Detection of Epigenetic Markers in Melanoma using DNA Methylation Analysis

Mirza S. Khan

Abstract

Unlike most other cancers, mortality due to melanoma has remained persistently high. Melanoma presently lacks a screening test and best-available treatment regimens are subject to near universal development of treatment resistance. This study evaluated the epigenome of melanoma and normal melanocytes to identify areas of differential mean methylation between both groups at individual cytosine-guanosine dinucleotides (CpG) loci and regions of CpG clusters. The most significant differentially methylated region in this study exhibited hypermethylation affecting the LY6G5C gene. Pathway analysis found that non-homologous end joining was significantly enriched in both the Gene Ontology and KEGG database. DNA methylation patterns affecting genes and processes as identified in this study may be useful as biomarkers and may suggest a role for epigenetic inhibitor therapy to prevent epigenetic modifications that may drive melanoma tumorogenesis and treatment resistance.

Introduction

Melanoma arises from the malignant transformation of pigment-producing cells called melanocytes. This malignancy has the highest mutation rate of all cancers[1]. Despite accounting for less than 5% of combined skin cancers, melanoma causes the majority of skin cancer-related deaths[2,3]. Epigenetic mechanisms, such as DNA methylation, alter gene expression independent of changes to the DNA sequence. These modifications are known to play a role in melanoma tumorogenesis[4]. Melanoma has many environmental risk factors, including ultraviolet (UV) radiation exposure and residing in regions with a depleted ozone layer[5], which may promote epigenetic modifications.

In the mammalian genome, DNA methylation occurs at the 5′ carbon of the cytosine ring in cytosine-guanosine dinucleotides (CpGs)[6]. Those regions of the genome that are rich in CpGs are referred to as CpG islands, and the majority of these islands are found in gene promoter regions[3,4,7,8]. Proper regulation of DNA methylation is vital to ensure that correct DNA transcription and chromatin remodeling occur at each cell of a specific tissue type[9]. Malignant cells often rely on aberrant DNA methylation to bypass normal cell function. Methylation of promoter regions has been shown to play a role in melanoma oncogenesis and progression[3,4]. Methylated promoter regions cannot be accessed by transcription factors resulting in silencing of associated gene expression, and vice versa. Aberrant methylation patterns include hypermethylation of tumor suppressor genes and hypomethylation of oncogenic genes[3,10]. Additionally, global hypomethylation is associated with increased genomic instability[4,8].

The "gold standard" for the study of DNA methylation is bisulfite-based DNA methylation[3]. Sodium bisulfite selectively converts unmethylated cytosine to uracil, leaving the methylated cytosine intact[9,11]. Amplification, fragmentation and hybridization of bisulfite-converted DNA can then be achieved using massively parallel next generation sequencing, such as with the Illumina® Infinium HumanMethylation 450K array[10,12]. Methylation values from the Illumina® 450K array have been shown to have excellent concordance with traditional bisulfite sequencing[13]. The 450K array contains 485,512 probes, which accounts for 19,755 unique CpG islands (96% of annotated CpG islands) and other gene regions. This coverage allows for analysis at single nucleotide

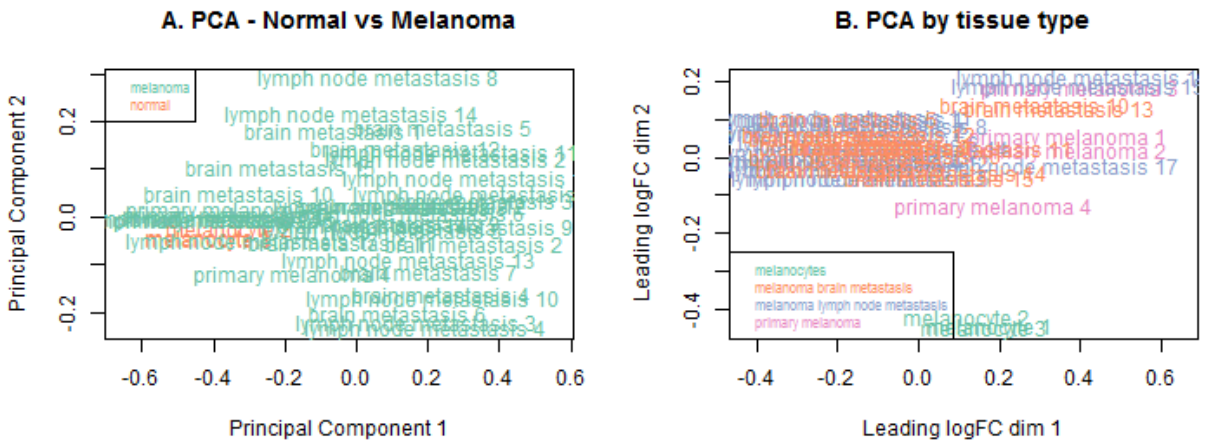

Figure 1: Principal Component Analysis.

