Refining the selection of patients with metastatic colorectal cancer for treatment with temozolomide using proteomic analysis of O6-methylguanine-DNA-methyltransferase

Title: **Refining the selection of patients with metastatic colorectal cancer for treatment with temozolomide using proteomic analysis of MGMT**

Running title: **Temozolomide: proteomic MGMT selection**

Sarit Schwartz[a], MD, Chris Szeto[b], PhD, Yuan Tian[a], PhD, Fabiola Cecchi[a], PhD, Salvatore Corallo[c], MD, Maria Alessandra Calegari[d], MD, Maria Di Bartolomeo[c], MD, Federica Morano[c], MD, Alessandra Raimondi[c], MD, Giovanni Fucà[c], MD, Antonia Martinetti[c], MD, Ivana De Pascalis[e], PhD, Maurizio Martini[e], MD, Antonino Belfiore[f], PhD, Massimo Milione[f], MD, Armando Orlandi[d], MD, Ludovic Barault[g,h], PhD, Carlo Barone[d], Prof, Filippo De Braud[c,i], Prof, Federica Di Nicolantonio[g,h], PhD, Steve Benz[b], PhD, Todd Hembrough[a*], PhD, Filippo Pietrantonio[c,i*], MD,

[a] NantOmics, LLC, Rockville, MD, United States; [b] NantOmics, LLC, Santa Cruz, CA, United States; [c] Medical Oncology Department, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; [d] Medical Oncology Unit, Fondazione Policlinico Universitario Agostino Gemelli, Rome, Italy; [e] Institute of Pathology, Università Cattolica del Sacro Cuore, Rome, Italy; [f] Department of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; [g]Department of Oncology, c/o Candiolo Cancer Institute IRCC, University of Turin, Candiolo (TO), Italy; [h] Candiolo Cancer Institute, FPO - IRCCS, Candiolo (TO), Italy; [i] Department of Oncology and Hemato-oncology, University of Milan, Milan, Italy

***Corresponding authors:** Filippo Pietrantonio, MD, Medical Oncology Department, Fondazione IRCCS Istituto Nazionale dei Tumouri, Via Venezian 1, Milan, Italy, email: filippo.pietrantonio@istitutotumori.mi.it, phone: +39 0223903807;

Todd Hembrough, PhD, President, Proteomics, NantOmics, LLC, 9600 Medical Center Drive, Rockville, MD 20850, email: Todd.Hembrough@nantomics.com, phone: 301-502-3655

HIGHLIGHTS

- 10% O6-methylguanine-DNA-methyltransferase (MGMT)-methylated metastatic colorectal cancers respond to temozolomide (TMZ).
- 50/50/35% responses for MGMT selection with mass spectrometry/methyl-BEAMing/RNA-seq.
- Mass spectrometry: 100% sensitivity/50% specificity/80% accuracy for response to TMZ.
- Low/negative MGMT protein expression was significantly associated with longer progression-free survival.
- Quantitative proteomic MGMT analysis could refine patients' selection for TMZ.

ABSTRACT

Background The repair enzyme O6-methyl-guanine-DNA-methyl-transferase (MGMT) is a validated predictor of benefit from temozolomide (TMZ) in glioblastoma. However, only 10% of patients with *MGMT*-methylated metastatic colorectal cancer (mCRC) respond to TMZ.

Methods Archived tumour samples (N=41) from 3 phase II TMZ trials carried out in *MGMT* methylated mCRC (assessed by methylation-specific polymerase-chain-reaction) were stratified by MGMT status as assessed by 3 different methods: mass spectrometry, PCR/methylBEAMing, and RNA-seq. The performance of each method was assessed in relation to overall response rate, progression-free survival (PFS) and overall survival (OS).

Results Overall, 9 of 41 patients responded to TMZ. Overall response rates were 50% (9/18), 50% (6/12) and 35% (8/23) among patients determined likely to respond to TMZ by mass spectrometry, methylBEAMing and RNA-seq, respectively. Low/negative MGMT protein expressors by mass spectrometry had longer PFS than high MGMT expressors (3.7 versus (vs) 1.8 months;HR=0.50, p=0.014). Results for OS were similar but statistically non-significant (8.7 vs 7.4 months; HR=0.55, p=0.077). No significant association between survival and MGMT status by methyl-BEAMing or RNA-seq could be demonstrated as

comparable subgroups survival could not be confirmed/excluded. Specifically, the association of high vs low MethlBEAMing MGMT hypermethylation with survival was HR=0.783, p=0.46 for PFS and 0.591, p=0.126 for OS, while association of low vs high RNA-seq MGMT level with survival was HR=0.697, p=0.159 for PFS and HR=0.697, p=0.266 for OS.

Conclusions Quantitative proteomic analysis of MGMT may be useful for refining the selection of patients eligible for salvage treatment with single-agent TMZ.

Key words: Colorectal cancer, MGMT, Temozolomide, Biomarker, Molecular diagnostics

INTRODUCTION

Alkylating agents such as temozolomide (TMZ) are used to treat tumors since their ability to alkylate DNA causes DNA damage leading to tumour cell death. The repair enzyme O6-methyl-guanine-DNA-methyl-transferase (MGMT) is involved in response to DNA damage caused by alkylating agents[1,2]. *MGMT* gene expression is epigenetically downregulated by hypermethylation of the promoter of CpG dinucleotides. This transcriptional silencing leads to absence of MGMT protein, thus impeding repair of chemotherapy-induced O6-alkylguanine adducts and potentially enhancing tumour susceptibility to alkylating drugs[2,3]. *MGMT* methylation status has been validated as predictor of benefit from TMZ in glioblastoma patients[2,4-6]. *MGMT* silencing occurs in around 38% of colorectal carcinomas[7]. *MGMT* status as qualitatively assessed by methylation-specific polymerase-chain-reaction (MSP) was used to select patients with refractory metastatic colorectal cancer (mCRC) for 5 clinical trials of alkylating agents[8-13]. However, the activity of TMZ in heavily pre-treated mCRC patients selected by MSP is limited, with overall response rates about 3-16%. In attempts to improve the selection of mCRC patients, the predictive value of *MGMT* "hyper"-methylation as quantitatively assessed by digital PCR/methyl-BEAMing (MB) was demonstrated[14,15]. This analysis corroborated reported discrepancies between MGMT protein expression by

immunohistochemistry and alterations of matching genes in mCRC[16] and in other solid tumours[17,18]. Protein quantitation by mass spectrometry (MS) is widely considered the gold standard for biomarker measurement in biological samples[19-21]. MS-based assays can objectively quantify MGMT protein in formalin-fixed paraffin-embedded (FFPE) tumour tissues in an antibody-independent manner. Just as quantitative methylation overcomes the limitations of MSP (eg, subjectivity of eye reading of the gel, lack of automation), MS-based protein quantitation avoids challenges inherent in immunohistochemical detection of MGMT protein such as high inter-observer variability and lack of standard antibody types and scoring methods.

We hypothesized that tumour protein expression of MGMT as measured by MS would be a biomarker of resistance to TMZ and correlate with MGMT status by MB and by RNA sequencing (RNA-seq). We tested our hypothesis in the archived tumour samples of patients with refractory mCRC enrolled in 3 trials of TMZ[10,12,13] and we used predictive modelling to test which MGMT assessment method or their combination would most accurately identify patients responsible to TMZ.

