

# **Nucleosome repositioning in glioblastoma**

Varvara Koraki Folli

A thesis submitted for the degree of MSD in Molecular  
Medicine

School of Life Sciences

University of Essex

November 2020

# Acknowledgements

First and foremost, I would like to thank my supervisor Dr Vladimir Teif for his consistent guidance and valuable help he has given me the past two years.

Additionally, I would like to recognize Dr Victor Zhurkin (NIH/NCI) for his help and feedback, as well as all our other collaborators.

Moreover, I would like to acknowledge Dr Christopher Clackson, Dr Hulkar Mamayusupova, Kristan Piroeva, and all the Teif Lab members for their advice and precious feedback on this project.

Finally, I would like to thank my mother, Mirela Qyqe, for her continuous support and words of encouragement throughout my university years.

# Abbreviations

**ChIP-seq:** Chromatin Immunoprecipitation followed by Sequencing

**cfDNA:** cell-free DNA

**ctDNA:** circulating tumour DNA

**CSF:** cerebrospinal fluid

**IDH:** isocitrate dehydrogenase

**GBM:** Glioblastoma Multiforme

**GSC:** Glioblastoma stem cell

**MES:** Mesenchymal

**MNase-seq:** Micrococcal Nuclease Sequencing

**NDR:** Nucleosome Depleted Region

**PCA:** Principal Component Analysis

**PTM:** Post-transcriptional Modification

**RTK:** receptor tyrosine kinase

**TF:** Transcription Factor

**TSS:** Transcription Start Sites

# Contents

<b>Abstract</b> .....	<b>1</b>
<b>1. Literature review</b> .....	<b>2</b>
1.1. Organisation of the eukaryotic genome in chromatin .....	2
1.2. Nucleosome positioning.....	3
1.3. Nucleosome positioning and cancer .....	10
1.4. Nucleosome positioning and cell-free DNA for cancer diagnostics .....	11
1.5. Glioblastoma Multiforme .....	15
1.6. Nucleosome positioning in GBM.....	20
1.7. Aims.....	21
<b>2. Methods</b> .....	<b>23</b>
2.1. MNase-seq and cfDNA datasets .....	23
2.2. External experimental datasets.....	24
2.4. DNA fragment size analysis.....	27
2.5. Calculation of average nucleosome occupancy profiles.....	27
2.6. Calculation of A/T/C/G nucleotide frequency .....	28
2.7. Lost, gained, more fuzzy and less fuzzy nucleosomes .....	28
2.8. Enrichment of lost/gained, more and less fuzzy nucleosomes.....	28
2.9. Alu repeats analysis.....	29
2.10. Gene ontology analysis.....	29
2.11. Principal Component analysis (PCA).....	30
<b>3. Results</b> .....	<b>31</b>
3.1. Distribution of DNA fragment lengths .....	31
3.2. Nucleosome occupancy profiles in GBM patients around TFs bound in GBM-related cell lines. ....	33
3.2.1. Average chromatin profiles differences in healthy and GBM samples based on MNase-seq data. ....	33
3.2.2. Average chromatin profiles differences in healthy and GBM samples based on cfDNA. ....	34
3.3. Nucleosome repositioning at genomic DNA repeats .....	37
3.3.1. Average nucleosome occupancy around ALU repeats .....	37
3.3.2. Average nucleosome occupancy profiles at size separated ALU repeats .....	37
3.3.3. Average nucleosome occupancy patterns around ALU repeats resemble those around TSS. ....	39

<b>3.4. Interplay of nucleosome repositioning and histone modifications in GBM.....</b>	<b>41</b>
3.4.1. <i>Nucleosome occupancy around domains with different histone modifications in glioblastoma .....</i>	<b>41</b>
3.4.2. <i>Chromatin differences between glioblastoma and neuroblastoma .....</i>	<b>43</b>
<b>3.5. Targeted loss and gain of nucleosomes from genomic regions .....</b>	<b>45</b>
3.5.1. <i>GBM-specific nucleosome loss, gain and fuzziness.....</i>	<b>45</b>
3.5.2. <i>Enrichment of lost and gained nucleosome regions at TFs.....</i>	<b>49</b>
3.5.3. <i>Enrichment of lost and gained nucleosome regions at different genomic regions.</i>	<b>49</b>
3.5.4 <i>Gene ontology analysis of promoters that gained and lost nucleosomes in GBM ..</i>	<b>51</b>
3.5.5. <i>Classification of samples using Principal Component Analysis based on nucleosome occupancy at GBM-sensitive genomic regions.....</i>	<b>52</b>
<b>3.6. Average nucleotide frequency profiles along nucleosomes from brain tissues.....</b>	<b>53</b>
<b>4. Discussion.....</b>	<b>56</b>
4.1. <b>Interplay of TFs and nucleosome positioning in GBM.....</b>	<b>56</b>
4.2. <b>Nucleosome occupancy changes around ALU repeats.....</b>	<b>62</b>
4.3. <b>Nucleosome occupancy changes in relation to histone modifications. ....</b>	<b>64</b>
4.4. <b>Nucleotide patterns along nucleosomes in MNase-seq and cfDNA .....</b>	<b>64</b>
4.5. <b>GBM-specific nucleosome gain and loss. ....</b>	<b>66</b>
4.6. <b>Diagnosis of patients based on PCA of nucleosome occupancy .....</b>	<b>70</b>
<b>References .....</b>	<b>72</b>

## Abstract

Glioblastoma multiforme (GBM) is the most malignant type of brain cancer in adults, with highly aggressive behaviour and low patient survival. The glial cells that are responsible for forming GBM show abnormalities in gene expression that can be caused by genetic or epigenetic causes. In particular, the positioning of nucleosomes along the genome determines the accessibility of DNA to regulatory proteins and therefore modulates gene expression. In this project I investigate nucleosome positioning in relation to other features, such as transcription factor (TF) binding, ALU repeats and histone modification changes. I show that the relationship between TF binding and nucleosome positioning in brain tissues, as well as cell-free DNA (cfDNA) from peripheral blood, is a highly promising system to investigate the deregulation of chromatin in cancer cells and diagnose GBM patients. The computational analysis performed here is based on experimental MNase-seq data in paired normal and tumour brain tissues from GBM patients as well as cfDNA from GBM patients and healthy individuals. I have determined GBM-specific changes in nucleosome occupancy profiles around binding sites of 20 glioblastoma-related TFs. Major changes of nucleosome positioning were found around binding sites of CTCF, RBPJ, MYC, KLF9 and JMJD6, which I propose for the role of new liquid biopsy markers. Moreover, I studied differences in nucleosome occupancy at different classes of genomic features such as ALU repeats, which showed unexpected resemblance to transcription start sites. Finally, I compiled a dataset of genomic regions that underwent GBM-specific nucleosome repositioning and showed that Principal Component Analysis (PCA) based on nucleosome occupancy values in these regions could be used for patient diagnosis. The results of my work have implications for future development of liquid biopsy assays for patient stratification based on nucleosomal DNA.

# 1. Literature review

## 1.1. Organisation of the eukaryotic genome in chromatin

The eukaryotic genome is compacted in the form of chromatin inside the cell nucleus. The fundamental structural unit of chromatin is the nucleosome which is composed of 147 base pairs of negatively charged DNA coiled around a positively charged histone octamer. The histone octamer contains two histone tetramers that both include two copies of histones H2A, H2B, H3 and H4 (Kornberg and Lorch, 1999; Fenley et al., 2018). Nucleosomes are further locked through the histone H1 that is bound to the nucleosome entry/exit, and the linker DNA connects nucleosome particles to each other. This creates the so-called “beads-on-a-string” pattern (Baldi, 2020). The constant dissociation and binding of nucleosomes and other proteins to DNA give rise to a very dynamic structure of chromatin (Kharerin et al., 2016).

Although initially nucleosomes were believed to provide a universal, nonspecific coating of genomic DNA, it has long been known that nucleosomes occupy favoured positions throughout the genome (Li et al., 2016). Nucleosomes can act as transcription repressors by occupying space and blocking the transcriptional machinery from binding to promoters. High-resolution, genome-wide analyses have revealed a common pattern: nucleosomes are depleted at many active enhancer, promoter, and terminator regions, and they typically occupy preferred positions in gene and non-gene regions. Nucleosome depletion has been detected upstream to transcription start sites (TSS), making it accessible to transcription factors and other proteins relating to gene expression activation (Li et al., 2016; Struhl and Segal, 2013). Nucleosomes have a few purposes in the genome. Firstly, through the high compacting and bending of DNA they hold  $\sim$ about 2 metres of DNA in the nucleus providing efficient storage and protection from hazards (Rudnizky et al., 2017). Secondly, the presence of nucleosomes can stop the interaction of DNA with other regulatory proteins. The DNA included in the nucleosome has

limited accessibility, and therefore nucleosomes can determine binding of DNA-binding proteins to regulatory regions. Moreover, nucleosomes often contain epigenetic marks, such as post-translational modified histones (example: methylation or acetylation), methylated DNA and/or histones variants (example: H2A.Z instead of the common H2A). These marks can indicate transcriptional status (Chereji and Clark, 2018). The above effects can indicate the importance of nucleosome positioning and will be thoroughly described in the context of positioning determinants below.

Regulatory proteins determine the way the genome is expressed and translated. Therefore, access to the DNA in desired regions is needed for the manipulation of gene regulation (Fenley et al., 2018). Distances from nucleosome to nucleosome can be different and their positioning along the DNA can change but is not at all random, since some location can be pre-determined (Teif and Clarkson, 2019). Different packaging of DNA can occur in different cell states and we can distinguish two functionally different territories, heterochromatin and euchromatin. Heterochromatin is characterised as extremely condensed and genes located there are transcriptionally silenced, while euchromatin is less condensed and contains genes that are transcriptionally more active. Different nucleosome modifications can differentiate between the chromatin states; for example, heterochromatin is associated with histone H3K9 methylation (H3K9me, H3K9me2, H3K9me3) while euchromatin shows enrichment of histones H3, H4 and H3K4 methylation (H3K4me) (Tamaru, 2010).

## **1.2. Nucleosome positioning**

The term “nucleosome positioning” describes the locations of nucleosomes in the genome. Nucleosome positioning is involved in epigenetic regulation of gene expression. As mentioned above, the genome-wide patterns of nucleosome positioning are not random and can be determined by several factors. That can be a combination of DNA sequence, ATP- dependent



nucleosome remodelling enzymes, competition of nucleosome and transcription factors, and histone modifications. All these determinants combined can influence nucleosome positioning and thus gene expression and most DNA-related processes (Struhl and Segal, 2013).

*1.2.1. DNA sequence affected nucleosome positioning.* Some positions of nucleosomes are naturally encoded in the DNA. The histone octamer's ability to bind DNA to form nucleosomes depends on the specific DNA sequence. It has been proven that positions of nucleosomes are affected during the cell cycle, which shows that the DNA sequence highly affects the organisation of nucleosomes. Current studies show that DNA sequence acquires effect over ~9% of constant nucleosome arrangement without the influence of other factors that may contribute to positioning (Segal et al., 2006; Teif and Clarkson, 2019).

Preferences in DNA sequence have been observed and certain motifs occurring in the nucleosome are proven to be essential for the formation and packaging of chromatin (Cui and Zhurkin, 2010). Certain sequences are nucleosome favourable or unfavourable by default. For instance, positioning of nucleosomes at genomic regions is an example of DNA-sequence favoured nucleosome positioning (Teif and Clarkson, 2019). A study by Tompitak et al (2017) proves the existence of nucleosome-favouring sequences in promoters showing the DNA-sequence favouring by default. Moreover, DNA sequence determinants of nucleosome formation depend on two main factors. First, the repetitions of specific dinucleotides along the nucleosomal DNA sequence give it the ability to bend around histones. This bending of DNA occurs at DNA helical repeats, at about every 10bp, that results in the optimal formation of nucleosomes (Segal et al., 2006; Struhl and Segal, 2013).

The second factor that can affect nucleosome positioning are longer stretches of specific sequences and in particular poly(dA:dT) and poly(dG:dC) tracts (Struhl and Segal, 2013). These tracts are intrinsically stiff and poly(dA:dT) is characterised as a nucleosome

disfavouring sequence compared to poly(dG:dC) (Raveh-Sadka et al., 2012; Zhang and Cui, 2014). Poly(dA:dT) tracts are abundant in eukaryotes and are associated with nucleosome-depletion. These sequences are frequent over promoter regions and often exhibit increased accessibility for transcription factor binding. Raveh-Sadka et al (2012) demonstrated the importance of the poly(dA:dT) sequences in nucleosome positioning. They showed that affected nucleosome positioning has significant effect on transcription by altering transcription factor binding.

*1.2.2. ATP-dependent remodelling enzyme-affected nucleosome positioning.* Although DNA sequence influences nucleosome positioning, ATP-dependent chromatin remodelers have a critical role in regulating nucleosome positioning and chromatin organisation. Nucleosome positioning is an active process that requires ATP-dependent enzymes. These enzymes use energy derived by ATP and can break and create bonds between histones and DNA (Piatti et al., 2011). Remodelers are responsible for chromatin reconstruction by movement of nucleosomes and transcriptional control, thus their knockdown can cause drastic changes on gene expression (Moshkin et al., 2007). In their research, Moshkin et al (2007) showed that knockdown of several remodelers such as SWI/SNF had important effects in gene regulation. Chromatin remodelers act through ATP-mediated nucleosome “sliding” and removal that drive nucleosome positioning. According to research conducted by Padinhateeri and Marko (2011), changes in remodelling activity alter nucleosome dynamics and influence promoter regions. After testing the remodeler-dependent nucleosome positioning mechanism, Rippe et al (2007) concluded that remodelers with binding preferences for specific nucleosome regions determine chromatin structures directly depended on the specific DNA sequence and specify accessibility of protein to DNA.

*1.2.3. Nucleosome and transcription factor competition and nucleosome positioning.* One of the major determinants of nucleosome positioning is the competition with Transcription Factors (TFs). TFs are DNA-binding proteins that regulate transcription by activation or inactivation of RNA polymerase function. Nucleosome stability in regions is set to be influenced by RNA polymerase expression (Schones et al., 2008). Therefore, these DNA-binding proteins have the ability to displace nucleosomes at regions that control expression and affect stability of nucleosomes (Zhu et al., 2018). Both nucleosomal histones as well as TFs are involved in gene regulation, and competition to bind on the DNA strand is observed when genes are switched on (Joseph et al., 2017).

A common example of nucleosome repositioning induced by TFs is PHO5 gene expression. The role of TF-mediated disassembly was demonstrated for PHO5 gene expression in yeast by Kharerin et al (2016). In most cases, nucleosomes form at high occupancy at the PHO5 promoter. However, presence of the TF Pho4p exhibits disruption of the nucleosome array and as a result, gene expression occurs. In the study, they proved that Pho4p does not just bind, but it establishes Pho4p-mediated “local remodeling activity” on promoter nucleosomes. This activity gives rise to active promoter states. In addition, research by Zhu et al (2018), found that most TFs have a dissociation effect on nucleosomes upon binding, yet, they identified TFs such as TBX2, that have stabilizing effects on nucleosomes. Overall, effects of TFs on the stability of nucleosomes can have major impact on nucleosome positioning. Thus, the dynamic binding of TFs has a crucial effect on gene expression since increasing or reducing access to TF binding sites can impact expression of genes and function of regulatory elements (Schones et al., 2008; Lövkvist et al., 2017).

*1.2.4. Histone and DNA modification-affected nucleosome positioning.* Histone modifications along with DNA methylation and nucleosome positioning are critical for the packaging of nucleosomes and regulation of gene expression (Portela and Esteller, 2010). DNA methylation

is a common determinant and contributor of nucleosome positioning in eukaryotes. Periodic DNA methylation patterns are common in the linkers of gene bodies and are a factor in the positioning of nucleosome. Moreover, modifications such as 5-methylcytosine and 5-formylcytosine affect flexibility and stability of nucleosomes (Chereji and Clark, 2018).

Entry to nucleosomal DNA is extremely hard while chromatin is condensed and that influences TFs. Access is only possible through histone modifications (Lövkvist et al., 2017). Histones can have modified conformation and function through biochemical modifications that occur on the proteins called post-transcriptional modifications (PTMs) (Onufriev and Schiessel, 2019). Amino acids of histone tails can be affected by PTMs, like methylation, acetylation, and phosphorylation. As an example, it was mentioned above that heterochromatin regions show low acetylation level. An explanation for this phenomenon is the typical histone modification H3K9me3 that recruits more modifying factor and spreads H3K9me3 modification (Wang et al., 2014b).

DNA methylation and histone PTMs control the formation of transcriptional protein complexes. They can cause alterations in structures and biological responses in chromatin which can lead to differential gene expression (Fenley et al., 2018; Berger, 2007). Moreover, histone modifications are crucial in chromatin organisation and repair. The position of DNA damage can be identified by histone modification to recruit remodelers to open chromatin and start the repairing process (Chereji and Clark, 2018; Williamson et al., 2012). Through modification of PTMs depending on the chromatin state (heterochromatin or euchromatin) we can predict transcription factor binding and interactions on nucleosomal DNA and histones, and therefore, nucleosome positioning.

*1.2.5. Nucleosome occupancy vs nucleosome positioning.* Nucleosome positioning is related to nucleosome occupancy, but they are both distinct. Nucleosome occupancy for an individual

cell is the probability for a given DNA base pair to be inside a nucleosome; for a bulk experiment, nucleosome occupancy is defined as the relative number of DNA sequence reads covering a given base pair in comparison with genome-average. Some regions like promoters and enhancers show depletion of nucleosomes or in other words low nucleosome occupancy (Struhl and Segal, 2013). Decrease in occupancy at nucleosome depleted regions (NDRs), such as promoters, is usually associated with the increase of gene expression. Increase or decrease of nucleosome occupancy in NDRs can indicate gene silencing or gene activation accordingly (Vasseur et al., 2016). Thus, nucleosome occupancy is as critical to biological functions such as nucleosome positioning (Struhl and Segal, 2013). For example, Lam et al. (2008) manipulated promoter sites to prove that the interplay between chromatin structure and binding-site affinity can allow promoters to respond to signals in a unique way and tailor gene expression.

*1.2.6. Nucleosome stability and fuzziness.* In addition to nucleosome occupancy, nucleosome positioning can be characterised by the stability of nucleosomes. Stability of nucleosomes is identified by the comparison nucleosome occupancy levels at a specific genomic region that is calculated at different phases of chromatin digestion. Highly similar levels of nucleosome occupancy around these regions indicate high nucleosome stability (Teif and Clarkson, 2019). Moreover, the term nucleosome fuzziness is used to describe the standard error of calculating nucleosome occupancy at a specific region-based averaging of replicate experiments (Teif and Clarkson, 2019; Vainshtein et al., 2017).

*1.2.7. Nucleosome positioning and protein interaction measurement.* Defining the positions that nucleosomes overtake and their interactions with proteins in the genome is very important in order to study and understand the regulation of transcription by chromatin.

1) Nucleosome positioning measurement: In this project we used two types of data to complete experimental analysis: MNase-seq and MNase-seq assisted H3.

- *MNase-seq*: Micrococcal nuclease digestion followed by high-throughput sequencing (MNase-seq) is the most popular technology used to profile nucleosome organisation in the genome. For this technique, Micrococcal nuclease (MNase), an endo/exonuclease that can cut DNA and digest RNA, is used to make incisions between nucleosomes creating small fragments (Hoeijmakers and Bártfai, 2018). MNase-seq makes use of Next-Generation Sequencing methods to obtain reads from the nucleosomal DNA fragments in order to define genome-wide nucleosome positioning and occupancy. It is extremely important to take into account the biases that are included when using MNase, such as the preference of the enzyme to AT-rich regions. As a result, this bias may wrongly interpret the effect of AT-rich sequences for nucleosome formation, therefore quality must be checked when using the MNase-seq technique (Hoeijmakers and Bártfai, 2018). According to Gutierrez et al. (2017), correcting this bias has been and still is a challenge, with no trustworthy method for sequence-bias corrections (Gutiérrez et al., 2017).
- *MNase-assisted histone H3 ChIP-seq*. During MNase-seq protocols protein-DNA interactions are “fixed” to avoid nucleosome movement. This procedure is not specific to nucleosomal protein and therefore, non-histone proteins maybe be picked up during MNase-seq digestion. To avoid this limitation, MNA-seq-assisted histone H3 ChIP-seq is used to specifically select nucleosomal DNA. In this case, during MNase-seq, chromatin is immunoprecipitated with antibodies against histone H3 and only DNA fragments connected to histone H3 are selected (Hoeijmakers and Bártfai, 2018).

- 2) Protein-DNA interactions measurement using ChIP-seq analysis: Chromatin immunoprecipitation Sequencing (ChIP-seq) is currently the method used to measure genome-wide protein location. ChIP-seq offers high resolution, genome-wide analysis of DNA-protein interactions. During the ChIP approach, cross-linked proteins and DNA elements create complexes that are enriched with antibody. The DNA fragments of interest are then sequenced and mapped to a genome (Park, 2009).

*1.2.7. Alu repeats and nucleosome positioning.* Alu repeats are one the most abundant short interspersed repeated sequences found in the human genome. These sequences were considered to be junk DNA since their function could not be related to any cell process but findings revealed their importance in the regulation of genes, developmental processes and overall genome integrity (Di Ruocco et al., 2018; Cordaux et al., 2010). Their wide spreading over the genome has been related to cancer. Mutations of the tumour suppressor gene p53 together with DNA methylation relates to increase Alu element transcription in colorectal, ovarian and breast cancer (Di Ruocco et al., 2018). In nucleosome organisation, Alu repeat elements have been reported to be related to nucleosome repositioning effects. Strong nucleosomes with preference for young Alu repeats show increased mutation rates and the combination of strong nucleosome positioning and hypermutation of Alu is facilitating the expansion of Alu across the genome in cancer (Li and Luscombe, 2020).

### **1.3. Nucleosome positioning and cancer**

The epigenetic dysregulation of cancer is a topic that has gained momentum among cancer researchers. Epigenetic mechanisms focused on transcriptional regulation remain uninvestigated thus, understanding of the importance of chromatin organisation and specifically nucleosome positioning in cancer is growing steadily (Chereji and Clark, 2018). Epigenetic features such as histone modifications, DNA methylation and nucleosome

positioning are found to be modified in all human cancers. Such modifications can alter gene expression during cancer initiation and progression (Li et al., 2016). For example, Shanmugam et al (2018) state reduction of histone H3K4me3 and H3K9ac mark levels are observed in promoters of target genes in melanoma cell lines. Similarly, they report H4K4me3 expression associated with tumour progression in MCF7 breast cancer cells.

Even though nucleosome positioning in the same cell type can be different from person to person, cancer cells can have different chromatin structure distinguishable from that in healthy people. Positioning of nucleosomes can be used as a cancer indicator in both a genetic and epigenetics level as it exhibits differences in chromatin and DNA methylation. A cancerous genome shows alternated nucleosome occupancy and positioning and epigenetic modification such as post-transcriptional modifications, chromatin-remodelers and TFs can be the reason for the development of cancer. According to Lay et al (2015), DNA methylation and histone modifications highly affected patterns of nucleosome positioning were identified in HCT111 colon cancer cells. Moreover, the importance of nucleosome positioning was demonstrated in bladder cancer where an aberrant function of remodelers and histone modifiers was observed (Gui et al., 2011). Specifically, in GBM modification of histones and chromatin remodelers are involved in the maintenance of stem cells and are major contributors to GBM treatment failure Romani et al (2018). These results can demonstrate the importance of nucleosome positioning during cancer developments and its role in different cancers will be discussed further in the report.

#### **1.4. Nucleosome positioning and cell-free DNA for cancer diagnostics**

Nucleosome positioning and genome-wide abundance of the size of cell-free DNA fragments have recently discovered to be correlated. Sequencing of plasma cfDNA can provide genome-



wide nucleosome occupancy that highly correlates with gene structure and expression observed in cells (Snyder et al., 2016).

This relationship is mostly contributed by the structure of cfDNA. cfDNA is found in the bloodstream and it consists of small double-stranded DNA fragments, less than 200bp long. It has been proposed to be secreted mostly by apoptotic cells with low normal cell tissue contribution. cfDNA shows a short life span in the blood indicating its ongoing secretion by cells and immediate degradation. cfDNA fragment sizes are close to the size of nucleosomes and chromatosomes (linker histone and nucleosome), at about 147bp and 167bp accordingly (Snyder et al., 2016). Moreover, it can be produced by normal as well as tumour cells during cell death and the tumour-rich fragments can contain important information of the properties of cancer. Through whole-genome sequencing of cfDNA, tumour-specific mutations, methylation and nucleosome positioning can be identified allowing scientists to identify a tumour's location and monitor its progression (An et al., 2019). For example, in lung cancer patients blood plasma can be used to detect mutation in the epidermal growth factor since tumour tissue is limited (Nakamura et al., 2012).