resolution[13–15]. The 450K array provides methylation values based on fluorescence intensity. Often, the degree of methylation at each probe is represented by β values or M (log-ratio) values. These values are then used to perform genome-wide DNA methylation analyses based on the phenotype or outcome of interest. The 450K array consists of two different probe types (Types I and II), which often have different distributions of methylation values[16]. The Infinium Type II probe is uniquely subject to dye bias[11]. Although Illumina® recommends analysis with the raw β values, pre-processing and normalization of the data prior to analysis is routinely performed to mitigate biases.

This study aims to identify significant CpG loci and gene regions that may be subject to epigenetic modification by aberrant DNA methylation in patients with melanoma. The results may serve as useful prognostic or diagnostic biomarkers. These findings may also support adjuvant treatment with epigenetic inhibitor therapy, such as *DNA methyltransferase* (DNMT) inhibitors.

Results

DNA methylation patterns were analyzed using 40 specimens: 3 normal melanocyte samples and 37 melanoma samples (4 primary melanoma, 17 stage III melanoma with lymph node metastases, and 16 stage IV melanoma with brain metastases). Multidimensional scaling techniques, such as Principal Component Analysis (PCA) and Redundancy analysis (RDA), using normalized β values from the assay revealed localized clustering of normal melanocytes (Figure 1, Supplemental Figure 1). This result was also found when using pairwise PCA among the four tissue samples available in this study.

Analysis of mean methylation at each CpG locus using a linear model and empirical Bayes framework identified 1,126 significantly hypomethylated and 2,265 significantly hypermethylated CpG loci among melanoma samples when compared to normal melanocytes. Significant differences in mean methylation were noted in CpG loci in each chromsome and visualized using a Manhattan plot (Supplemental Figure 2). A volcano plot was generated to demonstrate the distribution of statistical significance (negative log of the False-Discovery Rate adjusted p-values) against the

