p < 0.0001$, Tukey's multiple comparisons test). This finding was complemented by gene set enrichment analysis comparing tumour regression grade 3 vs tumour regression grade 1 resection transcriptional profiels which showed a statistically significant upregulation of the "GO_NATURAL_KILLER_CELL_ACTIVATION" gene set in tumour regression grade 3 resections (Figure 3B). Subsequently, validation of the natural killer cell specific

marker *NCR1* which emerged as a differentially expressed gene was performed using RT-PCR. There was observed to be a statistically significant 2.3 fold difference in expression between tumour regression grade 3 and tumour regression grade 1 patients (figure 3c, p= 0.04). The cell surface protein *NCAM1* also known as CD56 is considered one of the most specific proteomic markers for natural killer cells. A number of remaining resections with viable tumour from the discovery cohort were stained by immunohistochemistry for the natural killer cell marker CD56, confirming the presence of CD56 positive cells in a tumour regression grade 3 resection (Figure3D). Interestingly, *NCAM1* is considered a poor marker of natural killer cells at the transcript level due to its non-specific expression, as it is also strongly expressed by nerve ganglion and for this reason it is not included as a marker of natural killer cell quantification in microenvironment cell population counter [13]. Supplementary figure 2 supports this finding as it was observed that the expression of *NCAM1* is not tumour regression grade dependent in our cohort unlike the natural killer cell marker *NCR1*.

We next examined the transcriptional profiles obtained for the matched pre-treatment biopsies and assessed the abundance of natural killer cells, again using microenvironment cell population counter. It was observed that the presence of natural killer cells in pre-treatment biopsies was not tumour regression grade dependent and therefore did not have predictive value (figure 4a). A trend was observed that the abundance of natural killer cells tended to increase in tumour regression grade 3 patients in a matched pre and post analysis while tumour regression grade 1 patients experienced increases and decreases in the abundance of natural killer cells, however this did not reach statistical significance (figure 4b+c).

Next generation sequencing analysis of clinically relevant mutations (KRAS, TP53, BRAF, NRAS, APC, ERBB2, PIK3CA, and SMAD4) was performed to assess any

association with mutational profile and the presence of our natural killer-like gene expression signature. There was no observable association between our mutational data and the tumour regression grade dependent transcriptional subgroups identified, which is highlighted by a heatmap that depicts the mutations present in each tumour regression grade (supplementary figure 3). We also performed a comparative analysis to identify differences in the mutational landscape of the NIB13-0103 rectal cohort (n=38) when compared to the TCGA, Nature 2012 colorectal cancer dataset. [17,18]. The dataset was filtered for rectal cancer stage II-III (n=31) using cBioPortal. This comparison of mutations in the NIB13-0103 and TCGA cohorts demonstrated no statistical difference in the proportions of mutations of TP53 (73.68% vs 80.65%, p = 0.547), KRAS (55.26% vs 45.16.0% p= 0.472), APC (39.47% vs 77.42.0% p= 0.283), NRAS (0% vs 6.45% p = 0.198), BRAF (2.63% vs 0% p = 1.0), ERBB2 (0% vs 6.45% p= 0.198), PIK3CA (5.26% vs 9.68% p = 0.651), SMAD4 (7.89% vs 3.23% p =0.622) [17,18]. However, interpretation of these differences is difficult considering the small sample size of both cohorts. Similarly, we could not determine an association between MSI status in our patients and the natural killer-like response as no tumours were MSI high.

In order to confirm the gene expression findings using a diagnostically relevant methodology, we performed an immunohistochemistry-based assessment of resection tumour samples from an independent rectal cancer cohort (n=150) using the NK marker CD56. We observed a significantly higher total number of CD56+ cells per core in the tumour regression grade 3 resections compared to the number of CD56+ cells in tumour regression grade 1 resection specimens (p value = 0.008) (Figure 5a-d). The prognostic significance of CD56 positivity was investigated in this independent cohort, with patients categorised as CD56 positive when ≥4 positive cells per core

were observed. Using this threshold for positivity, patients with CD56 positivity were shown to have a significantly better overall survival (HR=0.282, 95%CI=0.109-0.729, χ^2=7.854, p=.005), which was independent of tumour regression grading (Figure 6A). Interestingly CD56 positivity conferred a trend towards better outcomes within all tumour regression grades, particularly within the tumour regression grade 1 samples within the independent cohort (Figure6b-d). Furthermore, CD56 positivity was shown to be independently prognostic through multivariate analysis p<.006 (supplementary table 5).