MATERIALS AND METHODS

Patients and samples

This was a pooled analysis of archived tumour samples and clinical data from patients of 3 clinical trials of TMZ in refractory mCRC (EudraCT 2012-002766-13, INT 20/13 #1, INT 20/13 #2)[10,12,13]. Patients met the following inclusion criteria: histologically confirmed mCRC; *MGMT* gene promoter methylation detected by MSP; at least one measurable lesion as defined by Response Evaluation Criteria in Solid Tumours (RECIST) version 1.1[22]; disease progression during or after treatment with standard chemotherapy and/or EGFR inhibitor therapy; and Eastern Cooperative Oncology Group (ECOG) performance status ≤2. Patients received a standard TMZ regimen (150 mg/m2/day for 5 consecutive days every 28 days) or a dose-dense regimen (75 mg/m2/day, 21 days on/7 days off). Radiological assessments were conducted approximately every 8 weeks.

The present *post hoc* analysis included patients with an available archived tumour sample and treatment

outcome data permitting evaluation of objective response rate (ORR), progression-free survival (PFS) and overall survival (OS). All samples and clinical data were anonymized and this study was approved by the ethics committee at Fondazione IRCCS Istituto Nazionale dei Tumori of Milan in accordance with the declaration of Helsinki. All patients had provided written informed consent to research use of their anonymized data.

Quantitative MGMT assessment

MGMT status was assessed by 3 methods: gene promoter methylation by MB, protein expression by MS, and messenger RNA (mRNA) expression by RNA-seq. Methylation status was performed at IRCC Candiolo, Turin, Italy as previously described.[14] Briefly, extracted and amplified DNA products from PCR were diluted and re-amplified with emulsion PCR. Following emulsion breaking and hybridization, fluorescence was assessed via flow cytometry; the percentage of methylation was calculated as the ratio of the fluorescence from the methylated probe over the sum of methylated and unmethylated probe signals. MGMT hyper-methylated status was defined as >63% cut-off[15].

MGMT protein was quantified with an MS assay as previously described[23]. Briefly, tumour areas of archived formalin-fixed, paraffin-embedded tissue sections were marked by a pathologist and microdissected using a non-contact laser method. The captured tumour cells were solubilized to tryptic peptides and the total protein concentration of each tryptic peptide mixture was measured. Each sample was subjected to triplicate proteomic analysis using stable isotope-labeled internal standard peptides for quantitation of analytical targets. Proteomic expression analysis was performed with a TSQ Quantiva™ triple quadrupole mass spectrometer (Thermo Scientific, San Jose, CA)[24]. Data analysis was performed using Pinpoint™ (Version 1.3; Thermo Scientific, San Jose, CA) and Pinnacle software (Optys Tech Corporation, PA).

Patients were stratified into groups of “low” and “high” MGMT protein expression using a pre-specified threshold for MGMT of 200 attomoles per microgram (amol/µg) of total protein, based on the proteomic assay’s limit of quantitation, which is determined from analyses of assay performance with respect to sensitivity and reproducibility.

RNA-seq was conducted by NantOmics as follows: RNA-Seq libraries were prepared for the tumour sample using KAPA Stranded RNA-Seq with RiboErase kit and sequenced on the Illumina sequencing platform. The resulting reads were aligned to refseq build 73 using BowTie2 v2.2.6, then processed by RSEM v.1.2.25[25] to estimate transcripts per kilobase million (TPM) and fragments per kilobase of exon per million fragments mapped (FPKM) for each isoform. Gene-level TPM and FPKM estimates are made using a weighted-average of the isoform estimates, weighted by an RSEM-estimated percentage of each isoform's expression among all isoforms in the sample.

As a predetermined threshold for mRNA expression that correlates with response to TMZ has not been established, mRNA expression levels of all TCGA samples of colon and rectal cancers were plotted to find a natural break in the expression pattern that would match with the proteomic cutoff of 200 amol/µg. The distribution of MGMT TPM appeared bimodal with a natural break at 3.5 log2(TPM+1). This threshold was highly associated with the proteomic threshold (Fisher's exact test $p<0.0008$) and was determined to be the optimal value for agreement between RNA-seq and proteomic values in Youden analysis. Expression levels of MGMT mRNA below this cutoff were considered indicative of likely response to TMZ.

IHC analysis for MGMT was performed and scored as previously described [12].

Geneset analysis

In attempts to annotate the MGMT observed in this cohort with functional biological pathways or ontologies, genes significantly associated with either MGMT protein subgroups or MGMT mRNA expression were analyzed using gene-set enrichment analysis (GSEA). A total of 5 curated geneset databases were used in the analysis: KEGG, GO Molecular Functions, GO Biological Processes, BioCarta, & Wikipathways. Genes associated with MGMT protein levels were identified by two-sample t-tests in gene expression between MGMT high (>200 amol/µg) vs. MGMT low (<200 amol/µg) subgroups. As no genes were significant after Bonferroni correction (adjusting for 19,270 hypotheses), a minimum p-value of 0.01 was used. Genes associated with MGMT mRNA expression were identified by correlation analysis, wherein the minimum R value for significance was 0.82 (one-sided, α=2.6e-6, β=0.99).

Statistical analysis

The performance of each MGMT assay was assessed using 3 patient endpoints: ORR according to RECIST version 1.1, PFS and OS. Cox proportional hazard modelling and the Mantel-Cox log-rank were used for survival comparisons. The Fisher's exact test was used to assess the relationship between MGMT status and patient response and the correlation between MGMT assessment methods.

Predictive modelling

The ability of the 3 MGMT assays (MS, MB and RNAseq) to predict patient response to TMZ was tested using leave-pair-out cross-validation. A predictive model was built using all samples except 2, and the model's performance was tested in one unseen positive sample and one unseen negative sample. This was repeated for all possible combinations of positive and negative samples. The average performance over all unseen test sets was the reported accuracy for a given predictive model.

RESULTS

Patients and samples

Tumour samples from 41 TMZ-treated patients were available for analysis. These patients had a median age of 69 years and had received a median of 3 chemotherapeutic regimens prior to TMZ. Most patients had an ECOG status of 0 or 1 (85%); and at least 2 metastatic sites (56%), with liver as the most frequent. As expected in mCRC, all patients eventually progressed on TMZ. Twenty-six patients (63%) had progressive disease, 9 (22%) partial response, 6 (15%) stable disease (Table 1).

MGMT status

All 41 archived samples were evaluable by MS and IHC; 35 were analyzed by MB and 39 were of sufficient quality for MGMT assessment by RNA-seq (Figure 1). Of patients assessed by MS-based proteomics, 18 (44%) had "low" tumour expression of MGMT protein (<200 amol/µg of tumour protein) and were therefore considered likely to respond to TMZ. The remainder (n=23) were "high" protein expressors prone to TMZ resistance. As expected in this population of patients enriched for the study of exceptional responders, low MGMT protein expression was relatively frequent (44%); by comparison, the prevalence of low MGMT expression among all samples of CRC analyzed in the authors' clinical laboratory during the past

year (n=104) was 14% (Table 2). Among MGMT "low" subgroup, no significant association with specific clinico-pathological features was observed when comparing responders versus non responders to TMZ (data not shown).

MGMT promoter methylation above the previously validated 63% cutoff was observed in 12 (34%) patients. In the 35 tumours analyzed by MB and MS, the agreement rate between methods was 77%; p=0.004. Using the experimental cutpoint for mRNA expression fit to the data (≤3.5 log2[TPM+1]), low mRNA expression was observed in the majority of samples (n=23; 59%) (Table 2). In the 39 patients analyzed for MS and RNA-seq, the agreement rate was 77%; p=0.0008.

