Appendix A – Remark

This document details references to check list items in the manuscript and supplementary information and tables whereby each of the REMARK guidelines have been addressed.

State the marker examined, the study objectives, and any pre-specified hypotheses.

Marker Examined

- "The DNA Damage Immune Response (DDIR) assay, formerly known as the DNA Damage Response Deficiency (DDIR) assay, was previously developed in breast cancer using an unsupervised hierarchical clustering approach. Biologically the DDIR assay indicates constitutive activation of the cyclic GMP-AMP synthase (cGAS)/Stimulator of interferon genes (STING) pathway in response to endogenous DNA damage. Deficiencies in DNA repair and the Fanconi Anaemia/BRCA pathway in particular, have been reported to activate this pathway. Importantly the 44 gene DDIR assay includes well known immune checkpoint targets, such as PD-L1 and Indoleamine 2,3-Dioxygenase 1 (IDO-1), as well as several inflammatory cytokines."

- The development of the DDIR assay is described in the following publication:
  Mulligan JM, Hill L a, Deharo S, Irwin G, Boyle D, Keating KE, et al. Identification and validation of an anthracycline/cyclophosphamide-based chemotherapy response assay in breast cancer. J Natl Cancer Inst. 2014 Jan;106(1):djt335.

Study objectives

- "To assess the ability of the DDIR assay to predict pathological response and prognosis following DNA-damaging neo-adjuvant chemotherapy in OAC."

Pre-specified hypotheses

- "We hypothesized that pathological tumour response and improved survival may be due to pre-existing deficiencies in DNA repair pathways with associated activation of an innate immune response."

Describe the characteristics (for example, disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria.

Patients

- "The OAC cohort comprised of 273 formalin fixed paraffin embedded collected at four centres in the UK (Belfast, Cambridge, Edinburgh and Southampton) as part of the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) consortium. All patients were treated with platinum-based neo-adjuvant chemotherapy prior to surgical resection with either an oesophagectomy or extended gastrectomy. Prior to platinum-based neo-adjuvant chemotherapy all patients were evaluated by computed tomography (CT) of the chest and abdomen, positron emission tomography (PET) and esophagogastroduodenoscopy with endoscopic ultrasound (EUS) where clinically indicated. Tumour site ranged from 33 (12.1%) oesophageal tumours to 130 (47.6%), 78 (28.6%) and 32 (11.7%) Siewert type 1, 2 and 3 type tumours respectively (Supplementary Table 2). Samples were obtained from the Northern Ireland Biobank and in collaboration with three further clinical sites, University of Cambridge, University of Southampton and University of Edinburgh."

Describe treatments received and how chosen (for example, randomized or rule-based).

- Samples were chosen from retrospective databases at four centres: Belfast, Cambridge, Edinburgh and Southampton. "FFPE pre-chemotherapy endoscopic biopsies from 273 patients with resectable OAC, treated with neo-adjuvant chemotherapy followed by surgical resection, were collected at four UK centres in the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) consortium between 2003 and 2014"

- Of note, all samples were randomized based on clinical factors prior to processing. “Post randomization, total RNA was extracted from 5x10 µm macrodissected FFPE tissue slides using the Recoverall™ Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Scientific, Waltham, MA) and amplified using the NuGen Ovation FFPE Amplification System v3 (NuGen San Carlos, CA).”

Describe type of biological material used (including control samples) and methods of preservation and storage.

- “FFPE pre-chemotherapy endoscopic biopsies from 273 patients with resectable OAC, treated with neo-adjuvant chemotherapy followed by surgical resection, were collected at four UK centres in the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) consortium between 2003 and 2014.”

- “Biopsies were reviewed for pathological subtype prior to marking for macrodissection and samples containing at least 50% adenocarcinoma tissue by area were taken forward.”

- “Matched FFPE OAC resection specimens were available for 126 patients who received neo-adjuvant chemotherapy prior to surgical resection at the Northern Ireland Cancer Centre.”

Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.

References to various aspects of this criterion are presented in individual sections below.

Laboratory protocols, QC and reagents/kits used

- Laboratory protocols, QC and reagents/kits used are detailed in the following sections of the Manuscript:

    - Gene Expression Profiling from FFPE Tissue

- Whole Genome Sequencing

Bioinformatic and biostatistical methods, Quantitation, scoring and QC

- **Microarray Quality Control**

"Microarray quality control analysis was performed using Robust Multi-array Average RMA pre-processed data, assessing the following metrics:

Array Image Analysis: analyses Affymetrix® CEL files (containing raw intensity data) and looks for large deviations in background values or unusual patterns in probe intensities that may indicate the presence of artefacts or problems with hybridization (such as leakage from the array, uneven washing)

GeneChip QC: examines a number of control parameters from the RPT files as supplied by Affymetrix®. Full descriptions of these control parameters are available from the Affymetrix® Statistical Algorithms Description Document.

