Identifying genetic variants regulating MGMT gene expression

A study in monozygotic Danish twins

Mohammadnejad, Afsaneh; Sørensen, Mette; Mengel-From, Jonas; Nygaard, Marianne; Timofeeva, Maria; He, Liang; Clemmensen, Signe Bedsted; Halekoh, Ulrich; Dahlrot, Rikke Hedegaard; Tan, Qihua; Hjelmborg, Jacob

Published in:
Genomics

DOI:
10.1016/j.ygeno.2023.110616

Publication date:
2023

Document version:
Final published version

Genomics 115 (2023) 110616

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Identifying genetic variants regulating *MGMT* gene expression - A study in monozygotic Danish twins

Afsaneh Mohammadnejad [a,*], Mette Soerensen [a,b], Jonas Mengel-From [a,b,c], Marianne Nygaard [a,b,c], Maria Timofeeva [a], Liang He [a], Signe Bedsted Clemmensen [a], Ulrich Halekoh [a], Rikke Hedegaard Dahlrot [d,e], Qihua Tan [a,b], Jacob v.B. Hjelmborg [a,c]

[a] Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, Denmark
[b] Department of Clinical Genetics, Odense University Hospital, Denmark
[c] The Danish Twin Registry, University of Southern Denmark, Denmark
[d] Clinical institute, University of Southern Denmark, Denmark
[e] Department of Oncology, Odense University Hospital, Denmark

ARTICLE INFO

Keywords:
MGMT
Genetic regulation, Glioblastoma
GWAS
Monozygotic twins
Genetic contribution

ABSTRACT

Identifying genetic factors affecting the regulation of the O-6-Methylguanine-DNA Methyltransferase (*MGMT*) gene and estimating the genetic contribution of the *MGMT* gene through within-pair correlation in monozygotic twin pairs is of particular importance in various types of cancer such as glioblastoma. We used gene expression data in whole blood from 448 monozygotic twins from the Middle Age Danish Twins (MADT) study to investigate genetic regulation of the *MGMT* gene by performing a genome-wide association study (GWAS) of the variation in *MGMT* expression. Additionally, we estimated within-pair dependence measures of the expression values looking for the genetic influence of significant identified genes. We identified 243 single nucleotide polymorphisms (SNPs) significantly ($p < 5e\text{-}8$) associated with expression of *MGMT*, all located on chromosome 10 near the *MGMT* gene. Of the 243 SNPs, 7 are novel cis-eQTLs. By further looking into the suggestively significant SNPs (increasing cutoff to $p = 1e\text{-}6$), we identified 11 suggestive trans-eQTLs located on chromosome 17. These variants were in or proximal to a total of seven genes, which may regulate *MGMT* expression. The within-pair correlation of the expression of *MGMT*, *TRIM37*, and *SEPT4* provided the upper bound genetic influence of these genes. Overall, identifying cis- or trans-acting genetic variations regulating the *MGMT* gene can pave the way for a better understanding of the *MGMT* gene function and ultimately in understanding the patient's sensitivity to therapeutic alkylating agents.

1. Introduction

O-6-Methylguanine-DNA Methyltransferase, known as *MGMT*, is located on chromosome 10q26.3 with a total length of 300,437 bp, and is a promising target for tumor treatment as it is involved in chemo-resistance to alkylating agents [1]. Alkylating agents act directly on DNA during all cell cycle processes, crosslinking the N-7-guanine residues, producing DNA strand breakage, aberrant base pairing, cell division suppression, and, eventually, cell death [2]. The protein encoded by *MGMT* removes the methyl group from the O6 position of guanine, resulting in resistance to alkylating agents, including temozolomide (TMZ) [3,4]. For decades, TMZ as standard chemotherapy has been used for the treatment of glioblastoma patients; yet at least half of TMZ-treated patients do not respond to the treatment. This may occur due to the over-expression of *MGMT* or the lack of DNA repair pathway in glioblastoma mechanism [5,6].

The loss of *MGMT* expression is common in various types of cancer [7]. Several studies have shown that genetic variation in *MGMT* impact

* Corresponding author at: Epidemiology and Biostatistics, Department of Public Health, Faculty of Health Science, University of Southern Denmark, J. B. Winsløws Vej 9B, DK-5000 Odense C, Denmark.

E-mail addresses: amohammadnejad@health.sdu.dk (A. Mohammadnejad), msoerensen@health.sdu.dk (M. Soerensen), jmengel-from@health.sdu.dk (J. Mengel-From), mnygaard@health.sdu.dk (M. Nygaard), mtimofeeva@health.sdu.dk (M. Timofeeva), lianghe@health.sdu.dk (L. He), sbclemmensen@health.sdu.dk (S.B. Clemmensen), uhalekoh@health.sdu.dk (U. Halekoh), rikke.dahlrot@rsyd.dk (R.H. Dahlrot), qtan@heath.sdu.dk (Q. Tan), jhjelmborg@health.sdu.dk (J.v.B. Hjelmborg).

https://doi.org/10.1016/j.ygeno.2023.110616
Received 29 December 2022; Received in revised form 4 March 2023; Accepted 17 March 2023
Available online 21 March 2023

gene expression and thus play a key role in carcinogenesis and development of several types of cancer [8–11]. In cancer patients, *MGMT* methylation or low *MGMT* expression levels in tumors might be beneficial as indicators of chemosensitivity to alkylating treatment. As a result, determining the factors that impact the expression or methylation status of *MGMT*, including genetic variants, would be of particular importance, as these could also influence the results of treatment with alkylating chemotherapy [11–13].

Genetic variants such as single nucleotide polymorphisms (SNPs) in the promoter or enhancer region of the *MGMT* gene can alter its transcription and downstream protein expression [12,14]. Genome-wide association studies (GWAS) can provide more insights into the role of SNPs underlying *MGMT* expression and identify expression quantitative trait loci (eQTLs) across the genome that regulate *MGMT* expression through trans- or cis-acting mechanisms. Moreover, the upper bound genetic contribution to *MGMT* expression levels is not well established. This can be estimated through the within monozygotic (MZ) twin pair dependence measures of the expression values. This study aimed to investigate the 1) genetic regulation of *MGMT* gene expression 2) gene-based analysis to identify significant mapped genes and their regulation in relation to *MGMT* expression and glioblastoma 3) taking advantage of the MZ twin pairs to determine the upper bound genetic contribution to *MGMT* expression.