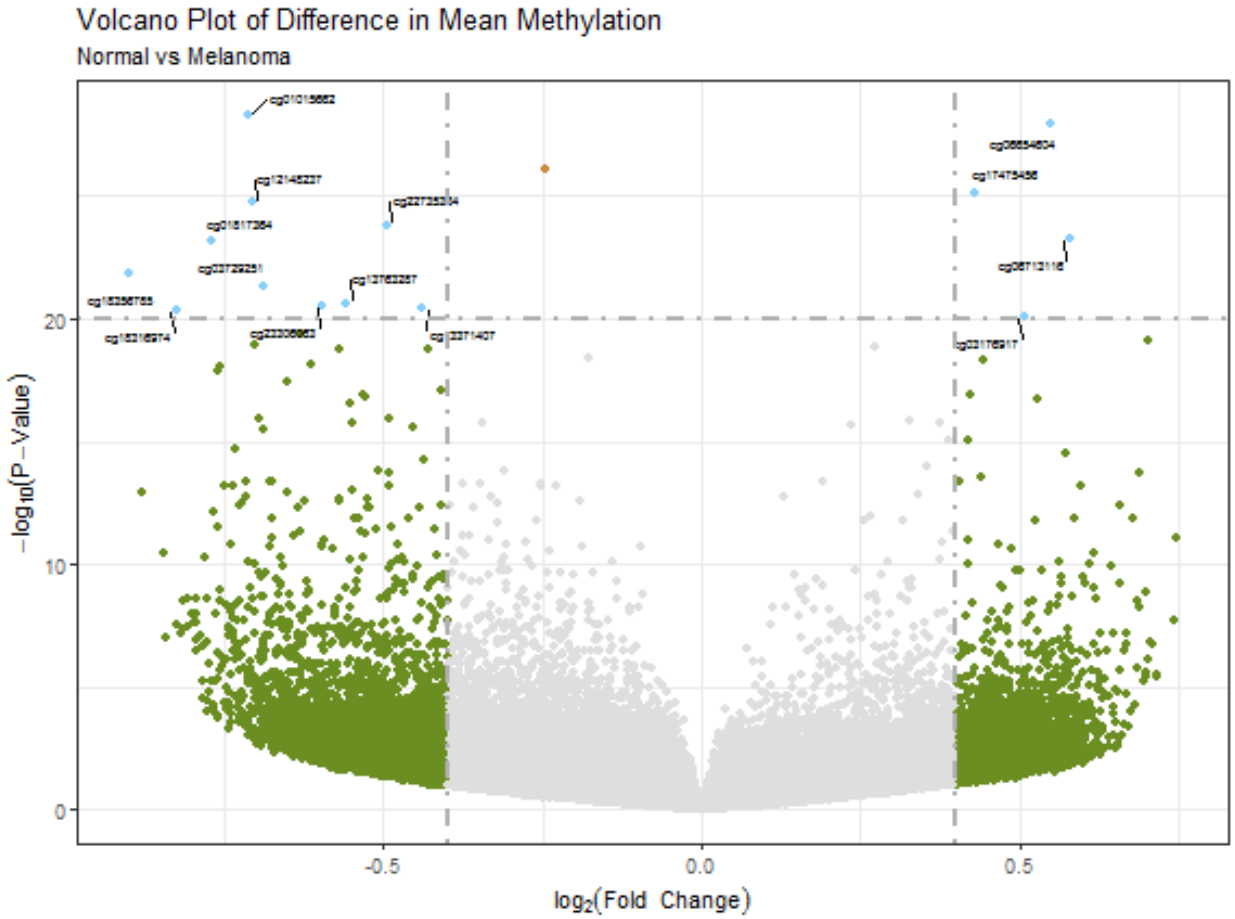

Figure 2: Volcano Plot.

difference of mean DNA methylation ($\log_2$ fold-change between methylation between melanoma and normal melanocyte samples). Those CpG loci with the greatest significant difference in mean methylation exceeding a negative log adjusted p-value threshold of 20 are labeled (Figure 2).

Two methods were used to identify differentially metyhlated regions (DMRs). Using a threshold of 0.3 as a meaningful mean difference in beta-values between the melanoma and normal melanocyte groups, the bumphunting algorithm identified 25,716 bumps (Supplemental Table 1). Using the DMRcate algorithm with a FDR adjusted p-value threshold of 0.05, 238 DMRs were identified (Supplemental Table 2). Among the most significant gene regions identified by both algorithms was a segment on Chromosome 6 from Megabase (Mb) position 31650735 to 31651362 Mb, which consists of a cluster of 21 CpGs (Figure 3). This particular region affects the LY6G5C gene.

After identifying the ten CpG sites with the most significant difference in mean methylation between melanoma and normal melanocyte groups, enrichment analysis was performed to identify over-representation of terms using Gene Ontology and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Analysis using the KEGG database only returned 2 pathways at a significance level of < 0.05 (Table 1), whereas many significant enriched terms were found using GO (Table 2).