IHC for MGMT was scored as negative in 4 (10%) samples, weakly positive in 8 (20%) and intense positive in the remaining 29 (70%). Even if IHC and MS analyses results showed a significant correlation (p=0.0003 by Chi-square test), 7 patients classified as MGMT negative by means of MS had intense MGMT expression by means of IHC.

Response and survival

Quantitative proteomics retrospectively identified 9 of 9 RECIST-defined responders to TMZ; all 9 responders had low MGMT protein levels by MS. Other 9 patients with low MGMT protein expression did not have RECIST-defined response on TMZ (ORR of low MGMT protein: 50%). None of the patients with high MGMT protein responded to TMZ (ORR of high MGMT protein: 0%; p=0.0001) (Table 2; Figure 2A**)**.

Positive MGMT methylation status by MB retrospectively identified 6 of 8 responders to TMZ; other 6 patients with positive MGMT status by MB were non-responders (ORR of *MGMT* hypermethylation: 50%). Two patients with negative methylation status responded to TMZ (ORR: 9%; p=0.011) (Table 2; Figure 2B).

Patients with low mRNA-expressing tumours by RNA-seq had a non-significantly higher ORR than higher mRNA expressors (35% vs 6%; p=0.115) (Table 2; Supplementary Figure 1).

In survival analyses, patients with low MGMT protein levels (<200 amol/µg) had longer median PFS (mPFS) than patients with high MGMT levels (3.7 vs 1.8 months; *p*=0.014) (Figure 3A). MGMT levels remained a statistically significant predictor of PFS when paired with other prognostic factors in Cox proportional hazards models; no other variable tested was more explanatory than MGMT protein level (Table 3).

Differences in OS by MGMT protein level were similar to PFS differences but did not reach statistical significance (8.7 vs 7.4 months, HR=0.55, *p*=0.077) (Figure 3B). There were no statistically significant differences in PFS or OS among patients stratified by MB or RNA-seq (Figure 3C, 3D, 3E and 3F).

Geneset analysis

RNA-seq identified 48 genes that were strong indicators of MGMT protein subgroups, and a further 30 genes that were significantly correlated with MGMT mRNA expression. In GSEA, neither of these gene sets were significant drivers of DNA damage response or other pathways associated with response to TMZ.

Predictive modelling

Predictive models of TMZ response were built using each of the MGMT assays (MS, MB and RNA-seq) and for combinations. These models were run using their established cutoffs as well as raw, continuous MGMT values with various experimental cutoffs. Ten datasets and 14 classification algorithms combined into 140 different modelling strategies; evaluating the predictive performance of these strategies in unseen samples required building an additional 2,772 unique predictive sub-models. The best modelling strategy included all three MGMT assessment methods (MB, MS and RNA-seq) with their established cutoffs. This model was 87% accurate in predicting TMZ response in unseen samples and performed better than that using MGMT protein quantity alone (80% accurate) (Figure 4). For each of the 3 assessment methods, experimental MGMT cutoffs (optimized using leave-out pair cross validation) did not perform better than the predefined cutoffs (Figure 5).

DISCUSSION

In this *post-hoc* pooled analysis of 3 phase II trials in refractory mCRC patients receiving TMZ, a proteomic test for MGMT protein had a 100% sensitivity and a 50% specificity when using clinical response as the gold standard. Although the sample size was too small to reach definitive conclusions, the proteomic test seemed to outperform both digital MB and RNA-seq in predicting response to TMZ. MGMT protein expression below a predefined threshold was significantly associated with longer mPFS, independently from other prognostic variables. Regarding MB and RNA-seq tests, no significant association with survival

could be demonstarted since a comparable survival of subgroups could not be confirmed or excluded, possibly for the limited study power. Patients with high MGMT protein expression had similar PFS to that reported for mCRC patients in clinical trials of TMZ. Therefore, the disappointing results of such trials may reflect the limited ability of standard *MGMT* assessment methods (e.g. MSP) to select the optimal candidates for TMZ.

In order to develop a hypothetical MGMT assay with maximal accuracy in identifying responders to TMZ, a predictive model was built using 3 different MGMT assays and their combinations. A combination of all 3 MGMT assessment methods was 87% accurate versus 80% accuracy using the proteomic test alone. Modelling confirmed that the thresholds for MGMT expression and MGMT methylation used to stratify patients in this study were more robust than other exploratory thresholds. These results point to the potential clinical value of MGMT protein quantitation, either alone or in combination with other methods. MS technique is also valuable for two reasons: first, it seems to outperform IHC since that the sensitivity of IHC may be not sufficient to categorize MGMT negativity. Second, MS may allow the selective detection of the active form of MGMT protein. In fact, the alkylated unactive form of MGMT is rapidly cleared by ubiquitine-mediated proteasomal degradation following conformational changes.

Concerning the discrepancies between RNA and MS, most of the discordant cases demonstrated a silencing at the RNA level while protein was found at high level. This could be explained by a slow turn-over of the protein in absence of DNA damage. In fact, in absence of DNA alkylation, the cells might switch off the transcription of MGMT, which will not affect the protein level already available. Additional process of transcription regulation might also be involved, such as deregulation of UBR1, a protein ligase E3 acting proved to affect MGMT transcription level [REF Leng Cancer Res 2015].

The importance of identifying potential responders to TMZ is emphasized by recently published findings of impairment of DNA mismatch repair or hypermutated status after the emergence of acquired resistance to TMZ in mCRC patients, thus becoming potentially eligible for immune-checkpoint inhibitors.[26] Studies in microsatellite stable mCRC are investigating the optimal duration of TMZ therapy prior to tumor mutational burden (TMB) testing, as well as “priming” treatment with TMZ followed, at the time of TMB-high-

associated disease progression, by sequential PD-1 blockade (ARETHUSA trial, NCT03519412) or short-term induction treatment with TMZ followed, in absence of disease progression, by its combination with CTLA-4 plus PD-1 blockade (MAYA trial). In parallel, our recent work suggested a potential synergy between TMZ and other active agents commonly used in mCRC, such as irinotecan (TEMIRI regimen), with novel translational data regarding molecular selection at both gene and protein level[27].

Methylation-mediated silencing of MGMT has been reported in 38% of mCRC[7] and the frequency of low MGMT protein expression in our study is similarly encouraging. Of 104 samples of CRC analyzed in our clinical laboratory during the past year, 15 (14%) underexpressed MGMT protein, thus likely to respond to TMZ, and this percentage is similar to response rates to alkylating agents in refractory mCRC[8-13].

In this study, about a half of patients with low MGMT expression failed to respond to TMZ, suggesting a role for other factors such as DNA damage repair, cell cycle, and immune profile. Indeed, other transcriptional and post-transcriptional processes might be involved in *MGMT* expression in CRC[28,29]. An analysis of 70 genes with known involvement in DNA damage repair and immune-mediated response failed to find differential gene expression in MGMT subgroups. Future studies may identify genetic signatures that could further refine predictions of response to TMZ.

This study has some clear limitations. The absence of a control group treated without TMZ leaves open the possibility that the investigated biomarkers may be prognostic rather than predictive. Moreover, there is evidence that MGMT status may change during the course of disease[8], limiting the reliability of data from tumour tissue obtained at the time of diagnosis. Finally, the absence of prospective validation of our results limits their current use outside a research setting. Of note, a phase II trial (NCT02414009) led by our Group is currently enrolling patients with *MGMT*-methylated – as assessed by MSP – and *RAS* mutated mCRC, who failed a previous oxaliplatin-based treatment, randomly allocated to either second-line FOLFIRI regimen or capecitabine plus TMZ (CAPTEM regimen). This trial has almost concluded its target enrollment and will give us the chance to validate the potential predictive utility of our MGMT-centered panel of biomarkers.