The ability to use blood plasma in order to diagnose and monitor diseases has been a great advantage in modern medicine. Traditional methods for cancer diagnostics such as tissue biopsies or X-rays can be avoided through the use of non-invasive liquid biopsies (Snyder et al., 2016). Liquid biopsy or fluid biopsy can be defined as the sampling and analysis of cfDNA found in biological fluids such as plasma, saliva, urine and the cerebrospinal fluid. Monitoring of cancers through cfDNA is a new alternative biopsy that is possible through the identification of mutations or aberrant chromosome numbers and it can be used to tackle limitations of tissue-base genetics (Snyder et al., 2016; McEwen et al., 2020).

The amount of cfDNA in plasma of cancer patients depends on many factors such as tumour stage, location and response to therapy. Quantification of overall cfDNA levels in cancer patients is not useful for diagnosis and monitoring since the concentrations of total cfDNA can be similar to healthy individuals (García-Olmo et al., 2013). This indicates that the usage of mutant DNA can be a better method for patient monitoring, a method that was demonstrated successfully by Diehl et al (2008) in metastatic colorectal cancer.

In cancer patients, cfDNA found in the circulation also contains cfDNA originating from tumour cells, known as circulating tumour DNA (ctDNA) (Murtaza and Caldas, 2016). ctDNA is shed into the bloodstream by cancer cells after apoptosis and small nucleosome fragments are secreted into the blood stream (Elazezy and Joosse, 2018). Analysis of these small fragments can occur through the mapping of the fragments to a genome through ChIP-seq for patient profiling and detecting of tumour specific changes in the DNA.

In their research, Snyder et al. (2016) presented genome-wide mapping of sequenced cfDNA fragments that promoted the understanding of the relationship between nucleosome positioning and cfDNA fragments. Similar nucleosome positioning landmarks were observed in cfDNA as well as direct footprint of TF occupancy. According to their research, this is a unique application to tumour classification by using epigenetic markers found in cfDNA fragment patterns.

Shorter fragment sized cfDNA has been associated with several types of cancers (Spindler et al., 2014; Singh et al., 2015) For example, Lapin et al (2018), demonstrated shorter cfDNA fragment sizes in pancreatic cancer and suggested increased levels of tumour DNA in patients with cancer but it cannot be used independently as a prognostic marker in contrast to cfDNA levels. Prognostic impact was calculated through Cox regression analysis and discovered that cfDNA level is a prognostic factor but this after testing cfDNA fragments sizes without the

presence of cfDNA levels and concluded that cfDNA cannot be used independently. The authors are pointing out that they did not test tumour-derived cfDNA as one of the limitations, and therefore analysis on tumour cfDNA is essential for the discovery of cfDNA fragment size and cancer prognosis. Moreover, increased quantity of fragments is not always cancer-specific, but it can be associated with conditions such as autoimmune diseases and exposure to surrounding environment. Adding to that, a small percentage of cfDNA in the plasma is derived from cancer cells, therefore the quantification of cfDNA cannot be specific (Underhill et al., 2016).

#### *1.4.1. Plasma cfDNA vs cerebrospinal fluid cfDNA*

In certain cancers such as nervous system tumours, cfDNA percentage in plasma can be low to absent due to obstacles such as mucin or the blood brain barrier (McEwen et al., 2020). In a 640-patient study, ctDNA was detected in more than 75% of patients with different cancer types but in less than 50% for primary brain tumours (Bettegowda et al., 2014). In such cases, cfDNA found in the cerebrospinal fluid (CSF) can be used instead of plasma cfDNA and has been proven to be more accurate when studying such tumours (McEwen et al., 2020; Miller et al., 2019). When comparing CSF and plasma cfDNA, CSF shows drastically lower levels of non-tumour cfDNA due to its pluricellular nature. It must be considered that generally, nervous system tumours show higher percentage of tumour-specific cfDNA, but this may be due to the physical obstacles that prevent the circulation of ctDNA. Moreover, tumour profiling for classification and therapy in brain cancers includes tissue extraction through brain surgery therefore sequencing of cfDNA from the cerebrospinal fluid can be an accurate and lower cost way of providing glioma genotypes (Miller et al., 2019).

## 1.5. Glioblastoma Multiforme

Glioblastoma Multiforme (GBM) is one of the most aggressive and lethal types of brain tumour. It derives from glial cells and represents the most common nervous system cancer in adults (Raucher, 2019). GBM is associated with one of the worst 5-year survival rates in all cancer and has an average survival of about 1 year after diagnosis (Philips et al, 2018). Brodbelt et al (2015) reports analysis of treatment and survival for more than 10 thousand cases in the England with a median survival of 6.1 months and rising to 14.9 with maximum treatment over the period 2007-2011. Global incidence of GBM is less than 10 per 100,000 persons and it is highly represented in patients aged more than 65 years (Silantsev et al, 2019). Gliomas are divided into 4 grades, I to IV, according to their clinical characteristics by the World Health Organisation (WHO) and they include tumour subtypes such as astrocytomas, oligodendromas and ependymomas. Grade I gliomas are usually benign and curable while grade II and III are more invasive and harder to treat. The most invasive grade of GBM is grade IV in which glioblastoma is included. High grade gliomas show mutations in most target genes, for example TP53, PTEN, CDKN2A and EGFR. (Yan et al., 2009). Patients are affected from grade IV (primary) GBM at a much higher occurrence at almost 90% of the GBM cases worldwide, while progression of the less aggressive grade I and grade II (secondary GBMs) affect 5-10% of patients (Szopa et al., 2017; Sasmita et al., 2018).

### 1.5.1. Treatment and prediction of GBM

Unlike other solid tumours, GBM can cover surrounding brain tissues but rarely metastasizes to other organs. Currently, the most common and successful therapy is surgical removal of tumour followed by radiation or chemotherapy (Raucher, 2019). Due to its highly invasive traits, it is challenging to remove all tumour tissue and as a result GBM reoccur in most cases. The median survival after reoccurrence is an overall of 6.2 months (Müller Bark et al., 2020).

The current treatments have low specificity and various side effects with no cure, but can provide patients additional survival time on the already low rate of 12-15 months (Raucher, 2019). Initial diagnosis happens by neuroimaging and biopsy of the tissue to define grade and characterise the tumour. There are several issues surrounding the resection or biopsy of brain tumour tissue, from the high risk of complications to lower quality of heterogeneity prediction of the whole tumour mass that can lead to inaccurate representation (Müller Bark et al., 2020). Using a treatment based on cfDNA is less invasive and could be proven beneficial in future therapies and diagnosis.

To avoid such risk, liquid biopsies can be used for GBM detection. As described in the “Plasma cfDNA vs cerebrospinal fluid DNA” section to used liquid biopsies in GBM, it has to be assumed that tumour-specific material crosses the blood-brain barrier. Disruption of the barrier has been reported in GBM and progression of the disease aggravates disruption and compromises the brain-blood barrier integrity. At present, liquid biopsies for GBM prediction can be a helpful tool to compliment current prediction and prognosis by providing a more intact view of the tumour characteristics (Müller Bark et al., 2020). That being said, there are no clinically validated cfDNA biomarkers for independent cfDNA GBM prediction due to similar limitations that were discussed in the “Nucleosome positioning and cell-free DNA for cancer diagnostics” section.

### *1.5.2. GBM subtypes*

GBM is highly characterised in a genomic level and its separated into groups relating to transcription profiles (proneural, neural, classical, and mesenchymal), genetics, such as IDH mutations, as well as epigenetics, for example CpG island methylator phenotype. Each of the profiles reflect altered pathways and genetic changes that affect progression and prognosis (Müller Bark et al., 2020). For example, patients that show higher survival are likely to be

patients carrying mutations of isocitrate dehydrogenase 1 (IDH1). Unfortunately, this genetic characterisation of the tumour has still not improved therapies (Lathia et al., 2015).

Glioblastoma is separated in 4 subtypes that correlate to the following genetic defects. First, IDH (proneural) subtype, that correlated to mutation of the IDH 1 gene, and focal copy number amplifications of the PDGFRA (RTKI subtype) or EGFR (RTKII subtype also called “classical”) genes that both code receptor tyrosine kinases (RTKs). Finally, the MES (mesenchymal) subtype shows lower cases of typical copy number amplifications in GBM (Hopp et al., 2015). In general, these subtypes moreover show differences in DNA methylation. Hopp et al (2015) calculated average methylation levels and variance for each subtype and they noticed that IDH, RTKII and to a lower extent MES were generally hypermethylated while RTKI were hypomethylated when compared to controls. In addition, methylation levels of the subtypes differ between glioma grades with Grade II and III astrocytoma having different methylation of RTKII and IDH genes when compared to GBM (Grade IV).

### 1.5.3. GBM biomarkers

Developments on next-generation sequencing methods have created a breakthrough on the identification of GBM-specific molecular characteristics that allow a more accurate and deep understanding of the molecular behaviour of the cancer. Therefore, there are a few types of biomarkers that have a key role in diagnosis, prognosis, and prediction of GBM (Szopa et al., 2017).

- 1) *Diagnostic biomarkers*: These biomarkers can provide more accurate classification of the cancer when diagnosis happens.
- 2) *Prognostic biomarkers*: prognostic biomarkers provide insight to a likely cancer outcome. For example, reoccurrence of GBM after treatment, can help occurrence of pseudoprognosis and avoidance of false treatment and calculated accurate survival time.

- 3) *Predictive biomarkers*: biomarker that can provide more accurate treatments for patient-specific strategies.

Some of the most common biomarkers known for GBM that are used as part of patient interrogation and diagnosis include: combined loss of nucleosome arms 1p and 19q, isocitrate dehydrogenase (IDH) mutations, mutations on or uncontrolled expression of p53, and epigenetic alterations for example, O6-methylguanine DNA methyltransferase (MGMT) hypermethylation (Müller Bark et al., 2020). More biomarkers for GBM also include: tumour suppressor protein TP53, Epidermal growth factor (EGFR), phosphatase and tensin homolog (PTEN) and others such as cancer stem cells and the p16INK4a gene (Szopa et al., 2017).

At this point, it is highly important to mention these molecular biomarkers have distinct presence at different grades of GBM and more specifically of primary and secondary GBM. Szopa et al. (2017) summarized the results for three studies (Ohgaki et al., 2004; Nakamura et al., 2001; Yan et al., 2009) and compared the molecular signatures of primary and secondary GBM. In their paper, they included frequency of the cases experiencing biomarker amplifications such as EGFR gene (36-60 % of primary and only 8% of secondary) and the PTEN mutation (25% of primary and 4% of secondary). Other common biomarkers that showed higher percentage of amplifications for primary GBM were the CDKN2A-p16<sup>INK4a</sup> deletion (31-78% of primary and 8% of secondary). In addition, they marked the mutations that are seen for secondary GBM that include TP53 mutation at 65% versus the 28% seen for primary as well as MGMT promoter methylation (75% of secondary and 26% of primary) and IDH1 mutation at 75% of secondary GBM that are not that common in primary at 5% (Szopa et al., 2017).

Besides heterogeneity between glioma grades there is differentiation in tumour progression and behaviour even between GBM subtypes and cell types. Cell types such as cancer cells with

stem-like properties can be found in GBM and they promote self-renewal and classification into different phenotypes (Garnier et al., 2019). GBM stem cells (GSCs) use surface markers and control specific signalling pathways, such as NOTCH, to promote GBM. Moreover, these stem cells can alternate between subtypes when being exposed to treatment, for example radiation, to provide a better environment for cell growth (Szopa et al., 2017).

#### *1.5.4. RTK signalling through the RAS/MAPK in GBM.*

Genetic aberrations of receptor tyrosine kinases (RTKs), including growth receptors EGFR and PDGFRA, have a present role in GBM (Ohgaki and Kleihues, 2009). RTKs are cell surface receptors that when activated drive two main pathways, the RAS/RAF/MAPK pathway that promotes proliferation migration and other stem -like properties and the P13K/AKT/mTOP that promotes proliferation and cell survival through the cell cycle (Pearson and Regad, 2017). Moreover, P13K is closely related to PTEN, which regulates the pathway and provides resistance to EGFR-based therapies (Szopa et al., 2017). This resistance makes these two pathways a promising GBM treatment (Pearson and Regad, 2017).

#### *1.5.5. RB tumour suppressor in GBM.*

The pRB pathway interacts with transcription factor E2F for the suppression of cell cycle progression. This leads to under-expression of genes related to the cell cycle, making reactivation of pRB a passable therapy of GBM. The pathway was found altered in 78-79% of GBM cases while the RB gene was deleted or mutated in 7.6-11% (Pearson and Regad, 2017).

#### *1.5.6. IDH1 and IDH2 mutations in GBM*

IDH1 mutations were discovered by Parsons et al. (2008). Advances in cancer genetics established the role of specific IDH mutations in human cancers such as malignant gliomas,



acute myeloid leukaemia and other carcinomas (Waitkus et al., 2018). Mutations of the IDH1 gene were found in 12% of glioblastomas and suggested that these mutations can occur after low-grade glioma formation and drive the progression to glioblastoma. The results of the study indicated that GBM patients carrying IDH2 or IDH2 mutation had increased overall survival at an average of 31 months compared to the 15-month survival of patients with wild-type IDH1 (Yan et al., 2009).

#### *1.5.7. Tumour suppressor P53 mutation in GBM*

The p53 pathway is critically involved in the progression of genes that are related to cell cycle response and TP53 is one of the most commonly altered genes in many cancers (Zhang et al., 2018). The p53-ARF-MDM2 pathway is aberrant in 84% of GBM patients and deregulation is highly connected to GBM proliferation, evasion of apoptosis and cancer cell stemness (Zhang et al., 2018). Events such as, stress signals and DNA damage trigger p53 response that inhibits protein complexes Cdk4/Cyclin F and Cdk2/Cyclin E to prevent cell cycle progression. In GBM, p53 signalling shows aberrant functions in 87% of the cases, with deletion or mutation of p53 at 28-35%. Many GBM treatments include the restoration of p53 for activation of its tumour suppressor properties (Ohgaki and Kleihues, 2009; Zhang et al., 2018).

### **1.6. Nucleosome positioning in GBM**

Due to the aggressive nature of GBM, it is a priority to understand the underlining mechanisms of it in order to produce efficient therapies. Highly important facts can be uncovered through the identification of nucleosome positioning. In their study, Malley et al (2011) predicted nucleosome positioning for sequences upstream and downstream on MGMT and several TFs. Through analysis of nucleosomes, they discovered association of positioning and methylation patterns and as well as nucleosome depletion around TF binding sites and Transcription Start

Sites (TSS). An interesting result was surfaced when nucleosome accessibility analysis revealed open chromatin structure was lost for MCMT methylated cells and as a result the TSS lost accessibility. The authors suggested blockage of the TSS may lead to transcriptional silencing while TFs, such as STAT3, were predicted to bind to regions are linked to gliomagenesis.

Chromatin accessibility has been analysed in GBM using the assay for transposase accessible chromatin with sequencing (ATAC-seq). ATAC-seq enables the study of open chromatin and can identify accessible regions of transcription (Tome-Garcia et al., 2018). In a recent study, Tome-Garcia et al. (2018), used ATAC-seq to capture distinct TF drivers of tumour progression in GSCs and identified TEAD1 to show overexpression. Other ATAC-seq studies, such as Bulstrode et al. (2017) identified high SOX2 and FOXG1 levels as a key factor on proliferation in GBM and moreover, analysed their interaction through the use of their ChIP-seq data. In addition, Guilhamon et al. (2018), identified GSC subtypes based on chromatin accessibility, DNA methylation and RNA-seq which offered new possibilities of discovering therapeutic compounds and improving therapeutic effectiveness. Through this research, it is clear that the identification of the positions that nucleosomes take is extremely important in order to characterise and discover new therapies for GBM. Finally, a particularly understudied area of analysis is the detection of genome-wide nucleosome positioning, as well as calculation of nucleosome occupancies at different genomic regions.

## **1.7. Aims**

The purpose of this study is to shed new light on the genome-wide nucleosome positioning in GBM through the identification of nucleosome positioning patterns in normal and tumour tissue from GBM patients. First, we aim to highlight any abnormalities in positioning found in

cancer compared to normal tissue around specific regions on the genome. We hope that by analysing nucleosome positioning we can provide GBM-specific targets for future therapies. Moreover, by analysis nucleosome positioning of cfDNA from healthy people and GBM patients we would like to provide insight on the diagnosis and monitoring of GBM without the use of traditional methods such as biopsy and X-rays. Using cfDNA fragments we hope to profile the patient's genome for the use of personalized medicine as well as monitor their progression. Finally, the results of the report will determine the possibility of nucleosome positioning as cancer biomarker.

## 2. Methods

### 2.1. MNase-seq and cfDNA datasets

We used eight samples from four different glioblastoma patient to which we referred to as G276, G125, G289 and G370. Patient information is described in Table 1. Samples were provided by Dr Paul Brennan (University of Edinburgh). For all four patients we used paired healthy and GBM tissue that was collected from tumour brain tissue and the healthy tissue at its periphery, extracted during the surgery. MNase-seq experiments were performed in these samples in the Teif lab as discussed previously (Shafiei, 2018). Sequencing data was checked by the “fastqc” program prior to nucleosome occupancy analysis. The project was conducted in the laboratory of Dr Vladimir Teif in collaboration with Dr Paul Brennan from the University of Edinburgh within the ethical approval by the East of Scotland Research Ethics Service (EoSRES) dated 14 July 2015, Ref: AG/15/ES/0094. Moreover, cell-free DNA (cfDNA) was extracted from blood plasma of one of these patients (patient G370).

**Table 1.** Summary of all information on the 5 patients used in the project.

File/sample names (internal)	New ID	Old ID	Patient information	# fragments 146-148 bp	# unique regions 146-148 bp
G N MNase-seq	G215	G	Female 60, GBM	31,568,029	23753429
G T MNase-seq	G215	G	Female 60, GBM	43,588,690	31069028
G T H3 Rep2	G215	G	Female 60, GBM	27,698,780	19879609
G370 cfDNA	G370	G370	Male 60, GBM	983,200	7282839
G370 N MNase-seq	G370	G370	Male 60, GBM	16,900,477	14146963
G370 T MNase-seq	G370	G370	Male 60, GBM	18,896,957	14395131
GG N MNase-seq	G276	GG	Female 40, Diffuse astrocytoma WHO grade II converted to GBM	6,439,815	5337656
GG T MNase-seq	G276	GG	Female 40, Diffuse astrocytoma WHO grade II converted to GBM	6,709,874	5485448
JJ N MNase-seq	G237	JJ	Female 40, Anaplastic astrocytoma WHO grade II converted to GBM	11,288,991	9001558

JJ T MNase-seq	G237	JJ	Female 40, Anaplastic astrocytoma WHO grade II converted to GBM	4,203,114	3655699
QQ T MNase-seq	G289	QQ	Male 40, GBM (samples taken after first surgery/chemo/radiotherapy)	13,098,188	11891749
QQ T H3 ChIP-seq	G289	QQ	Male 40, GBM (samples taken after first surgery/chemo/radiotherapy)	11,382,689	6546446

## 2.2. External experimental datasets

For the analysis of nucleosome occupancy in GBM we have used TF dataset, cfDNA patient datasets as well as histone modifications. The TF datasets were obtained from the three databases pre-processed experimental ChIP-seq genomic data: ReMap (Griffon et al., 2015), ChIP-Atlas (Oki et al., 2018) and Gene Expression Omnibus (GEO) (Barrett et al., 2013). A total of 20 TFs shown in Table 2, where it demonstrates all details as well as accession numbers for the TFs. Next, investigation of cfDNA was based on data reported in Song et al. (2017). cfDNA bed files were downloaded through the GEO database using the same commands as described above. Finally, histone modifications in isocitrate dehydrogenase (IDH), mesenchymal (MES), receptor tyrosine kinase I (RTKI) and receptor tyrosine kinase II (RTKII) GBM subtypes found in GEO under GSE121723. Accession numbers for histone modifications are as following for (a) IDH subtype: H3K27ac (GSM3444433), H3K27me3 (GSM3444434), H3K4me1 (GSM3444435) and H3K3me3 (GSM3444436), (b) MES subtype: H3K27ac (GSM3444438), H3K27me3 (GSM3444439), H3K4me1 (GSM3444441) and H3K3me3 (GSM3444441), (c) RTKI: H3K27ac (GSM3444480), H3K27me3 (GSM3444481), H3K4me1 (GSM3444483) and H3K3me3 (GSM3444484), and (d) RTKII: H3K27ac (GSM3444450), H3K27me3 (GSM3444451), H3K4me1 (GSM3444453) and H3K3me3 (GSM3444454).

**Table 2.** Datasets of TF binding sites in GBM-related cell lines used in the analysis.

<b>TF name</b>	<b>Cell type</b>	<b>Reference</b>
ARNT2	Glioblastoma Stem Cell (GSC)	GSE98330 (Bogreas et al., 2018)
ASCL1	G523NS cells treated with doxycycline	GSM2335531, GSM2335532, GSM2335533 (Park et al., 2017)
BMAL1	Glioma stem cells	GSM1306364
BRD4	U87MG (Uppsala 87 Malignant Glioma)	Remap GSM1038284 (Lovén et al., 2013)
CHD4	0308 cell line, glioblastoma tumour initiating cells	ChIP-seq Atlas GSM1265802 (Chudnovsky et al., 2014)
CREBBP	T98G human glioblastoma cells	ChIP-seq Atlas GSM525265 (Ramos et al., 2010)
CTCF	H54 (D54) glioblastoma cell line	GSM822303 ENCODE project
CTCF	Neuro-blastoma cells in continuous culture	GSM803333 (Gertz et al., 2013)
EP300	T98G human glioblastoma cells	ChIP-seq Atlas GSM525266 (Ramos et al., 2010)
JMJD6	3565 cells	GSM2360990 (Miller et al., 2017)
JMJD6	528 cells	GSM1922076 (Miller et al., 2017)
KLF4	U87	GSE97632 (Wan et al., 2017)
KLF9	Human GBM stem-like cells	remap2018 GSM1522563 (Ying et al., 2014)
MAX	U87MG (Uppsala 87 Malignant Glioma)	remap2018 GSM894077 (Lin et al., 2012)
MED1	U87MG (Uppsala 87 Malignant Glioma)	remap2018 GSM894082 (Lin et al., 2012)
MYC	U87MG	remap2018 GSM894061 (Lin et al., 2012)
Olig2	MGG8	GSM1306365 (Suvà et al., 2014)
Olig2	MGG8	GSM1306367 (Suvà et al., 2014)
RBPJ	brain tumour initiating cells	GSM2101764 (Xie et al., 2016)
RBPJ	glioblastoma initiating cells	GSM2101765 (Xie et al., 2016)

SALL2	MGG8 Glioblastoma cell line	GSM1306364 (Suvà et al., 2014)
SOX2	LN229 glioma cancer cells	GSM586971 (Fang et al., 2011)
SOX2	NOD-SCID mouse harbouring HOT1 tumour derived from human glioblastoma	GSE58345 (Singh et al., 2017)
SOX2	MGG8	GSM1306360 GSM1306362 (Suvà et al., 2014)
POU3F2	MGG8 human glioblastoma stem cell line	GSM1306358 (Suvà et al., 2014)

### **2.3. Computational analysis setup**

The data was analysed computationally using the University of Essex high-throughput cluster. All computational analysis needed for this study was performed in the Linux environment through the terminal emulator PuTTY (<https://www.putty.org/>). Bash scripts were run using the command line and the containing command were created using WinSCP (<https://winscp.net/eng/download.php>), an open-source client that offers basic file managing. Finally, graphic display and data analysis was done via the program OriginPro 2020 (<https://www.originlab.com/origin>).

### **2.4. DNA fragment size analysis**

In order to investigate the distribution of nucleosomal DNA fragments in tissue and cfDNA we used an R script that calculates the frequency of nucleosome fragments in bed files pre-processed with NucTools (Vainshtein et al., 2017)(see Appendix, Script 1). After the investigation of nucleosomal fragment length in all patients we decided to only select two different ranges of nucleosome fragment sizes for two different analysis to increase probability of the fragments having nucleosomal DNA ancestry. To extract specific fragment lengths of 120-180 bp and 146-148bp we used the perl script “extract\_nuc\_sizes.pl” (see Appendix, Script 2).

### **2.5. Calculation of average nucleosome occupancy profiles**

To determine nucleosome occupancy profiles around genomic regions, we used the HOMER package and the command “annotatePeaks.pl” that uses a bed file of regions and the HOMER tag directories files and outputs average nucleosome occupancy around the regions. Separate HOMER tag directories were created for each MNase-seq and cfDNA samples using DNA fragments with sizes 120-180 bp for genome assembly hg19.



## 2.6. Calculation of A/T/C/G nucleotide frequency

Nucleotide profiles were calculated along the nucleosome fragments of sizes 146-148 bp using HOMER package, command “annotatePeaks.pl” with parameters -CpG -size 200 -hist 1.