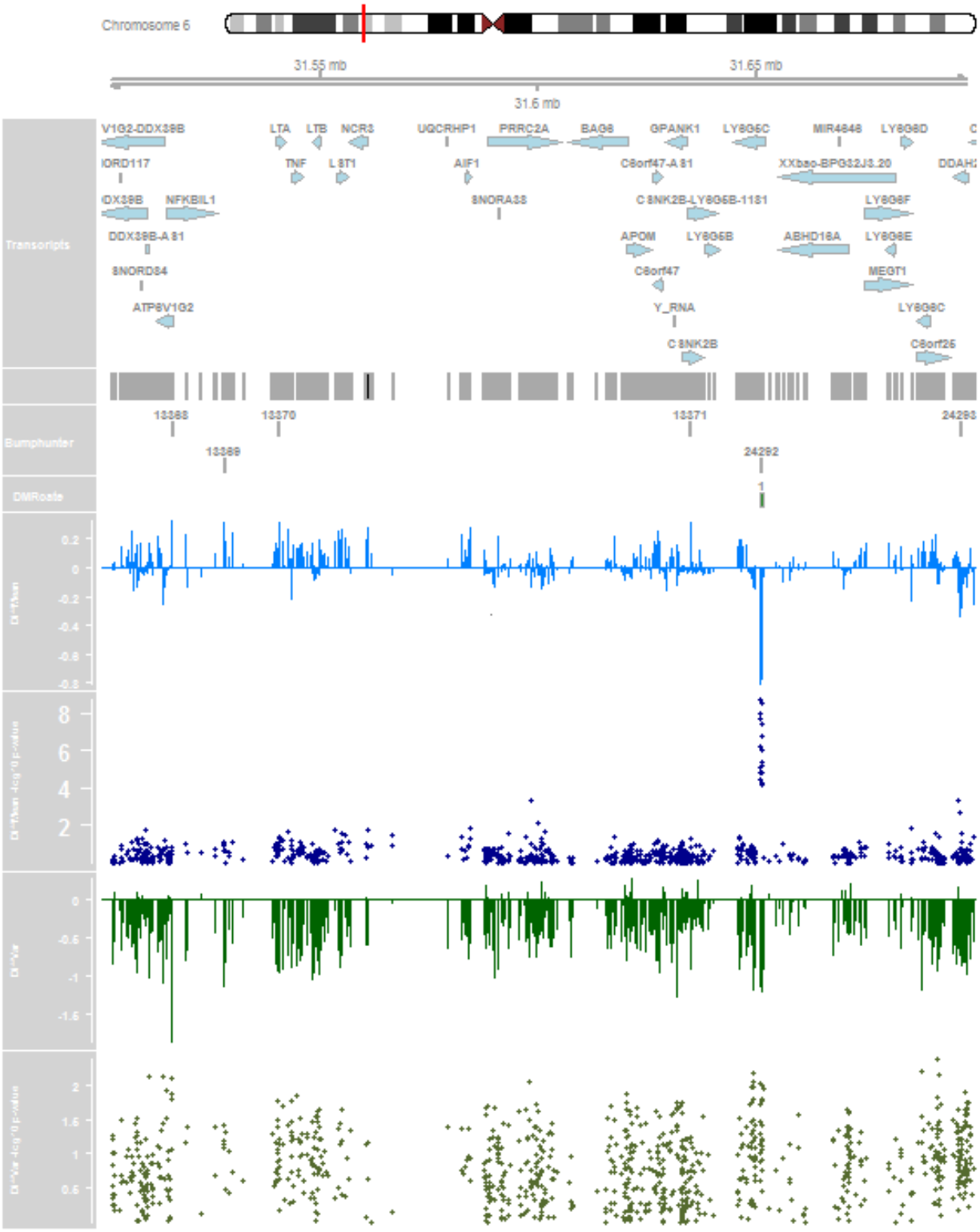

Figure 3: Chromosome 6, 31.5 Mb to 31.7 Mb with related differential methylation information.

Table 1: Top GO entries from Enrichment Analysis

	Term	Ont	No. Genes	DM Genes	P-value	FDR
GO:2001229	negative regulation of response to gamma radiation	BP	1	1	0.0006389	1
GO:0097681	double-strand break repair via alternative nonhomologous end joining	BP	2	1	0.0006389	1
GO:0070105	positive regulation of interleukin-6-mediated signaling pathway	BP	2	1	0.0011660	1
GO:0070103	regulation of interleukin-6-mediated signaling pathway	BP	3	1	0.0013330	1
GO:2001228	regulation of response to gamma radiation	BP	2	1	0.0014914	1
GO:0005958	DNA-dependent protein kinase-DNA ligase 4 complex	CC	4	1	0.0017893	1
GO:0035234	ectopic germ cell programmed cell death	BP	7	1	0.0019251	1
GO:0033152	immunoglobulin V(D)J recombination	BP	4	1	0.0021819	1
GO:0070419	nonhomologous end joining complex	CC	9	1	0.0023365	1
GO:0002326	B cell lineage commitment	BP	4	1	0.0023796	1

Table 2: Top KEGG Pathways from Enrichment Analysis

	Pathway	No. Genes	DM Genes	P-value	FDR
path:hsa03450	Non-homologous end-joining	13	1	0.0023588	0.7783931
path:hsa04110	Cell cycle	124	1	0.0314906	1.0000000

Discussion

This study assessed differential methylation of CpG loci and CpG-rich gene regions by comparing genome-wide methylation patterns of melanoma and normal melanocyte tissue specimens. These findings helped to identify genes and associated pathways affected by abberant metyhlation and supports the role that each may play in melanoma tumorogenesis.