Despite these limitations, the results of this preliminary study support the ability of a proteomic MGMT assay to refine the selection of TMZ responders and suggest that quantitated MGMT protein may be a useful biomarker in clinical settings.

Conflict of interest: Sarit Schwartz is a former employee of NantOmics. Chris Szeto, Fabiola Cecchi, Yuan Tian, Steve Benz and Todd Hembrough are employees of NantOmics. Filippo Pietrantonio has received consultant/advisory board fees from Roche, Amgen, Eli Lilly, Sanofi, Merck-Serono and Bayer. Maria Di Bartolomeo has received consultant/advisory board fees from Eli Lilly. Filippo De Braud has received consultant/advisory board fees from Roche, Novartis, Amgen, and Celgene. The remaining authors declare no conflict of interest.

Funding: Part of this work was supported by grants AIRC IG n. 17707 (FDN), and AIRC Special Program 5 per mille Metastases Project n 21091 (FDN); Fondo per la Ricerca Locale (ex 60%), Università di Torino, 2017 (FDN); the Fondazione Piemontese per la Ricerca sul Cancro –ONLUS 5 per mille 2015 Ministero della Salute (FDN); grant Fondazione Piemontese per la Ricerca sul Cancro-ONLUS 5 per mille 2011 Ministero della Salute (FDN); RC 2017 Ministero della Salute (FDN and LB). Ludovic Barault was the recipient of a MIUR-cofunded postdoctoral 'Assegno di Ricerca' from the University of Torino in 2018.

Acknowledgements: Ellen Wertheimer, an employee of NantOmics, provided medical writing assistance.

Availability of data and material: Data supporting the results of this article are available from the authors upon request.

REFERENCES

[1] Pegg AE, Mammalian O6-alkylguanine-DNA alkyltransferase: regulation and importance in response to alkylating carcinogenic and therapeutic agents. Cancer Res 1990;50(19):6119-29.

[2] van Nifterik KA, van den Berg J, van der Meide WF, et al., Absence of the MGMT protein as well as methylation of the MGMT promoter predict the sensitivity for temozolomide. Br J Cancer 2010;103(1):29-35.

[3] Esteller M, Toyota M, Sanchez-Cespedes M, et al., Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is associated with G to A mutations in K-ras in colorectal tumorigenesis. Cancer Res 2000;60(9):2368-71.

[4] Dunn J, Baborie A, Alam F, et al., Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. Br J Cancer 2009;101(1):124-31.

[5] Hegi ME, Diserens AC, Gorlia T, et al., MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 2005;352(10):997-1003.

[6] Weller M, Tabatabai G, Kastner B, et al., MGMT Promoter Methylation Is a Strong Prognostic Biomarker for Benefit from Dose-Intensified Temozolomide Rechallenge in Progressive Glioblastoma: The DIRECTOR Trial. Clin Cancer Res 2015;21(9):2057-64.

[7] Esteller M, Hamilton SR, Burger PC, Baylin SB, Herman JG, Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. Cancer Res 1999;59(4):793-7.

[8] Amatu A, Barault L, Moutinho C, et al., Tumor MGMT promoter hypermethylation changes over time limit temozolomide efficacy in a phase II trial for metastatic colorectal cancer. Ann Oncol 2016;27(6):1062-7.

[9] Amatu A, Sartore-Bianchi A, Moutinho C, et al., Promoter CpG island hypermethylation of the DNA repair enzyme MGMT predicts clinical response to dacarbazine in a phase II study for metastatic colorectal cancer. Clin Cancer Res 2013;19(8):2265-72.

[10] Calegari MA, Inno A, Monterisi S, et al., A phase 2 study of temozolomide in pretreated metastatic colorectal cancer with MGMT promoter methylation. Br J Cancer 2017;116(10):1279-86.

[11] Hochhauser D, Glynne-Jones R, Potter V, et al., A phase II study of temozolomide in patients with advanced aerodigestive tract and colorectal cancers and methylation of the O6-methylguanine-DNA methyltransferase promoter. Mol Cancer Ther 2013;12(5):809-18.

[12] Pietrantonio F, de Braud F, Milione M, et al., Dose-Dense Temozolomide in Patients with MGMT-Silenced Chemorefractory Colorectal Cancer. Target Oncol 2016;11(3):337-43.

[13] Pietrantonio F, Perrone F, de Braud F, et al., Activity of temozolomide in patients with advanced chemorefractory colorectal cancer and MGMT promoter methylation. Ann Oncol 2014;25(2):404-8.

[14] Barault L, Amatu A, Bleeker FE, et al., Digital PCR quantification of MGMT methylation refines prediction of clinical benefit from alkylating agents in glioblastoma and metastatic colorectal cancer. Ann Oncol 2015;26(9):1994-9.

[15] Sartore-Bianchi A, Pietrantonio F, Amatu A, et al., Digital PCR assessment of MGMT promoter methylation coupled with reduced protein expression optimises prediction of response to alkylating agents in metastatic colorectal cancer patients. Eur J Cancer 2017;71:43-50.

[16] Shen L, Kondo Y, Rosner GL, et al., MGMT promoter methylation and field defect in sporadic colorectal cancer. J Natl Cancer Inst 2005;97(18):1330-8.

[17] Castillo-Martin M, Thin TH, Collazo Lorduy A, Cordon-Cardo C, Immunopathologic Assessment of PTEN Expression. Methods Mol Biol 2016;1388:23-37.

[18] Ilie MI, Bence C, Hofman V, et al., Discrepancies between FISH and immunohistochemistry for assessment of the ALK status are associated with ALK 'borderline'-positive rearrangements or a high copy number: a potential major issue for anti-ALK therapeutic strategies. Ann Oncol 2015;26(1):238-44.

[19] Jemal M, High-throughput quantitative bioanalysis by LC/MS/MS. Biomed Chromatogr 2000;14(6):422-9.

[20] Ong SE, Foster LJ, Mann M, Mass spectrometric-based approaches in quantitative proteomics. Methods 2003;29(2):124-30.

[21] Picotti P, Aebersold R, Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 2012;9(6):555-66.

[22] Eisenhauer EA, Therasse P, Bogaerts J, et al., New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45(2):228-47.

[23] Hembrough T, Thyparambil S, Liao WL, et al., Application of selected reaction monitoring for multiplex quantification of clinically validated biomarkers in formalin-fixed, paraffin-embedded tumor tissue. J Mol Diagn 2013;15(4):454-65.

[24] Catenacci DV, Liao WL, Thyparambil S, et al., Absolute quantitation of Met using mass spectrometry for clinical application: assay precision, stability, and correlation with MET gene amplification in FFPE tumor tissue. PLoS One 2014;9(7):e100586.

[25] Li B, Dewey CN, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323,2105-12-323.

[26] Germano G, Lamba S, Rospo G, et al., Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. Nature 2017;552(7683):116-20.

[27] Morano F, Corallo S, Niger M, et al., Temozolomide and irinotecan (TEMIRI regimen) as salvage treatment of irinotecan-sensitive advanced colorectal cancer patients bearing MGMT methylation. Ann Oncol 2018.

[28] Zhang L, Zeng J, Zeng Z, et al., MGMT in colorectal cancer: a promising component of personalized treatment. Tumour Biol 2016;37(8):11443-56.