## 2.7. Lost, gained, more fuzzy and less fuzzy nucleosomes

The regions that lost and gained nucleosomes and showed more and less fuzzy nucleosome in cancer were determined using NucTools. First, for lost and gained nucleosomes we calculated average occupancy separately for healthy and tumour for each of the four patients using script “bed2occupancy\_average.pl” for a window of 100bp and 1000bp. Next, using the average occupancy files generated at the previous step we identified stable nucleosomes using the script “table\_nucs\_replicates.pl” with parameter -StableThreshold 0.5. Then, we compared the two conditions, healthy and tumour, using the “compare\_two\_conditions.pl” script with “threshold1” 1.4 and “threshold2” 1.4 for 100-bp window and “threshold1” 0.35 and “threshold2” 0.35 for 1000-bp window.

Regions of fuzzy nucleosome were identified for 100-bp windows using “stable\_nucs\_replicates.pl” with a -FuzzyThreshold” of 1 followed by the script “compare\_two\_conditions.pl” with “threshold1” 1.4 and “threshold2” 1.4.

## 2.8. Enrichment of lost/gained, more and less fuzzy nucleosomes

The enrichment of lost, gained, more and less fuzzy nucleosomes at different genomic features was calculated using BedTools (Quinlan and Hall, 2010). Using the 1000bp regions that have been already calculated, we applied the command “bedTools shuffle” to create files of lost, gained, fuzzy and stable nucleosomes that are randomly shuffled in the human genome. Next, we used both shuffled regions and non-shuffled regions and intersect them with genomic

regions using the “intersectBed” command. The number of sites within the intersected output files were then counted using the command “wc” and both values of non-shuffled and shuffled regions were divided, to get fold enrichment.

## **2.9. Alu repeats analysis**

We analysed nucleosome occupancy in all ALU repeats binding sites using the method explained in the Methods section 2.5. We further split the ALU repeats in 2 size groups 50-200bp indicating the monomeric ALU repeats and 250-400bp for Dimeric ALU repeats. To extract the specific base pair lengths of 50-200 bp and 250-400bp we used the perl script “extract\_nuc\_sizes.pl” (see Appendix, Script 2) and adjusted the -min and -max parameters. Moreover, further calculations were made by aligning the signal around the left end (start) of ALU repeats and the splitting into sizes. We then calculated nucleosome occupancy as described above.

## **2.10. Gene ontology analysis**

In order to proceed with the Gene Ontology analysis, we decided to intersect the 100bp long lost and gained nucleosome regions with several genomic regions that contain the genes included in the regions. We once again used the “intersectBed” command, using the genomic region files a parameter -a and the lost and gained nucleosome regions as parameter -b and used parameter -u to only choose unique regions. The output files include genes names unique for lost and gained nucleosomes. We then used the output genes in DAVID and proceeded with analysing the Biological Process terms.

## 2.11. Principal Component analysis (PCA)

To start the analysis, we used the chr1 occupancy “.occ” files for healthy and tumour for each individual patient and the lost and gained nucleosome regions with a window of 1000bp that were calculated already (see section 2.7). First, using the individual healthy and tumour occupancy files for each patient, we converted the files from “.occ” to “.bed” format, by formatting the files using the command “awk” with parameters “-v = OFS='\t' ” to choose specific columns and “perl -p -e 's/^/chr1\t/{ print \$1, \$1, \$2}”’. A total of 8 files were formatted (4 healthy and 4 GBM) and was then individually intersected with the 1000bp-long lost and gained nucleosome regions using “intersectBed”. Next, we compiled all occupancies into a single file using the command “paste” and making use we use the parameter “column -s \$'\t' -t” to make use the columns are separated. Finally, performed PCA of the two output files in the R environment using Script 3, Appendix.

## 3. Results

### 3.1. Distribution of DNA fragment lengths

To begin our investigation of nucleosome positioning, we started with the analysis of DNA fragments length for all samples used in this study (Figure 1). Overall, Figure 1 shows that there is a wide distribution of fragment lengths for all patients. From Figure 1A and B, it is clear that all tissue samples show peak of fragment lengths at around 146 base pairs (bp) (which corresponds to a slightly over-digested mononucleosome) whereas for patient G370 the peak was at 153 bp (which corresponds to a slightly under-digested mononucleosome). The distribution of DNA fragment lengths based on cfDNA from patient G370 has a peak at 165 bp, which corresponds to the chromatosome particle (the nucleosome plus linker DNA). On the other hand, this distribution is similar to that for the previously published cfDNA from blood plasma from GBM patients and healthy individual datasets (Song et al., 2017) (Figure 1C and D). It is interesting to note that cfDNA from the latter study is more enriched with small pieces of sub-nucleosomal sizes in GBM in comparison with healthy people. To highlight the difference on fragments lengths in cfDNA, we plotted MNase-seq and cfDNA samples for patient G370 separately in chromosome Y (Figure 1E). As expected, after looking at Figure 1A-D, in Figure 1E cfDNA shows a peak at 167bp while MNase-seq has a clearly lower fragment length at 155bp for both GBM and healthy. In the following calculations we have narrowed down the distribution of DNA fragments to unify these size distributions. For example, Figure S1 (Appendix) shows how average nucleosome occupancy profiles around CTCF calculated based on all DNA fragments obtained from MNase-seq differ from the profiles calculated based on the fraction of DNA fragments with sizes within 120-180bp. When comparing the two, the main difference is that the overall nucleosome occupancy is higher for all patients but patient G237 shows higher occupancy in GBM than healthy as well as showing

a nucleosome peak at the centre of CTCF. This difference only appeared after filtering nucleosomal DNA fragment sizes and indicates that size-filtering can reveal more detailed nucleosome occupancy patterns.

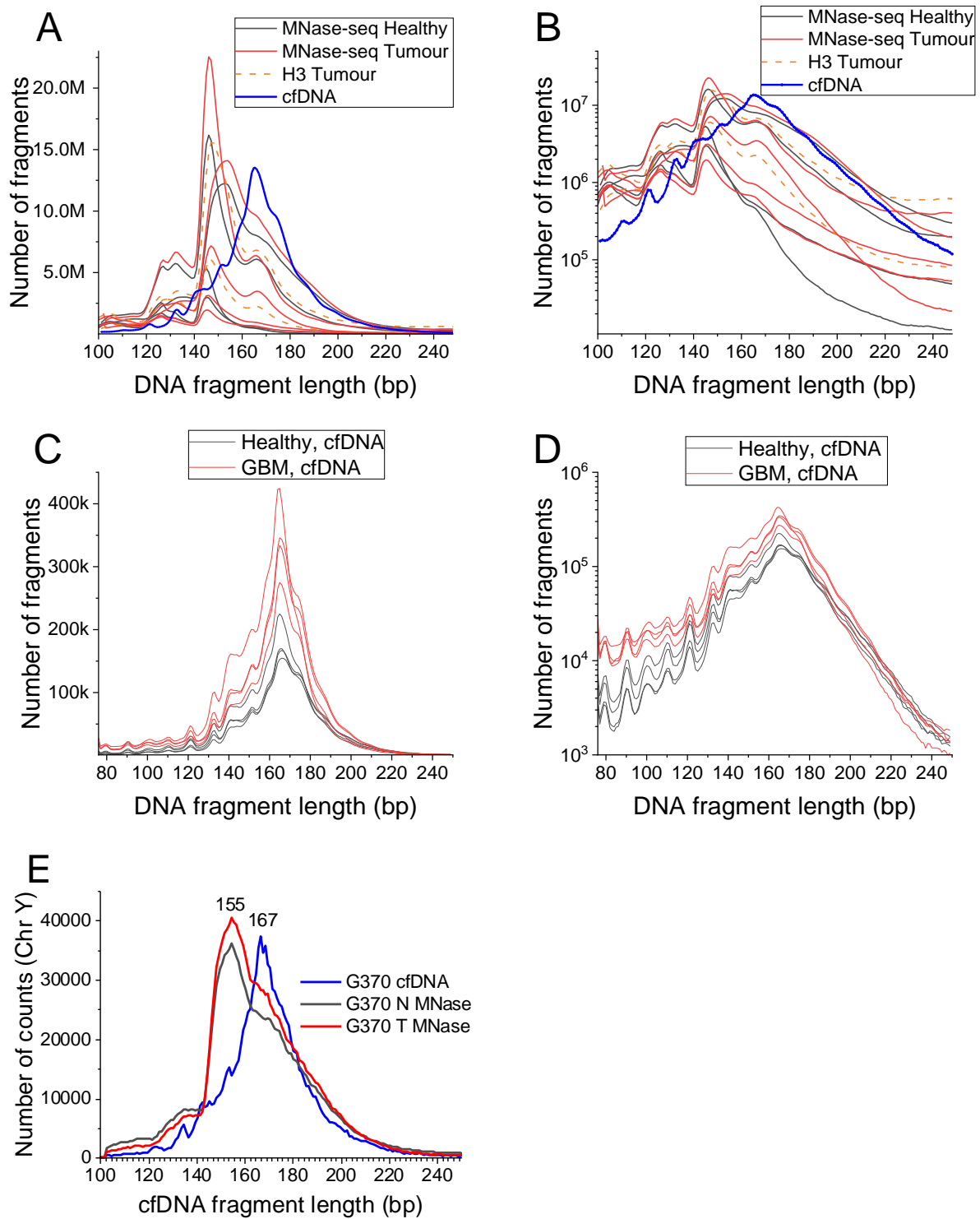


Figure 1. Distribution of DNA fragment lengths for all samples used in this study. (A) Length distribution of nucleosomal DNA fragments across GBM and healthy MNase-seq samples

from four GBM patients G125, G237, G276 and G370 and one cfDNA sample (blue points) from patient G370 from our study. (B) The same as (A) in logarithmic scale. (C) Distribution of cfDNA fragment lengths across all samples from 4 GBM patients and 4 healthy individuals from Song et al (2017). Healthy individuals are shown in black and GBM patients in red. (D) The same as (C) in logarithmic scale. (E) Distributions of DNA fragment sizes on chromosome Y for the three matched samples for patient G370: MNase-seq in healthy brain sample is shown in black, MNase-seq in tumour brain sample in red and cfDNA extracted from blood of the same patient in blue.

### **3.2. Nucleosome occupancy profiles in GBM patients around TFs bound in GBM-related cell lines.**

We acquired TF binding datasets in GBM-related cell lines as detailed in Methods. After quality check using the “fastqc” program, we determined average nucleosome occupancy around several TFs from GBM-related cell lines (Table 2) in MNase-seq and H3 assisted MNase-seq samples and cfDNA samples for both healthy and GBM-affected individuals. In order to study changes in nucleosome occupancy profiles around TF binding sites, we made separate calculations for healthy and GBM tissue from 4 patients G125, G237, G276 and G370, and calculations for cfDNA from blood plasma for patient G370 and datasets reported in Song et al (2017). TFs were selected due to their direct relation and contribution to GBM progression which will be explained thoroughly for each TF in the “Discussion” section.

#### *3.2.1. Average chromatin profiles differences in healthy and GBM samples based on MNase-seq data.*

Figure 2 shows average nucleosome occupancy profiles around TF binding sites in GBM-related cell lines (Figure 2). In each case occupancy profiles were averaged over the 4 GBM patients separately for non-malignant and tumour brain tissues. Overall, occupancy of nucleosomes around binding sites of TFs is lower in tumour than normal samples but there is large variation between patients. For example, on average patient G237 shows higher occupancy of nucleosomes around TFs. Nucleosome occupancy profiles around CTCF (Figure

2A) show lower variation between healthy samples in comparison with tumour. Nucleosome occupancy profiles in patients G276 (Figure S3A, Appendix), G237 (Figure S3B, Appendix) and G125 (Figure S3C, Appendix) form peaks at CTCF sites but, unlike other profiles, for patient G237 nucleosome occupancy in GBM is higher than in healthy tissue. It is important to mention that average nucleosome profile around CTCF in GBM tissues is very fuzzy (Figure 1A), the profiles of individual patients show distinct shape changes between normal and GBM tissues (Figure S3). Similar patterns are seen around RBPJ (Figure 2C). Interestingly, G237 does not have higher occupancy for GBM tissue around POU3F2 (Figure S7, Appendix), SOX2 (Figure S8, Appendix) and OLIG2 (Figure S10, Appendix), three of the four TFs involved in the POU3F2, SOX2, SALL2 and OLIG2 complex. Nucleosomes around both SOX2 and POU3F2 profiles are depleted at the centre of the binding sites for three out of four patients. On the other hand, nucleosome profiles around OLIG2 (Figure S10, Appendix) show similar behaviour in all patients, with nucleosome peaks at the centre of the TF binding sites. In addition, nucleosome occupancy profiles at JMJD6 (Figure 2J) and KLF9 (Figure 2N) showed interesting changes. On average, both TFs show lower and highly variant nucleosome occupancy in GBM. For JMJD6, nucleosomes at the centre of show less defined depletion for patients G276 and G215 in GBM while for other patients' profiles are similar between the two groups. Finally, KLF9 (Figure S17, Appendix) shows nucleosomes at the centre of the binding sites form peaks for patient G276 (Figure S17A, Appendix) and G237 (Figure S17B, Appendix) in both healthy and GBM tissue. Finally, patient G125 (Figure S17C, Appendix) profiles changed from depletion to augmentation of nucleosomes at the centre.

### *3.2.2. Average chromatin profiles differences in healthy and GBM samples based on cfDNA.*

We also calculated nucleosome occupancy profiles around TF binding sites reconstructed based on cfDNA taken from the plasma of 4 healthy individuals and 4 GBM patients (Song et al. 2017) (Figure S24, Appendix). These nucleosome occupancy profiles do not show great

differences from healthy to GBM. When compared to MNase-seq sample profiles, average nucleosome occupancy is always higher in GBM, while in MNase-seq samples occupancy in GBM is always lower. Other than that, patterns of the cfDNA are highly similar to the ones observed in MNase-seq of brain tissues. Cases that show different cfDNA profiles at the centre of TF binding sites include ASCL1 (Figure S24H, Appendix) and MAX (Figure S24L, Appendix).



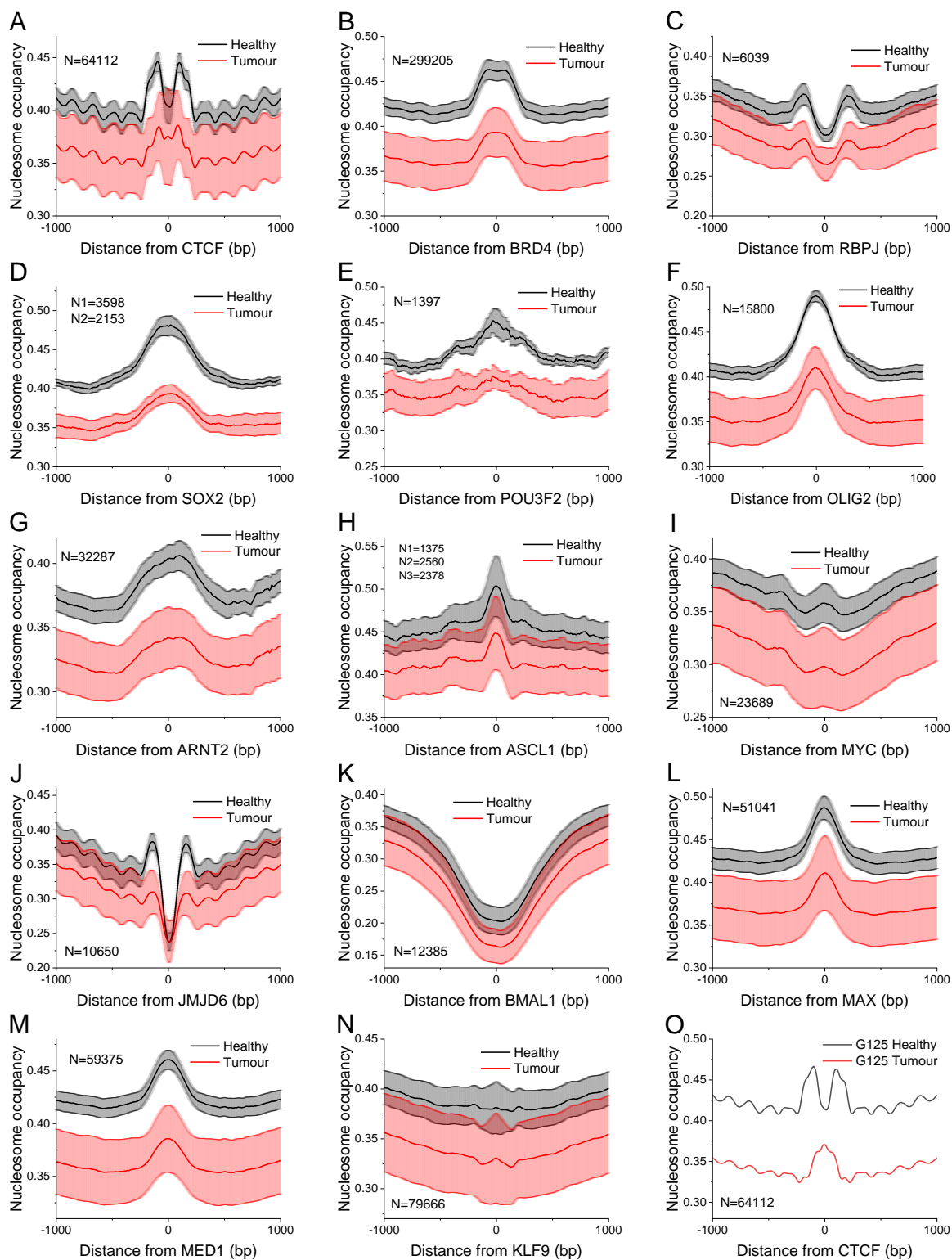


Figure 2. (A-N) Average profiles of MNase-seq nucleosome occupancy around binding sites of 14 transcription factors averaged over 4 GBM patients for samples taken from the tumour periphery marked as “healthy” (black lines) and samples from the bulk of the tumour marked as GBM (red lines). Lighter areas correspond to the standard errors of averaging. The number of regions (N) for each TF file is indicated on the graph. O) Nucleosome occupancy profiles around binding sites of CTCF for paired tumour and non-tumour brain tissues from an individual patient G125.

### 3.3. Nucleosome repositioning at genomic DNA repeats

We further investigated nucleosome occupancy around ALU repeats and TSS for healthy and cancer tissue and cfDNA from healthy people and GBM patients and we expanded our analysis of ALU sequence repeats by calculating nucleosome occupancy for two different size groups 50-200bp and 250-400bp (Figure 3). Moreover, we decided to compare nucleosome occupancy patterns at ALU repeats and TSS for healthy and GBM tissue as well as cfDNA from one of our patients (Figure S25A, Appendix).

#### 3.3.1. Average nucleosome occupancy around ALU repeats

First, we measured occupancy of nucleosomes around ALU repeats. Figure 3 shows occupancy around ALU in healthy and GBM tissue and cfDNA accordingly. In the MNase-seq tissue samples, occupancy of nucleosomes seems to be fuzzier in cancer, but the pattern and peak positioning are very similar to healthy (Figure 3A and B). On the other hand, patterns of the cfDNA samples have significant differences between healthy and GBM (Figure 3C and D). When comparing the two figures we can see that MNase-seq ALU profiles are unexpectedly very different from the cfDNA profiles, with MNase-seq revealing two depletions of nucleosome occupancy, while in cfDNA we can only observe only one at the centre of the ALU binding sites.

#### 3.3.2 Average nucleosome occupancy profiles at size separated ALU repeats

After studying nucleosome occupancy of ALU repeats we realised that ALU have a great number of binding sites, so we decided to split ALU regions based on the size of the ALU. Figures 3A and 3B shows nucleosome occupancy for two different ALU size regions, 50-200bp and 250-400bp averaged over all MNase-seq samples accordingly. Overall, while patterns are similar for both size groups, it is clear that sizes 50-200 have clearer patterns than sizes 250-400. Furthermore, we calculated nucleosome occupancy around the two ALU repeat size

groups for the cfDNA from Song et al (2017). We observe that occupancy around ALU repeats has more defined peaks for 50-200bp (Figure 3C) than 250-500bp (Figure 3D) but less fuzziness in patterns overall. Additionally, when comparing MNase-seq and cfDNA nucleosome occupancy patterns we can still observe difference in the nucleosome profiles at the centre even after the size separation.

We further analysed nucleosome occupancy calculations by aligning the signal around the left end (start) of ALU repeat binding sites shown in Appendix S25. Overall, the patterns are similar to those in Figure 3, with the difference that occupancy seems to be slightly fuzzier at the start sites in MNase-seq samples. Moreover, patterns that were calculated around the left end of ALU are shifted on the right.

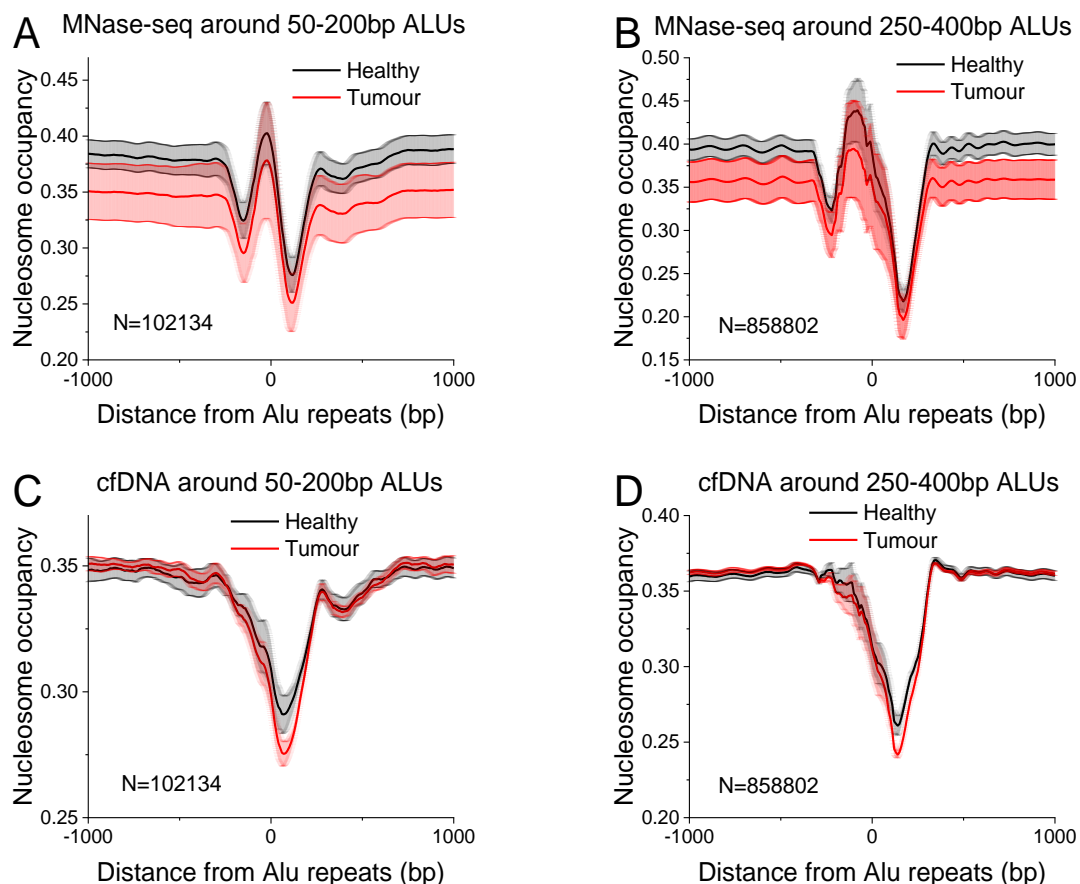


Figure 3. Average nucleosome occupancy around ALU repeat elements. (A-B) Nucleosome occupancy around ALU repeats averaged over the 4 GBM patients' healthy tissue MNase (black) and GBM tissue MNase (red). (A) 50-200bp and (B) 250-400bp. Nucleosome occupancy around ALU repeats averaged over the 4 samples of cfDNA from healthy people

(black) and 4 samples of cfDNA from GBM patients (red) from Song et al (2017) for ALU sizes 50-200bp (C) and 250-400bp (D). All calculations were made by aligning the signal around centres of ALU repeats. Grey areas correspond to the standard errors of averaging.

### *3.3.3 Average nucleosome occupancy patterns around ALU repeats resemble those around TSS.*

Next, we calculated nucleosome occupancy around TSS regions in healthy and GBM MNase-seq tissue (Figure 4B) and cfDNA (Figure 4D). Nucleosome occupancy patterns are very similar between MNase-seq and cfDNA and the profiles shapes look as expected, showing a depletion of nucleosomes followed by a peak of occupancy downstream the TSS that annotates the “+1” nucleosome. Moreover, nucleosome occupancy is fuzzier in cancer for the MNase-seq samples.