Multi-dimensional scaling techniques, namely PCA and RDA, both revealed that the normal melanocytes tended to cluster together. Pairwise PCA was also performed to identify patterns among each of the four tissue types, i.e. normal, primary melanoma, stage III and stage IV melanoma. This approach demonstrated clustering among the normal melanocyte group and a possible cluster among the primary melanoma tissue samples; however, the clustering was more heterogenous among the stage III and stage IV melanoma cohorts. There may be other clinical and phenotypic features that our data presently lacks that may allow for improved clustering analysis.

Plots of the distribution of β and M values showed that β values had a less variable distribution for characterization of methylated and unymethylated probe data (Supplemental Figure 3). Hence, analysis of differential methylation was performed using β values. Of note, for DNA methylation studies, Zhuang et al. found that M values provided more reliable true positive identification for studies containing small sample sizes, and that both values provided similar reliability for large sample sizes[17].

Analysis of differences in mean methylation between melanoma and normal melanocyte samples revealed 1,126 significantly hypomethylated and 2,265 significantly hypermethylated CpG loci in the melanoma group compared to normal melanocytes. These significantly different CpG loci were present on each chromosome. However, using a more rigorous adjusted p-value threshold provides a smaller range of the most significant DMPs for further study.

Differences in variances among DNA methylation between both groups, melanoma and normal melanocytes, was inconclusive. This may be due to the small sample size of normal melanocytes

and primary melanoma specimens. Alternatively, this could represent that more advanced stages of melanoma contain greater heterogeneity in their DNA methylation patterns.

Given that CpG dinucleotides tend to cluster in CpG islands, most of which occupy promoter regions, region-based differential methylation analysis helps to minimize multiple comparisons and is valuable to identify contiguous genomic regions that differ between both groups. Two different algorithms were employed to identify significant DMRs. These algorithms perform well when DNA methylation levels are correlated across CpG loci[6], as one may expect to find in CpG islands. The bump hunting method, which uses smoothed methylation values and Storey's optimal discovery procedure to control FDR, identified 25,716 bumps using a β value mean difference threshold of 0.3 (Supplemental Table 1). By contrast, the DMRcate algorithm found 238 significant DMRs (Supplemental Table 2). There is strong concordance between the results of both algorithms. The bump hunting algorithm performs poorly where CpG coverage is sparse, such as in non-coding regions, whereas DMRcate will treat even a single CpG as a DMR. Other algorithms, such as Probe Lasso and comb-p, function like DMRcate in this respect[13].

The most significant DMR identified using the DMRcate algorithm was also among the most significant bumps. This region is located on Chromosome 6, from the 31650735 Mb to the 31651362 Mb position. As shown in Figure 3, this gene region, which comprises 21 unique CpG loci, affects the LY6G5C gene, which is a member of the leukocyte antigen-6 (LY6) gene family in the major histocompatibility complex (MHC) class III region on chromosome 6[18]. Each of the 21 CpGs in this DMR is significantly hypermethylated when compared to each locus in the normal melanocyte cohort (Figure 4). LY6G5C, and other members of the LY6 family, have been found to be associated with or drivers of tumorogensis in other malignancies[19]. By contrast, LY6G5C is a favorable prognostic marker for endometrial cancer[21]. Network analysis of the LY6 gene family by Luo et al. showed that LY6 plays a role in many cellular processes, including cell growth, apoptosis, autophagy and immune response[19].

Enrichment analysis was conducted using the most significant CpG loci to identify over-represented terms from the KEGG database and Gene Ontology. Only two significant KEGG pathways - non-homologous end-joining and cell cycle - were found when using a p-value for over-representation of 0.05. By contrast, there were many significantly over-represented terms when using GO, with two terms accounting for non-homologous end joining. These findings suggest that processes related to non-homologous end joining are enriched and play an important role in melanoma. Increased sun and UV exposure are well-documented environmental risk factors for melanoma. Epigenetic processes are also mediated by environmental factors, such as UV exposure. One consequence of environmental stresses is that it may lead to DNA double-strand breaks, which are normally repaired by end joining. Further studies are required to determine if aberrant non-homologous end-joining predisposes to melanoma or if it plays a significant downstream role in driving melanoma oncogenesis. There are also other associated genes and pathways that may influence non-homologous end-joining. For example, the tumor suppressor gene PTEN has been shown to have a direct regulatory role in non-homologous end joining[22].