This assessment facilitates the monitoring of profile quality and allows for the evaluation of assay and hybridization performance. Affymetrix® has specified absolute thresholds (lower limit, upper limit or both limits) for a number of these parameters. In addition, it is expected that for the majority of studies data should be comparable, therefore an assessment of overall profile similarity is performed using thresholds based on median absolute deviation (MAD). Any values outside median +/- 3.5 sigma (sigma defined as 1.4826 times the median absolute deviation [MAD]) for that metric will be flagged as potential QC outliers.

Principal Component Analysis (PCA): a well-established exploratory analysis method for high-dimensional data and provides low-dimensional summaries of sample and variable properties.

PCA is used in this context to detect outliers and (known or unknown) systematic structures using pre-processed expression data

Intensity Distribution Analysis: a collection of graphical distribution assessment methods, employed to assess the spread of the data and examine the position of individual profiles, relative to the full profile set. Histogram plots of normalized expression data are constructed to visualize profile spread and differences and expression box-plots are provided to graphically examine median, upper and lower quartile ranges for each profile.

Within a study, it is generally expected that all samples should show a comparable distribution. The program assesses distribution similarity using the Kolmogorov-Smirnov (KS) Test and samples are deemed as outliers based on a predefined KS statistic thresholds."

Functional Analysis

- "A moderated t-test (SAMR) was applied to identify differentially expressed genes between the DDIR-positive and negative groups. Using an FDR of <0.05 a list of 707 genes were derived and used as input into Database for Annotation, Visualization and Integrated Discovery (DAVID). The enrichment of Gene Ontology Biological Process terms was analysed. Terms with p-values less than 0.05 were retained and fold enrichment and FDR reported."

Generation of DDIR Assay Scores

- "Pre-processing of the OAC independent validation dataset was performed using a refRMA model applied to RMA background corrected data, which applied pre-defined normalization and summarization parameters to the probe level data specific for the Xcel™ platform. Prior to DDIR score calculation, datasets were median summarized from probe to probe set levels, followed by a median summary to gene level. DDIR scores were calculated using the DDIR assay parameters (derived from a regression based model) applied as previously described.

  An optimized DDIR threshold (identified in an independent technical study of n=45 samples prior to profiling of the independent OAC validation cohort) for DDIR on Xcel™ microarray platform of 0.3403 was applied to dichotomize DDIR assay scores of the OAC independent validation samples to classify samples as DDIR positive (>threshold) otherwise DDIR negative (<=threshold)."

Statistical Assessment

- "All tests of significance were two-sided and performed at a 5% alpha significance level. The p-value for the test was calculated for original, log-transformed and rank time scales. Clinical factor association with DDIR status was performed using chi-

squared test. Cox proportional hazards regression was used to investigate the prognostic effects of the DDIR signature on relapse-free (RFS) and overall survival (OS) defined as the time from surgical resection to relapse of disease or death from any cause, respectively. The estimated effect of the signature was adjusted for factors available at the time of diagnosis (clinical tumour status, clinical nodal status and tumour grade) by fitting a multivariate model. The proportional hazard assumption was verified for the Cox model using a formal statistical test based on the Schoenfeld residuals.[3] The p value for the test was calculated for original, log-transformed and rank time scales. Samples with unknown clinical factors were excluded. All tests of significance were two-sided and performed at a 5% alpha level. One way ANOVA was used to test the association between DDIR score and pathological response status defined by TRG."

- Further details of the Bioinformatics and Biostatistical methods are provided in Supplementary Methods within the Sections:
- Microarray Quality Control (QC)
- Analysis of Gene Expression Data
- Whole Genome Sequencing

State the method of case selection, including whether prospective or retrospective and whether stratification or matching (for example, by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.

References to various aspects of this criterion are presented in individual sections below.

Method of case selection, including whether prospective or retrospective

- "The OAC cohort comprised of 273 formalin fixed paraffin embedded collected at four centres in the UK (Belfast, Cambridge, Edinburgh and Southampton) as part of the Oesophageal Cancer Clinical and Molecular Stratification consortium."
- "FFPE pre-chemotherapy endoscopic biopsies from 273 patients with resectable OAC, treated with neo-adjuvant chemotherapy followed by surgical resection, were

collected at four UK centres in the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) consortium between 2003 and 2014"

- "For independent in silico validation a publically available dataset of 57 OAC resections which did not receive DNA-damaging chemotherapy (GSE19417) was assessed."

Whether stratification or matching (for example, by stage of disease or age) was used

- The cohort of cases were selected based on the following inclusion/exclusion criteria as opposed to a population based cohort. Inclusion criteria were T2-T3 NX M0 oesophageal adenocarcinomas treated by platinum-based neo-adjuvant chemotherapy and oesophagectomy and at least 2 years follow-up. Exclusion criteria were adenocarcinoma of the body/distal stomach and treatment with surgery alone.

- Supplementary Table 2 provides Demographic & Clinical characteristics of the Oesophageal Adenocarcinoma cohort.

Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.

- The cohort of cases were selected from 2003 to 2014 and had at least two years follow up.

Precisely define all clinical endpoints examined.