Furthermore, we decided to compare ALU repeat and TSS nucleosome profiles from Figure 4. The nucleosome occupancy profiles around both ALU repeats (Figure 4A), as well as TSS (Figure 4C) show lower occupancy in GBM tissue, while patient variation is higher in GBM. A very interesting but unexpected similarity between nucleosome occupancy profiles of ALU repeats and TSS is their profile shapes. It seems that nucleosome profiles show similar depletions and augmentations of nucleosomes around the centre of both regions, but the profiles are inverted.

Next, we plotted nucleosome occupancy profiles around ALU repeats and TSS for healthy and GBM tissue as well as cfDNA only from patient G370, for which we have data both for the paired tumour/normal brain MNase-seq and cfDNA from peripheral blood. The corresponding profiles are demonstrated in Figure 4E and F. Both graphs highlight differences between brain tissue and cfDNA at ALU repeats and TSS. In Figure 4E, high occupancy of nucleosomes is seen for cfDNA and well as healthy tissue, while GBM tissue shows significantly lower occupancy at ALU repeats. For the cfDNA samples, nucleosomes at the centre of ALU show

higher augmentation when compared to brain tissue. Finally, in Figure 4F profile shapes are similar across all samples, but cfDNA shows the highest occupancy around TSS.

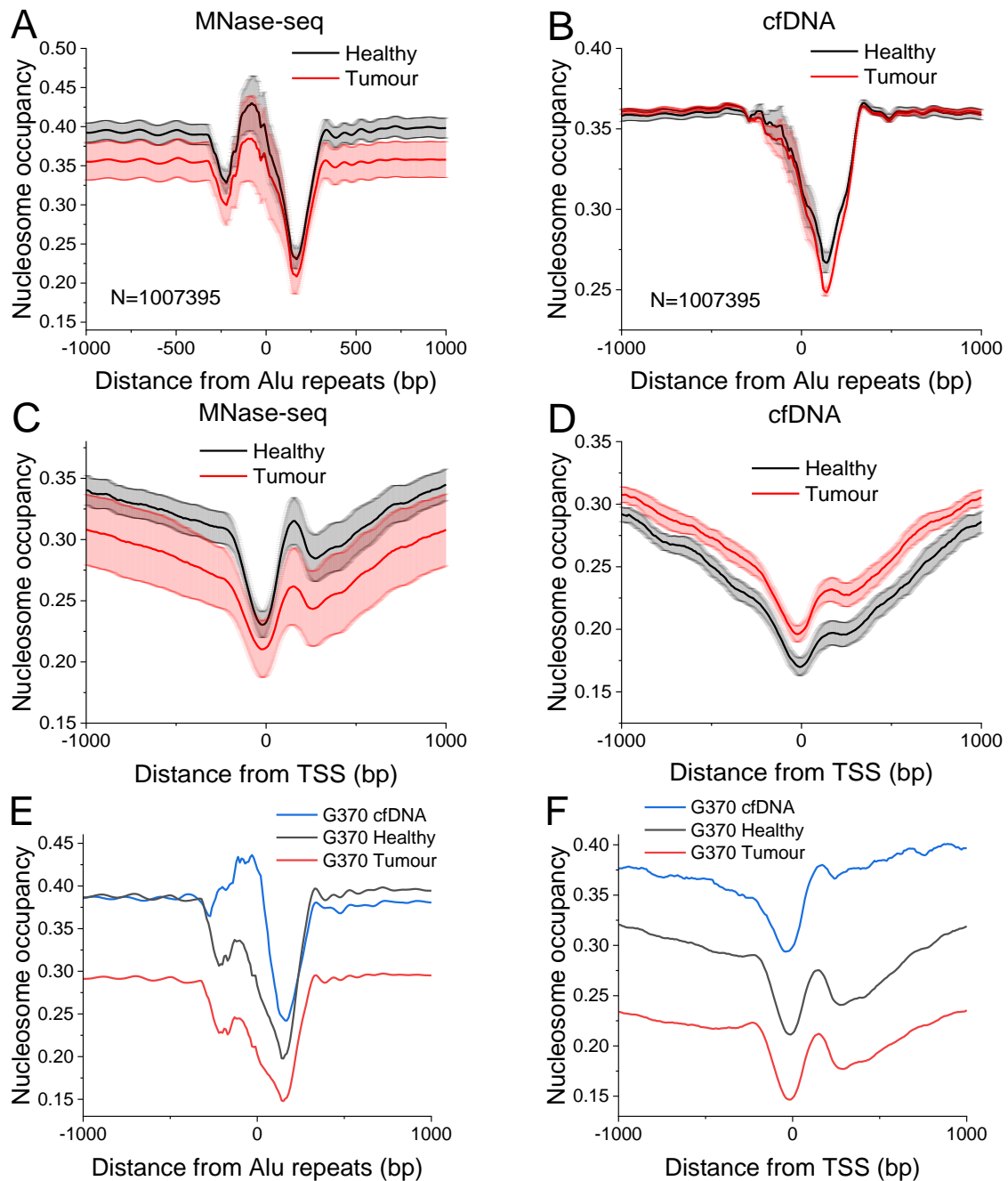


Figure 4. Comparison of average nucleosome occupancy around ALU repeats and transcription start sites (TSS) based on MNase-seq and cfDNA. (A) Nucleosome occupancy around TSS averaged over 4 GBM patients, separately for healthy (black) and GBM tissue (red). (B) cfDNA density profiles around TSS averaged over 4 samples from healthy people (black) and 4 samples from GBM patients (red) from Song et al (2017). (C-D) The same as (A-B) but profiled around TSSs. Grey areas correspond to the standard errors of averaging. (E-F). Nucleosome occupancy profiles for an individual GBM patient based on cfDNA (blue) and MNase-seq from healthy (black) and tumour (red) areas of the brain.

### 3.4. Interplay of nucleosome repositioning and histone modifications in GBM

For our next analysis, we calculated nucleosome occupancy around regions enriched with 4 histone modifications (H3K27ac, H3K27me3, H3K4me1 and H3K4me3) found in normal and tumour brain samples (Figure 5). These profiles were calculated for histone marks found in each of the 4 subtypes of GBM: IDH, MES, RTKI and RTKII. Moreover, we investigated nucleosome occupancy around CTCF binding sites for subclasses of CTCF sites that are enriched with 4 histone marks (Figure 6). Nucleosome profiles in Figure 6 are intersected with histone modifications H3K27ac, H3K27me3, H3K4me1 and H3K4me3 only for IDH GBM subtype.

#### *3.4.1. Nucleosome occupancy around domains with different histone modifications in glioblastoma*

Nucleosome occupancy profiles were calculated around 4 histone modifications: H3K27ac (Figure 5A), H3K27me3 (Figure 5B), H3K4me1 (Figure 5C) and H3K4me3 (Figure 5D) specific for the MES GBM subtype. Higher nucleosome occupancy is observed around all histone marks for non-malignant samples in comparison to tumour brain tissues. Slight differences can be seen within the MES subtype for nucleosome occupancy around H3K27ac, where nucleosome occupancy variation is high and nucleosome occupancy level is clearly lower for GBM (Figure 5A). For H3K27me3, nucleosomes at the centre share a similar nucleosomal peak for both healthy and GBM tissue (Figure 5B). Moreover, the same profiles were calculated for the other 3 GBM subtypes, IDH (Figure S27, Appendix), RTKI (Figure S28, Appendix) and RTKII (Figure S29, Appendix). Nucleosome occupancy profiles across all subtypes have similar profiles. Figure S28B and Figure S29B shows similar nucleosome occupancy around H3K27me3 for RTKI and RTKII subtype accordingly. The two figures show

smooth nucleosome peaks and higher variance of nucleosome occupancy between GBM patient compared to healthy individuals. Profiles around H3K27me3 for IDH subtype show nucleosomes at the centre create a sharper peak.

When comparing H3K27me3 profiles in all subtypes, RTKI and RTKII show the most resemblance while IDH and MES have sharper nucleosome peaks. It is important to mention that the number of sites noted on each figure is very different between subtypes and can affect quality of peaks. Finally, after looking at nucleosome occupancy profiles for all individual patients (Figure S30-S45, Appendix), it is worth noting that patient G237 was the only patient that always showed higher nucleosome occupancy in GBM versus non-malignant brain tissues for all GBM subtypes.

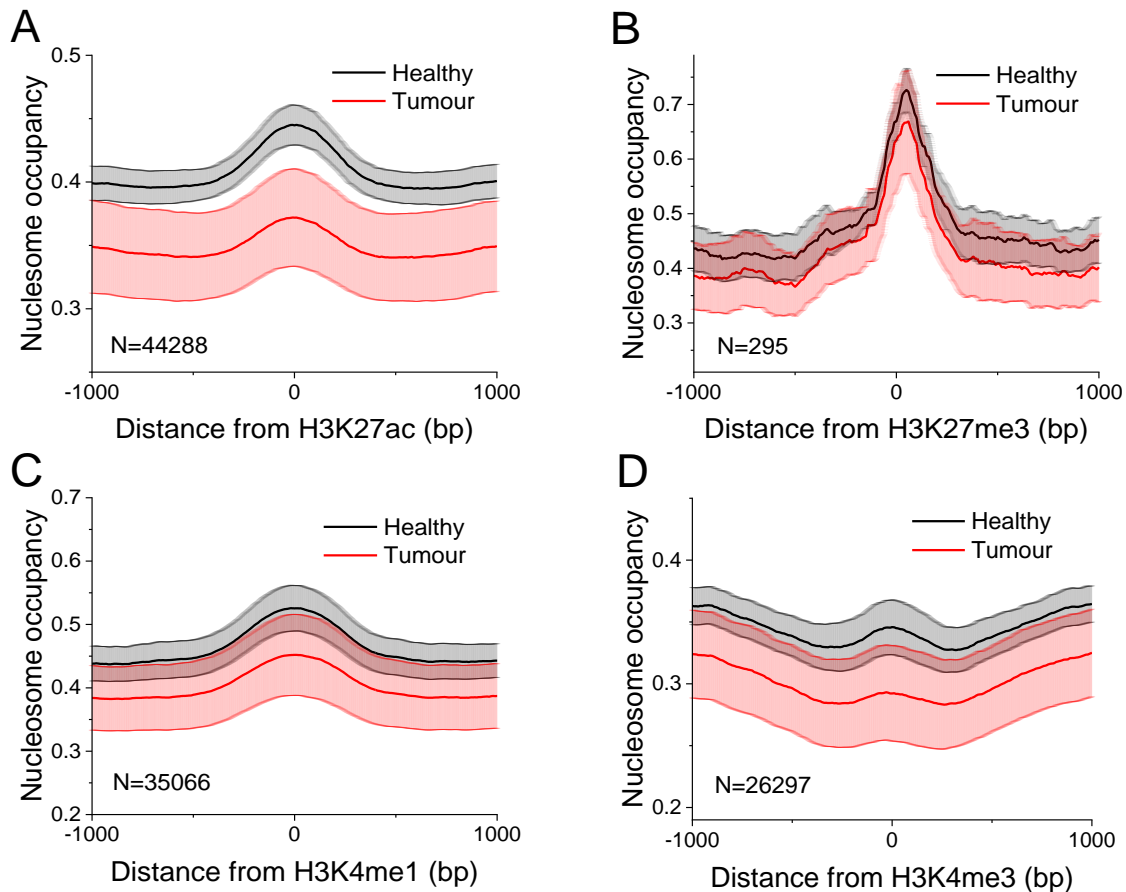


Figure 5. Average nucleosome occupancy profiles around regions enriched with different histone modifications. (A) H3K27ac (GSM3444438), (B), H3K27me3 (GSM3444439), (C) H3K4me1 (GSM3444441) and (D) H3K4me3 (GSM3444441) binding sites found in mesenchymal (MES) GBM subtype averaged over the 4 GBM patients from the experiments performed in the University of Essex. Grey areas correspond to the standard errors of averaging.

### 3.4.2. Chromatin differences between glioblastoma and neuroblastoma

Next, nucleosome occupancy profiles around CTCF after enrichment of patient nucleosomes with 4 histone marks found in IDH GBM subtype were calculated for all patients and are illustrated in Figure 6. Nucleosome occupancy profiles calculated around CTCF derived from a neuroblastoma cell line form a peak at the centre and nucleosome occupancy is slightly higher at the same region. It is important to highlight the results in Figures 6A and 6B. Figure 6A shows nucleosome occupancy around CTCF and H3K27ac binding sites. Occupancy is low at the centre of the binding sites that indicates presence of CTCF at H3K27ac binding sites.



Moreover, Figure 6B show high nucleosome occupancy levels around common binding sites of CTCF and H3K27me3. Both figures highlight the role of CTCF in active (H3K27ac) and inactive (H3K27me3) chromatin in GBM.

Furthermore, we calculated the same nucleosome occupancy profiles for CTCF derived from a GBM cell line (Figure S48, Appendix). Nucleosome occupancy patterns appear similar but distinctively less smooth. When comparing GBM and neuroblastoma, nucleosomal profiles for H3K27ac, H3K4me1 and H3K4me3 show similarity, but H3K27me3 profiles are completely different. In GBM, occupancy patterns are less clear while neuroblastoma shows a nucleosome peak at the centre. Even with a smaller number of sites, neuroblastoma demonstrates clearer profiles around CTCF.

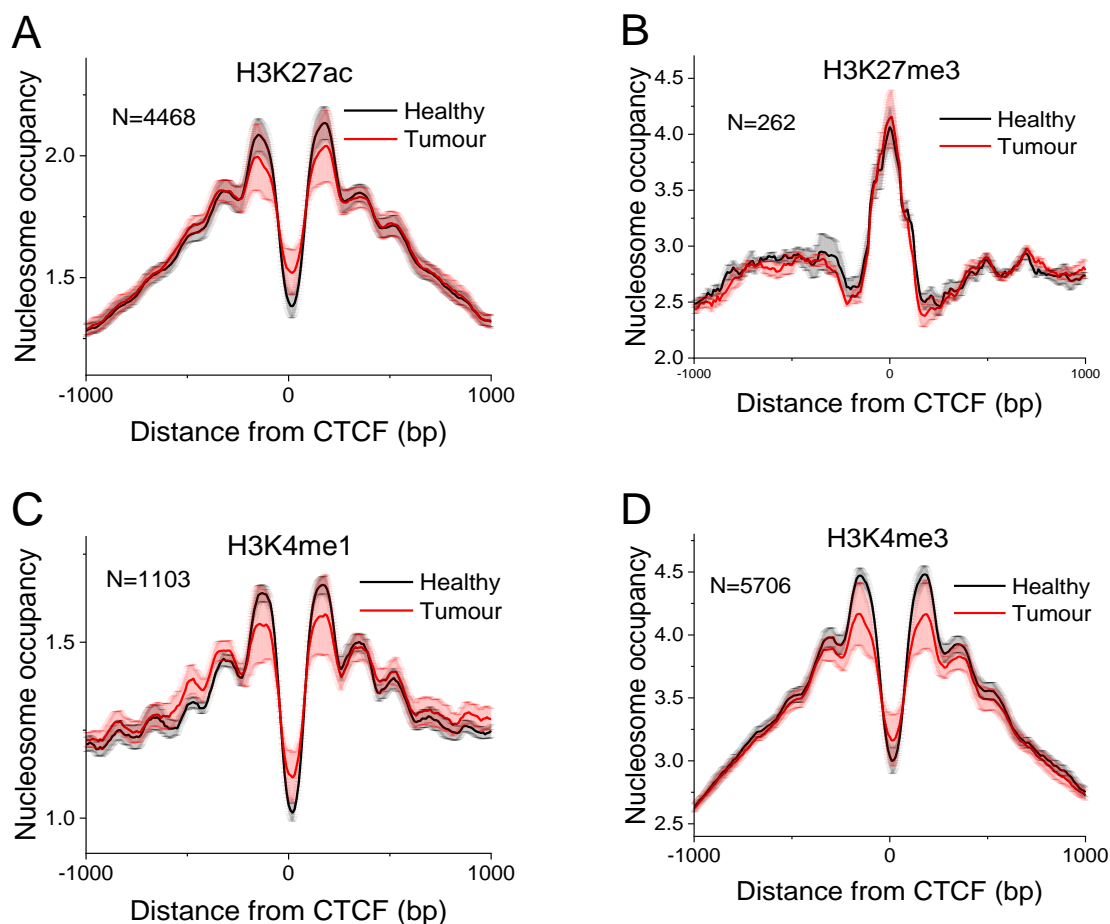


Figure 6. Nucleosome occupancy profiles around CTCF in regions enriched with different histone modifications from proneural IDH GBM subtype around CTCF bound in neuroblastoma (Gertz et al., 2013) in tumour for healthy (black) and tumour brain tissue (red) from the 4 GBM patients. (A) H3K27ac, (B) H3K27me3, (C) H3K4me1, (D) H3K4me3. Lighter areas correspond to the standard error of averaging.

### 3.5. Targeted loss and gain of nucleosomes from genomic regions

#### 3.5.1. GBM-specific nucleosome loss, gain and fuzziness

After determining 100bp and 1000bp regions that gained and lost nucleosomes in cancer using NucTools, average nucleosome occupancy profiles around these regions were calculated (Figure 7). Figure 7A shows nucleosome profiles of genomic regions that gained nucleosomes in GBM. Higher occupancy of nucleosomes is shown for GBM tissue at the centre of the regions with the nucleosome depletion decreasing. Individual patient profiles are provided in Appendix (Figure S49). Interestingly, occupancy levels are higher for patients G276 and G237

but lower for G125 and G370. Moreover, for patient G237 occupancy at the centre of nucleosome gained regions change from depletion to augmentation in GBM.

Profiles of genomic regions that lost nucleosomes in GBM (Figure 7B) show lower occupancy of nucleosomes in GBM compared to normal and high depletion at the centre. Healthy tissue shows high nucleosome occupancy overall and small depletion on nucleosomes at the centre. In GBM, nucleosome occupancy is lower and depletion is more defined. The individual graphs (Figure 8) for lost nucleosome regions show similarities between patients. Patients G125 and G370 have very similar occupancies with GBM lower than healthy and similar depletion of nucleosomes at the centre. Moreover, patient G276 shows similar occupancy overall, but extremely low occupancy at the centre from GBM tissue. On the other hand, nucleosome occupancy for patient G237 was very different at the centre since it changes from 'peak' to major depletion. This is a very interesting change that is related to only this one patient. Overall, when comparing regions of lost and gained nucleosomes in healthy and GBM, 16,815 and 29,436 sites were identified for each accordingly.

Next, we decided to calculate occupancy for 1000bp-long gained and lost regions (Figure 7C and D). In Figure 7C, the effect on nucleosome occupancy between healthy and GBM nucleosome a lot clearer when compared to the 100bp regions. For genomic regions that gained nucleosomes in GBM (Figure 7C), placement of nucleosomes is indicated in the centre of the gained nucleosome binding sites in GBM that endorses the gained nucleosome region calculations. Similar consistency can be seen for genomic regions that lost nucleosomes in GBM, where nucleosomes seem to be released around lost nucleosome binding site regions, a result that agrees with our NucTools calculations.

Next, we calculated nucleosome occupancy around sites that were calculated to be less or more fuzzy in GBM for 100bp regions in Figures 7E and 7F accordingly. In Figure 7C, we can see

occupancy of nucleosomes that exhibit more fuzzy behaviour. In GBM, occupancy levels are lower compared to normal and occupancy at the centre of the binding sites is lower indicating a higher number of nucleosomes at the centre. For Figure 7D, occupancy is similar between healthy and normal tissue around less fuzzy nucleosomes. At the centre of the bunding sites healthy tissue shows more nucleosomes than GBM.

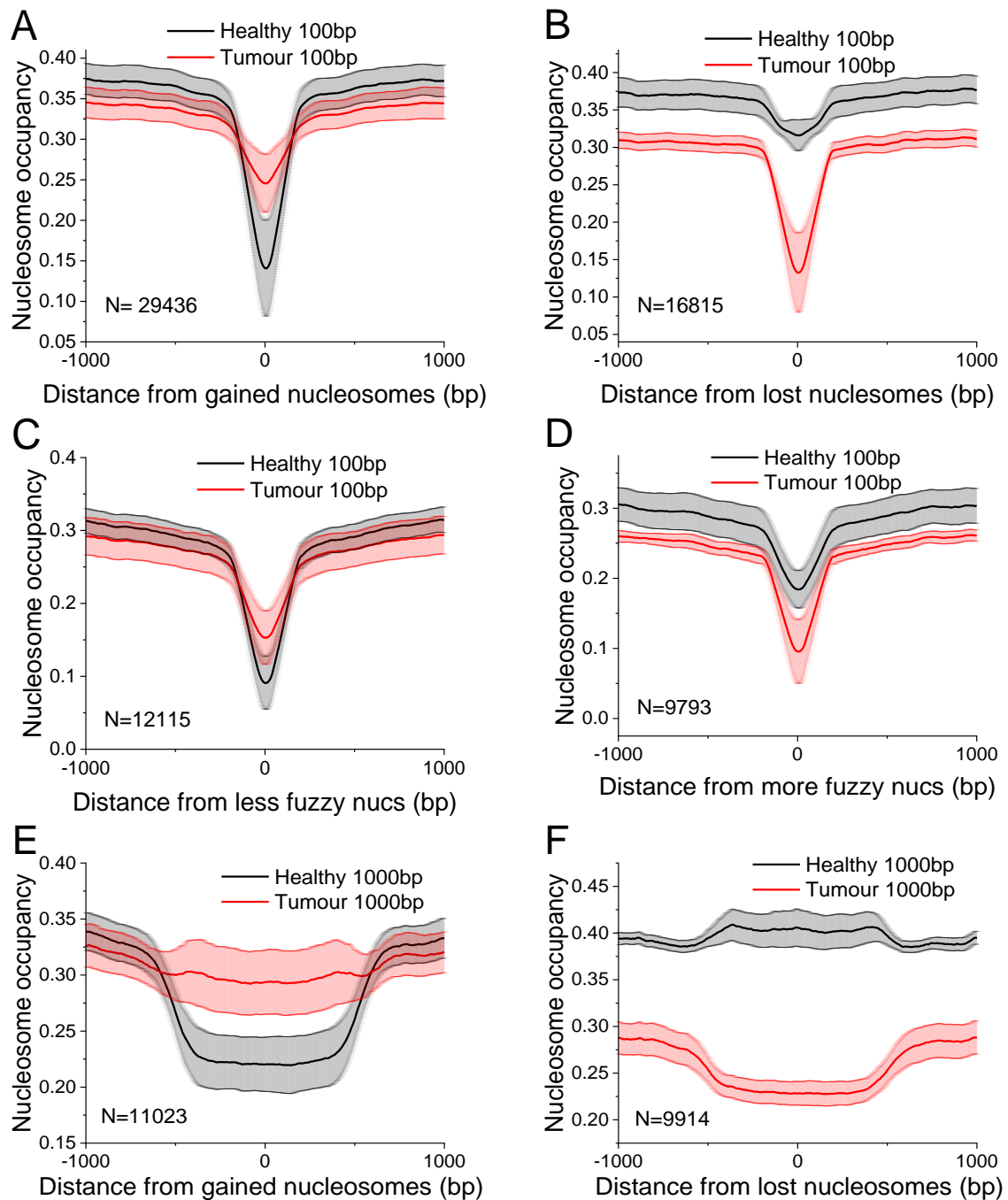


Figure 7. Average nucleosome occupancy profiles around regions that lost or gained nucleosomes or where nucleosome occupancy became more and less fuzzy in tumour versus healthy brain tissues from 4 GBM patients. Averaged profiles are shown separately for healthy (black) and tumour brain tissues (red). Lighter areas correspond to the standard errors of averaging. The number of regions (N) for each analysis is indicated on the graph. (A) 100-bp regions that gained nucleosomes in GBM versus healthy brain tissues, (B) 100-bp regions that lost nucleosomes in GBM versus healthy brain tissues, (C) 100-bp regions where nucleosome occupancy became less fuzzy in GBM versus healthy brain tissues, (D) 100-bp regions where nucleosome occupancy became more fuzzy in GBM versus healthy brain tissues (E-F) The same as (A-B) but for 1000-bp genomic regions.