There is evidence of bias affecting the validity of pathway analysis using genome-wide methylation assays, such as with the 450K array data used in this study. This method assumes that each gene has the same probability of appearing on the pathway list; however, the array contains a varying number of probes for each gene. The Illumina 450K array has 1 to 1288 probes per gene. Thus, those genes with a large number of probes have a far greater probability of falsely appearing enriched[23,24]. Appropriate clinical and biological understanding should be used when performing pathway analysis using genome-wide methylation assay data. Other processing methods to mitigate bias

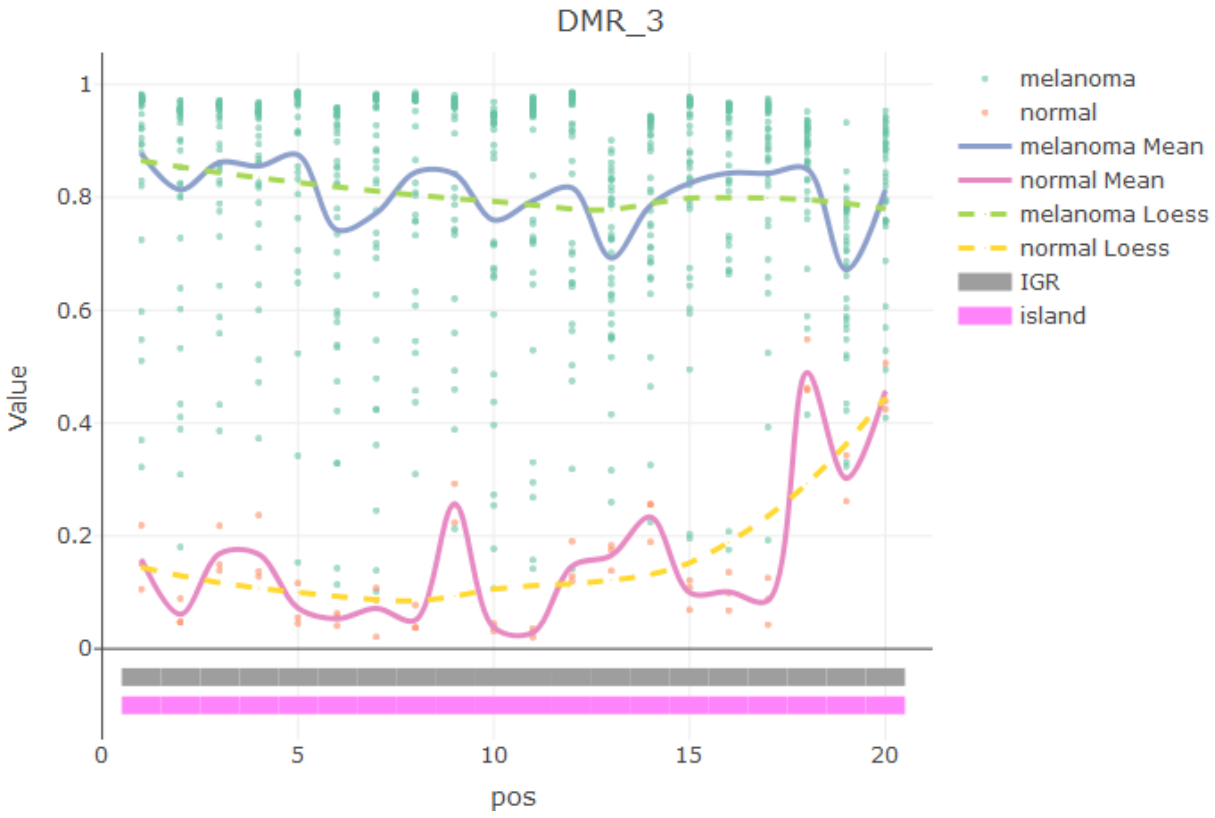

Figure 4: Methylation at the 21 CpGs at DMR located on Chr6:31650735-31651362

should also be employed.