- "Pathological response was assessed in the matched resection specimens according to the method described by Mandard et al with a responder defined as Tumour Regression Grade (TRG) ≤ 2."

- “Relapse-free (RFS) and overall survival (OS) defined as the time from surgical resection to relapse of disease or death from any cause, respectively.”

List all candidate variables initially examined or considered for inclusion in models.

- “The estimated effect of the signature was adjusted for factors available at the time of diagnosis (clinical tumour status, clinical nodal status and tumour grade) by fitting a multivariate model”

Give rationale for samples size

- “Assuming a marker positive rate of 21% (estimated from preliminary data) a sample set of 273 patients had an 80% power to detect a Hazard Ratio (HR) of 2.”

Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.

- Samples with missing clinical factors were excluded from analysis.

The methods sections of main manuscript and Supplementary Methods detail the statistical methods

Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.

- “A threshold of 0.3403 was optimized in an independent technical study of n=45 OAC samples and applied independently to the validation cohort dichotomizing patients as DDIR positive (>0.3403) or DDIR negative (≤0.3403).”

- “DDIR scores were calculated using the DDIR assay parameters (derived from a regression based model) applied as previously described.”

Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the number of patients and the number of events.

- Supplementary Figure 1 includes the REMARK Study Design Flow Diagram outlining the flow of patients (including numbers) and the inclusion of patients for either analysis and/or QC failures at each critical stage of the process. "Comparison of the reporting of the DDIR assay as a predictive marker in oesophageal adenocarcinoma with the REMARK guidelines" (Supplemental Table 1 & Appendix A) and REMARK study design diagram (Supplementary Figure 1)."

Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumour marker, including numbers of missing values.

- Supplementary Tables 2-4 detail the distributions of demographic and clinical variables, for all datasets assessed in this study.

- Further description of the patient characteristics is provided within Supplementary Methods in the Section entitled 'Patient Samples'

Show the relation of the marker to standard prognostic variables.

- Table 1 & Supplementary Tables 9 & 10 show the relation and performance of the marker compared to standard prognostic variables.

- "Multivariable analysis was performed to test the association between DDIR status and each survival endpoint following adjustment for factors available at diagnosis (Table 2). DDIR positive patients had improved RFS relative to DDIR negative patients (HR 0.61, 95%CI 0.38-0.98; p= 0.042) and assay positivity was also independently associated with improved OS (HR 0.52, 95%CI 0.31-0.88; p= 0.015)."

Present univariable analysis showing the relation between the marker and outcome, with the estimated effect (for example, hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analysed. For the effect of a tumour marker on a time-to-event outcome, a Kaplan-Meier plot is recommended.

- Kaplan-Meier plots for univariate survival are presented in Figure 1 along with the associated Hazard ratios and 95% confidence intervals.

- Supplementary Table 5 also provides supporting univariate analysis of the marker used as a continuous predictor in the independent validation cohort.

For key multivariable analyses, report estimated effects (for example, hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.

- Multivariable analysis for the marker is reported in Table 2 with hazard ratios and confidence intervals clearly stated within.

Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.

- All Tables and Supplementary Tables included provide estimated effects and confidence intervals of the marker including standard prognostic factors, regardless of significance.

If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.

- "To assess whether the DDIR assay was prognostic, independent of DNA-damaging chemotherapy treatment, it was applied to a publically available dataset of 57 OAC resections which did not receive neo-adjuvant chemotherapy"

Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.

- "We have demonstrated that the DDIR assay is predictive of response and independently prognostic following DNA-damaging neo-adjuvant chemotherapy and surgical resection in OAC. DDIR assay positivity was associated with improved survival following chemotherapy and surgery and identified those patients with a higher probability of obtaining a pathological response, reduced nodal burden and clear resection margins. When assessed alongside clinical factors available at the time of diagnosis DDIR status demonstrated superior prognostic ability compared to standard clinicopathological factors. Application of the DDIR assay to a cohort of patients who did not receive neo-adjuvant therapy demonstrated no difference in survival according to DDIR status indicating that the DDIR assay may not be prognostic in its own right but only in the context of DNA-damaging therapy."

- "Limitations of the study include the use of a retrospective clinical cohort which may influence survival outcomes due to the absence of standardised follow-up procedures and so the DDIR assay will require further validation in a randomised controlled trial dataset and by a prospective study."

Discuss implications for future research and clinical value.

- "We have developed an array-based classifier using pre-treatment FFPE biopsies to predict benefit from, and response to, neo-adjuvant therapy in resectable OAC. The assay is readily applicable to routine pathological samples with potential for rapid translation into clinical use. The identification of a subgroup of tumours with deficiencies in their DNA repair mechanisms will enable these patients to be selected for more effective therapy and improve survival outcomes. Also, knowing the underlying biology of these tumours allows the possibility of further enhancing response to therapy through combinations with novel inhibitors of DNA repair and immunotherapy. Overall the DDIR assay enables treatment selection and patient stratification in esophago-gastric adenocarcinoma and may improve response to therapy, resection rates and survival in this poor prognostic disease."