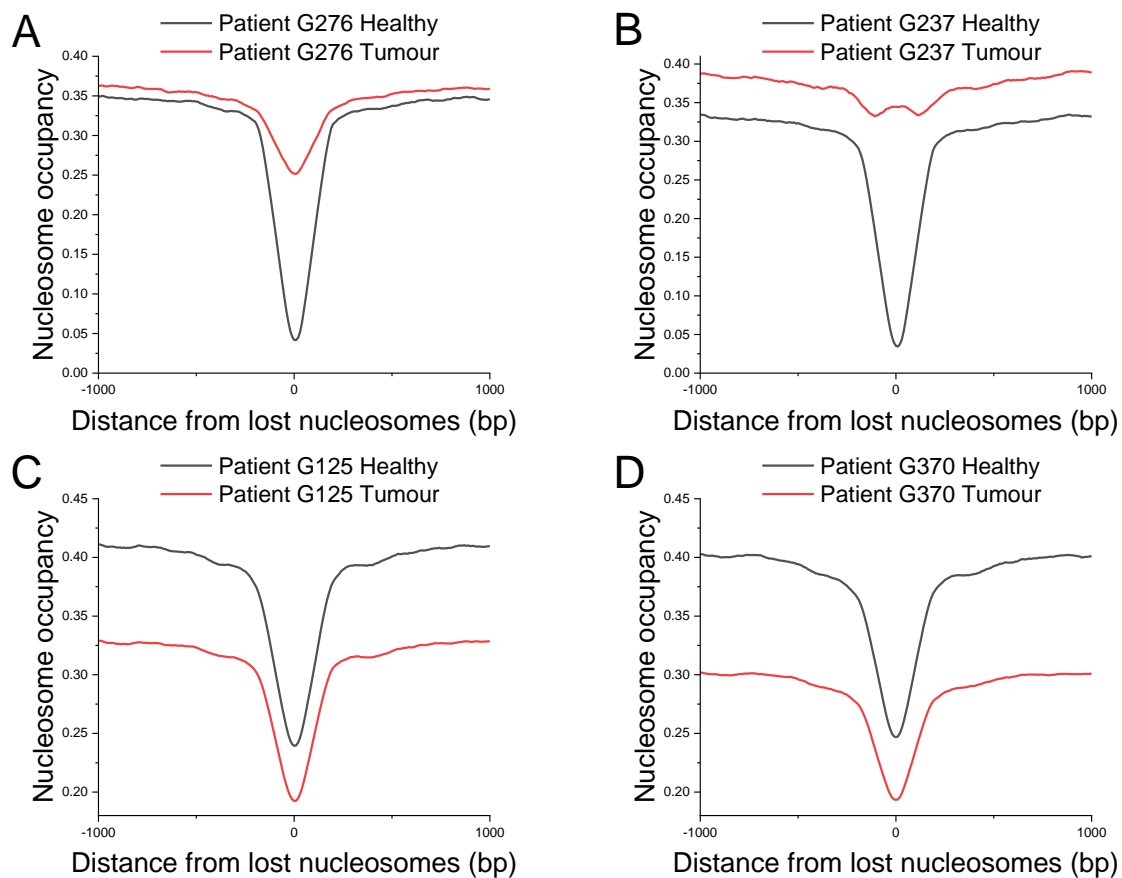


Figure 8. Average nucleosome occupancy profiles around genomic regions that lost nucleosomes in tumour versus healthy brain tissues, plotted individually for 4 GBM patients. (A) Patient G276 (B) Patient G237 (C) Patient G125 (D) Patient G370. Non-malignant brain tissue samples are shown in black and GBM profiles in red.

### *3.5.2. Enrichment of lost and gained nucleosome regions at TFs.*

To further investigate nucleosome positioning in GBM, enrichment of the calculated lost and gained nucleosome regions was measured for all TFs and presented in Figure 9. Figure 9 demonstrates fold enrichment analysis of gained and lost nucleosome regions at TF binding sites. Overall, enrichment changes according to the TF and no specific pattern can be observed. For example, all CTCF binding sites seem to be enriched with gained nucleosome regions, a result that indicates the loss of CTCF binding sites in cancer. We further investigated CTCF by using different quantiles of binding strength in Figure 9B. It is clear that CTCF quantile 3 has the highest enrichment of gained nucleosome regions at more than 3.5. Interestingly, less strong quantiles 1 and 2 seem to have enrichment of lost nucleosome regions at their binding sites. Finally, in Figure 9B we calculated enrichment at the top 10% of CTCF sites. It seems that regions gaining nucleosomes tend to be more enriched at higher quality CTCF regions.

In addition to CTCF, enrichment of lost nucleosome regions can be seen for TFs like ARNT2, BMAL1, POU3F2, RBPJ, CHD4, MYC, MAX and JMJD6 in Figure 9A. Enrichment of regions losing nucleosomes at binding sites indicates “nucleosome-free” regions where production of the TF can increase. For TFs such as JMJD6, MYC, KLF4 and BMAL1 it is seen that there is enrichment for both lost and gained nucleosome regions but higher for lost is observed.

### *3.5.3. Enrichment of lost and gained nucleosome regions at different genomic regions*

We decided to go beyond TF binding sites and look at enrichment at different genomic regions (Figure 9B). First, we can see that overall, there is a gain in nucleosomes at all regions except CpG islands where higher lost nucleosome region enrichment can be seen. Main genomic regions such as promoters, enhancers and ALU repeats show small enrichment of gained

nucleosomes regions indicating the loss of binding sites for each region. CpG island show gain and loss of nucleosomes at regions.

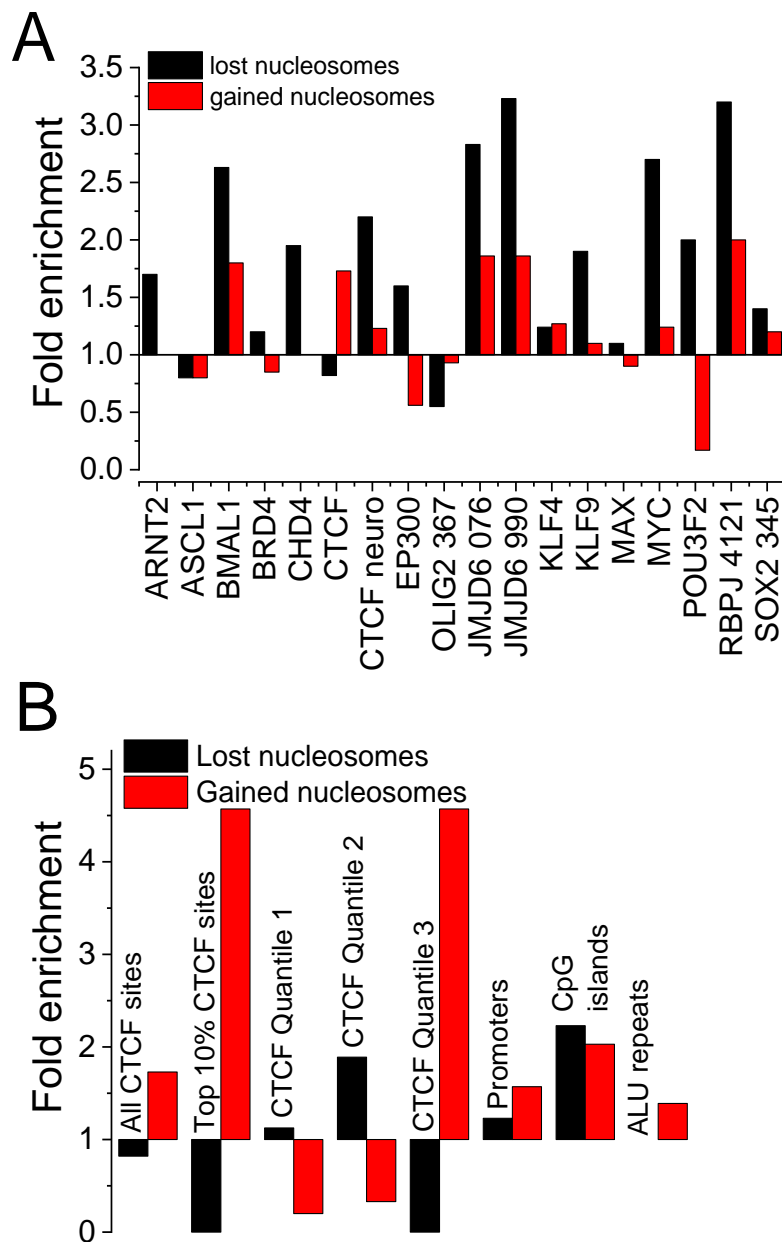


Figure 9. Fold enrichment of binding sites of TFs and other genomic regions that gained/lost nucleosome regions in GBM versus healthy brain tissues. (A) Enrichment of 100bp regions of lost and gained nucleosomes in GBM at binding sites of 19 TFs. (B) Enrichment of 100bp regions of lost and gained nucleosomes in GBM at genomic regions and CTCF binding sites including different binding strength quantiles, 1 being the weakest and 3 the strongest. Black bars indicate enrichment in genomic regions that lost nucleosomes and red bars indicate enrichment in regions that gained nucleosomes in GBM.

### 3.5.4 Gene ontology analysis of promoters that gained and lost nucleosomes in GBM

Next, we performed Gene Ontology (GO) analysis of promoters that lost and gained nucleosomes in GBM. For this analysis we decided to use 1000bp-long regions of the gained and lost nucleosome region datasets. First, in Figure 10A we can see processes related biological processes of genes that are expressed in promoters that gained nucleosomes. First, we can see GO such as “blood coagulation” and “cell-cell signalling” having lower p-values and therefore indicating significance for gained nucleosome regions. Other GO terms that show significance are “B cell differentiation,” “Type I interferon-mediated signalling”, “humoral immune response”, “B cell proliferation” and “natural killer cell” and well as “T cell activation involved in immune response”. It is clear that promoters with gained nucleosome regions that are expressed in GBM are highly involved in the expression of immune responses. This is a result that highlights the immune-system-corrected nature of GBM and will be discussed in section 4 of the thesis. Lastly, we can see terms such as “response to exogenous dsRNA” and “peptidyl-serine phosphorylation of STAT protein” showing significant p-value. Connection of these pathways with GBM deregulation will be discussed in the “Discussion” section.

For lost nucleosome regions, very different GO terms are enriched based on p-value (Figure 10B). The top terms that we found are analysis were “viral transcription”, “intercellular signal transduction” and “regulation of protein ADP-ribosylation”. Moreover, other terms such as “stress activated MARP cascade”, “GDP metabolic process”, “rRNA processing”, “RNA splicing”, “G2/M transition of mitotic cell cycle”, “cell cycle” and “negative regulation of protein ubiquitination”. Generally, GO terms involved in gene regulation and the cell cycle seem to be link to the gene expressed at promoters that are losing nucleosomes. The calculated gene ontology for both gaining and losing nucleosome regions can be linked to glioblastoma initiation and development will be further analysed in the “Discussion” section.



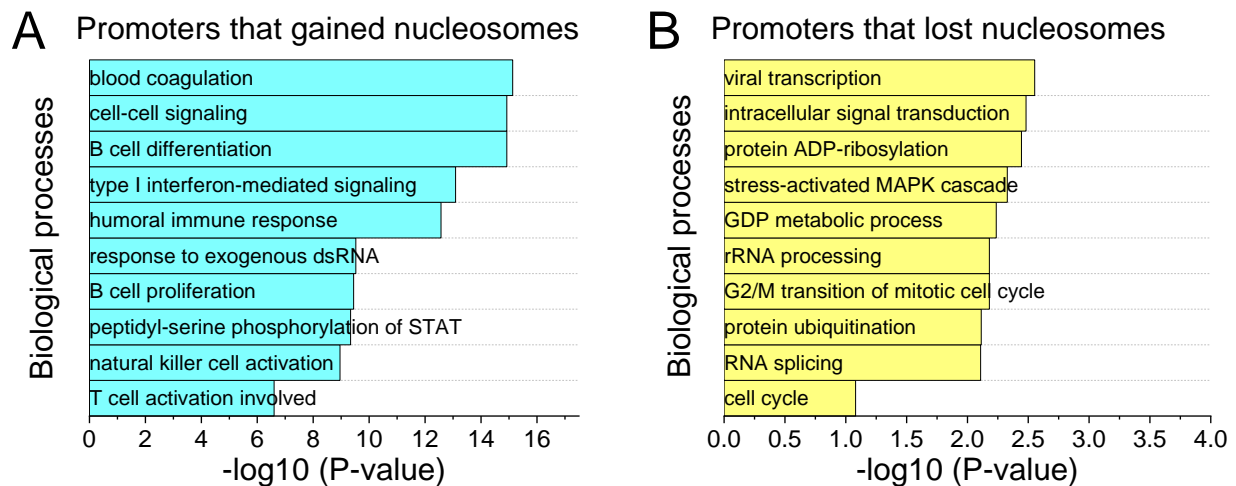


Figure 10. Gene ontology analysis of promoters that lost and gained nucleosomes in GBM. (A) Biological processes of genes expressed in promoters after intersection with 1000bp-long regions of nucleosomes that are gained and (B) lost nucleosomes in GBM using online functional annotation tool DAVID.

### 3.5.5. Classification of samples using Principal Component Analysis based on nucleosome occupancy at GBM-sensitive genomic regions

In order to provide a proof that it is possible to classify samples into tumour and non-malignant based on nucleosome occupancy, we continued with Principal Component Analysis (PCA) for four pairs of malignant/non-malignant brain tissue samples. We used for this analysis the set of “GBM-sensitive” regions that gained or lost nucleosomes in GBM. Using this set of regions, we performed PCA of nucleosome occupancy at regions gaining nucleosomes (A) and losing nucleosomes (B). First, it is clear that nucleosome positioning is different between healthy and tumour and can be easily grouped. For regions that gained nucleosomes, the healthy patients form a distinct cluster and are characterized by high values. GBM patients show lower values and are more clustered than normal. On the other hand, regions that lost nucleosomes, both clusters are less tightly grouped. Moreover, two healthy patients G370 and G125 show similarity, while when comparing G125 healthy and tumour nucleosome occupancy they are fairly similar. To conclude, after performing a PCA on lost and gained regions in GBM there is a clear separation between healthy and tumour along a principal component that is related to

nucleosome occupancy. This provides as with the proof of principal that it is possible to use nucleosome occupancy at a set of GBM-sensitive regions to distinguish malignant/non-malignant samples.

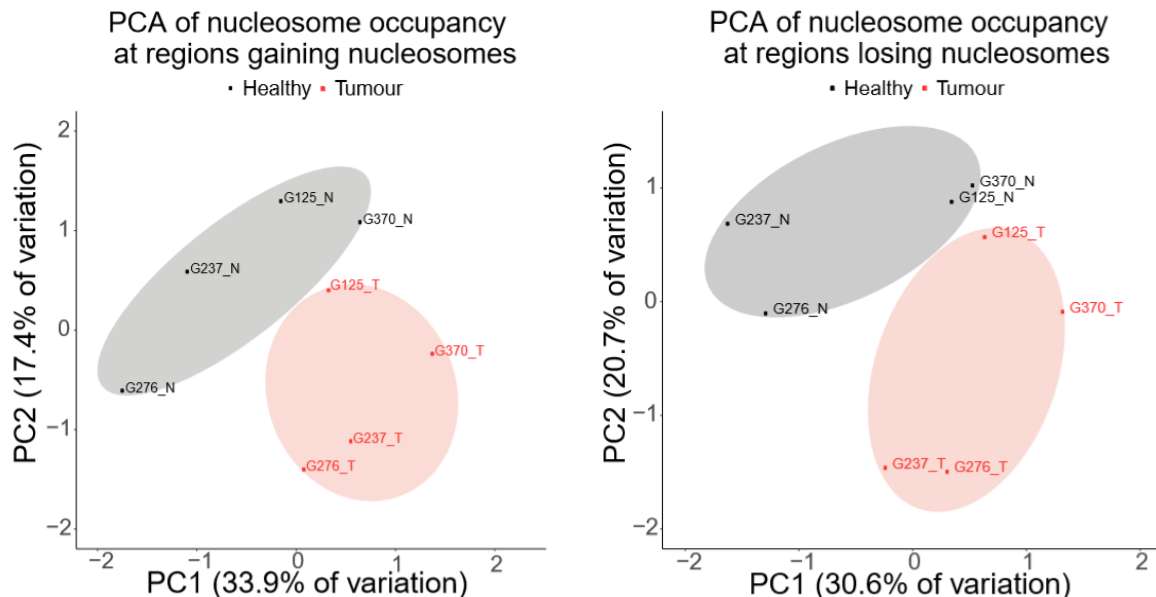


Figure 11. Principal Component analysis (PCA) allows classifying healthy versus tumour samples based on nucleosome occupancy. Four healthy (black) and four GBM (red) MNase-seq tissue samples were used for the analysis. A) PCA based on the average values of nucleosome occupancy in the set of 9,914 regions of length 1000-bp that were characterised by increased nucleosome occupancy in GBM. B) PCA based on the average values of nucleosome occupancy in the set of 11,023 regions of length 1000-bp that were characterised by the decrease of nucleosome occupancy in GBM. At both types of regions gaining and losing nucleosomes, PC1 and PC2 explain 51.3% of the variance.

### 3.6. Average nucleotide frequency profiles along nucleosomes from brain tissues

Next, we examined the average DNA nucleotide frequency along the nucleosome based on MNase-seq in malignant and non-malignant brain tissues from GBM patients. Often, nucleotide frequencies within the nucleosome dyad in cancer can highly influence the mutation rate (Yazdi et al., (2015)). For this analysis we filtered DNA fragments to only contain fragments sizes 146-148bp. Figure 12 shows the frequency of nucleotides ‘A’, ‘T’, ‘C’ and ‘G’ as a function of the distance from the nucleosome centre (dyad) in healthy and GBM tissue.

Frequency of A nucleotide shows similar frequency patterns between healthy and GBM for each patient, apart from patient G276 which show major differences at positions -60 to -50 and 10 to 20 between GBM and healthy. Moreover, similar differences can be observed in 'T' frequency -20 to -10 and 50 to 60, as it is expected. In addition, both 'A' and 'T' show lower frequency for healthy near the nucleosome dyad at 3 and -3 accordingly. Frequency of 'C' shows similar results. Most patients have closely parallel patterns in healthy and GBM. Patient G276 shows the most differential patterns from healthy to GBM. At 25bp, tumour frequency pattern is different from healthy tissue. In addition, frequency of 'G' nucleotide shows differences around -55bp, -3bp and 57bp with patient G276 showcasing the biggest frequency difference between the 4 individuals.

Furthermore, we looked at "A" nucleotide frequency around the nucleosome dyad for four different healthy and four cfDNA GBM samples from Song et al. (2017) in Appendix, Figure S51. Overall, in cfDNA we found that average patterns did not have major changes, beside the change expressed at 60bp in GBM patients.

Next, we looked at average nucleotide patterns in healthy and GBM MNase-seq calculated by aligning the signal at the start of the nucleosome dyad to check for differences specifically at the start sites of the nucleosome dyad (Figure S57). While there are no major changes overall, nucleotides "A" (Figure S57A) and "C" (Figure S57C) show some differences in GBM. "A" shows higher frequency at 20bp, 40bp and 80bp and "T" frequency is higher at 85bp. The fuzziest frequency is expressed by "C" and can be seen throughout the nucleotide pattern in GBM, while levels of "G" showed not differences between healthy and GBM.

Moreover, the major differences nucleotide patterns presented in cancer can be shown in Figure S56 in Appendix. We can see that in the top panel nucleotide patterns between the two patients in healthy tissue are very similar to each other with very small insignificant differences. On the

contrary, in GBM MNase-seq shows different nucleotide patterns in GBM indicating high variance of A/T/C/G nucleotide placement in cancer. Overall, through these results we prove that nucleotides patterns can be affected in GBM, as well as pinpoint specific places along the nucleosome where we can see pattern differentiation. However, at least for the limited patient cohort considered in this project, the changes of nucleosomal nucleotide frequencies in tumour were not as pronounced as other changes discussed in the previous chapters.

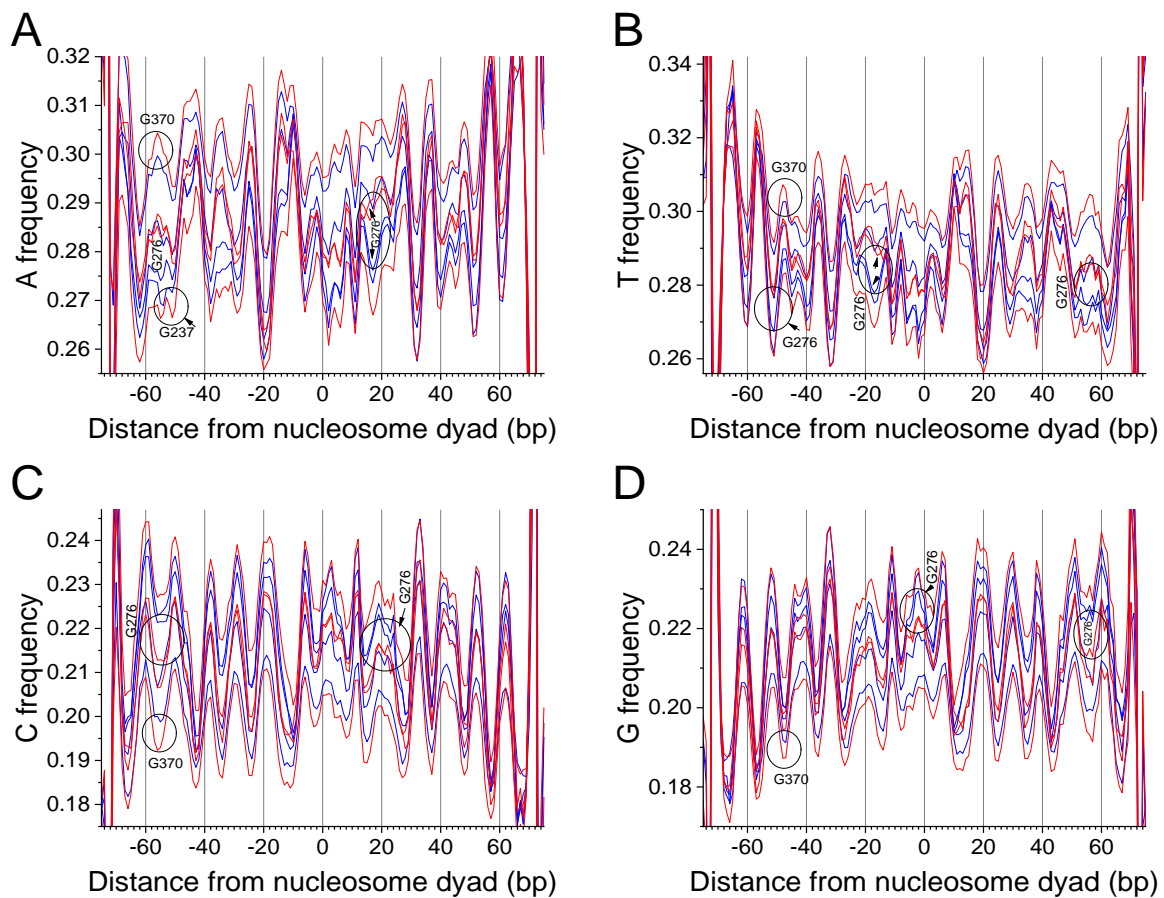


Figure 12. Nucleotide frequencies around nucleosomes (146-148bp) based on MNase-seq in paired brain samples from GBM patients (blue – normal; red – tumour). A – adenine, B – thymine, C – cytosine, D – guanine.

## 4. Discussion

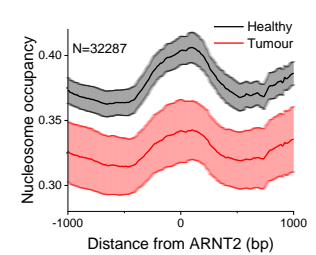
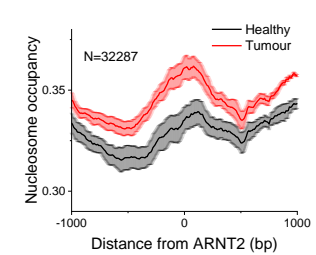
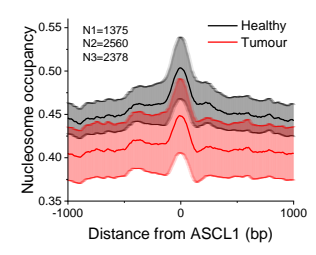
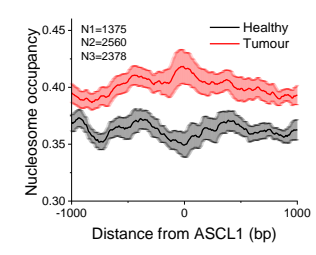
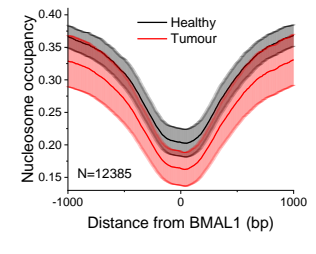
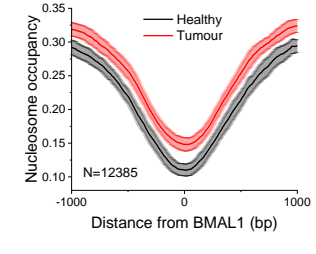
Glioblastoma is defined on a genetic as well as an epigenetic level, with the latter becoming increasingly important as it offers vital information about cancer pathogenesis and insight that can help with cancer treatments. The analysis of nucleosome positioning is a new direction that we show here to be a promising way of deciphering the molecular details of gene regulation and mis-regulation in cancer. In this project, we studied for the first time, nucleosome occupancy changes in brain tissues from glioblastoma patients in order to identify changes in epigenetic regulation. Moreover, we have reconstructed nucleosome positioning based on cfDNA from GBM patients and healthy people with the goal to make current diagnostic methods more efficient. We identified a set of cancer-sensitive nucleosome positioning changes that can serve as new GBM markers that can be used for prognosis and therapy and will help in making cfDNA-based liquid biopsy more effective.