2. Materials and methods

2.1. Study population

The study sample comprises 448 MZ twins aged 56–80 (mean age = 66.21 years, SD = 6.01 years), including 248 male and 200 female twins recruited by The Danish Twin Registry as a part of the Middle Age Danish Twins (MADT) study [15]. The study population included MZ twins who participated in the MADT 2008–2011 follow-up study and for whom genome-wide genotype and gene expression data were generated (Table 1). Zygosity was determined using a questionnaire procedure that includes four questions about physical similarities and correctly classifies zygosity in more than 95% of the pairs [16]. Cell counts of blood leukocyte subtypes (monocytes, lymphocytes, basophils, neutrophils, and eosinophils) were counted using a Coulter LH 750 Hematology Analyzer (Beckman Coulter). The MADT study was approved by the Regional Committees on Health Research Ethics for Southern Denmark (S-VF-19980072), and written informed consents were obtained from all participants.

2.2. Genotyping and SNP imputation

Genome-wide SNP genotyping of DNA extracted from whole blood was conducted in one twin from each MZ twin pair using the Illumina PsychArray (Illumina, San Diego, California, USA). Genotyping was conducted by the SNP&SEQ Technology Platform, Science for Life Laboratory, Uppsala, Sweden (http://snpseq.medsci.uu.se/genotyping/snp-services/). Pre-imputation quality control (QC) removed SNPs based on genotype call rate < 98%, Hardy-Weinberg equilibrium (HWE) $p < 1e\text{-}6$, and minor allele frequency (MAF) = 0, and individuals on sample call rate < 99%, relatedness and gender mismatch. Pre-phasing and imputation to the 1000 Genomes phase 3 reference panel were performed using IMPUTE2 [17]. In post-imputation QC, non-autosomal SNPs, and SNPs with MAF ≤ 0.05 and imputation INFO score (information metric) < 0.6 were filtered out. This resulted in 5,779,266 genotyped and imputed autosomal SNPs being included in the statistical analysis [18].

Table 1
Descriptive statistics of the 448 monozygotic twins from the Middle Age Danish Twins (MADT) study included in the study.

Variables	Male	Female	Total
MADT	248	200	448
Mean age±sd (min, max)	66.53 ± 6.16 (57.00,79.88)	65.82 ± 5.82 (55.94,79.23)	66.21 ± 6.01 (55.94,79.88)

2.3. RNA extraction, gene expression analysis and expression data preprocessing

RNA extraction, gene expression analysis and expression data preprocessing were performed as described in Nygaard et al. (2019) [19]. In short, whole blood was collected in PAXgene Blood RNA Tubes (PreAnalytiX GmbH, Hombrechtikon, Switzerland), and total RNA was extracted using the PAXgene Blood miRNA kit (QIAGEN) according to the protocol of the manufacturer. The extracted RNA concentration was determined using a NanoDrop spectrophotometer ND-8000 (NanoDrop Technologies), and the quality was assessed by the Agilent 2100 Bioanalyzer (Agilent Technologies). Gene expression profiling was performed using the Agilent SurePrint G3 Human GE v2 8 × 60 K Microarray. Sample labeling and array hybridization were carried out in accordance with the 'Two-Color Microarray-Based Gene Expression Analysis – Low Input Quick Amp Labeling'-protocol. Samples were labeled with Cy5, and the reference consisting of a pool of 16 samples was labeled with Cy3. Hybridization, washing, scanning, and quantification were performed according to the array manufacturer's recommendations [20,21].

The R package *limma* was used for QC of the data [22]. Low expression values were detected based on detection values, which include *p*-values for testing whether each probe is more intense than the negative control probes. Detection p-values<0.05 were included as evidence that the probe corresponded to a truly expressed genes. Additionally, we calculated the coefficient of variation (CV) (i.e., a relative measure of variability calculated by dividing the standard deviation by the mean) for each probe and excluded probes with CV < 0.1. Background correction using the *normexp* method was done on the raw intensity data, within-array normalization using loess normalization to intensity measurements of two colors (cy3/cy5), and between-array normalization based on the quantile normalization method to make data from the different arrays comparable [23]. The missing expression values were imputed using the k-nearest neighbor algorithm, and replicated probes were summarized, calculating their median. Expression values for the *MGMT* gene for analysis were extracted from this processed data set.

2.4. Statistical analysis

We first adjusted the *MGMT* gene expression for covariates including age, sex, cell type composition. We then performed the genome-wide association study of *MGMT* gene expression values with imputed SNP dosages using the linear mixed model (LME) from the *lme4* package [24] while adjusting for the twin correlation in the sample by including random effect in the model. Next, we checked if there were any novel eQTLs among the top GWAS significant SNPs ($p < 1e\text{-}6$) by comparing our results to already identified eQTLs reported for *MGMT* in the GTEx Portal (https://gtexportal.org/home/). Finally, we investigated the chromatin states using the Roadmap Epigenomics Consortium ("Expanded 18-state model") [25] to capture known classes of genomic elements, such as promoters, enhancers, transcribed, repressed, and repetitive regions. We further prioritized the top candidate SNPs based on scores provided in the RegulomeDB database (https://www.regulomedb.org/regulome-search/) to identify variants that are more likely to play a functional role [26].

2.4.1. Gene-based analysis

A gene-based analysis was performed using the MAGMA tool developed by de Leeuw et al. 2015 [27] which is integrated in FUMA [28] for all SNPs ($p < 0.05$) based on the summary results of the GWAS

analysis. MAGMA maps SNPs to genes with a default window of 10 kb using 1000 Genomes Phase3 as a reference and calculates linkage disequilibrium (LD) within and between genes, also based on the 1000 genome Phase 3 reference. A default window of 10 kb means that SNPs are assigned to a gene if they located within 10 kb of a gene, either upstream or downstream, defined by the transcription start and stop sites of the gene. To obtain *P*-values for the gene-based analysis, MAGMA employs multiple linear regression. The Benjamini & Hochberg's false discovery rate (FDR) correction approach [29] was applied for multiple testing adjustment and FDR < 0.05 was used as the significance threshold.