Discussion

It has been proposed that common alterations to immune/inflammatory pathways may confer sensitivity to long course preoperative chemoradiotherapy in locally advanced rectal cancer patients [19]. Understanding the nature of the immune response that occurs following long course preoperative chemoradiotherapy may be problematic in pre-therapeutic biopsies. Therefore, we have interrogated post-therapeutic locally advanced rectal cancer resections using a combination of modern high-throughput genomic technologies and traditional pathology techniques, to characterise the immune response that has occurred and allow us to hypothesise its use a predictive biomarker in on-treatment biopsies but also how harness it for therapeutic benefit.

Both radiation and chemotherapy can augment immune response possibly through increased expression of antigens [8], these antigens may be responsible for engaging the cytotoxic cells of the adaptive immune system [9-11]. The increased presence of CD8+ cytotoxic lymphocytes are well defined in colorectal cancer particularly in microsatellite unstable tumours and are associated with an improved prognosis [12]. Based upon the expression of immune checkpoints such as PDL1 there is rationale

for the use of immune checkpoint inhibitors in MSI-H colorectal cancers and although most immune checkpoint inhibitors execute their effects through CD8+ cytotoxic lymphocytes, therapeutic modulation of other immune cell types such as macrophages, myeloid-derived suppressor cells and natural killer cells could have similar effects in the rectal paradigm [13-15]. Natural killer cells are cytotoxic lymphocytes of the innate immune system and act as the body's first line of defence against tumour cells. Natural killer cells have been observed to be critical for response to gemcitabine in pancreatic mouse models [16]. Furthermore, the presence and activity of natural killer cells may be an indicator of favourable outcome in breast cancer and oropharyngeal squamous cell carcinoma [17,18].

In this study we observed significantly increased expression of genes associated with a natural killer-like phenotype after long course preoperative chemoradiotherapy. This, to our knowledge, is a novel observation following treatment of locally advanced rectal cancer. While the overall frequency of these cells may not be very high, they do appear to be associated with a strong biological response underpinning their potential biological significance and implications in tumour regression. The presence of natural killer cells was initially defined *in silico* using the microenvironment cell population counter tool which gives a robust estimation to the abundance of NK cells [13]. The abundance of natural killer cells was also assessed in the pre-therapeutic biopsies, however increased natural killer cells was shown to be therapy related and therefore not a predictive biomarkers of response. Validation of the microarray results was performed by employing RT-PCR using *NCR1*. The expression of *NCR1* is considered one of the more specific markers of natural killer cells, however natural cytotoxicity receptors can also be expressed on innate lymphoid cells [20,21]. The delineation of classical natural killer cell populations from iNKT cells, mucosal associated invariant

T cells, ILC1s, ILC3s and gamma delta T cells remains challenging in formalin fixed paraffin embedded tissue as there is considerable overlap in the receptor repertoires, and as such cell specific detailing still relies on flow cytometry approaches [22-26]. Therefore, the general expression of immune derived CD56 still remains the gold standard for detecting natural killer cells in formalin fixed paraffin embedded tissue. As a result of our gene expression profiling we investigated the expression of the natural killer cell marker CD56 using immunohistochemistry. A comprehensive analysis of CD56 natural killer cell expression of post-therapeutic resection specimens could not be performed in the discovery cohort due to tumour exhaustion in many of the tumour regression grade 3 specimens through serial sectioning of the formalin fixed paraffin embedded blocks. However CD56 immunohistochemistry was successfully employed in the validation cohort and we observed this cell type to be more frequent in tumour regression grade 3 responders than tumour regression grade 1 after therapy. Our evaluation of CD56 immunohistochemistry was performed on tissue microarray cores which is a limitation of this study due to potential regional bias, therefore, we feel that further adjusting our scoring criterion to enable CD56 assessment in full resection specimens deserves further investigation. The presence of CD56+ natural killer cells by immunohistochemistry in treatment naïve colorectal cancer has been described as infrequent [27,28]; on this basis we did not discern between those in the adjacent-stroma and intraepithelial natural killer cells, however, in a larger dataset of full face resection specimens we would prefer to assess the influence of different regional infiltration.