Another limitation of the current study is the small sample size of normal melanocyte specimens. Future studies should include a larger cohort of the healthy tissue population. The array instrument and its output are prone to bias, such as two-color dye bias, fragment length bias, and other probe- and array-specific effects[11]. Methods were used during data processing and analysis to minimize the effect of biases on the results, yet some bias may have persisted and affected the results. Additionally, the empirical Bayes method used for identification of differential methylation at individual CpG loci assumes independence, which may affect some of the results in regions where DNA methylation levels are correlated across CpG loci[6]. However, many of the findings from this study are supported elsewhere in the literature or are otherwise in agreement with the current understanding of melanoma risk factors and its underlying pathophysiology.

Future studies will incorporate other analytic approaches, such as ProbeLasso and comb-p, to identify differentially methylated areas. Given the risk of bias due to the nature of probe design in the Illumina 450K array, best evidence approaches for normalization should be used to provide more robust results. Network analysis and systems biology tools will also be useful to model and evaluate the effects of the involved genes and their downstream molecular processes. Incorporating other phenotypic and clinical features for analysis may also be useful. For example, comparison of each stage of melanoma to identify genes and pathways that may influence disease progression or severity. Furthermore, analysis using clinical data related to treatment response may help reveal if and how epigenetic modifications confer treatment resistance.

Despite recent advancements in melanoma therapy, mortality rates for those with advanced melanoma remain persistently high. With the introduction of more novel therapies, survival has improved; however, nearly all will ultimately develop treatment resistance. Aberrant DNA methylation is commonly seen in melanoma as with other cancers. This mechanism allows malignant cells to bypass normal cell function by influencing transcription regulation and chromatin remodeling. This study identifies the LY6G5C gene of the LY6 gene family and processes involving non-homologous end joining as potentially useful biomarkers and pathways for further study of melanoma pathophysiology and mechanisms of treatment resistance.

Methods

The epigenomic data used in this study was obtained from the National Center for Biotechnology Information's Gene Expression Omnibus (data accessible at NCBI GEO database, GEO accession "GSE44661")[25,26]. This dataset included bisulphite converted DNA from 40 tissue samples ($n_{melanoma} = 37, n_{normal} = 3$) that was assayed using Illumina's Infinium 450K Human Methylation BeadChip. Intra-array normalization procedures to mitigate the biases introduced by the 450K array's two probe design, such as two-color dye-bias adjustment, were not performed. The data had already undergone pre-processing from the raw assay output by the experimenters. Per the study's documentation, pre-processing was conducted by the original study authors using the methylumi package in Bioconductor. Evaluation of the distribution of β and M values was conducted to assess data quality for downstream analysis. Probes measuring single nucleotide polymorphisms (SNPs) and those loci with missing information were excluded as they are insufficient to measure DNA methylation level[15]. Annotation was performed using the hg19 reference genome.

To identify differences between melanoma and normal melanocyte samples, dimensionality reduction and ordination techniques were applied to the processed data. These methods included PCA

of melanoma and non-melanomatous samples, Redundancy analysis and PCA of pairwise comparisons of normal melanocytes, primary melanoma, stage III melanoma and stage IV melanoma involving brain metastases[27–29].

Analyis of methylation of both individual CpG sites and clustered CpG regions was performed; β values were used for analysis. Differentially methylated positions (DMPs) were identified using differences in the means and differences in variance of methylation at CpG sites[27,30]. DMP analysis was performed using techniques commonly used to detect differential gene expression. This two-step process begins by fitting a linear model to estimate β_i^* between both groups (melanoma vs normal melanocytes) at each ith locus with a null hypothesis that the estimated β_i^* for both groups is the same. This is followed by an empirical Bayes approach to decrease the estimated sample variances to a pooled estimate, which improves inference with small sample sizes[6,31]. To control False Discovery Rate, p-values were adjusted using the Benjamini-Hochberg method.

Differentially methylated regions were analyzed using two methods: bump hunting[32,33] and DMRcate[13]. The most significant differentially methylated CpG locations were identified and pathway analysis using GO and KEGG was performed for the associated gene regions[24].

Code for the aforementioned analysis is available online at https://github.com/mirzask/melanoma-epigenetics.

References

1. Gallagher, S. J., Tiffen, J. C. & Hersey, P. Histone Modifications, Modifiers and Readers in Melanoma Resistance to Targeted and Immune Therapy. *Cancers* **7**, 1959–1982 (2015).