2.4.2. Characterization of genomic loci and gene/positional mapping in FUMA

The positional mapping of SNPs to genes is done by using the GWAS summary statistics to define independent SNPs, lead SNPs, candidate SNPs, and genomic loci. FUMA defines independent SNPs as those with $p <$ 5e-8 (GWAS genome-wide significant cutoff) and independent from each other at linkage disequilibrium r2 < 0.6, and lead SNPs as the subset of independent SNPs that are independent of each other at linkage disequilibrium r2 < 0.1. Candidate SNPs including those with MAF ≥ 0.05 and in LD (r2 ≥ 0.6) with other independent significant SNPs, identified from the 1000 Genome phase 3 EUR reference panel, meaning that they are not available in the original GWAS. The genomic risk loci are defined by combining the lead SNPs that are close to each other ($<$ 250 kb) and the borders of loci defined by candidate SNPs. The 1000 Genome phase 3 EUR reference panel was used as a reference to calculate LD. The identified candidate SNPs were mapped to genes if 1) the SNP is located within ±10 kb of a gene (±10 kb is the default window in FUMA), 2) the SNP has a significant eQTL association (FDR < 0.05)[28].

2.4.3. Gauging familial relationship of significant genes

To gauge the upper bound genetic influence or heritability of significant genes, we estimated two within MZ pair dependence measures of the expression values; 1) The informational correlation coefficient (generalized correlation coefficient or Linfoot) [18,30,31], which measures the magnitude in functional relationship of the expression of genes within the pair ranging from total independence at value zero to an exact functional relationship at value one and 2) The classic Pearson product-moment correlation that measures the degree of linear dependence of the gene's expression within the pair ranging from −1 to 1. We note that the measure in 1) determines information conveyed between the gene expression values within the pair; in particular, a zero value is a strong indication of independence of genes. Also, if the two expression values are bivariate Gaussian distributed, then the numeric value of the measure in 2) equals the value of the measure in 1). The linear correlation measured in 2) also provides a direction of the association.

3. Results

A total of 243 SNPs significantly ($p <$ 5e-8) associated with *MGMT* expression were identified (Table S1), all of which are located on chromosome 10 from position 131,246,957-131,465,477 bp (Build 37) near or in *MGMT*. The GWAS QQ and Manhattan plots are shown in Fig. 1a and b. Fig. 2 shows a regional plot of the GWAS result in the

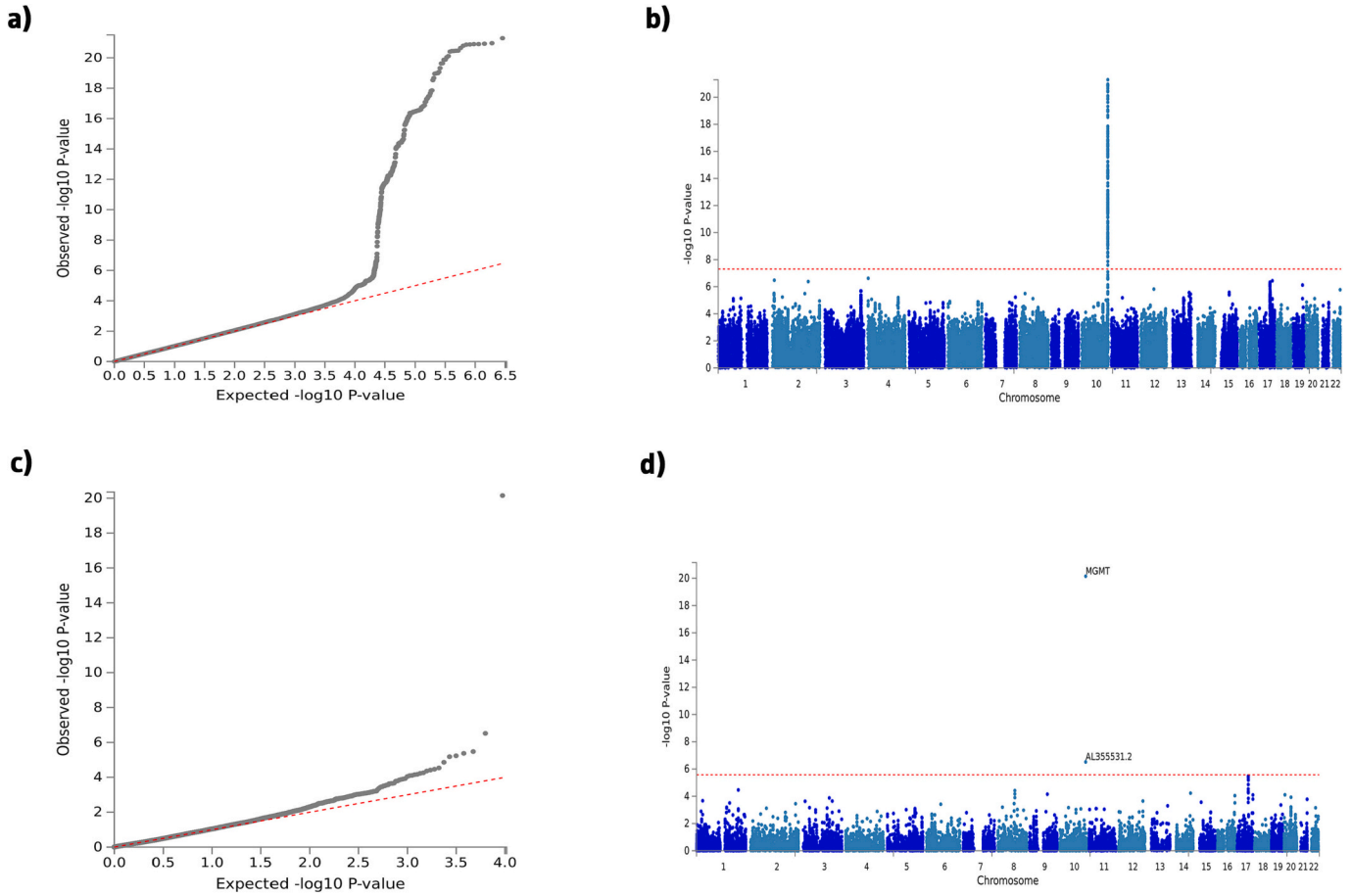

Fig. 1. Summary result of the GWAS and Gene-based analysis. a) The QQ plot of SNPs from GWAS summary statistic. b) The Manhattan plot of SNPs from GWAS summary statistics. c) The QQ plot of the genes identified by gene-based analysis in MAGMA. d) The Manhattan plot of the gene-based analysis as computed by MAGMA based on the input of GWAS summary statistics.