A recent study demonstrated the clinical importance of natural killer cells in colorectal cancer, Sconocchia *et al* 2014 observed that an enrichment of both natural killer and CD8 T cells improved the outcome of colorectal cancer patients [10]. With regards to

post therapeutic infiltration of immune cells in rectal cancer, the presence of natural killer cells has not previously been observed to be significant in post therapy rectal specimens by CD56+ immunohistochemistry [7]. However, it is worth noting in this study, positive infiltration of CD56 natural killer cells was assessed as 4>/high powered field rather than ≥4 positive cells per core which may account for different observations. Post-therapeutic infiltration of adaptive immune cells in locally advanced rectal cancer has been shown to increase after therapy, Teng *et al* 2015 observed significantly higher CD3 and CD8 TILs and in a further study also CD4 tumour infiltrating lymphocytes [6,7].

It is evident that the clinical utility of measuring natural killer cell density in pre-treatment locally advanced rectal cancer biopsies may be limited. However, our observation of tumour regression grade dependent natural killer cell abundance in post-therapeutic resections warrants the collection and analysis of on-treatment or post-therapeutic biopsies in order to assess natural killer cell biology that may have the potential to provide real-time clinical guidance on the efficacy of the therapeutic intervention. We observed that patients with CD56 positivity had a significantly better outcome to chemo-radiation than those with lower levels of CD56 staining with multivariate analysis showing CD56 positivity to be a significant independent prognostic factor. As a result, we hypothesise that encouraging this natural killer cell biology with immune modulating compounds may be a novel therapeutic strategy for locally advanced rectal cancer patients.

By using QUADrATiC software to carry out connectivity mapping analysis against the subset of FDA-approved compounds in the LINCs database, we identified Lenalidomide and Bortezomib as two FDA approved compounds with proven anti-tumour and immunomodulatory effects that may induce the expression of the 27 genes

identified in **Supplementary Table 4**. Lenalidomide and proteosomal inhibitors (Bortezomib and Carfilzomib) are already in clinical use for the treatment of haematological malignancies and have demonstrated the ability regulate natural killer cell function. [15,16,29-33]. By analysing the connections further using the Contribution Fraction capabilities of QUADrATiC, both compounds were shown to have an association with the increased expression of cancer testis/antigens which are potentially activators of T cell receptor mediated cytotoxicity, and thus may play a role in therapy-related immunomodulation [4,5].

We hypothesise that using immune modulating compounds such as lenalidomide and low dose bortezomib (or the second generation proteosomal inhibitor Carfilzomib) as single agents or even in combination could prime the natural killer cell-related immune system, improving the efficacy of standard chemoradiotherapy. Aside from immune modulating compounds such as bortezomib and lenalidomide, there are currently three other methods of stimulating or harnessing natural killer cell function i) endogenous interleukins such as IL-2 and IL-15, can stimulate natural killer cell activity [34], ii) monoclonal antibodies that act as natural killer cell immune checkpoint inhibitors such as Lirilumab [35] iii) autologous and allogeneic transfer of wild type and genetically engineered natural killer cells are also being tested for efficacy in both haematological and solid tumours [36-38].

In conclusion, we have provided independently validated evidence that the presence of a CD56+ natural killer-like immune response after therapy confers better overall survival for locally advanced rectal cancer patients, which may also be independent of tumour regression. We can conclude that natural killer cells are an important element of the immune landscape of locally advanced rectal cancer and on this basis we have hypothesised that harnessing a natural killer cell response using immune modulating

compounds could further improve outcomes for patients with locally advanced rectal cancer. Finally, real-time longitudinal assessment of the natural killer cell response in on-treatment biopsy material may inform clinical decision-making for duration of treatment or surgical decision to remove tumour if no response to long course preoperative chemoradiotherapy is evident.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1. Gollins S, Sebag-Montefiore D. Neoadjuvant Treatment Strategies for Locally Advanced Rectal Cancer. *Clin Oncol (R Coll Radiol)* 2016;**28:**146-151.