2. Schinke, C. *et al.* Aberrant DNA methylation in malignant melanoma. *Melanoma Research* **20**, 253–265 (2010).

3. Greenberg, E. S., Chong, K. K., Huynh, K. T., Tanaka, R. & Hoon, D. S. B. Epigenetic biomarkers in skin cancer. *Cancer Letters* **342**, 170–177 (2014).

4. Besaratinia, A. & Tommasi, S. Epigenetics of human melanoma: Promises and challenges. *Journal of Molecular Cell Biology* **6**, 356–367 (2014).

5. Henriksen, T., Dahlback, A., Larsen, S. H. & Moan, J. Ultraviolet-radiation and skin cancer. Effect of an ozone layer depletion. *Photochemistry and Photobiology* **51**, 579–582 (1990).

6. Li, D., Xie, Z., Le Pape, M. & Dye, T. An evaluation of statistical methods for DNA methylation microarray data analysis. *BMC Bioinformatics* **16**, 217 (2015).

7. Jones, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**, 484–492 (2012).

8. Kuan, P. F., Wang, S., Zhou, X. & Chu, H. A statistical framework for Illumina DNA methylation arrays. *Bioinformatics* **26**, 2849–2855 (2010).

9. Aryee, M. J. *et al.* Accurate genome-scale percentage DNA methylation estimates from microarray data. *Biostatistics* **12**, 197–210 (2011).

10. Micevic, G., Theodosakis, N. & Bosenberg, M. Aberrant DNA methylation in melanoma: Biomarker and therapeutic opportunities. *Clinical Epigenetics* **9**, 34 (2017).

11. Siegmund, K. D. Statistical approaches for the analysis of DNA methylation microarray data. *Human Genetics* **129**, 585–595 (2011).

12. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).

13. Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics & Chromatin* **8**, 6 (2015).

14. Morris, T. J. & Beck, S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods* **72**, 3–8 (2015).

15. Wang, D. *et al.* IMA: An R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics* **28**, 729–730 (2012).

16. Wang, T. *et al.* A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. *Epigenetics* **10**, 662–669 (2015).

17. Zhuang, J., Widschwendter, M. & Teschendorff, A. E. A comparison of feature selection and classification methods in DNA methylation studies using the Illumina Infinium platform. *BMC Bioinformatics* **13**, 59 (2012).

18. *LY6G5C lymphocyte antigen 6 family member G5C*. (NCBI - Gene).

19. Luo, L. *et al.* Distinct lymphocyte antigens 6 (Ly6) family members Ly6D, Ly6E, Ly6K and Ly6H drive tumorigenesis and clinical outcome. *Oncotarget* **7**, 11165–11193 (2016).

20. Dai, W. *et al.* Comparative methylome analysis in solid tumors reveals aberrant methylation at chromosome 6p in nasopharyngeal carcinoma. *Cancer Medicine* **4**, 1079–1090 (2015).

21. *LY6G5C - Endometrial cancer*. (Human Protein Atlas).

22. Sulkowski, P. L., Scanlon, S. E., Oeck, S. & Glazer, P. M. PTEN Regulates Nonhomologous End Joining By Epigenetic Induction of NHEJ1/XLF. *Molecular cancer research: MCR* **16**, 1241–1254 (2018).

23. Geeleher, P. *et al.* Gene-set analysis is severely biased when applied to genome-wide methylation data. *Bioinformatics (Oxford, England)* **29**, 1851–1857 (2013).

24. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics (Oxford, England)* **32**, 286–288 (2016).

25. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* **30**, 207–210 (2002).

26. Marzese, D. M. *et al.* Epigenome-wide DNA methylation landscape of melanoma progression to brain metastasis reveals aberrations on homeobox D cluster associated with prognosis. *Human Molecular Genetics* **23**, 226–238 (2014).

27. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47 (2015).

28. Oksanen, J. *et al.* Vegan: Community Ecology Package. (2018).

29. Ruiz-Arenas, C. & Gonzalez, J. R. MEAL: Perform methylation analysis. (2018).

30. Phipson, B. & Oshlack, A. DiffVar: A new method for detecting differential variability with application to methylation in cancer and aging. *Genome Biology* **15**, 465 (2014).

31. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, Article3 (2004).

32. Jaffe, A. E. *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *International Journal of Epidemiology* **41**, 200–209 (2012).

33. Jean-Philippe Fortin & Hansen, K. D. Minfi Tutorial. (2014).