[29] Zhao W, Soejima H, Higashimoto K, et al., The essential role of histone H3 Lys9 di-methylation and MeCP2 binding in MGMT silencing with poor DNA methylation of the promoter CpG island. J Biochem 2005;137(3):431-40.

FIGURE LEGENDS

Figure 1. Consort diagram of the translational study. Archived FFPE tissue sections were obtained from patients who had received TMZ in one of 3 Phase II clinical trials. 41 samples were evaluable by proteomics, 35 were analyzed by digital MB and 39 were analyzed by RNA-seq.

Figure 2. Percent change in sum of longest RECIST diameters (from baseline) among TMZ-treated patients (n=41) by (A) MGMT protein status and (n=35) (B) *MGMT* promoter hypermethylation status.

Figure 3. PFS (A) and OS (B) of TMZ-treated patients with metastatic colorectal cancer, by MGMT protein expression level, PFS (C) and OS (D) stratified by MGMT methylation status and PFS (E) and OS (F), by MGMT mRNA level (RNA-seq).

Figure 4. Average accuracy of predictive models per leave-pair-out cross-validation. Two classification strategies were employed: predefined cutoffs and exploratory cutoffs determined as optimum in a training set. Predefined and exploratory cutoffs were assessed in the exact same training and testing sets for direct comparison.

Figure 5. Average predictive accuracy in unseen samples for 58 predictive modelling strategies, by MGMT assessment method group. Groups are ordered left-to-right by average accuracy. Groups labeled "predefined" are discretized by their predefined cutoffs prior to predictive modelling. Each point represents a different predictive modelling strategy (i.e., combination of MGMT assessment method group and classification algorithm). Univariate datasets were analyzed using only Youden analysis and predefined cutoffs. Multivariate datasets were used as input for all other classification algorithms shown. Although

prediction strategies that use all three MGMT assessment methods outperformed the univariate proteomic cutoff, the accuracy in the proteomic data is the most robust (lowest data dispersion) in this small cohort.

Supplementary Figure 1

Percent change in sum of longest RECIST diameters (from baseline) among TMZ-treated patients (n=39) by MGMT RNA level.

Table 1. Characteristics of TMZ-treated patients with metastatic colorectal cancer (n=41)

Characteristic	**No (%)**
Sex	
M	20 (49)
F	21 (51)
Age	
Median (range)	69 (48-85)
Clinical Trial	
INT 20/13 #1	13 (27)
INT 20/13 #2	11 (32)
EudraCT 2012-002766-13	17 (41)
ECOG performance status	
0	15 (37)
1	20 (49)
2	6 (14)
RAS and BRAF mutational status	
Wild type (RAS and BRAF)	18 (44)
KRAS mutated	19 (46)
BRAF mutated	4 (10)
Primary tumor location	
Right-sided colon	18 (44)
Left-sided colon	20 (49)
Rectum	3 (7)
No. of metastatic sites	
1 metastatic site	18 (44)
≥ 2 metastatic sites	23 (56)
Sites of metastases	
Liver	32 (78)
Lung	24 (58)
Peritoneum	6 (15)
No. of previous treatments	
Median (range)	3 (2-5)
Objective best response rate (RECIST)	
PR	9 (22)
SD	6 (15)
PD	26 (63)

Abbreviations: TMZ, temozolomide; M, male; F, female; No, number; RECIST, Response Evaluation Criteria in Solid Tumors; ECOG, Eastern Cooperative Oncology Group; PR, partial response; SD, stable

disease; PD, progressive disease.

Table 2. Overall response rate (ORR) of TMZ-treated patients by MGMT status as assessed by 3 methods: mass spectrometry-based proteomics, methylBEAMing, and RNA-seq.

Assessment method/status	**N (%)**	**ORR**	***p****
MGMT protein (N=41)			
<200 amol/ug	18 (44)	50%	*0.0001*
≥200 amol/ug	23 (56)	0%	
MGMT hypermethylation (N=35)			
>63%	12 (34)	50%	*0.011*
≤63%	23 (66)	9%	
MGMT RNA-seq (N=39)			
<3.5 log2[TPM+1]	23 (59)	35%	*0.115*
>3.5 log2[TPM+1]	16 (41)	6%	

Abbreviations: MGMT, O6-methylguanine-DNA-methyltransferase; TPM, transcripts per million; N, number; ORR, overall response rate.

p* two-tailed Fisher's exact test

Table 3: Analyzing potential confounders to MGMT association with PFS: the p-values are associated with the explanatory coefficient for each potential confounder in the presence of MS MGMT status in a bivariate Cox-proportional hazard model for PFS. Neutrophil/lymphocyte at baseline and LDH were explored as continuous variables. Age is defined as the years elapsed between birth and date at histological diagnosis. The table is sorted in order of likelihood to be a confounder.

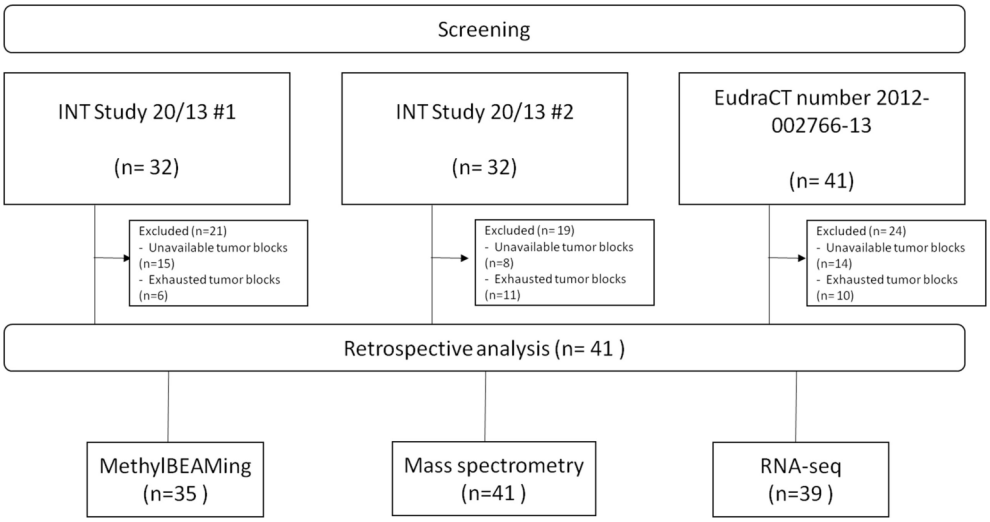
	Confounder p-value	**MGMT adjusted p-value**
BRAF mutation	0.137	0.018
KRAS mutation	0.203	0.020
Gender	0.424	0.032
ECOG	0.202	0.038
n° of previous treatment	0.910	0.014
LDH baseline level	0.017	0.005
n° metastatic sites	0.083	0.004
Neutrophil/lymphocyte at baseline	0.149	0.046
Peritoneal disease	0.219	0.012
Primary tumor location	0.381	0.026
Site of the archived tissue	0.631	0.019
Age	0.855	0.022