2. Trakarnsanga A, Gonen M, Shia J*, et al.* Comparison of tumor regression grade systems for locally advanced rectal cancer after multimodality treatment. *J Natl Cancer Inst* 2014;**106:**10.1093/jnci/dju248. Print 2014 Oct.

3. Rodel C, Martus P, Papadoupolos T, *et al.* Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *J Clin Oncol* 2005;**23:**8688-8696.

4. Sharma A, Bode B, Wenger RH, *et al.* gamma-Radiation promotes immunological recognition of cancer cells through increased expression of cancer-testis antigens in vitro and in vivo. *PLoS One* 2011;**6:**e28217.

5. Sharma A, Bode B, Studer G, *et al.* Radiotherapy of human sarcoma promotes an intratumoral immune effector signature. *Clin Cancer Res* 2013;**19:**4843-4853.

6. Teng F, Mu D, Meng X, *et al.* Tumor infiltrating lymphocytes (TILs) before and after neoadjuvant chemoradiotherapy and its clinical utility for rectal cancer. *Am J Cancer Res* 2015;**5:**2064-2074.

7. Teng F, Meng X, Kong L, *et al.* Tumor-infiltrating lymphocytes, forkhead box P3, programmed death ligand-1, and cytotoxic T lymphocyte-associated antigen-4 expressions before and after neoadjuvant chemoradiation in rectal cancer. *Transl Res* 2015;**166:**721-732.e1.

8. Alvi MA, McArt DG, Kelly P, *et al.* Comprehensive molecular pathology analysis of small bowel adenocarcinoma reveals novel targets with potential for clinical utility. *Oncotarget* 2015;**6:**20863-20874.

9. Luu-The V, Paquet N, Calvo E, *et al.* Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction. *BioTechniques* 2005;**38:**287-293.

10. Sconocchia G, Eppenberger S, Spagnoli GC, *et al.* NK cells and T cells cooperate during the clinical course of colorectal cancer. *Oncoimmunology* 2014;**3:**e952197.

11. Brown GT, Cash B, Alnabulsi A, *et al.* The expression and prognostic significance of bcl-2-associated transcription factor 1 in rectal cancer following neoadjuvant therapy. *Histopathology* 2016;**68:**556-566.

12. McCourt CM, McArt DG, Mills K, *et al.* Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS One* 2013;**8:**e69604.

13. 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.

14. O'Reilly PG, Wen Q, Bankhead P, *et al.* QUADrATiC: scalable gene expression connectivity mapping for repurposing FDA-approved therapeutics. *BMC Bioinformatics* 2016;**17:**198-016-1062-1.

15. Totani H, Ri M, Kato C, *et al.* Phase I study of once weekly treatment with bortezomib in combination with lenalidomide and dexamethasone for relapsed or refractory multiple myeloma. *Int J Hematol* 2016;**103:**316-321.

16. Wang M, Martin T, Bensinger W, *et al.* Phase 2 dose-expansion study (PX-171-006) of carfilzomib, lenalidomide, and low-dose dexamethasone in relapsed or progressive multiple myeloma. *Blood* 2013;**122:**3122-3128.

17. Cerami E, Gao J, Dogrusoz U, *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;**2:**401-404.

18. Gao J, Aksoy BA, Dogrusoz U, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;**6:**pl1.

19. Agostini M, Janssen KP, Kim IJ, *et al.* An integrative approach for the identification of prognostic and predictive biomarkers in rectal cancer. *Oncotarget* 2015;**6:**32561-32574.

20. Artis D, Spits H. The biology of innate lymphoid cells. *Nature* 2015;**517:**293-301.

21. Morvan MG, Lanier LL. NK cells and cancer: you can teach innate cells new tricks. *Nat Rev Cancer* 2015;**16:**7-19.

22. Bjorkstrom NK, Ljunggren HG, Sandberg JK. CD56 negative NK cells: origin, function, and role in chronic viral disease. *Trends Immunol* 2010;**31:**401-406.