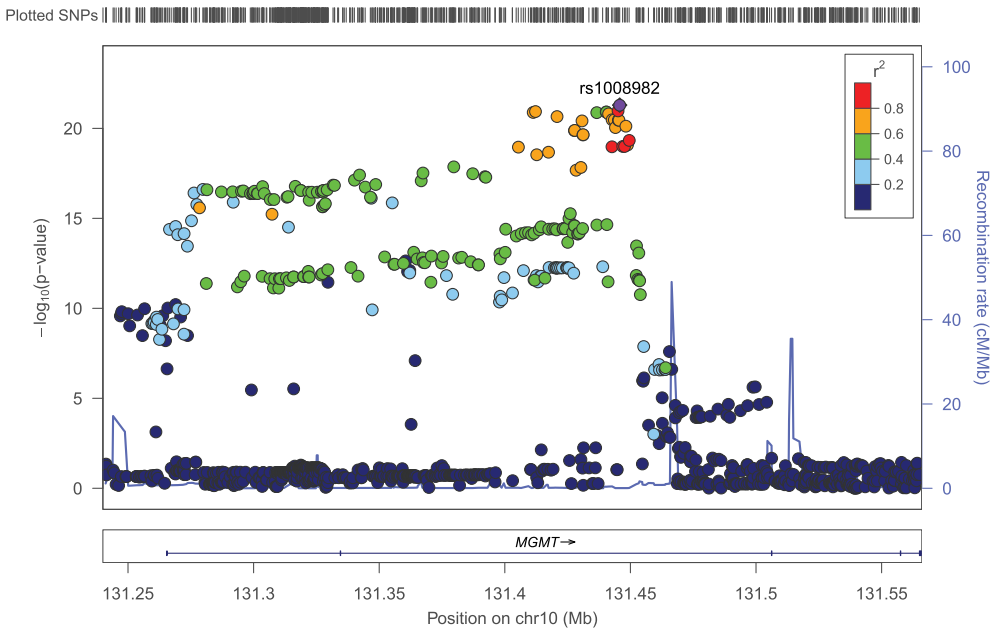

Fig. 2. Regional plot showing the association *P*-values on the -log10 scale on the y-axis and the chromosomal position along the x-axis near to MGMT gene. The label in different colors illustrates the pairwise LD pattern with the most strongly associated SNP.

MGMT gene region (chr10:131,240,000-chr10:131,566,000 bp, cf. Build 37). Of the 243 significantly associated SNPs, seven were identified as novel cis-eQTLs (see Table 2) that have not been identified previously as significantly associated with MGMT expression (based on information listed in the GTEx Portal, https://gtexportal.org/home/). We also identified eleven suggestive ($p < 1e\text{-}6$) trans-eQTLs positioned on chromosome 17 (see Table 3). Furthermore, we investigated whether there is any physical interaction between chromosome 10 MGMT/SNP loci and chromosome 17 loci. We used Hi-C database (http://3dgenome.fsm.northwestern.edu/) [32], chose the human hg19 assembly and GM12878 tissue, which is a lymphoblastoid cell line. Next, we used the region chr10:131246957–131,565,261 (representing the chromosome10 SNP loci, the MGMT coding region, and the MGMT promotor) as location 1, and the region chr17:56459408–68,650,399 (representing the chromosome 17 loci) as location 2. The MAGMA gene-based analysis (see QQ and Manhattan plots in Fig. 1c and d) based on the GWAS result of all included SNPs resulted in the identification of seven genes ($p < 1.38e\text{-}5$, FDR < 0.05) (see Table S2), including *MGMT* ($p = 1.08e\text{-}19$, FDR $= 2.05e\text{-}15$), *AL355531.2* ($p = 6.09e\text{-}07$, FDR $= 0.01$), *TRIM37* ($p = 4.96e\text{-}06$, FDR $= 0.03$), *SEPT4* ($p = 6.69e\text{-}06$, FDR $= 0.03$), *C17orf47* ($p = 8.18e\text{-}06$, FDR $= 0.03$), *PPM1E* ($p = 1.004e\text{-}05$, FDR $= 0.03$), and *TEX14* ($p = 1.77e\text{-}05$, FDR $= 0.04$). Furthermore, the within-pair dependency measure of the expression levels for the protein coding genes among these seven genes based on both Linfoot and Pearson correlations is shown in Fig. 3.

From the positional mapping in FUMA, we identified 2 lead SNPs, 15 independent significant SNPs with $P < 5e\text{-}8$, 15 mapped genes (Figs. S1a, b, and c), 287 candidate SNPs, as well as 255 SNPs that are in LD ($r^2 < 0.6$) with the significant independent SNPs. Most of the SNPs are in intronic or in intergenic regions.

4. Discussion

By performing a GWAS of *MGMT* gene expression in 448 Danish MZ twins we investigated the genetic regulation of the *MGMT* gene. Additionally, we gauged the genetic influence of the expression of significant genes by calculating the Informational Correlation Coefficient, Linfoot, and Pearson correlation of expression levels within MZ pairs.

We identified 243 significant SNPs located on chromosome 10 within the *MGMT* gene region, showing the genetic regulation underlying the *MGMT* gene. Importantly, 7 of these SNPs are novel cis-eQTLs that have not already been detected. We found that the 243 SNPs are distributed

Table 2
The list of seven novel significant cis-eQTLs identified by GWAS.