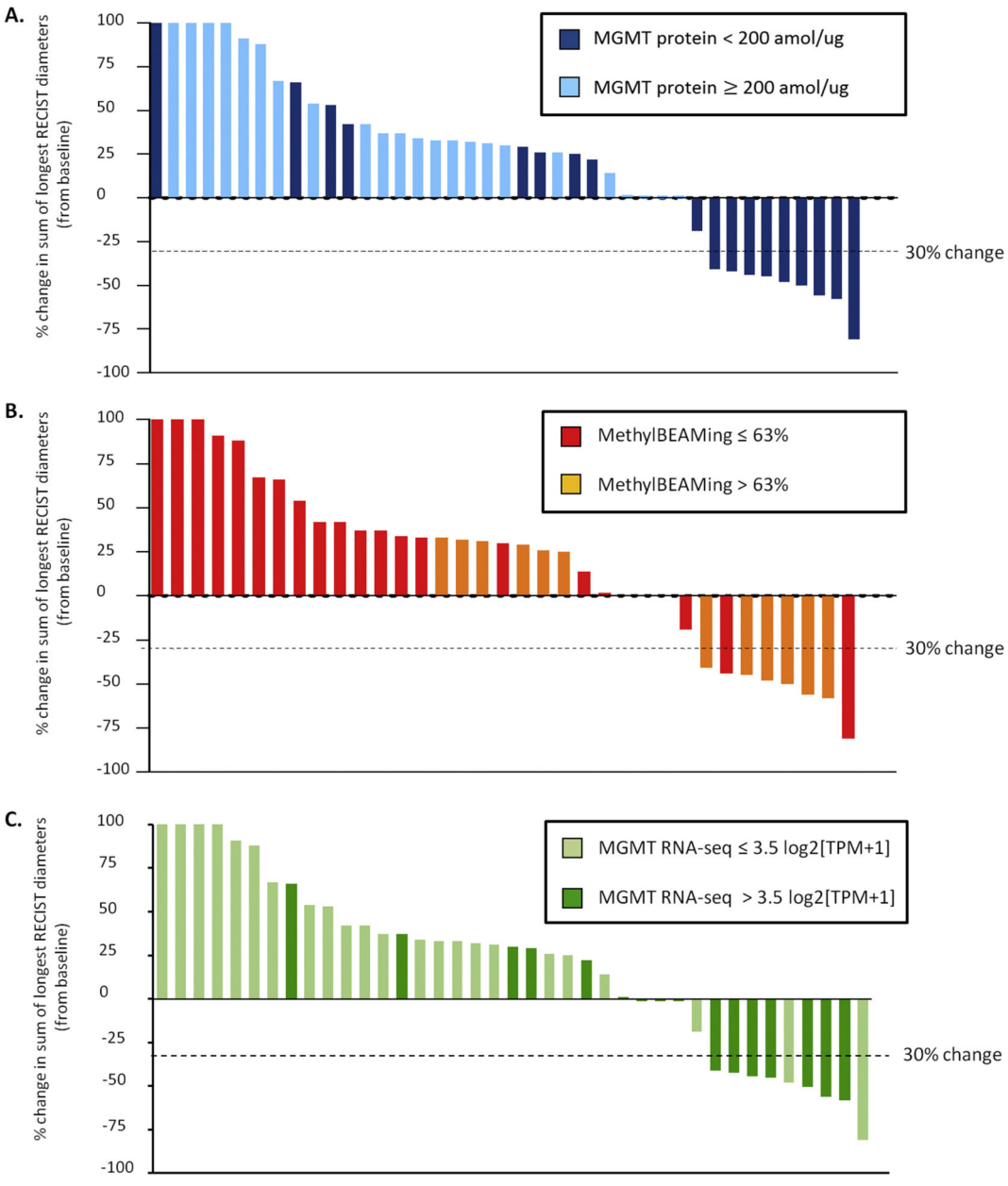
Abbreviations: MGMT, O6-methylguanine-DNA-methyltransferase; ECOG, Eastern Cooperative Oncology Group; n°, number.