23. Godfrey DI, Uldrich AP, McCluskey J, *et al.* The burgeoning family of unconventional T cells. *Nat Immunol* 2015;**16:**1114-1123.

24. Gapin L. Where do MAIT cells fit in the family of unconventional T cells? *PLoS Biol* 2009;**7:**e70.

25. Spits H, Bernink JH, Lanier L. NK cells and type 1 innate lymphoid cells: partners in host defense. *Nat Immunol* 2016;**17:**758-764.

26. Killig M, Glatzer T, Romagnani C. Recognition strategies of group 3 innate lymphoid cells. *Front Immunol* 2014;**5:**142.

27. Sandel MH, Speetjens FM, Menon AG, *et al.* Natural killer cells infiltrating colorectal cancer and MHC class I expression. *Mol Immunol* 2005;**42:**541-546.

28. Halama N, Braun M, Kahlert C, *et al.* Natural killer cells are scarce in colorectal carcinoma tissue despite high levels of chemokines and cytokines. *Clin Cancer Res* 2011;**17:**678-689.

29. Armeanu S, Krusch M, Baltz KM, *et al.* Direct and natural killer cell-mediated antitumor effects of low-dose bortezomib in hepatocellular carcinoma. *Clin Cancer Res* 2008;**14:**3520-3528.

30. Hu W, Gao ZY, Wang W. Bortezomib enhances the sensitivity of prostate cancer cells to natural killer cell-mediated cytotoxicity. *Zhonghua Nan Ke Xue* 2014;**20:**218-224.

31. Wu L, Parton A, Lu L, *et al.* Lenalidomide enhances antibody-dependent cellular cytotoxicity of solid tumor cells in vitro: influence of host immune and tumor markers. *Cancer Immunol Immunother* 2011;**60:**61-73.

32. Xu Y, Sun J, Sheard MA, *et al.* Lenalidomide overcomes suppression of human natural killer cell anti-tumor functions by neuroblastoma microenvironment-associated IL-6 and TGFbeta1. *Cancer Immunol Immunother* 2013;**62:**1637-1648.

33. Krieg S, Ullrich E. Novel immune modulators used in hematology: impact on NK cells. *Front Immunol* 2013;**3:**388.

34. Childs RW, Carlsten M. Therapeutic approaches to enhance natural killer cell cytotoxicity against cancer: the force awakens. *Nat Rev Drug Discov* 2015;**14:**487-498.

35. Kohrt HE, Thielens A, Marabelle A*, et al.* Anti-KIR antibody enhancement of anti-lymphoma activity of natural killer cells as monotherapy and in combination with anti-CD20 antibodies. *Blood* 2014;**123:**678-686.

36. Lim O, Jung MY, Hwang YK*, et al.* Present and Future of Allogeneic Natural Killer Cell Therapy. *Front Immunol* 2015;**6:**286.

37. Rezvani K, Rouce RH. The Application of Natural Killer Cell Immunotherapy for the Treatment of Cancer. *Front Immunol* 2015;**6:**578.

38. Burga RA, Nguyen T, Zulovich J*, et al.* Improving efficacy of cancer immunotherapy by genetic modification of natural killer cells. *Cytotherapy* 2016;**18:**1410-1421.

Figure Legends

Figure 1. Consort diagram outlining the inclusion and exclusion process for discovery (matched pre-treatment biospies and post-therapeutic resections n=40) and validation cohort (post-therapeutic tissue microarray n=150) prior to data generation and analysis

Figure 2. Differential gene expression analysis and Pathway Analysis of locally advanced rectal cancer resections in discovery cohort (n=40). (A) Principal Component Analysis of locally advanced rectal cancer resections by tumour regression grade shows separation of tumour regression grade 1 gene expression profiles and tumour regression grade 3 profiles. (B) Heatmap showing hierarchical clustering analysis of resections by 27 most differentially expressed genes, (blue denotes down-regulated expression of genes and red upregulation), the heatmap depicts an upregulation of these genes in tumour regression grade 3 resections (C) Ingenuity Pathway Analysis of 27 genes, Bar chart shows natural killer cell signalling to be the most statistically enriched pathway. Legend - tumour regression grade 1 (red), tumour regression grade 2 (yellow) and tumour regression grade 3 (blue).