SNP	CHR*	BP*	A1*	A2	RDB**	Chromatin state	Function	P
rs12241533	10	131,445,027	G	A	5	Weak transcription	Intronic	1.09e-21
rs1762416	10	131,303,513	A	G	6	Heterochromatin	Intronic	1.75e-17
rs12573708	10	131,318,710	G	A	4	Weak transcription	Intronic	2.84e-17
rs1807213	10	131,295,111	A	T	6	Weak transcription	Intronic	3.17e-17
rs519690	10	131,424,033	G	A	3a	Quiescent/Low	Intronic	3.82e-15
rs4751106	10	131,424,653	G	T	5	Quiescent/Low	Intronic	5.73e-13
rs1627492	10	131,403,007	T	C	2b	Weak Transcription	Intronic	1.43e-11

* CHR: Chromosome; BP: Base Pair; A1: Affected allele.
** RegulomeDB score: 5 = TF binding or DNase peak; 6 = Motif hit; 4 = TF binding + DNase peak; 6 = Motif hit; 3a = TF binding + any motif + DNase peak; 2b = TF binding + any motif + DNase Footprint + DNase peak.

Table 3
The list of eleven suggestive trans-eQTLs identified by GWAS.

SNP	CHR*	BP*	A1*	A2	RDB**	Chromatin state	P
rs11867678	17	68,650,399	C	T	4	Quiescent/Low	3.67e-07
rs9889419	17	56,473,259	T	C	1f	Quiescent/Low	4.62e-07
rs9788975	17	56,460,800	G	A	4	Quiescent/Low	5.64e-07
rs9904993	17	56,468,569	G	T	4	Quiescent/Low	6.49e-07
rs4793585	17	56,470,511	G	A	5	Quiescent/Low	6.53e-07
rs4793941	17	56,472,366	G	A	4	Quiescent/Low	6.59e-07
rs1990008	17	56,459,408	C	G	4	Weak transcription	6.60e-07
rs2301867	17	56,480,074	T	A	4	Quiescent/Low	7.61e-07
rs723991	17	56,482,039	G	A	5	Quiescent/Low	7.98e-07
rs58647683	17	56,483,194	T	C	4	Quiescent/Low	8.78e-07
rs9303398	17	56,484,257	T	C	5	Quiescent/Low	9.33e-07

* CHR: Chromosome; BP: Base Pair; A1: Affected Allele.
** RegulomDB score: 1f = eQTL + TF binding / DNase pea; 4 = TF binding + DNase peak; 5 = TF binding or DNase peak.

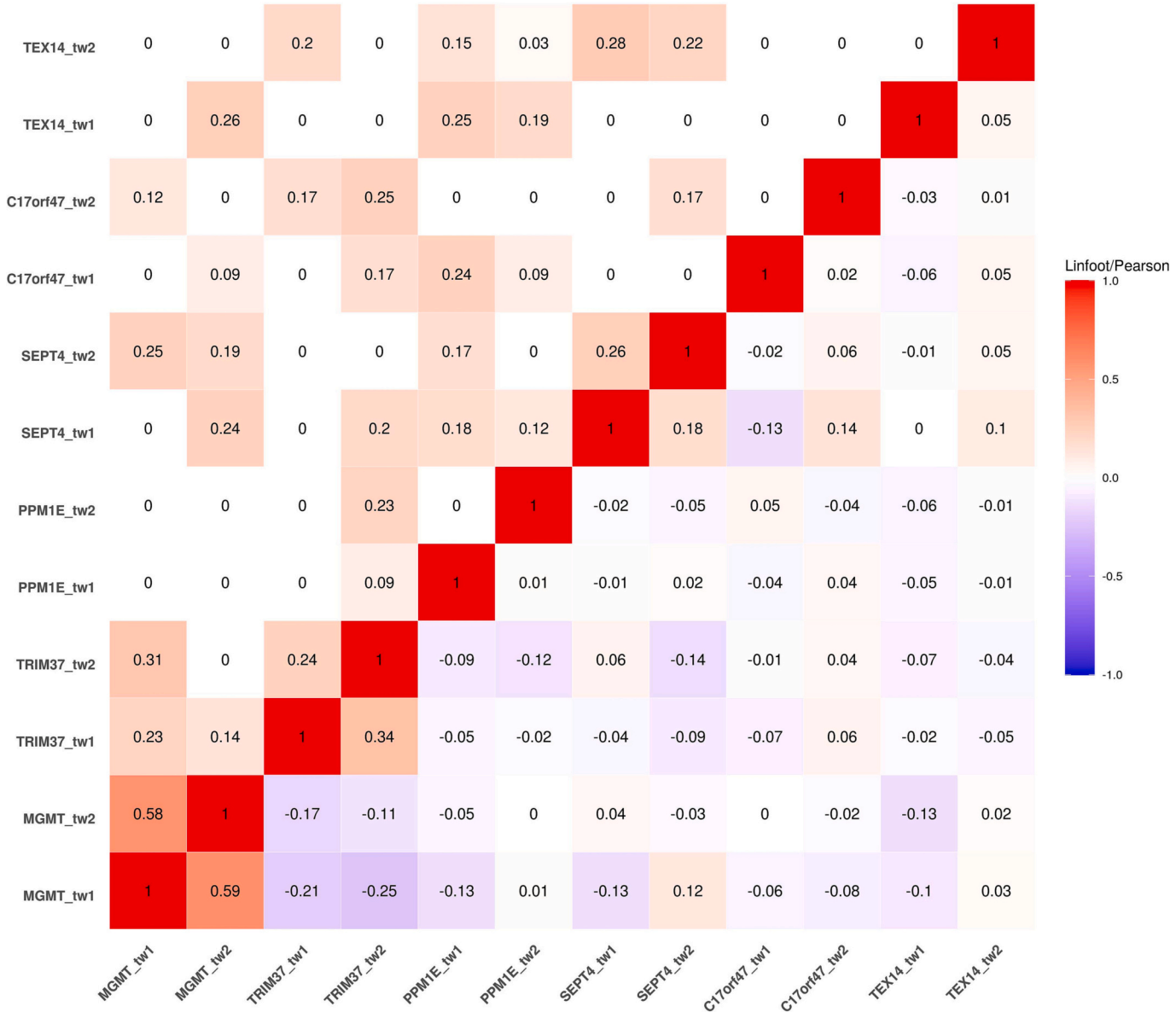

Fig. 3. Within MZ pairs measures of dependency of expressions for significant genes based on Linfoot (upper triangle) and Pearson (lower triangle) correlations.