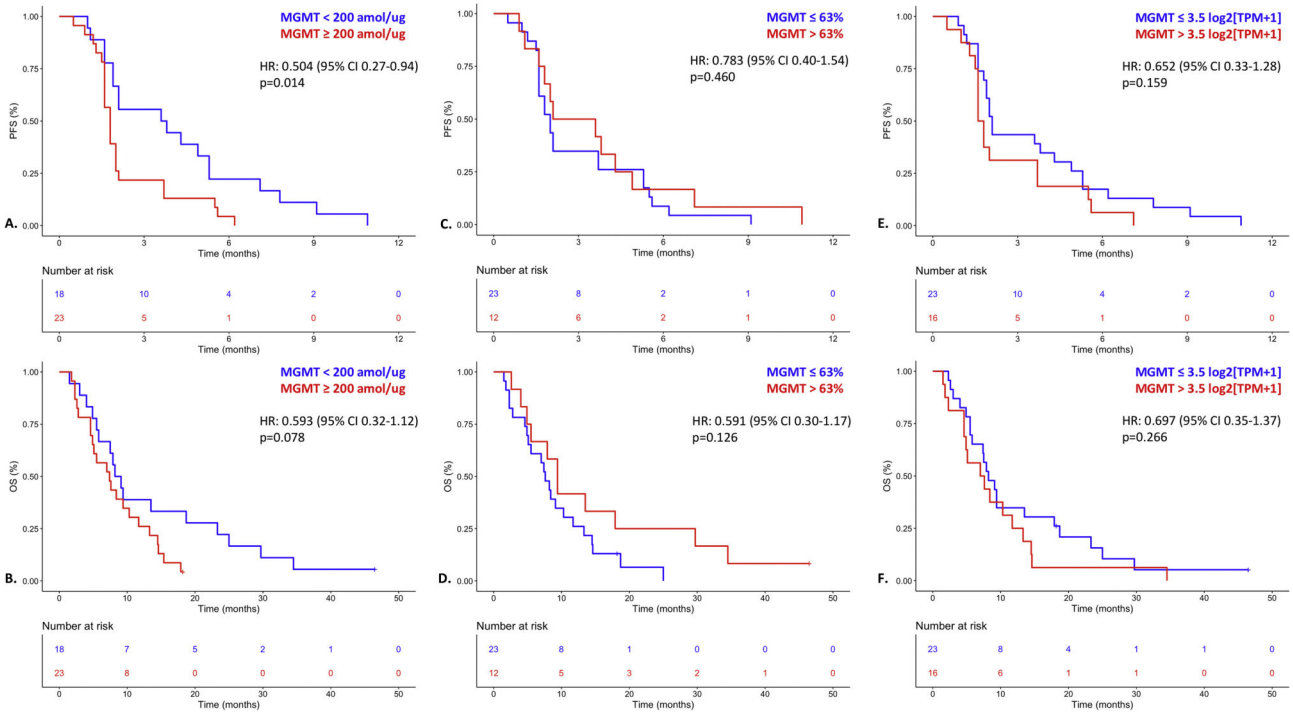
1

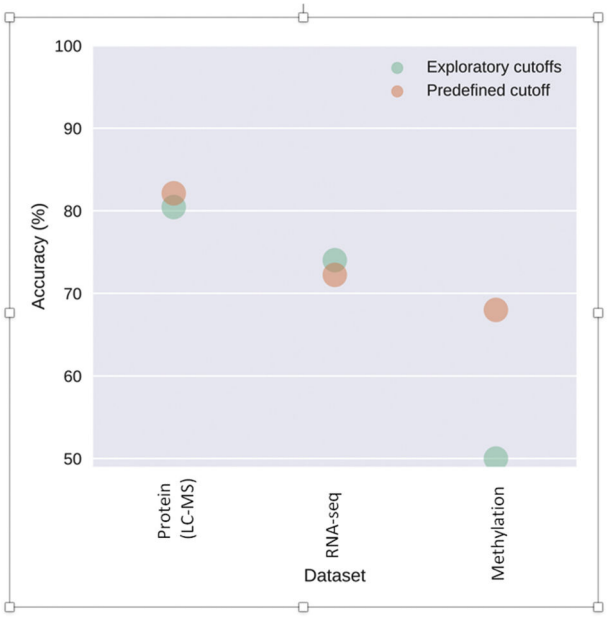
2

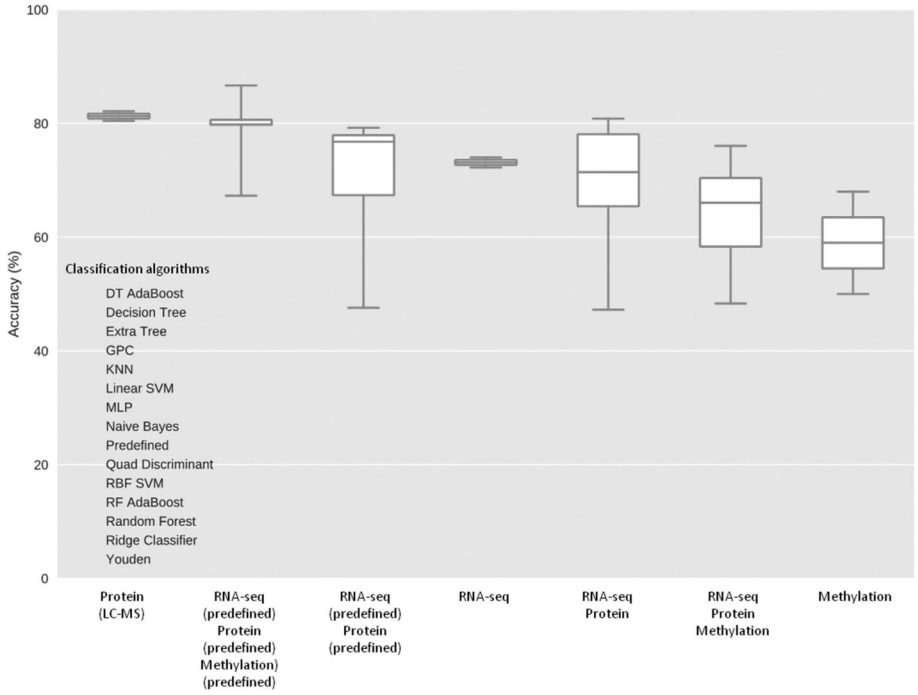
3
4


Screening
INT Study 20/13 #1
(n= 32)
INT Study 20/13 #2
(n= 32)
EudraCT number 2012-002766-13
(n= 41)
Excluded (n=21)
- Unavailable tumor blocks (n=15)
- Exhausted tumor blocks (n=6)
Excluded (n= 19)
- Unavailable tumor blocks (n=8)
- Exhausted tumor blocks (n=11)
Excluded (n= 24)
- Unavailable tumor blocks (n=14)
- Exhausted tumor blocks (n= 10)
Retrospective analysis (n= 41)
MethylBEAMing
(n=35)
Mass spectrometry
(n=41)
RNA-seq
(n=39)


A.
% change in sum of longest RECIST diameters (from baseline)
100
75
50
25
0
-25
-50
-75
-100
MGMT protein < 200 amol/ug
MGMT protein ≥ 200 amol/ug
30% change
B.
% change in sum of longest RECIST diameters (from baseline)
MethylBEAMing ≤ 63%
MethylBEAMing > 63%
30% change
C.
% change in sum of longest RECIST diameters (from baseline)
MGMT RNA-seq ≤ 3.5 log2[TPM+1]
MGMT RNA-seq > 3.5 log2[TPM+1]
30% change


A.
MGMT < 200 amol/ug
MGMT ≥ 200 amol/ug
HR: 0.504 (95% CI 0.27-0.94)
p=0.014
PFS (%)
Time (months)
Number at risk
18 10 4 2 0
23 5 1 0 0
B.
MGMT < 200 amol/ug
MGMT ≥ 200 amol/ug
HR: 0.593 (95% CI 0.32-1.12)
p=0.078
OS (%)
Time (months)
Number at risk
18 7 5 2 1 0
23 8 0 0 0 0
C.
MGMT ≤ 63%
MGMT > 63%
HR: 0.783 (95% CI 0.40-1.54)
p=0.460
PFS (%)
Time (months)
Number at risk
23 8 2 1 0
12 6 2 1 0
D.
MGMT ≤ 63%
MGMT > 63%
HR: 0.591 (95% CI 0.30-1.17)
p=0.126
OS (%)
Time (months)
Number at risk
23 8 1 0 0 0
12 5 3 2 1 0
E.
MGMT ≤ 3.5 log2[TPM+1]
MGMT > 3.5 log2[TPM+1]
HR: 0.652 (95% CI 0.33-1.28)
p=0.159
PFS (%)
Time (months)
Number at risk
23 10 4 2 0
16 5 1 0 0
F.
MGMT ≤ 3.5 log2[TPM+1]
MGMT > 3.5 log2[TPM+1]
HR: 0.697 (95% CI 0.35-1.37)
p=0.266
OS (%)
Time (months)
Number at risk
23 8 4 1 1 0
16 6 1 1 0 0


100
90
80
70
60
50
Accuracy (%)
Exploratory cutoffs
Predefined cutoff
Protein (LC-MS)
RNA-seq
Methylation
Dataset


100
80
60
40
20
0
Accuracy (%)
Classification algorithms
DT AdaBoost
Decision Tree
Extra Tree
GPC
KNN
Linear SVM
MLP
Naive Bayes
Predefined
Quad Discriminant
RBF SVM
RF AdaBoost
Random Forest
Ridge Classifier
Youden
Protein (LC-MS)
RNA-seq (predefined) Protein (predefined) Methylation) (predefined)
RNA-seq (predefined) Protein (predefined)
RNA-seq
RNA-seq Protein
RNA-seq Protein Methylation
Methylation

Supplementary Table 1. The 50 most differentially expressed genes in three different subgroupings: responders versus non-responders, MGMT high versus low by RNA-seq predefined cutoff, and MGMT high versus low by MS predefined cutoff. None of these genes were significantly differentially expressed after multiple-hypothesis correction, but were used as the basis for GSEA to annotate function.

Rank	Response geneset	RNA-seq geneset	MS geneset
1	KIR3DS1	TNN	FRG2C
2	FFAR3	KCND3	MGMT
3	FOXL2	HAPLN4	IRS2
4	CALB1	SLC2A9	GJC3
5	SSX3	MAN1C1	RTBDN
6	CRB2	TRABD2B	RLN2
7	GABRB2	KDM8	KCNB1
8	OR5AR1	UGT2B17	PER1
9	HSPA4L	EXOC3L4	TMEM156
10	CAPN14	PAOX	MROH8
11	NPPC	IGSF23	ZNF556
12	TFF3	FRG2C	DUSP1
13	URGCP-MRPS24	SLC16A11	CHN2
14	MAGEA12	RXRA	KLF9
15	PRLH	SLITRK3	NDST3
16	ODF4	ADH6	ERN1
17	MCHR1	GNAO1	RAD9B
18	KRT74	LILRB5	SHD
19	PKD1L1	PDE6G	NR1I2
20	GOLGA8J	GREM2	ADAMTS17
21	IL36B	DCANP1	ATP6V0D2
22	ARSH	SHC2	AFMID
23	TRIM69	ABCG2	CYP4F12
24	FCAMR	SLC16A2	ZNF43
25	PAX6	ADRA1A	IL23R
26	KRT18	FITM1	TEF
27	KRTAP12-4	GOLGA6A	TNN
28	OR5C1	PNPLA7	TSC22D3
29	PSG11	HAGH	XKR5
30	APELA	RUNDC3B	TMEM170B
31	GJC3	IL6R	KLC3
32	SGPP2	DTX1	STAC2
33	RPS10-NUDT3	CTNNA3	TREH