Figure 3. Assessing natural killer cell biology in locally advanced rectal cancer resections of discovery cohort (n=40). (A) Microenvironment cell population counter NK cell estimate scores assessed in post therapeutic locally advanced rectal cancer resections (tumour regression grade 1 vs tumour regression grade 3, p <0.0001) demonstrates a higher abundance of natural killer cells in tumour regression grade 3 resections when compared to tumour regression grade 1. (B) Gene Set Enrichment Analysis comparing tumour regression grade 3 resections (phenotype 1) and tumour regression grade 1 resections (phenotype 0) using the GO_NATURAL KILLER_CELL_ACTIVATION gene set (FDR < 0.0001), shows an upregulation of genes associated with natural killer cell activity in tumour regression grade 3 resections compared to tumour regression grade 1 (C) RT-PCR validation of mRNA expression of NCR1 in tumour regression grade 1 + tumour regression grade 3 rectal resections. (p = 0.04, students t test). (D) Image depicting CD56 positive natural killer cells by immunohistochemistry in tumour regression grade 3 post-therapeutic resection from the discovery cohort (x40 magnification).

Figure 4. Assessing the abundance of NK cells in biopsies and matched resections in discovery cohort (n=40) using microenvironment cell population counter. (A) Microenvironment cell population counter NK cell estimate scores assessed in pre therapeutic locally advanced rectal cancer biopsies shows no difference in abundance of natural killer cells between tumour regression grades (B) Normalised microenvironment cell population counter NK cell estimate scores assessed in pre therapeutic tumour regression grade 1 locally advanced rectal cancer biopsies and matched resections shows changes after long course preoperative chemoradiotherapy (C) Normalised microenvironment cell population counter NK cell estimate scores assessed in pre therapeutic tumour regression

grade 3 locally advanced rectal cancer biopsies and matched resections shows changes after long course preoperative chemoradiotherapy.

Figure 5. The association between CD56 positive natural killer cell**s and tumour regression grade in post-therapeutic validation locally advanced rectal cancer tissue microarray (n=150) using immunohistochemistry.** (A) Dot plot showing total counts of CD56 positive natural killer cells per tissue microarray core grouped by tumour regression grading (tumour regression grade 3 vs tumour regression grade 1). There was observed to be significantly more CD56 positive natural killer cells in tumour regression grade 3 resections compared to tumour regression grade 1 resections by students T test p = 0.0083. (B) Representative image of a CD56 positive immunohistochemistry in tissue microarray core of a tumour regression grade 3 responder with CD56 natural killer cells highlighted in yellow (x40 magnification). (C) Representative image of a CD56 positive tissue microarray core by immunohistochemistry of a tumour regression grade 1 responder with CD56 natural killer cells highlighted in yellow (x40 magnification).

Figure 6. Investigating the influence of CD56 immunohistochemistry positivity on overall survival in post-therapeutic locally advanced rectal cancer validation tissue microarray (n=149)*. (A) Kaplan-Meier Curves showing overall survival of all patients (tumour regression grade 1-3) based upon CD56 positivity (HR=0.282, 95%CI=0.109-0.729, $\chi2$=7.854, Log-rank p= 0.005). (B) Kaplan-Meier Curves showing overall survival of CD56 positive patients within tumour regression grade 3 subgroup (HR=0.404, 95%CI=0.081-2.004, $\chi2$=1.317, Log-rank p= 0.251). (C) Kaplan-Meier Curves showing overall survival of CD56 positive patients within

tumour regression grade 2 subgroup (HR=0.323, 95%CI=0.094-1.109, χ2=3.621, log-rank p= 0.057) (D) Kaplan-Meier Curves showing overall survival of CD56 positive patients within tumour regression grade 1 subgroup (HR=0.036, 95%CI=0.000-116.410, χ2=1.677, log-rank p= 0.195). * **One patient from CD56** immunohistochemistry **validation tissue microarray did not have survival data.**