across seven chromatin states including weak enhancer, ZNF genes and repeats, flanking TSS upstream, strong transcription, weak transcription, heterochromatin, and quiescent/low using the Roadmap Epigenomics Consortium [25]. Half of the SNPs are found in the chromatin states weak and strong transcriptions which makes sense as many of them are within the gene body. One SNP (rs12262382) is in weak

enhancer and three in flanking TSS upstream (promoters) (Table S1). Of the 7 new cis-eQTLs, 4 are in weak transcription chromatin states and one is in heterochromatin (Table 2). Furthermore, the lower scores provided by RegulomeDB database indicate more evidence for variants to play a functional role. In total, 60 SNPs (25%) are with score 1, suggesting that they are likely to affect transcription factor binding and are linked to expression of their target gene. Among the 7-novel cis-eQTLs, rs1627492 has a RegulomeDB score of 2b, suggesting that is likely to affect the transcription factor binding, while the 6 other SNPs have scores of 3a and below (Table 2). Of the eleven suggestive trans-eQTLs, the highest scoring SNP (rs9889419 with the target gene *TRIM37*) has a score of 1 f. Fig. 2 depicts a regional plot of SNPs mapping to the *MGMT* gene, with the strongest associated SNP being rs1008982 ($p = 5.1e-22$). The QQ plot (Fig. 1a) shows a clear significant deviation of top SNPs in the upper tail, and the Manhattan plot (Fig. 1b) confirms this with a significant peak on chromosome 10. In addition, a cluster of suggestively significant SNPs ($p < 1e-6$) located on chromosome 17 is evident in the Manhattan plot, suggesting that these SNPs might be suggestive trans-eQTLs for *MGMT* expression (Fig. 1b). When exploring the physical interaction between identified loci, we found no evidence of interaction between the chromosome 10 MGMT/SNP loci and the chromosome 17 loci.

The MAGMA gene-based analysis identified seven significant genes (FDR < 0.05), including *MGMT*, *AL355531.2*, *TRIM37*, *SEPT4*, *C17orf47*, *PPM1E*, and *TEX14* (Table S2). These genes are also depicted in the Manhattan plot (Fig. 1d). The *MGMT* and *AL355531.2* genes are located on chromosome 10, whereas the rest are located on chromosome 17. As expected, the *MGMT* gene appeared as the top significant gene and thereby confirms our top SNPs identified in the GWAS. Studies have shown that the upregulation of *MGMT* expression contributes to TMZ resistance, causing problems in treating malignant brain tumors [5,33]. Additionally, there is a significant correlation between *MGMT* expression and the nuclear factor-κB (NF-κB) pathway in glioma cells. Inhibition of NF-κB can reduce *MGMT* activity and, when combined with TMZ treatment, may provide clinical benefits for glioma patients [34]. Overexpression of *TRIM37* gives resistance to the DNA-damaging anticancer drug cisplatin by activating the NF-B pathway [34]. In human glioma, *TRIM37* is significantly upregulated, and high levels of *TRIM37* expression were found to be associated with gliomas grade [35]. Several studies have shown that the chromosome 17q region harboring *TRIM37*, is amplified approximately 50–60% of neuroblastomas and ~ 10% of breast cancers [36–40]. *TRIM37* is a member of the tripartite motif or TRIM subfamily of zinc finger proteins, including RING, B-Box, and coiled-coil domains [41]. A study has investigated the expression of *TRIM37* in human glioma tissue from 124 patients with primary glioma who underwent surgical therapy. They reached the conclusion that this gene is highly upregulated in human gliomas, and that the expression level of *TRIM37* is closely correlated with glioma grade [35]. Another study examined tissue from 41 human osteosarcoma patients and found that *TRIM37* might be a potential therapeutic target for the treatment of pediatric osteosarcoma [42]. Interestingly, *MGMT* has also been linked to osteosarcoma in that the abnormal methylation of the *MGMT* gene in osteosarcoma tissue is shown to be related to the expression level of *MGMT* [43]. One of the other significant genes, *SEPT4*, is a member of the septin gene family, and this gene is highly expressed in the brain and heart. One of the isoforms of the *SEPT4* gene, *ARTS*, is localized in mitochondria and is a pro-apoptotic tumor suppressor [44,45]. Additionally, changes in septin expression have been reported to be associated with glioblastomas, squamous cell carcinomas, melanomas, renal cell carcinomas, colorectal cancer, and lung cancer [46–48]. Chen et al. (2017) studied *PPM1E* expression in gastric cancer tissue from six individuals and compared them with six normal gastric tissue samples. They discovered *PPM1E* expression to be upregulated, which is considered critical for shutting off AMPK signaling and increasing cancer cell growth TMZ [49].

The candidate SNPs identified as part of the positional mapping of SNPs, mapped to 15 coding and non-coding genes locating on 10q26.2 and 10q26.3 in the genome. This further supports the importance of SNPs in the vicinity of the *MGMT* genomic region for *MGMT* expression. Previous studies have discovered an *MGMT* enhancer that is associated with *MGMT* promoter methylation and *MGMT* expression regulation [50,51]. However, glioblastomas have different levels of *MGMT* promoter methylation and *MGMT* expression [52]. Consequently, the *MGMT* enhancer may be involved in the regulation of genes other than the *MGMT* gene. Chen et al. (2021) reported that *MKI67* downregulation following enhancer deletion increases glioma cell susceptibility to TMZ, suggesting that *MKI67* is regulated by the *MGMT* enhancer [53,54]. Another gene potentially coregulated with *MGMT* is *EBF3*, which is close to *MGMT* and widely inactivated in gliomas, colon cancer, and leukemia. The *EBF3* gene is co-methylated with *MGMT* implying a possible regulatory link [53,55,56].

The within-pair correlation of gene expression values of significant genes estimating the upper bound genetic influence on the expression of the genes is shown in Fig. 3. The genetic influence on *MGMT* (Linfoot = 0.58, Pearson = 0.59), *TRIM37* (Linfoot = 0.24, Pearson = 0.34), and *SEPT4* (Linfoot = 0.26, Pearson = 0.18) are estimated to be 0.59, 0.34 and 0.26, respectively indicating that a large proportion of the expression of these genes is regulated by the shared genetic component within families. The familial relationships of the expression values of the other genes indicated either independence at zero or a weak relationship. It is important to note that further genetic contribution or heritability can be estimated through within-pair dependencies in both MZ, and dizygotic (DZ) twins based on biometric twin modelling [57,58]. On top of the genetic effects, environmental effects that influence the expression of the genes is interesting to consider.