34	LBHD1	KIAA0408	CHPT1
35	CALCB	RTN4RL2	FSD1
36	SDR9C7	GPD1	ALKBH7
37	HES6	OXER1	IGF2BP1
38	PSG4	LCAT	SMO
39	P3H2	SLC7A9	TDRD6
40	TMEM184A	ENPP7	PLCG2
41	PSG1	TEF	NPHS1
42	BARHL2	TM6SF2	TBC1D32
43	FGF11	SPDYC	C8orf44
44	UTS2	AASS	FCAR
45	LY6D	FBP1	FAM228A
46	XG	C8orf46	C1QTNF4
47	SSX1	ADH1C	PDLIM3
48	FAM156B	ECHDC2	ANKRD60
49	TXNL4A	OSGIN1	PARD6G
50	GPR87	CYP1A2	NANOG

Supplementary Table 2. Significant GSEA results using top 50 differentially expressed genes between RNA-seq-based MGMT subgroups.

Term	Overlap	P-value	Adjusted P-value	Genes	Database
Glycolysis / Gluconeogenesis_Homo sapiens_hsa0...	3/67	0.000629	0.018797	ADH1C;FBP1;ADH6	KEGG_2016
Retinol metabolism_Homo sapiens_hsa00830	3/65	0.000576	0.018797	ADH1C;UGT2B17;ADH6	KEGG_2016
Drug metabolism - cytochrome P450_Homo sapiens...	3/69	0.000686	0.018797	ADH1C;UGT2B17;ADH6	KEGG_2016
Metabolism of xenobiotics by cytochrome P450_H...	3/73	0.000808	0.018797	ADH1C;UGT2B17;ADH6	KEGG_2016
Chemical carcinogenesis_Homo sapiens_hsa05204	3/82	0.001133	0.021076	ADH1C;UGT2B17;ADH6	KEGG_2016
Glycerophospholipid metabolism_Homo sapiens_hs...	3/95	0.001731	0.026832	PNPLA7;GPD1;LCAT	KEGG_2016
alcohol dehydrogenase (NAD) activity (GO:0004022)	2/8	0.00017	0.029557	ADH1C;ADH6	GO_Molecular_Function_2015
Tyrosine metabolism_Homo sapiens_hsa00350	2/35	0.003457	0.045934	ADH1C;ADH6	KEGG_2016

Supplementary Table 3. Genes significantly associated with continuous MGMT RNA-seq expression values

Gene	Correlation*
HAGH	0.857435
BHMT2	0.854641
SLC16A2	0.847256
KDM8	0.846308
SAT2	0.838964
SLC2A9	0.83796
HSD17B13	0.832825
RXRA	0.83108
ACOT2	0.828849

* Significance was defined as having a correlation coefficient > 0.823, which corresponds to a p-value < 2.6×10^{-6} (*i.e.* Bonferroni adjustment of 0.05 threshold for testing 19270 hypotheses within 39 samples at 99% power to detect).

Only one pathway in KEGG (hsa_04919) was found to be enriched for these genes; a thyroid hormone signaling pathway with 2/118 genes overlapping (adjusted p=0.03), however, this result was not considered sufficiently strong or related to mention with regards to temozolomide response prediction.

Supplementary Table 4. Observed accuracies and corresponding p-values of predictive models (given the background distribution). Leave-pair-out cross-validation was used so that the expected accuracy from random classifications was fixed at 50% even when the ratio of responders to non-responders differed (as between datasets). 1000 random classifications for each cross-validation fold were performed to define background distributions for random classification.

Dataset	Algorithm	Accuracy	p-value
MS	**Predefined**	0.821212	5.46E-59
	Youden	0.804545	3.11E-53
RNA-seq	**Predefined**	0.722403	6.71E-28
	Youden	0.74026	3.23E-32
MB	**Predefined**	0.68	2.70E-16
	Youden	0.5	4.90E-01

p-values of the observed accuracies given the background distribution

Supplementary Table 5. Values of probability of achieving the observed accuracies of the predictive modelling strategies, considering the three different MGMT assessment methods.

Dataset	Algorithm	Accuracy	p-value
Exp(subgroups) + Prot(subgroups) + Meth(subgroups)	KNN	0.866667	5.45E-63
	Quad Discriminant	0.833333	2.03E-52
	DT AdaBoost	0.814583	6.53E-47
	RBF SVM	0.80625	1.45E-44
	Naive Bayes	0.802083	2.05E-43
	Decision Tree	0.797917	2.79E-42
	Extra Tree	0.797917	2.79E-42
	Linear SVM	0.797917	2.79E-42
	RF AdaBoost	0.797917	2.79E-42
	Ridge Classifier	0.797917	2.79E-42
	MLP	0.79375	3.66E-41
	Random Forest	0.716667	2.79E-23
	GPC	0.672917	1.70E-15
Exp(subgroups) + Prot(subgroups)	Quad Discriminant	0.792208	3.74E-47
	DT AdaBoost	0.782468	3.39E-44
	MLP	0.782468	3.39E-44
	Ridge Classifier	0.782468	3.39E-44
	Naive Bayes	0.772727	2.44E-41
	KNN	0.771104	7.15E-41
	Decision Tree	0.767857	6.02E-40
	Extra Tree	0.767857	6.02E-40
	Random Forest	0.753247	6.39E-36
	RF AdaBoost	0.673701	6.56E-18
	GPC	0.477273	1.31E-01
	Linear SVM	0.475649	1.14E-01
	RBF SVM	0.475649	1.14E-01
Exp + Prot	Random Forest	0.808442	2.66E-52
	DT AdaBoost	0.795455	3.68E-48
	Naive Bayes	0.788961	3.71E-46
	RF AdaBoost	0.780844	1.03E-43
	KNN	0.779221	3.12E-43
	Decision Tree	0.777597	9.36E-43
	MLP	0.727273	2.58E-29
	RBF SVM	0.701299	2.07E-23
	GPC	0.689935	4.75E-21
	Extra Tree	0.654221	1.67E-14
	Quad Discriminant	0.618506	2.83E-09
	Linear SVM	0.491883	3.43E-01

	Ridge Classifier	0.472403	8.64E-02
Exp + Prot + Meth	Linear SVM	0.760417	8.93E-33
	Naive Bayes	0.71875	1.08E-23
	KNN	0.714583	7.17E-23
	Ridge Classifier	0.704167	6.99E-21
	RF AdaBoost	0.689583	2.93E-18
	Quad Discriminant	0.68125	7.57E-17
	RBF SVM	0.629167	2.05E-09
	MLP	0.61875	3.24E-08
	Random Forest	0.604167	1.07E-06
	DT AdaBoost	0.583333	7.53E-05
	Decision Tree	0.50625	3.92E-01
	Extra Tree	0.485417	2.50E-01
	GPC	0.483333	2.21E-01

Legend: Exp, *MGMT* gene expression with RNA-seq; prot, MGMT protein expression level assessed with MS; Meth, *MGMT* methylation assessed with MB