### 4.1. Interplay of TFs and nucleosome positioning in GBM

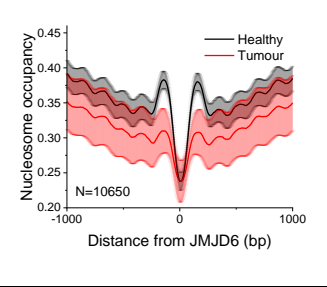
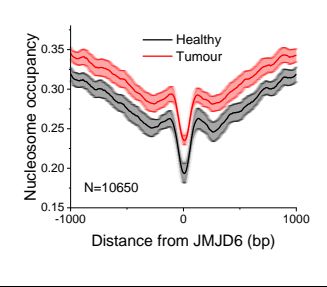
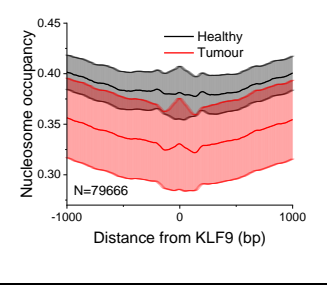
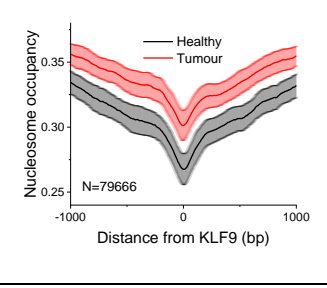
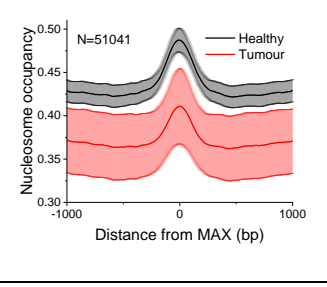
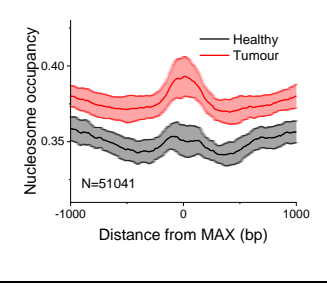
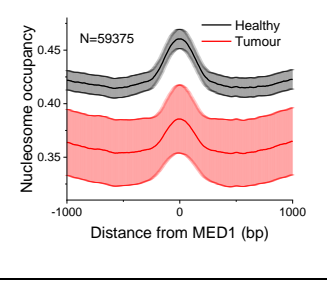
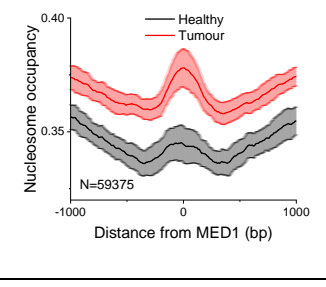
During our analysis, we identified several TFs involved in GBM formation, whose binding sites were characterised by differential nucleosome positioning in GBM. We found that on average, nucleosome occupancy around TFs is less well-defined in GBM than in healthy tissue. Moreover, we calculated nucleosome occupancy profiles around TFs in cfDNA. Average nucleosome occupancy patterns allowed us to define which binding sites can be used as GBM markers. A summary of all analysed TF datasets is shown in Table 3.

One of TFs whose binding sites were deregulated in terms of chromatin rearrangement in GBM was the chromatin insulator CCCTC-binding factor (CTCF) (Figure 2A and 2O, Supplemental Figures S1 and S3). Note that while the average nucleosome profile around CTCF in GBM tissues is very fuzzy (Figure 1A), the profiles of individual patients show distinct shape changes between normal and GBM tissues (Figure 2O, Figure S3).

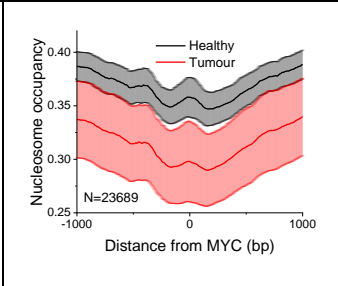
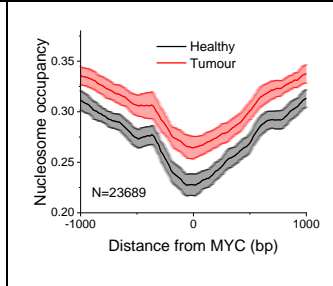
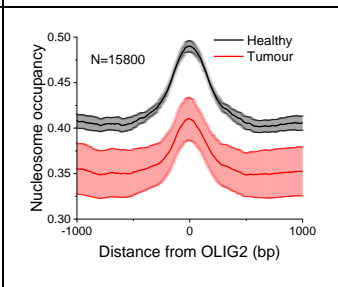
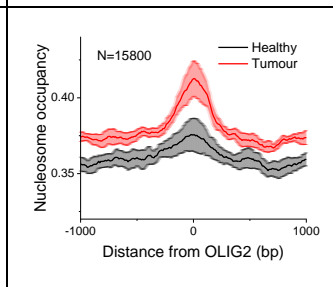
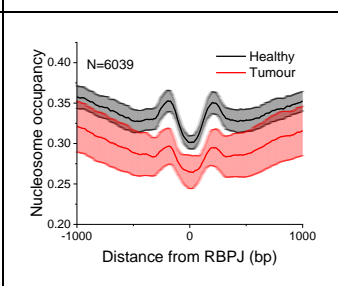
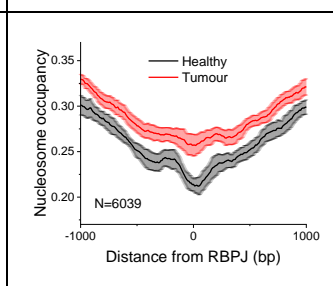
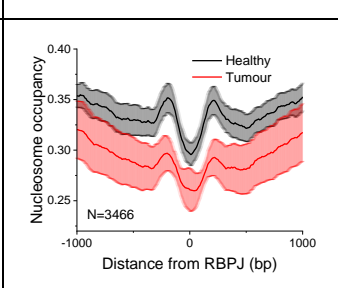
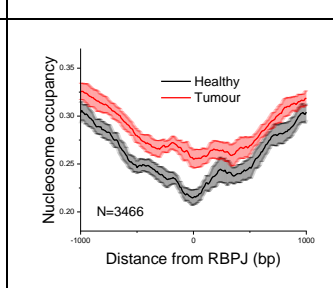
**Table 3.** Summary of results for all analysed TF datasets.

TF name	Cell type	TF dataset reference	Gene expression fold change	Lost/gained nucleosomes fold enrichment	MNase-seq in healthy (black line) vs tumour brain tissues (red)	cfDNA in healthy people (black line) vs GBM patients (red)
ARNT2	Glioblastoma Stem Cell (GSC)	GSE98330 (Bogea et al., 2018)	0.9	lost: 1.7; gained: 1		
ASCL1	G523NS cells treated with doxycycline	GSM2335531, GSM2335532, GSM2335533 (Park et al., 2017)	3.3	lost: 1.3; gained: 1  lost: 0.7; gained: 0.6  lost: 0.4; gained: 0.8		
BMAL1	Glioma stem cells	GSM1306364	0.83	lost: 2.63; gained: 1.8		

BRD4	U87MG (Uppsala 87 Malignant Glioma)	Remap GSM1038284 (Lovén et al., 2013)	0.83	lost: 1.2; gained: 0.85		
CTCF	H54 (D54) glioblastoma cell line	GSM822303 ENCODE project	0.64	lost: 0.82; gained: 1.73		
CTCF	Neuro- blastoma cells in continuous culture	GSM803333 (Gertz et al., 2013)	0.64	lost: 2.2; gained: 1.23		
JMJD6	3565 cells	GSM2360990 (Miller et al., 2017)	1.1	lost: 3.23; gained: 1.86		

JMJD6	528 cells	GSM1922076 (Miller et al., 2017)	1.1	lost: 2.83; gained: 0.91		
KLF9	Human GBM stem-like cells	remap2018 GSM1522563 (Ying et al., 2014)	2.75	lost: 1.9; gained: 1.1		
MAX	U87MG (Uppsala 87 Malignant Glioma)	remap2018 GSM894077 (Lin et al., 2012)	0.25	lost: 1.4; gained: 0.93		
MED1	U87MG (Uppsala 87 Malignant Glioma)	remap2018 GSM894082 (Lin et al., 2012)	1.1	lost: 1.1; gained: 0.9		



MYC	U87MG	remap2018 GSM894061 (Lin et al., 2012)	11.3	lost: 2.7; gained: 1.24		
Olig2	MGG8	GSM1306365 (Suvà et al., 2014)	2.4	lost: 0.66; gained: 0.93		
RBPJ	brain tumour initiating cells	GSM2101764 (Xie et al., 2016)	1.02	lost: 3.13; gained: 1.1		
RBPJ	glioblastoma initiating cells	GSM2101765 (Xie et al., 2016)	1.02	lost: 3.2; gained: 2		

SOX2	NOD-SCID mouse harbouring HOT1 tumour derived from human glioblastoma	GSE58345 (Singh et al., 2017)	3.1	lost: 1.4; gained: 1.2		
SOX2	MGG8	GSM1306360 GSM1306362 (Suvà et al., 2014)	3.1	lost: 0.7; gained: 0  lost: 0.5; gained: 0.8		
POU3F2	MGG8 human glioblastoma stem cell line	GSM1306358 (Suvà et al., 2014)	2.2	lost: 2; gained: 0.17		

CTCF plays an important role in gene regulation through the recruitment of other proteins to promoters of target genes as well as contributing to the creation of three-dimensional chromatin structure (Oh et al., 2017). We found that nucleosome occupancy profiles around CTCF in healthy vs. tumour tissue showed a consistent change at the centre of CTCF binding sites, indicating the average loss of CTCF. A recent study by Damaschke et al (2020) determined that the decrease of CTCF expression was associated with hypermethylation of CTCF binding sites in prostate cancer. In IDH mutant gliomas, CTCF binding sites exhibit hypermethylation, compromising the binding of CTCF protein. This reduction of CTCF binding has been associated with loss of insulation between topological domains and aberrant gene activation (Flavahan et al., 2016). In the latter study, Flavahan et al demonstrated that loss of CTCF permits the continued interaction of a constitutive enhancer and the known glioma oncogene PDGFRA. They revealed strong expression of the PDGFRA oncogene in IDH gliomas and associated this with poorer outcome in gliomas. They directly proved the implication of CTCF in gliomas by allowing CTCF to bind and restore PDGFRA insulation by treating glioma cell lines with DNA methyltransferase inhibitor 5-azacytidine. This treatment reduced the hypermethylation of CTCF motifs by ~2.5-fold, increased occupancy by ~1.7 fold and down-regulation PDGFRA expression by ~5-fold. The results clearly implicated the consequences of loss of CTCF binding. In our study, clear loss of CTCF binding has been identified after discussing our nucleosome occupancy profiles. The epigenetic disruption of CTCF can show multiple oncogenic effects on a transcriptional level and can provide an advantage to GBM and other cancers (Fang et al., 2020).

Our analysis also showed that CTCF binding sites are characterised by high enrichment of gained nucleosomes in GBM (Figure 9), indicating average loss of CTCF at these locations. Moreover, we found that CTCF binding sites that have the strongest motifs are among the most disrupted in GBM in terms of their nucleosome occlusion (Figure 9B). It is known that DNA

methylation can affect CTCF binding by regulating the affinity of CTCF to DNA (Hashimoto et al., 2017). Therefore, it is possible that in cancer, strong CTCF binding sites could have lost CTCF due to aberrant DNA methylation. Consequently, further research can be done on the interplay of nucleosome positioning and DNA methylation around CTCF.

We next focused on cfDNA nucleosome fragment positioning around CTCF binding sites. Other studies have found that cfDNA fragments can footprint CTCF and other TF binding sites (Snyder et al., 2016). We found that cfDNA occupancy around CTCF was increased in tumour. Increased presence of cfDNA in cancer has been observed before and is detected due to higher cellular turnover and cell death due to treatment (McAnena et al., 2017; Kuroi et al., 1999). However, current approaches don't allow to accurately determine the accessibility or changes of TF binding sites in tissues based on cfDNA (Ulz et al., 2019). Our results show that the CTCF binding site nucleosome profiles reconstructed based on MNase-seq and cfDNA are closely similar to each other, which may offer a good start for GBM detection without the use of traditional methods of biopsy.

In addition to CTCF, our analysis detected loss of or lower accessibility to binding sites of several other TFs, with most prominent cases for RBPJ, BRD4, JMJD6 and KLF9. These TFs showed differential patterns of nucleosomes at the centre of the binding sites that indicated loss of the TFs in GBM tissue.

For the case of RBPJ, nucleosome patterns around its binding sites varied in patients, with half of the patients showing a significant increase of nucleosome occupancy at RBPJ binding sites (with the shape of the occupancy profile changing from a local minimum to a peak, Figure S5). This variation can be contributed to different GBM subtypes. RBPJ is the main regulator of the NOTCH pathway activity, a pathway that has been linked to cancer due to its self-renewal and proliferation characteristics. NOTCH has been targeted in GBM therapies but was proven to

have limited effects in clinical trials. In a previous study, Xie et al (2016) reported NOTCH inhibitors to be ineffective when RBPJ is present. In the absence of NOTCH, RBPJ regulated genes critical to tumour initiating cell stemness and cell cycle progression. Moreover, it was preferentially expressed by brain tumour initiating cells and critical for their self-renewal and growth. The latter paper suggested that RBPJ would be an effective target against tumour growth. A recent study by Zhang et al. (2020a) demonstrated that RBPJ levels can fluctuate depending on GBM subtype such as proneural-like glioblastoma stem cells (GSG) GSCs and mesenchymal-like GSCs. Suppression of RBPJ was shown to upregulate proneural markers and downregulate mesenchymal markers. After knockdown of RBPJ, transition of proneural to mesenchymal subtype GSCs was halted and mesenchymal phenotypes were transitions back to proneural leading to slower progression of GBM (Zhang et al., 2020a). Since the mesenchymal subtype shows the most aggressive behaviour, knockdown of RBPJ accompanied with other therapies can be a promising target for GBM regulation. While RBPJ could be used as a GBM therapy option, it can be proven to be a good GBM marker considering cfDNA liquid biopsy. However, cfDNA patterns around RBPJ have not been investigated before. Here we performed this for the first time and concluded that cfDNA profiles in GBM patients are similar to MNase-seq profiles in brain tissues. To support our findings, the enrichment analysis for RBPJ showed up to 3-fold enrichment at genomic regions that lost nucleosomes in GBM. Finally, even though cfDNA profiles are distinctly less detailed, it is clear that further analysis of RBPJ can uncover a very useful new marker for diagnosis using cfDNA.

In the case of KLF9, nucleosome occupancy profiles around binding sites of this TF also showed a drastic change from a local minimum to a peak for half of the patients (Figure S17). KLF9 controls GSC differentiation by repressing the NOTCH pathway and CD151, a protein that has been associated with GBM malignancies (Tilghman et al., 2016). KLF9 has previously been shown to be downregulated in other cancers (Zhong et al., 2018; Li et al., 2019). In their

study, Tung et al. (2018) demonstrated how upregulated KLF9 expression along with epigenetic modulators can promote cancer cell death and proved KLF9 to benefit GBM therapies. Moreover, Ying et al. (2014) demonstrated that after KLF9 induction in GBM, several GBM markers, such as OLIG2, SOX2, Nestin and BMI1 were decreased. Therefore, our nucleosome positioning profiles are consistent with the downregulation of KLF9 in GBM and could be used as GBM marker in the future.

In the case of BRD4, the tumour-associated change of the shape nucleosome occupancy profiles around its binding sites was not as drastic as for CTCF, RBPJ and KLF9, but still detectable. Dysfunction of BRD4 is known to lead to incorrect transcriptional activation of oncogenes, such as c-Myc and BCL2 in many cancers, including neuroblastoma, breast cancer and glioblastoma (Shi et al., 2020). Previously, Cheng et al. (2013) suggested that epigenetic mechanisms controlled by BET proteins, including BRD4, are commonly used by GBM. Moreover, Pastori et al. (2014) demonstrated the importance of BRD4 in glioblastoma by disrupting expression in GBM cells. Their results not only indicated overexpression of BRD4, but also showed that after BRD4 knockdown, GBM cells reduced cell cycle progression and proliferation. Considering the ability of BRD4 to inhibit c-Myc expression it makes it an attractive target of GBM therapy. Taking into account the previous studies, it is clear the overexpression of BRD4 is a common symptom of GBM and our results relate to its function. Indeed, our analysis showed enrichment of lost nucleosomes at BRD4, suggesting higher expression of this protein. Finally, our analysis showed that cfDNA nucleosome profiles have highly similar patterns to those based on MNase-seq in brain tissues (Figure S24B, Appendix). These findings are consistent with the literature mentioned above, showing that BRD4 could be used as a marker in cfDNA biopsies.

There is a known connection between BRD4 and JMJD6 in cancer. BRD4 is used by JMJD6 by forming a protein complex to activator of enhancer-mediated-pause-release of RNA

polymerase II, leading to aberrant gene expression in GBM (Wong et al., 2019; Miller et al., 2017). JMJD6 is known to regulate multiple cancer-related signalling, such as downregulation of p53 activity, MAPK signalling and suppression of Myc-induced apoptosis, making it a major target against cancer (Yang et al., 2020). Overexpression of JMJD6 is often seen in several cancers, including GBM, a behaviour that correlates with our current nucleosome positioning results (Wang et al., 2014a). Our analysis of nucleosome occupancy patterns shows reduced positioning of nucleosomes at the centre on JMJD6 binding sites in GBM tumours as opposed to paired non-tumour brain samples (Figures S15 and S16). These nucleosome patterns, along with our analysis that indicated enrichment of lost nucleosomes are JMJD6 sites (Figure 9), suggest overexpression of JMJD6, in line with previous literature. Our reconstruction of nucleosome occupancy based on the corresponding cfDNA shows similarities with the profiles reconstructed based on MNase-seq. Additional cancer-specific cfDNA analysis could give us more information for the use of JMJD6 as a new marker in liquid biopsies.

Finally, it is important to mention the cluster of neurodevelopmental TFs (POU3F2, SOX2, SALL2, OLIG2) that are directly involved in GBM propagation across the genome (Suvà et al., 2014). Suvà et al. (2014) discovered for the first time that induction of this set of TFs is sufficient to reprogram cell differentiation and used it to identify GBM-reproductive cells and distinguish key regulatory targets of the four TFs. Individually, the set of TFs are still connected to GBM and general brain activity. In their research Zhang et al (2020b) established a novel marker TRIM25 that is solely expressed through activation of SOX2 in GSCs. They reported that activation of SOX2 promoted the stemness and high invasiveness in GBM and inactivation of the TF showed reduced expression of GSCs. In the same context, upregulated SOX2 levels have been seen by Wang et al. (2019), where high SOX2 expression reinforced NOTCH1 expression in GSCs. Moreover, OLIG2 and POU3F2 promote neural differentiation, playing a key role in the rapid reproduction of GBM. Finally, SALL2 is expressed in brain tissue as a

promoter of neural development (Fiscon et al., 2018). Due to the nature of this complex in GBM, it is expected of them to have higher expression in cancer patients. While our nucleosome positioning graphs did not show specific changes in nucleosome occupancy patterns for the available TFS, SOX2, POU3F2 and OLIG2, the enrichment analysis provides us with some interesting results. Both SOX2 and POU3F2 showed high enrichment of lost nucleosomes that indicates higher expression of both TFs. For OLIG2, we did not have any significant enrichment and therefore, further analysis can be beneficial. Overall, “fuzzier” nucleosome positioning is presented for all these three TFs.

#### **4.2. Nucleosome occupancy changes around ALU repeats.**

During this project we also analysed nucleosome occupancy around ALU repeat elements and discovered that their nucleosome occupancy is influenced in GBM (Figures 3 and 4). ALU repeats are widely hypomethylated in cancer cells (Jordà et al., 2017). We separated ALU repeat into two groups: roughly dimeric and monomeric based on their length. Overall, we saw that nucleosome occupancy levels around ALU repeats were lower in GBM. ALU repeats are characterised by CG dinucleotide periodicities which affect the distribution of nucleosomes that they harbour (Bettecken et al., 2011). ALUs have been reported to be associated with well-phased nucleosomes that are enriched with enhancer-like active modifications (Su et al., 2014). The nucleosome occupancy profiles that we calculated round ALUs contain two peaks at the centre, which may correspond to phased nucleosomes. The hypomethylation of ALUs in GBM may suggest that higher expression of ALU repeats would be expected, in line with our observation of lower nucleosome occupancy in GBM around ALUs.

Recently, Chen et al. (2016) showed concentration of ALU repeats in cerebrospinal fluid can be used as a biomarker for glioma diagnosis based on levels of methylation. Unexpectedly, nucleosome occupancy patterns around ALU repeats that we have reconstructed based on



cfDNA are very different from those based on MNase-seq, losing that “two-peak” structure that we saw before. This can indicate that while ALUs contain two well-phased nucleosomes in brain tissues, these may be lost in cfDNA.

Since nucleosome positioning of ALU elements resembles the inactivation of promoters, abnormal positioning due to cancer can highly affect gene expression (Tanaka et al., 2010). For example, the effect of ALU on TFs has been studied by Rozenberg et al. (2018). In their research, they identified motifs related to RBPJ binding were sequences within ALU repeats, giving RBPJ a chance to interact with more than just regions containing canonical RBPJ binding sites. Our results show differential positioning of nucleosomes in GBM and while there is lack of literature regarding the effect of nucleosomes in ALU elements in cancer these results can be important for future research.

Unexpectedly, we found a striking resemblance of nucleosome profiles around ALU repeats and TSS. The ALU nucleosome patterns seem to mirror that of TSS. Promoters of active genes are known to contain a nucleosome-free region (NFR) upstream of TSS, followed by the +1 nucleosome, which displays histone variants and histone tail modification important for nucleosome eviction (Jiang and Pugh, 2009). This positioning is clear in our nucleosome occupancy patterns around TSS, with the NFR at the centre and the appearance of +1 nucleosome downstream the TSS. The presence of the two nucleosomes in both cases can be the cause of the similarity observed in the ALU nucleosome profiles.

One can speculate that the close relationship ALUs exhibit with TSS could be related to their common structure. Alu elements located immediately upstream of TSS are most likely to be functional (Su et al., 2014). Moreover, Bouttier et al. (2016) found four motifs that were present in AluJ and AluS element that were similarly distributed in TSS. Furthermore, CpG

methylation of ALUs and their interaction with TFs have been reported to act similarly to promoters (Chen and Yang, 2017).

### **4.3. Nucleosome occupancy changes in relation to histone modifications.**

Histone modifications such as methylation and acetylation play a vital role in GBM. Histone acetylation of H2A, H2B, H3 and H4 are highly involved in gene transcription, and methylation can activate (H3K4, H3K36 and H3K79) or repress (H3K9, H3K27 and H4K20) gene transcription (Dong and Cui, 2019). In a previous analysis, Dong and Cui (2019) showed the interplay of GBM-related TFs and histone modifications through the enzyme LSD1, which decreases trimethylation of H3K4 and elevates MYC expression. In turn, elevated MYC levels regulate OLIG2, SOX2 and POU3F2 levels, therefore showing elevated expression of GBM stemness.

In our analysis, we looked at nucleosome positioning differences around regions enriched with different histone modifications (Figures 5 and 6). It is clear that nucleosome occupancy around H3K27ac is lower in GBM, which indicates the activity of H3K27ac. These results are in line with other literature data. For example, Mack et al. (2019) investigated H3K27ac levels around TFs binding sites. In SOX2 expression, they found presence of H3K27ac at promoters and enhancers. Moreover, they revealed that high activity of enhancer H3K27ac could be seen at GSCs and GBM samples, that made them differentiate from other brain tumours and normal brain samples. In another publication, Rheinbay et al. (2013) showed that expression of H3K4me1 can be connected with ASCL1 binding. In GSCs, ASCL1 was binding to enhancers marked by H3K4me1, expressing several genes involved in Wnt regulation. Our results show changes in nucleosome occupancy levels in GBM but no specific nucleosome changes for H3K4me1.

In another study Lin et al. (2015), showed that H3K27me3 gene activity was twice larger than H3K4me3 in GSCs when compared to astrocytes, a result that is expected since stem cells are thought to have more actively expressed genes. We have also shown higher levels of H3K27me3 in GBM. While there were no H3K4me3-specific changes, lower nucleosome occupancy can be seen around the binding sites, while peaks are clearly less defined in GBM.