5. Conclusions

In summary, by associating *MGMT* expression with SNP genotypes in MZ twins, this study identified seven novel *MGMT* cis-eQTLs on chromosome 10 and eleven suggestive trans-eQTLs on chromosome 17. Additionally, the within MZ pair correlations could detect the upper bound genetic influence of the identified genes emphasizing the advantage of twin studies. Findings from this study merit further verification in cancer tissues as to the usefulness of the identified eQTLs as biomarkers and predictors for improving treatment response in cancer patients.

Financial support

This work was supported by The Academy of Geriatric Cancer Research (AgeCare), grant number 71778, Odense University Hospital, Denmark.

Author contribution

Conceptualization, A.M., J.v.B.H., Q.T. and R.H.D; methodology, A. M. and J.v.B.H; formal analysis, A.M.; Writing manuscript, A.M.; project administration, A.M and J.v.B.H; Review and editing A.M., M.S., J.M., M.N., M.T., L.H., S.B.C., U.H., R.H.D, Q.T., and J.v.B.H.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Data availability

According to Danish and EU legislations, transfer and sharing of individual-level data require prior approval from the Danish Data Protection Agency and require that data sharing requests are dealt with on a case-by-case basis. However, we welcome any enquiries regarding

collaboration and individual requests for data sharing. Requests can be directed to J.v.B.H, jhjelmborg@health.sdu.dk.

Acknowledgments

MADT was supported by grants from The National Program for Research Infrastructure 2007 from the Danish Agency for Science, Technology, and Innovation (09-063256) and the US National Institutes of Health (P01 AG08761). Genotyping was supported by NIH R01 AG037985 (Pedersen).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110616.

References

[1] W. Yu, et al., O6-methylguanine-DNA methyltransferase (MGMT): challenges and new opportunities in glioma chemotherapy, Front. Oncol. 9 (2020) 1547.
[2] R. Ralhan, J. Kaur, Alkylating agents and cancer therapy, Expert Opin. Ther. Patents 17 (9) (2007) 1061–1075.
[3] J. Brennand, G.P. Margison, Reduction of the toxicity and mutagenicity of alkylating-agents in mammalian-cells harboring the escherichia-coli alkyltransferase gene, Proc. Natl. Acad. Sci. U. S. A. 83 (17) (1986) 6292–6296.
[4] K.A. van Nifterik, et al., Absence of the MGMT protein as well as methylation of the MGMT promoter predict the sensitivity for temozolomide, Br. J. Cancer 103 (1) (2010) 29–35.
[5] S.Y. Lee, Temozolomide resistance in glioblastoma multiforme, Genes Dis 3 (3) (2016) 198–210.
[6] L. Wang, et al., Genetic and environmental determinants of O6-methylguanine DNA-methyltransferase (MGMT) gene methylation: a 10-year longitudinal study of Danish twins, Clin. Epigenetics 13 (1) (2021) 35.
[7] N.J. Hawkins, et al., MGMT methylation is associated primarily with the germline C>T SNP (rs16906252) in colorectal cancer and normal colonic mucosa, Mod. Pathol. 22 (12) (2009) 1588–1599.
[8] S.L. Gerson, MGMT: its role in cancer aetiology and cancer therapeutics, Nat. Rev. Cancer 4 (4) (2004) 296–307.
[9] L.M. Huang, et al., An intronic genetic variation of MGMT affects enhancer activity and is associated with glioma susceptibility, Cancer Manag. Res. 10 (2018) 3995–4003.
[10] C.Y. Hsu, et al., The MGMT promoter single-nucleotide polymorphism rs1625649 had prognostic impact on patients with MGMT methylated glioblastoma, PLoS One 12 (10) (2017).
[11] J. Kuroiwa-Trzmielina, et al., SNP rs16906252C > T Is an expression and methylation quantitative trait locus associated with an increased risk of developing MGMT-methylated colorectal cancer, Clin. Cancer Res. 22 (24) (2016) 6266–6277.
[12] R.W. Rapkins, et al., The MGMT promoter SNP rs16906252 is a risk factor for MGMT methylation in glioblastoma and is predictive of response to temozolomide, Neuro-Oncology 17 (12) (2015) 1589–1598.
[13] C.C. Brown, et al., A genome-wide association analysis of temozolomide response using lymphoblastoid cell lines shows a clinically relevant association with MGMT, Pharmacogenet. Genomics 22 (11) (2012) 796–802.
[14] M. Xu, et al., Influence of promoter/enhancer region haplotypes on MGMT transcriptional regulation: a potential biomarker for human sensitivity to alkylating agents, Carcinogenesis 35 (3) (2014) 564–571.
[15] D.A. Pedersen, et al., The Danish twin registry: an updated overview, Twin Res Hum Genet 22 (6) (2019) 499–507.
[16] L. Christiansen, et al., Age- and sex-differences in the validity of questionnaire-based zygosity in twins, Twin Res. 6 (4) (2003) 275–278.
[17] E.M. van Leeuwen, et al., Population-specific genotype imputations using minimac or IMPUTE2, Nat. Protoc. 10 (9) (2015) 1285–1296.
[18] A. Mohammadnejad, et al., Generalized correlation coefficient for genome-wide association analysis of cognitive ability in twins, Aging (Albany NY) 12 (22) (2020) 22457–22494.
[19] M. Nygaard, et al., Global expression profiling of cognitive level and decline in middle-aged monozygotic twins, Neurobiol. Aging 84 (2019) 141–147.
[20] A. Mohammadnejad, et al., Global gene expression profiling and transcription factor network analysis of cognitive aging in monozygotic twins, Front. Genet. 12 (2021), 675587.
[21] A. Mohammadnejad, et al., Differential lncRNA expression profiling of cognitive function in middle and old aged monozygotic twins using generalized association analysis, J. Psychiatr. Res. 140 (2021) 197–204.
[22] M.E. Ritchie, et al., A comparison of background correction methods for two-colour microarrays, Bioinformatics 23 (20) (2007) 2700–2707.
[23] Y.H. Yang, N.P. Thorne, Normalization for two-color cDNA microarray data, in: Lecture Notes-Monograph Series, 2003, pp. 403–418.
[24] D. Bates, et al., Package Lme4: Linear Mixed-Effects Models Using Eigen and S4 67, 2014.
[25] A. Kundaje, et al., Integrative analysis of 111 reference human epigenomes, Nature 518 (7539) (2015) 317–330.
[26] A.P. Boyle, et al., Annotation of functional variation in personal genomes using RegulomeDB, Genome Res. 22 (9) (2012) 1790–1797.
[27] C.A. de Leeuw, et al., MAGMA: generalized gene-set analysis of GWAS data, PLoS Comput. Biol. 11 (4) (2015), e1004219.
[28] K. Watanabe, et al., Functional mapping and annotation of genetic associations with FUMA, Nat. Commun. 8 (1) (2017) 1826.
[29] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate - a practical and powerful approach to multiple testing, J. Royal Stat. Soc. Ser. B-Stat. Methodol. 57 (1) (1995) 289–300.
[30] E.H. Linfoot, An informational measure of correlation, Inf. Control. 1 (1) (1957) 85–89.
[31] A. Mohammadnejad, et al., Novel DNA methylation marker discovery by assumption-free genome-wide association analysis of cognitive function in twins, Aging Cell 20 (2) (2021), e13293.
[32] Y. Wang, et al., The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions, Genome Biol. 19 (1) (2018) 151.
[33] C.H. Fan, et al., O6-methylguanine DNA methyltransferase as a promising target for the treatment of temozolomide-resistant gliomas, Cell Death Dis. 4 (10) (2013), e876.
[34] A. Valletti, et al., Targeting chemoresistant tumors: could TRIM proteins-p53 axis be a possible answer? Int. J. Mol. Sci. 20 (7) (2019).
[35] C.Y. Chang, et al., Tripartite motif-containing protein 37 is overexpressed in human glioma and its downregulation inhibits human glioma cell growth in vitro, Int. J. Clin. Exp. Pathol. 11 (2) (2018) 502–511.
[36] F. Meitinger, et al., TRIM37 controls cancer-specific vulnerability to PLK4 inhibition, Nature 585 (7825) (2020) 440–446.
[37] N. Ho, et al., Delineation of the frequency and boundary of chromosomal copy number variations in paediatric neuroblastoma, Cell Cycle 17 (6) (2018) 749–758.
[38] J. Li, et al., Oncogenic properties of PPM1D located within a breast cancer amplification epicenter at 17q23, Nat. Genet. 31 (2) (2002) 133–134.
[39] Y. Liu, et al., Targeting 17q23 amplicon to overcome the resistance to anti-HER2 therapy in HER2+ breast cancer, Nat. Commun. 9 (1) (2018) 4718.
[40] D.V. Bulavin, et al., Amplification of PPM1D in human tumors abrogates p53 tumor-suppressor activity, Nat. Genet. 31 (2) (2002) 210–215.
[41] J. Kallijarvi, et al., TRIM37 defective in mulibrey nanism is a novel RING finger ubiquitin E3 ligase, Exp. Cell Res. 308 (1) (2005) 146–155.
[42] Y. Tao, et al., TRIM37 promotes tumor cell proliferation and drug resistance in pediatric osteosarcoma, Oncol. Lett. 14 (6) (2017) 6365–6372.
[43] Q. Cui, et al., Relationship between hypermethylated MGMT gene and osteosarcoma necrosis rate after chemotherapy, Pathol. Oncol. Res. 17 (3) (2011) 587–591.
[44] Y. Gottfried, et al., The mitochondrial ARTS protein promotes apoptosis through targeting XIAP, EMBO J. 23 (7) (2004) 1627–1635.
[45] M. García-Fernández, et al., Sept4/ARTS is required for stem cell apoptosis and tumor suppression, Genes Dev. 24 (20) (2010) 2282–2293.
[46] P.A. Hall, S.E. Russell, The pathobiology of the septin gene family, J. Pathol. 204 (4) (2004) 489–505.
[47] L. Dolat, Q.C. Hu, E.T. Spiliotis, Septin functions in organ system physiology and pathology, Biol. Chem. 395 (2) (2014) 123–141.
[48] D. Angelis, E.T. Spiliotis, Septin mutations in human cancers, Front Cell Dev Biol 4 (2016) 122.
[49] X.Y. Chen, et al., A novel enhancer regulates MGMT expression and promotes temozolomide resistance in glioblastoma, Nat. Commun. (2018) 9.
[50] L.C. Harris, J.S. Remack, T.P. Brent, Identification of a 59 bp enhancer located at the first exon/intron boundary of the human O6-methylguanine DNA methyltransferase gene, Nucleic Acids Res. 22 (22) (1994) 4614–4619.
[51] K.L. McDonald, et al., The T genotype of the MGMT C>T (rs16906252) enhancer single-nucleotide polymorphism (SNP) is associated with promoter methylation and longer survival in glioblastoma patients, Eur. J. Cancer 49 (2) (2013) 360–368.
[52] R.H. Dahlrot, et al., Posttreatment effect of MGMT methylation level on glioblastoma survival, J. Neuropathol. Exp. Neurol. 78 (7) (2019) 633–640.
[53] D.W. Parsons, et al., An integrated genomic analysis of human glioblastoma multiforme, Science 321 (5897) (2008) 1807–1812.
[54] B. Chen, et al., Long-range gene regulation network of the MGMT enhancer modulates glioma cell sensitivity to temozolomide, J Genet Genom. 48 (10) (2021) 946–949.
[55] M.A. Ibrahim Al-Obaide, et al., Genomic Space of MGMT in human glioma revisited: novel motifs, regulatory RNAs, NRF1, 2, and CTCF involvement in gene expression, Int. J. Mol. Sci. 22 (5) (2021).
[56] E.J. Rodger, et al., Characterisation of DNA methylation changes in EBF3 and TBC1D16 associated with tumour progression and metastasis in multiple cancer types, Clin. Epigenetics 11 (1) (2019) 114.
[57] J.B. Hjelmborg, et al., The heritability of prostate cancer in the Nordic twin study of cancer, Cancer Epidemiol. Biomark. Prev. 23 (11) (2014) 2303–2310.
[58] Q. Tan, et al., Twin methodology in epigenetic studies, J. Exp. Biol. 218 (Pt 1) (2015) 134–139.