We have investigated nucleosome occupancy patterns around clusters of histone modifications in different GBM subtypes. For MES, H3K27ac and H3K27me3 show the most interesting results, showing decrease and increase of nucleosome occupancy in GBM vs normal brain tissues, accordingly. Other GBM subtypes did not show as interesting results because enhancers in MES subtypes drive gene expression in a more aggressive manner than other subtypes in GBM (Hall et al., 2018).

Next, we examined several histone modification changes that are directly related to TF activity. After looking at nucleosome occupancy around CTCF we concluded that this TF shows lower activity around histone modification on average. We concluded above that CTCF shows depletion in expression and other studies have shown that CTCF depletion shows gained in repressive marks. This is happening due to the absence of CTCF insulation allowing repressive histone to spread upon active chromatin domain (Weth et al., 2014). Moreover, Weth et al. (2014) explained that loss of CTCF dramatically increased H3K27me3. We have reported here higher increase of nucleosomes related to H3K27me3 at the centre of CTCF, which highlight the absence of CTCF at regions enriched with H3K27me3 leading to gene expression inhibition. Additionally, our results suggest that CTCF is present in all other histone modification we studied, but nucleosome occupancy patterns are different between patients in GBM. In their research Ren et al (2017), observed a positive correlation between CTCF binding and activity of H3K27ac. Their results indicated that CTCF significantly influences H3K27ac activity and can directly impact its activity across gene regions leading to aberrant expression.

Our H3K27ac patterns show presence of CTCF at H3K27ac binding sites which correlates with the above research. Moreover, the H3K4me1 is found within the CTCF loops and knockdown of CTCF shows decreased levels of the histone mark (Oti et al., 2016). Furthermore, nucleosomes associated with CTCF binding have been reported to be marked by H3K4me3 (Herold et al., 2012). Therefore, our results are consistent with the literature. These results could reveal GBM's epigenetic and transcriptional profile as well as details chromatin state in GBM, which can provide valuable information about future therapeutic targets.

#### **4.4. Nucleotide patterns along nucleosomes in MNase-seq and cfDNA**

After calculating the frequency of A/T/C/G nucleotides along the nucleosome dyad in GBM, we found that G and C nucleotide frequency was higher in GBM and A and T nucleotide lower (Figure S56, Appendix). Rubin and Green (2009) have observed high transition of A to G and T to C linked to gene expression. Moreover, recurrent mutations of T to C nucleotide at active promoters are observed at specific nucleotide context "CTTCCG" inducing the ETS-family TF expression and extending mutation production in melanomas (Gonzalez-Perez et al., 2019). Our results show higher frequency of C and G and lower frequency of A and T nucleotides and they correlate with previous literature in other cancers. Yazdi et al. (2015) confirmed increases in nucleosome occupancies were associated with changes consistent with those favourable for nucleotides selected within nucleosomes. They demonstrated that nucleotide frequencies within the nucleosome dyad in cancer can highly influence the mutation rate as a function of nucleosome occupancy. Therefore, our analysis of changes of nucleotide frequencies within the nucleosome core can highlight the positions in which those mutations are triggered and the reason for which they occur.

#### **4.5. GBM-specific nucleosome gain and loss.**

To complement the hypothesis-driven analyses described above, we have also compiled a systematic dataset composed of genomic regions that lost or gained nucleosomes, as well as the regions where nucleosome occupancy became less or more fuzzy in GBM vs paired non-tumour brain tissues. This dataset of GBM-sensitive regions can be used for an unbiased diagnostics based on nucleosome positioning reconstructed from MNase-seq or cfDNA. It's important to note that we detected significant heterogeneity between GBM patients. Occupancy of nucleosomes at the centre of lost nucleosome regions showed that patient G237 had significantly higher loss of nucleosomes from such sites. This nucleosome loss can lead to transcriptional up-regulation and genomic instability (Hu et al., 2014). If we take into account the abnormalities the patient G237 has shown in nucleosome occupancy compared to other patients, we can easily connect this to the higher loss of nucleosomes genome wide. Similarly, gained nucleosome regions for patient G237 were significantly more nucleosome free in healthy tissue and showed very different results in GBM, while other patients did not show such sharp difference between healthy and GBM. Nucleosome gain around regions relevant to access of DNA repair machinery and DNA transcription can prevent DNA repair and increase aberrant transcription leading to increase mutation rate (Yazdi et al., 2015). Therefore, these results could indicate that patient G237 has very different behaviours due to the increased loss and gain of nucleosome at specific regions compared to other patients. Further research on nucleosome occupancy around lost and gained nucleosome regions at specific locations can provide us with important information of how nucleosome occupation can affect overall progression of GBM.

Enrichment of lost and gained regions at TFs was discussed in section 4.1 "Interplay of TFs and nucleosome positioning in GBM". In addition, we calculated enrichment of lost/gained nucleosomes at other genomic features such as promoters, enhancers and CpG islands (Figure

9B). Enrichment of regions that lost nucleosomes at CpG islands is quite important since the change in methylation level of CpG islands in gene promoters usually is associated with repression of transcription (Fenouil et al., 2012).

Our Gene Ontology (GO) analysis performed for promoters with gained nucleosome regions revealed enrichment of biological processes related to immune responses (Figure 10). In addition, these promoters were associated with GO term “blood coagulation” (GO:0007596), that is considered to be an important aspect of brain tumour evolution. GBM is associated with hypercoagulability, where it releases substances that induce abnormal activation of blood coagulation and forms thrombi inside blood vessels and reduces patient survival (Navone et al., 2019). In addition, significant enrichment of “cell-cell signalling” (GO:0007267) was observed. Signalling between tumour cells are important for tumour progression and communication of GBM cells and neurons is relevant the development of cell proliferation (Portela et al., 2020). Portela et al. (2020) reported that in GBM, tumour epithelial cell and mesenchymal cell signalling it important for the activation of NOTCH signalling that is critical for tumour progression. We also found the enrichment of terms “natural killer cell mediated immunity” (GO:0002228), “T cell activation involved in immune response” (GO:0002286), “negative regulation of type I interferon-mediated signalling pathway” (GO:0060339), as well as other immune responses, highlighting possible targets for GBM immunotherapies. Glioblastoma expresses immunosuppressive cytokines that are secreted by tumour cells, microglia and tumour-associated macrophages. Activation of T cells and natural killer cells is suppressed by immunosuppressive pathways such as STAT3. Moreover, glioblastoma cells can escape immune recognition by downregulation expression of the major histocompatibility complex molecules (Brown et al., 2018).

Malfunctions of the type I interferon-mediated signalling pathway in GBM can be responsible for lack of immune response and resistance to cancer treatments. Mediation of the non-

canonical type I interferon-mediated signalling pathway by STAT proteins is involved, and its negative regulation by STAT activation is highly correlated with cancer survival (Budhwani et al., 2018). In their research, Piperi et al. (2019) highlighted the role of STAT3 as an oncogenic TF and its role in anti-tumour immunity. After inhibition of STAT3, they revealed block of self-renewal and cell growth in GBM. Interestingly, enrichment of the term “peptidyl-serine phosphorylation of STAT protein” (GO:0033139) was significant in our GO analysis. Phosphorylation of STAT3 at serine 727 has been proven to contribute to tumorigenesis and is related to GBM resistance (Lin et al., 2014).

These results include different relevant immune responses connected to our patients. This allows us to conclude that involvement of TFs in immune responses is an important target for GBM immunotherapy. The contribution of STAT proteins and particularly STAT3 can be a good prognostic marker and target for future GBM therapies. Moreover, the profiling of nucleosome positioning around STAT3 could be an interesting approach for future research.

Next, GO for genes involved in promoters enriched with lost nucleosome regions showed high significance of “viral transcription” (GO:0019083), “intracellular signal transduction” (GO:0035556) and “regulation of protein ADP-ribosylation” (GO:0010835). Previous research has shown the involvement of viruses in the proliferation of GBM (Wang et al., 2017). In their study, Wang et al (2017) identified two viruses, SV40 and WMSV, that can be used as prognostic markers in GBM. Next, significant enrichment of intercellular signal transduction can be an important process involved in GBM progression. In their research, Mao et al. (2012) highlighted importance of the phosphatidylinositol 3-kinase (PI3K) complex, an intercellular member of the PI3K-PTEN-Akt-mTOR pathway that regulates normal cellular functions. They showed that the PI3K pathway was altered in 70% of GBMs due to overexpression of EGFR receptor, the receptor that is responsible for the regulation of the complex. Their data suggested the targeting of EGFR could be a viable target for future GBM therapies. Interestingly, “stress

activated MARP cascade” (GO:0051403), another GO term that showed enrichment in our data, has been proven to act as a repression of the EGFR pathway, by altering the GSC cell cycle state (Soeda et al., 2017). Additionally, many other GO terms related to cell cycle progression were “rRNA processing” (GO:0006364), “RNA splicing” (GO:0008380), “G2/M transition of mitotic cell cycle” (GO:0000086) and “cell cycle” (GO:0007049). Cell cycle progression in GSCs has shown abnormalities and checkpoint defects and inhibition of G2/M regulators has shown positive results in GBM therapies (Tachon et al., 2018; Castro-Gamero et al., 2018). Regarding RNA GO terms, rRNA contains modifications, such as RNA m<sup>6</sup>A, that was enriched in single nucleosome polymorphisms (SNPs) that affect RNA stability, processing and splicing (Dong and Cui, 2020). In recent research, Dong and Cui (2020) mentioned that m<sup>6</sup>A plays an essential role in cell fate determination, stem cell maintenance and neuronal functions, and proved that such modifications increased GBM malignancy. Overall, promoters with lost nucleosome regions in GBM promoted anti-tumour processes as well as pathways that can be related to GBM progression. In both GO analyses, our results showed that there is involvement of lost and gained nucleosome regions in GBM at promoters at the above biological processes and identified new targets for future therapies.

#### **4.6. Diagnosis of patients based on PCA of nucleosome occupancy**

Last but not least, in order to provide evidence for the use of nucleosome occupancy as a marker in GBM, we performed principal component analysis (PCA) using nucleosome occupancy values in the dataset of GBM-sensitive regions that lost or gained nucleosomes in GBM (Figure 11). For both cases, our data was divided into two main clusters, healthy and GBM. This analysis confirmed that based on nucleosome occupancy data alone, the datasets could form clusters and differentiate between healthy and GBM nucleosome occupancy. To conclude, our results show that it is feasible to demonstrate the differences between healthy and GBM tissue and reveal the functional significance of nucleosome occupancy in GBM and in cancer overall.



Finally, it is important to discuss the limitations of this study. Due to the small samples size of patient groups, immediate conclusions on our data cannot be taken to the clinic. This pilot study showed nucleosome occupancy as a new potential marker in cancer diagnostics and in monitoring GBM progression. However, this analysis needs to be extended in the future in a larger cohort to improve overall statistics and further highlight the importance of the results. Moreover, in this study, healthy control samples were considered to be the brain tissues from patients from the tumour periphery under assumption that that these were not affected by the tumour, which may not always be the case. Additionally, GBM tumour sub-types were not identified in the patients' samples for both tissue and cfDNA. Finally, there is no record of patient treatment prior to collection of samples taken from cfDNA patients, which may have affected the outcome of the analysis. In the future, it would be also interesting to look at DNA methylation datasets in GBM in order to make better assertions about the overall nucleosome positioning in GBM. Furthermore, future work can include extended analysis of cfDNA from a larger patient cohort.

In conclusion, our investigations on the effect of nucleosome positioning in GBM highlighted many pathways that could be used as GBM-related therapy targets. Overall, we have proven for the first time that nucleosome occupancy patterns can be distinguished between tumour brain tissues and paired non-malignant brain tissues from the same patients, and that these patterns can be used as a diagnostic GBM marker. Moreover, the analysis of cfDNA-based reconstruction of nucleosome occupancy shows promising results for the use of cfDNA in GBM patient diagnosis in liquid biopsies.

## References

An, Q., Hu, Y., Li, Q., Chen, X., Huang, J., Pellegrini, M., Zhou, X. J., Rettig, M. and Fan, G. (2019) 'The size of cell-free mitochondrial DNA in blood is inversely correlated with tumor burden in cancer patients', *Precis Clin Med*, 2(3), pp. 131-139.

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. and Komorowski, J. (2009) 'Nucleosomes are well positioned in exons and carry characteristic histone modifications', *Genome Res*, 19(10), pp. 1732-41.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S. and Soboleva, A. (2013) 'NCBI GEO: archive for functional genomics data sets--update', *Nucleic Acids Res*, 41(Database issue), pp. D991-5.

Berger, S. L. (2007) 'The complex language of chromatin regulation during transcription', *Nature*, 447(7143), pp. 407-12.

Bettecken, T., Frenkel, Z. M. and Trifonov, E. N. (2011) 'Human nucleosomes: special role of CG dinucleotides and Alu-nucleosomes', *BMC Genomics*, 12(1), pp. 273.

Bettegowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B. R., Wang, H., Luber, B., Alani, R. M., Antonarakis, E. S., Azad, N. S., Bardelli, A., Brem, H., Cameron, J. L., Lee, C. C., Fecher, L. A., Gallia, G. L., Gibbs, P., Le, D., Giuntoli, R. L., Goggins, M., Hogarty, M. D., Holdhoff, M., Hong, S. M., Jiao, Y., Juhl, H. H., Kim, J. J., Siravegna, G., Laheru, D. A., Lauricella, C., Lim, M., Lipson, E. J., Marie, S. K., Netto, G. J., Oliner, K. S., Olivi, A., Olsson, L., Riggins, G. J., Sartore-Bianchi, A., Schmidt, K., Shih, I., Oba-Shinjo, S. M., Siena, S., Theodorescu, D., Tie, J., Harkins, T. T., Veronese, S., Wang, T. L., Weingart, J. D., Wolfgang, C. L., Wood, L. D., Xing, D., Hruban, R. H., Wu, J., Allen, P. J., Schmidt, C. M., Choti, M. A., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Papadopoulos, N. and Diaz, L. A. (2014) 'Detection of circulating tumor DNA in early- and late-stage human malignancies', *Sci Transl Med*, 6(224), pp. 224ra24.

Bogea, A., Morvan-Dubois, G., El-Habr, E. A., Lejeune, F. X., Defrance, M., Narayanan, A., Kuranda, K., Burel-Vandenbos, F., Sayd, S., Delaunay, V., Dubois, L. G., Parrinello, H., Rialle, S., Fabrega, S., Idbaih, A., Haiech, J., Bièche, I., Virolle, T., Goodhardt, M., Chneiweiss, H. and Junier, M. P. (2018) 'Changes in chromatin state reveal ARNT2 at a node of a tumorigenic transcription factor signature driving glioblastoma cell aggressiveness', *Acta Neuropathol*, 135(2), pp. 267-283.

Bouttier, M., Laperriere, D., Memari, B., Mangiapane, J., Fiore, A., Mitchell, E., Verway, M., Behr, M. A., Sladek, R., Barreiro, L. B., Mader, S. and White, J. H. (2016) 'Alu repeats as transcriptional regulatory platforms in macrophage responses to M. tuberculosis infection', *Nucleic Acids Res*, 44(22), pp. 10571-10587.

Brodbeck, A., Greenberg, D., Winters, T., Williams, M., Vernon, S., Collins, V. P. & Group, U. N. C. I. N. B. T. 2015. Glioblastoma in England: 2007-2011. *Eur J Cancer*, 51(4), pp 533-542.

Brown, N. F., Carter, T. J., Ottaviani, D. and Mulholland, P. (2018) 'Harnessing the immune system in glioblastoma', *Br J Cancer*, 119(10), pp. 1171-1181.

Budhwani, M., Mazzieri, R. and Dolcetti, R. (2018) 'Plasticity of Type I Interferon-Mediated Responses in Cancer Therapy: From Anti-tumor Immunity to Resistance', *Front Oncol*, 8, pp. 322.

Bulstrode, H., Johnstone, E., Marques-Torrejon, M. A., Ferguson, K. M., Bressan, R. B., Blin, C., Grant, V., Gogolok, S., Gangoso, E., Gargica, S., Ender, C., Fotaki, V., Sproul, D., Bertone, P. and Pollard, S. M. (2017) 'Elevated FOXG1 and SOX2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators', *Genes Dev*, 31(8), pp. 757-773.

Castro-Gamero, A. M., Pezuk, J. A., Brassesco, M. S. and Tone, L. G. (2018) 'G2/M inhibitors as pharmacotherapeutic opportunities for glioblastoma: the old, the new, and the future', *Cancer Biol Med*, 15(4), pp. 354-374.

Cordaux, R., Sen, S. K., Konkol, M. K. and Batzer, M. A. (2010) 'Computational methods for the analysis of primate mobile elements', *Methods Mol Biol*, 628, pp. 137-51.

Chen, J., Huan, W., Zuo, H., Zhao, L., Huang, C., Liu, X., Hou, S., Qi, J. and Shi, W. (2016) 'Alu methylation serves as a biomarker for non-invasive diagnosis of glioma', *Oncotarget*, 7(18), pp. 26099-106.

Chen, L. L. and Yang, L. (2017) 'ALU alternative Regulation for Gene Expression', *Trends Cell Biol*, 27(7), pp. 480-490.

Cheng, Z., Gong, Y., Ma, Y., Lu, K., Lu, X., Pierce, L. A., Thompson, R. C., Muller, S., Knapp, S. and Wang, J. (2013) 'Inhibition of BET bromodomain targets genetically diverse glioblastoma', *Clin Cancer Res*, 19(7), pp. 1748-59.

Chereji, R. V. and Clark, D. J. (2018) 'Major Determinants of Nucleosome Positioning', *Biophys J*, 114(10), pp. 2279-2289.

Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V., Chen, P. Y., Stroud, H., Yu, Y., Hetzel, J. A., Kuo, F., Kim, J., Cokus, S. J., Casero, D., Bernal, M., Huijser, P., Clark, A. T., Krämer, U., Merchant, S. S., Zhang, X., Jacobsen, S. E. and Pellegrini, M. (2010) 'Relationship between nucleosome positioning and DNA methylation', *Nature*, 466(7304), pp. 388-92.

Chudnovsky, Y., Kim, D., Zheng, S., Whyte, W. A., Bansal, M., Bray, M. A., Gopal, S., Theisen, M. A., Bilodeau, S., Thiru, P., Muffat, J., Yilmaz, O. H., Mitalipova, M., Woolard, K., Lee, J., Nishimura, R., Sakata, N., Fine, H. A., Carpenter, A. E., Silver, S. J., Verhaak, R. G., Califano, A., Young, R. A., Ligon, K. L., Mellinghoff, I. K., Root, D. E., Sabatini, D. M., Hahn, W. C. and Chheda, M. G. (2014) 'ZFX4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state', *Cell Rep*, 6(2), pp. 313-24.

Cui, F. and Zhurkin, V. B. (2010) 'Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro', *J Biomol Struct Dyn*, 27(6), pp. 821-41.

Damaschke, N. A., Gawdzik, J., Avilla, M., Yang, B., Svaren, J., Roopra, A., Luo, J. H., Yu, Y. P., Keles, S. and Jarrard, D. F. (2020) 'CTCF loss mediates unique DNA hypermethylation landscapes in human cancers', *Clin Epigenetics*, 12(1), pp. 80.

Diehl, F., Schmidt, K., Choti, M. A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S. A., Kinzler, K. W., Vogelstein, B. and Diaz, L. A. (2008) 'Circulating mutant DNA to assess tumor dynamics', *Nat Med*, 14(9), pp. 985-9

Di Ruocco, F., Basso, V., Rivoire, M., Mehlen, P., Ambati, J., De Falco, S. and Tarallo, V. (2018) 'Alu RNA accumulation induces epithelial-to-mesenchymal transition by modulating miR-566 and is associated with cancer progression', *Oncogene*, 37(5), pp. 627-637.

Dong, Z. and Cui, H. (2019) 'Epigenetic modulation of metabolism in glioblastoma', *Semin Cancer Biol*, 57, pp. 45-51.

Dong, Z. and Cui, H. (2020) 'The Emerging Roles of RNA Modifications in Glioblastoma', *Cancers (Basel)*, 12(3).

Elazezy, M. and Joosse, S. A. (2018) 'Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management', *Comput Struct Biotechnol J*, 16, pp. 370-378.

Fang, C., Wang, Z., Han, C., Safgren, S. L., Helmin, K. A., Adelman, E. R., Serafin, V., Basso, G., Eagen, K. P., Gaspar-Maia, A., Figueroa, M. E., Singer, B. D., Ratan, A., Ntziachristos, P. and Zang, C. (2020) 'Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation', *Genome Biol*, 21(1), pp. 247.

Fang, X., Yoon, J. G., Li, L., Yu, W., Shao, J., Hua, D., Zheng, S., Hood, L., Goodlett, D. R., Foltz, G. and Lin, B. (2011) 'The SOX2 response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis', *BMC Genomics*, 12, pp. 11.

Fenley, A. T., Anandkrishnan, R., Kidane, Y. H. and Onufriev, A. V. (2018) 'Modulation of nucleosomal DNA accessibility via charge-altering post-translational modifications in histone core', *Epigenetics Chromatin*, 11(1), pp. 11.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J. Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I. and Andrau, J. C. (2012) 'CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters', *Genome Res*, 22(12), pp. 2399-408.

Fiscon, G., Conte, F., Licursi, V., Nasi, S. and Paci, P. (2018) 'Computational identification of specific genes for glioblastoma stem-like cells identity', *Sci Rep*, 8(1), pp. 7769.

Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., Suvà, M. L. and Bernstein, B. E. (2016) 'Insulator dysfunction and oncogene activation in IDH mutant gliomas', *Nature*, 529(7584), pp. 110-4.

García-Olmo, D. C., Picazo, M. G., Toboso, I., Asensio, A. I. and García-Olmo, D. (2013a) 'Quantitation of cell-free DNA and RNA in plasma during tumor progression in rats', *Mol Cancer*, 12, pp. 8.

Garnier, D., Renoult, O., Alves-Guerra, M. C., Paris, F. and Pecqueur, C. (2019) 'Glioblastoma Stem-', *Front Oncol*, 9, pp. 118.

Gertz, J., Savic, D., Varley, K. E., Partridge, E. C., Safi, A., Jain, P., Cooper, G. M., Reddy, T. E., Crawford, G. E. and Myers, R. M. (2013) 'Distinct properties of cell-type-specific and shared transcription factor binding sites', *Mol Cell*, 52(1), pp. 25-36.

Gonzalez-Perez, A., Sabarinathan, R. and Lopez-Bigas, N. (2019) 'Local Determinants of the Mutational Landscape of the Human Genome', *Cell*, 177(1), pp. 101-114.

Griffon, A., Barbier, Q., Dalino, J., van Helden, J., Spicuglia, S. and Ballester, B. (2015) 'Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape', *Nucleic Acids Res*, 43(4), pp. e27.

Gui, Y., Guo, G., Huang, Y., Hu, X., Tang, A., Gao, S., Wu, R., Chen, C., Li, X., Zhou, L., He, M., Li, Z., Sun, X., Jia, W., Chen, J., Yang, S., Zhou, F., Zhao, X., Wan, S., Ye, R., Liang, C., Liu, Z., Huang, P., Liu, C., Jiang, H., Wang, Y., Zheng, H., Sun, L., Liu, X., Jiang, Z., Feng, D., Wu, S., Zou, J., Zhang, Z., Yang, R., Zhao, J., Xu, C., Yin, W., Guan, Z., Ye, J., Zhang, H., Li, J., Kristiansen, K., Nickerson, M. L., Theodorescu, D., Li, Y., Zhang, X., Li, S., Wang, J., Yang, H. and Cai, Z. (2011) 'Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder', *Nat Genet*, 43(9), pp. 875-8.

Gutiérrez, G., Millán-Zambrano, G., Medina, D. A., Jordán-Pla, A., Pérez-Ortín, J. E., Peñate, X. and Chávez, S. (2017) 'Subtracting the sequence bias from partially digested MNase-seq data reveals a general contribution of TFIS to nucleosome positioning', *Epigenetics Chromatin*, 10(1), pp. 58.

Hall, A. W., Battenhouse, A. M., Shivram, H., Morris, A. R., Cowperthwaite, M. C., Shpak, M. and Iyer, V. R. (2018) 'Bivalent Chromatin Domains in Glioblastoma Reveal a Subtype-Specific Signature of Glioma Stem Cells', *Cancer Res*, 78(10), pp. 2463-2474.

Hashimoto, H., Wang, D., Horton, J. R., Zhang, X., Corces, V. G. and Cheng, X. (2017) 'Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA', *Mol Cell*, 66(5), pp. 711-720.e3.

Herold, M., Bartkuhn, M. and Renkawitz, R. (2012) 'CTCF: insights into insulator function during development', *Development*, 139(6), pp. 1045-57.

Hoeijmakers, W. A. M. and Bártfai, R. (2018) 'Characterization of the Nucleosome Landscape by Micrococcal Nuclease-Sequencing (MNase-seq)', *Methods Mol Biol*, 1689, pp. 83-101.

Hu, Z., Chen, K., Xia, Z., Chavez, M., Pal, S., Seol, J. H., Chen, C. C., Li, W. and Tyler, J. K. (2014) 'Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging', *Genes Dev*, 28(4), pp. 396-408.

Jiang, C. and Pugh, B. F. (2009) 'Nucleosome positioning and gene regulation: advances through genomics', *Nat Rev Genet*, 10(3), pp. 161-72.

Jordà, M., Díez-Villanueva, A., Mallona, I., Martín, B., Lois, S., Barrera, V., Esteller, M., Vavouri, T. and Peinado, M. A. (2017) 'The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells', *Genome Res*, 27(1), pp. 118-132.

Joseph, S. R., Pálffy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A. and Vastenhouw, N. L. (2017a) 'Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos', *Elife*, 6.

Joseph, S. R., Pálffy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A. and Vastenhouw, N. L. (2017b) 'Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos', *Elife*, 6.

Kharerin, H., Bhat, P. J., Marko, J. F. and Padinhateeri, R. (2016) 'Role of transcription factor-mediated nucleosome disassembly in PHO5 gene expression', *Sci Rep*, 6, pp. 20319.

Kornberg, R. D. and Lorch, Y. (1999) 'Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome', *Cell*, 98(3), pp. 285-94.

Kulic, I., Robertson, G., Chang, L., Baker, J. H., Lockwood, W. W., Mok, W., Fuller, M., Fournier, M., Wong, N., Chou, V., Robinson, M. D., Chun, H. J., Gilks, B., Kempkes, B., Thomson, T. A., Hirst, M., Minchinton, A. I., Lam, W. L., Jones, S., Marra, M. and Karsan, A. (2015) 'Loss of the Notch effector RBPJ promotes tumorigenesis', *J Exp Med*, 212(1), pp. 37-52.

Kuroi, K., Tanaka, C. and Toi, M. (1999) 'Plasma Nucleosome Levels in Node-Negative Breast Cancer Patients', *Breast Cancer*, 6(4), pp. 361-364.

Lam, F. H., Steger, D. J. and O'Shea, E. K. (2008a) 'Chromatin decouples promoter threshold from dynamic range', *Nature*, 453(7192), pp. 246-50.

Lam, F. H., Steger, D. J. and O'Shea, E. K. (2008b) 'Chromatin decouples promoter threshold from dynamic range', *Nature*, 453(7192), pp. 246-50.

Lapin, M., Oltedal, S., Tjensvoll, K., Buhl, T., Smaaland, R., Garresori, H., Javle, M., Glenjen, N. I., Abelseth, B. K., Gilje, B. and Nordgård, O. (2018) 'Fragment size and level of cell-free DNA provide prognostic information in patients with advanced pancreatic cancer', *J Transl Med*, 16(1), pp. 300.

Lathia, J. D., Mack, S. C., Mulkearns-Hubert, E. E., Valentim, C. L. and Rich, J. N. (2015) 'Cancer stem cells in glioblastoma', *Genes Dev*, 29(12), pp. 1203-17.

Lay, F. D., Liu, Y., Kelly, T. K., Witt, H., Farnham, P. J., Jones, P. A. and Berman, B. P. (2015) 'The role of DNA methylation in directing the functional organization of the cancer epigenome', *Genome Res*, 25(4), pp. 467-77.

Li, C. & Luscombe, N. M. 2020. Nucleosome positioning stability is a modulator of germline mutation rate variation across the human genome. *Nat Commun*, 11(1), pp 1363.

Li, H. T., Duymich, C. E., Weisenberger, D. J. and Liang, G. (2016) 'Genetic and Epigenetic Alterations in Bladder Cancer', *Int Neurourol J*, 20(Suppl 2), pp. S84-94.

Li, Y., Sun, Q., Jiang, M., Li, S., Zhang, J., Xu, Z., Guo, D., Gu, T., Wang, B., Xiao, L., Zhou, T. and Zhuo, W. (2019) 'KLF9 suppresses gastric cancer cell invasion and metastasis through transcriptional inhibition of MMP28', *FASEB J*, 33(7), pp. 7915-7928.

Lin, B., Lee, H., Yoon, J. G., Madan, A., Wayner, E., Tønning, S., Hothi, P., Schroeder, B., Ulasov, I., Foltz, G., Hood, L. and Cobbs, C. (2015) 'Global analysis of H3K4me3 and H3K27me3 profiles in glioblastoma stem cells and identification of SLC17A7 as a bivalent tumor suppressor gene', *Oncotarget*, 6(7), pp. 5369-81.

Lin, C. Y., Lovén, J., Rahl, P. B., Paranal, R. M., Burge, C. B., Bradner, J. E., Lee, T. I. and Young, R. A. (2012) 'Transcriptional amplification in tumor cells with elevated c-Myc', *Cell*, 151(1), pp. 56-67.

Lin, G. S., Chen, Y. P., Lin, Z. X., Wang, X. F., Zheng, Z. Q. and Chen, L. (2014) 'STAT3 serine 727 phosphorylation influences clinical outcome in glioblastoma', *Int J Clin Exp Pathol*, 7(6), pp. 3141-9.

Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., Bradner, J. E., Lee, T. I. and Young, R. A. (2013) 'Selective inhibition of tumor oncogenes by disruption of super-enhancers', *Cell*, 153(2), pp. 320-34.

Lövkvist, C., Sneppen, K. and Haerter, J. O. (2017) 'Exploring the Link between Nucleosome Occupancy and DNA Methylation', *Front Genet*, 8, pp. 232.

Mack, S. C., Singh, I., Wang, X., Hirsch, R., Wu, Q., Villagomez, R., Bernatchez, J. A., Zhu, Z., Gimple, R. C., Kim, L. J. Y., Morton, A., Lai, S., Qiu, Z., Prager, B. C., Bertrand, K. C., Mah, C., Zhou, W., Lee, C., Barnett, G. H., Vogelbaum, M. A., Sloan, A. E., Chavez, L., Bao, S., Scacheri, P. C., Siqueira-Neto, J. L., Lin, C. Y. and Rich, J. N. (2019) 'Chromatin landscapes reveal developmentally encoded transcriptional states that define human glioblastoma', *J Exp Med*, 216(5), pp. 1071-1090.

Malley, D. S., Hamoudi, R. A., Kocialkowski, S., Pearson, D. M., Collins, V. P. and Ichimura, K. (2011) 'A distinct region of the MGMT CpG island critical for transcriptional regulation is preferentially methylated in glioblastoma cells and xenografts', *Acta Neuropathol*, 121(5), pp. 651-61.

Mao, H., Lebrun, D. G., Yang, J., Zhu, V. F. and Li, M. (2012) 'Deregulated signaling pathways in glioblastoma multiforme: molecular mechanisms and therapeutic targets', *Cancer Invest*, 30(1), pp. 48-56.

McAnena, P., Brown, J. A. and Kerin, M. J. (2017) 'Circulating Nucleosomes and Nucleosome Modifications as Biomarkers in Cancer', *Cancers (Basel)*, 9(1).

McEwen, A. E., Leary, S. E. S. and Lockwood, C. M. (2020) 'Beyond the Blood: CSF-Derived cfDNA for Diagnosis and Characterization of CNS Tumors', *Front Cell Dev Biol*, 8, pp. 45.

Miller, A. M., Shah, R. H., Pentsova, E. I., Pourmaleki, M., Briggs, S., Distefano, N., Zheng, Y., Skakodub, A., Mehta, S. A., Campos, C., Hsieh, W. Y., Selcuklu, S. D., Ling, L., Meng, F., Jing, X., Samoila, A., Bale, T. A., Tsui, D. W. Y., Grommes, C., Viale, A., Souweidane, M. M., Tabar, V., Brennan, C. W., Reiner, A. S., Rosenblum, M., Panageas, K. S., DeAngelis, L. M., Young, R. J., Berger, M. F. and Mellingshoff, I. K. (2019) 'Tracking tumour evolution in glioma through liquid biopsies of cerebrospinal fluid', *Nature*, 565(7741), pp. 654-658.

Miller, T. E., Liao, B. B., Wallace, L. C., Morton, A. R., Xie, Q., Dixit, D., Factor, D. C., Kim, L. J. Y., Morrow, J. J., Wu, Q., Mack, S. C., Hubert, C. G., Gillespie, S. M., Flavahan, W. A., Hoffmann, T., Thummalaipalli, R., Hemann, M. T., Paddison, P. J., Horbinski, C. M., Zuber, J., Scacheri, P. C., Bernstein, B. E., Tesar, P. J. and Rich, J. N. (2017) 'Transcription elongation factors represent in vivo cancer dependencies in glioblastoma', *Nature*, 547(7663), pp. 355-359.

Moshkin, Y. M., Mohrmann, L., van Ijcken, W. F. and Verrijzer, C. P. (2007a) 'Functional differentiation of SWI/SNF remodelers in transcription and cell cycle control', *Mol Cell Biol*, 27(2), pp. 651-61.

Moshkin, Y. M., Mohrmann, L., van Ijcken, W. F. and Verrijzer, C. P. (2007b) 'Functional differentiation of SWI/SNF remodelers in transcription and cell cycle control', *Mol Cell Biol*, 27(2), pp. 651-61.

Murtaza, M. and Caldas, C. (2016) 'Nucleosome mapping in plasma DNA predicts cancer gene expression', *Nat Genet*, 48(10), pp. 1105-6.

Müller Bark, J., Kulasinghe, A., Chua, B., Day, B. W. and Punyadeera, C. (2020) 'Circulating biomarkers in patients with glioblastoma', *Br J Cancer*, 122(3), pp. 295-305.

Nakamura, M., Watanabe, T., Yonekawa, Y., Kleihues, P. and Ohgaki, H. (2001) 'Promoter methylation of the DNA repair gene MGMT in astrocytomas is frequently associated with G:C --> A:T mutations of the TP53 tumor suppressor gene', *Carcinogenesis*, 22(10), pp. 1715-9.

Nakamura, T., Sueoka-Aragane, N., Iwanaga, K., Sato, A., Komiya, K., Kobayashi, N., Hayashi, S., Hosomi, T., Hirai, M., Sueoka, E. and Kimura, S. (2012) 'Application of a highly sensitive detection system for epidermal growth factor receptor mutations in plasma DNA', *J Thorac Oncol*, 7(9), pp. 1369-81.

Navone, S. E., Guarnaccia, L., Locatelli, M., Rampini, P., Caroli, M., La Verde, N., Gaudino, C., Bettinardi, N., Riboni, L., Marfia, G. and Campanella, R. (2019) 'Significance and Prognostic Value of The Coagulation Profile in Patients with Glioblastoma: Implications for Personalized Therapy', *World Neurosurg*, 121, pp. e621-e629.

Oh, S., Oh, C. and Yoo, K. H. (2017) 'Functional roles of CTCF in breast cancer', *BMB Rep*, 50(9), pp. 445-453.

Ohgaki, H., Dessen, P., Jourde, B., Horstmann, S., Nishikawa, T., Di Patre, P. L., Burkhard, C., Schüler, D., Probst-Hensch, N. M., Maiorka, P. C., Baeza, N., Pisani, P., Yonekawa, Y., Yasargil, M. G., Lütolf, U. M. and Kleihues, P. (2004) 'Genetic pathways to glioblastoma: a population-based study', *Cancer Res*, 64(19), pp. 6892-9.



- Ohgaki, H. and Kleihues, P. (2009a) 'Genetic alterations and signaling pathways in the evolution of gliomas', *Cancer Sci*, 100(12), pp. 2235-41.
- Ohgaki, H. and Kleihues, P. (2009b) 'Genetic alterations and signaling pathways in the evolution of gliomas', *Cancer Sci*, 100(12), pp. 2235-41.
- Ohgaki, H. and Kleihues, P. (2013) 'The definition of primary and secondary glioblastoma', *Clin Cancer Res*, 19(4), pp. 764-72.
- Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R., Sese, J. and Meno, C. (2018) 'ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data', *EMBO Rep*, 19(12).
- Onufriev, A. V. and Schiessel, H. (2019) 'The nucleosome: from structure to function through physics', *Curr Opin Struct Biol*, 56, pp. 119-130.
- Oti, M., Falck, J., Huynen, M. A. and Zhou, H. (2016) 'CTCF-mediated chromatin loops enclose inducible gene regulatory domains', *BMC Genomics*, 17, pp. 252.
- Padinhateeri, R. and Marko, J. F. (2011) 'Nucleosome positioning in a model of active chromatin remodeling enzymes', *Proc Natl Acad Sci U S A*, 108(19), pp. 7799-803.
- Park, N. I., Guilhamon, P., Desai, K., McAdam, R. F., Langille, E., O'Connor, M., Lan, X., Whetstone, H., Coutinho, F. J., Vanner, R. J., Ling, E., Prinos, P., Lee, L., Selvadurai, H., Atwal, G., Kushida, M., Clarke, I. D., Voisin, V., Cusimano, M. D., Bernstein, M., Das, S., Bader, G., Arrowsmith, C. H., Angers, S., Huang, X., Lupien, M. and Dirks, P. B. (2017) 'ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells', *Cell Stem Cell*, 21(3), pp. 411.
- Park, P. J. (2009) 'ChIP-seq: advantages and challenges of a maturing technology', *Nat Rev Genet*, 10(10), pp. 669-80.
- Parsons, D. W., Jones, S., Zhang, X., Lin, J. C., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Siu, I. M., Gallia, G. L., Olivi, A., McLendon, R., Rasheed, B. A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D. A., Tekleab, H., Diaz, L. A., Hartigan, J., Smith, D. R., Strausberg, R. L., Marie, S. K., Shinjo, S. M., Yan, H., Riggins, G. J., Bigner, D. D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V. E. and Kinzler, K. W. (2008) 'An integrated genomic analysis of human glioblastoma multiforme', *Science*, 321(5897), pp. 1807-12.
- Pastori, C., Daniel, M., Penas, C., Volmar, C. H., Johnstone, A. L., Brothers, S. P., Graham, R. M., Allen, B., Sarkaria, J. N., Komotar, R. J., Wahlestedt, C. and Ayad, N. G. (2014) 'BET bromodomain proteins are required for glioblastoma cell proliferation', *Epigenetics*, 9(4), pp. 611-20.
- Pearson, J. R. D. and Regad, T. (2017) 'Targeting cellular pathways in glioblastoma multiforme', *Signal Transduct Target Ther*, 2, pp. 17040.

- Philips, A., Henshaw, D. L., Lamburn, G. & O'Carroll, M. J. 2018. Brain Tumours: Rise in Glioblastoma Multiforme Incidence in England 1995-2015 Suggests an Adverse Environmental or Lifestyle Factor. *J Environ Public Health*, 2018(7910754).
- Piatti, P., Zeilner, A. and Lusser, A. (2011) 'ATP-dependent chromatin remodeling factors and their roles in affecting nucleosome fiber composition', *Int J Mol Sci*, 12(10), pp. 6544-65.
- Piccioni, D. E., Achrol, A. S., Kiedrowski, L. A., Banks, K. C., Boucher, N., Barkhoudarian, G., Kelly, D. F., Juarez, T., Lanman, R. B., Raymond, V. M., Nguyen, M., Truong, J. D., Heng, A., Gill, J., Saria, M., Pingle, S. C. and Kesari, S. (2019) 'Analysis of cell-free circulating tumor DNA in 419 patients with glioblastoma and other primary brain tumors', *CNS Oncol*, 8(2), pp. CNS34.
- Piperi, C., Papavassiliou, K. A. and Papavassiliou, A. G. (2019) 'Pivotal Role of STAT3 in Shaping Glioblastoma Immune Microenvironment', *Cells*, 8(11).
- Portela, A. and Esteller, M. (2010) 'Epigenetic modifications and human disease', *Nat Biotechnol*, 28(10), pp. 1057-68.
- Portela, M., Mitchell, T. and Casas-Tintó, S. (2020) 'Cell-to-cell communication mediates glioblastoma progression in', *Biol Open*, 9(9).
- Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841-2.
- Ramos, Y. F., Hestand, M. S., Verlaan, M., Krabbendam, E., Ariyurek, Y., van Galen, M., van Dam, H., van Ommen, G. J., den Dunnen, J. T., Zantema, A. and 't Hoen, P. A. (2010) 'Genome-wide assessment of differential roles for p300 and CBP in transcription regulation', *Nucleic Acids Res*, 38(16), pp. 5396-408.
- Raucher, D. (2019) 'Tumor targeting peptides: novel therapeutic strategies in glioblastoma', *Curr Opin Pharmacol*, 47, pp. 14-19.
- Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A. and Segal, E. (2012) 'Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast', *Nat Genet*, 44(7), pp. 743-50.
- Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D. R. and Zhao, K. (2017) 'CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression', *Mol Cell*, 67(6), pp. 1049-1058.e6.
- Rheinbay, E., Suvà, M. L., Gillespie, S. M., Wakimoto, H., Patel, A. P., Shahid, M., Oksuz, O., Rabkin, S. D., Martuza, R. L., Rivera, M. N., Louis, D. N., Kasif, S., Chi, A. S. and Bernstein, B. E. (2013) 'An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma', *Cell Rep*, 3(5), pp. 1567-79.

Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E. and Längst, G. (2007) 'DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes', *Proc Natl Acad Sci U S A*, 104(40), pp. 15635-40.

Romani, M., Pistillo, M. P. and Banelli, B. (2018) 'Epigenetic Targeting of Glioblastoma', *Front Oncol*, 8, pp. 448.

Rozenberg, J. M., Taylor, J. M. and Mack, C. P. (2018) 'RBPJ binds to consensus and methylated cis elements within phased nucleosomes and controls gene expression in human aortic smooth muscle cells in cooperation with SRF', *Nucleic Acids Res*, 46(16), pp. 8232-8244.

Rubin, A. F. and Green, P. (2009) 'Mutation patterns in cancer genomes', *Proc Natl Acad Sci U S A*, 106(51), pp. 21766-70.

Rudnizky, S., Malik, O., Bavly, A., Pnueli, L., Melamed, P. and Kaplan, A. (2017) 'Nucleosome mobility and the regulation of gene expression: Insights from single-molecule studies', *Protein Sci*, 26(7), pp. 1266-1277.

Sasmita, A. O., Wong, Y. P. and Ling, A. P. K. (2018) 'Biomarkers and therapeutic advances in glioblastoma multiforme', *Asia Pac J Clin Oncol*, 14(1), pp. 40-51.

Schones, D. E., Cui, K., Cuddapah, S., Roh, T. Y., Barski, A., Wang, Z., Wei, G. and Zhao, K. (2008) 'Dynamic regulation of nucleosome positioning in the human genome', *Cell*, 132(5), pp. 887-98.

Shafiei N. (2018) Nucleosome repositioning in cancer, MSc Thesis, University of Essex, 2018.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., Wang, J. P. and Widom, J. (2006) 'A genomic code for nucleosome positioning', *Nature*, 442(7104), pp. 772-8.

Shanmugam, M. K., Arfuso, F., Arumugam, S., Chinnathambi, A., Jinsong, B., Warriar, S., Wang, L. Z., Kumar, A. P., Ahn, K. S., Sethi, G. and Lakshmanan, M. (2018) 'Role of novel histone modifications in cancer', *Oncotarget*, 9(13), pp. 11414-11426.

Shi, C., Ye, Z., Han, J., Ye, X., Lu, W., Ji, C., Li, Z., Ma, Z., Zhang, Q., Zhang, Y., He, W., Chen, Z., Cao, X., Shou, X., Zhou, X., Wang, Y., Zhang, Z., Li, Y., Ye, H., He, M., Chen, H., Cheng, H., Sun, J., Cai, J., Huang, C., Ye, F., Luo, C., Zhou, B., Ding, H. and Zhao, Y. (2020) 'BRD4 as a therapeutic target for nonfunctioning and growth hormone pituitary adenoma', *Neuro Oncol*, 22(8), pp. 1114-1125.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M. and Iyer, V. R. (2008) 'Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation', *PLoS Biol*, 6(3), pp. e65.

Singh, D. K., Kollipara, R. K., Vemireddy, V., Yang, X. L., Sun, Y., Regmi, N., Klingler, S., Hatanpaa, K. J., Raisanen, J., Cho, S. K., Sirasanagandla, S., Nannepaga, S., Piccirillo, S., Mashimo, T., Wang, S., Humphries, C. G., Mickey, B., Maher, E. A., Zheng, H., Kim, R. S., Kittler, R. and Bachoo, R. M. (2017) 'Oncogenes Activate an Autonomous Transcriptional Regulatory Circuit That Drives Glioblastoma', *Cell Rep*, 18(4), pp. 961-976.

Singh, N., Gupta, S., Pandey, R. M., Chauhan, S. S. and Saraya, A. (2015) 'High levels of cell-free circulating nucleic acids in pancreatic cancer are associated with vascular encasement, metastasis and poor survival', *Cancer Invest*, 33(3), pp. 78-85.

Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. and Shendure, J. (2016) 'Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin', *Cell*, 164(1-2), pp. 57-68.

Soeda, A., Lathia, J., Williams, B. J., Wu, Q., Gallagher, J., Androutsellis-Theotokis, A., Giles, A. J., Yang, C., Zhuang, Z., Gilbert, M. R., Rich, J. N. and Park, D. M. (2017) 'The p38 signaling pathway mediates quiescence of glioma stem cells by regulating epidermal growth factor receptor trafficking', *Oncotarget*, 8(20), pp. 33316-33328.

Song, C. X., Yin, S., Ma, L., Wheeler, A., Chen, Y., Zhang, Y., Liu, B., Xiong, J., Zhang, W., Hu, J., Zhou, Z., Dong, B., Tian, Z., Jeffrey, S. S., Chua, M. S., So, S., Li, W., Wei, Y., Diao, J., Xie, D. and Quake, S. R. (2017) '5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages', *Cell Res*, 27(10), pp. 1231-1242.

Spindler, K. L., Appelt, A. L., Pallisgaard, N., Andersen, R. F., Brandslund, I. and Jakobsen, A. (2014) 'Cell-free DNA in healthy individuals, noncancerous disease and strong prognostic value in colorectal cancer', *Int J Cancer*, 135(12), pp. 2984-91.

Struhl, K. and Segal, E. (2013a) 'Determinants of nucleosome positioning', *Nat Struct Mol Biol*, 20(3), pp. 267-73.

Struhl, K. and Segal, E. (2013b) 'Determinants of nucleosome positioning', *Nat Struct Mol Biol*, 20(3), pp. 267-73.

Su, M., Han, D., Boyd-Kirkup, J., Yu, X. and Han, J. J. (2014) 'Evolution of Alu elements toward enhancers', *Cell Rep*, 7(2), pp. 376-385.

Suvà, M. L., Rheinbay, E., Gillespie, S. M., Patel, A. P., Wakimoto, H., Rabkin, S. D., Riggi, N., Chi, A. S., Cahill, D. P., Nahed, B. V., Curry, W. T., Martuza, R. L., Rivera, M. N., Rossetti, N., Kasif, S., Beik, S., Kadri, S., Tirosh, I., Wortman, I., Shalek, A. K., Rozenblatt-Rosen, O., Regev, A., Louis, D. N. and Bernstein, B. E. (2014) 'Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells', *Cell*, 157(3), pp. 580-94.

Szopa, W., Burley, T. A., Kramer-Marek, G. and Kaspera, W. (2017) 'Diagnostic and Therapeutic Biomarkers in Glioblastoma: Current Status and Future Perspectives', *Biomed Res Int*, 2017, pp. 8013575.

Tachon, G., Cortes, U., Guichet, P. O., Rivet, P., Balbous, A., Masliantsev, K., Berger, A., Boissonnade, O., Wager, M. and Karayan-Tapon, L. (2018) 'Cell Cycle Changes after Glioblastoma Stem Cell Irradiation: The Major Role of RAD51', *Int J Mol Sci*, 19(10).

Tamaru, H. (2010a) 'Confining euchromatin/heterochromatin territory: jumonji crosses the line', *Genes Dev*, 24(14), pp. 1465-78.

Tamaru, H. (2010b) 'Confining euchromatin/heterochromatin territory: jumonji crosses the line', *Genes Dev*, 24(14), pp. 1465-78.

Tanaka, Y., Yamashita, R., Suzuki, Y. and Nakai, K. (2010) 'Effects of Alu elements on global nucleosome positioning in the human genome', *BMC Genomics*, 11, pp. 309.

Teif, V. B. (2016) 'Nucleosome positioning: resources and tools online', *Brief Bioinform*, 17(5), pp. 745-57.

Teif V.B. and Cherstvy A.G. (2016). Chromatin and epigenetics: current biophysical views. *AIMS Biophysics* 3, 88-98.

Teif, V. B. and Rippe, K. (2009) 'Predicting nucleosome positions on the DNA: combining intrinsic sequence preferences and remodeler activities', *Nucleic Acids Res*, 37(17), pp. 5641-55.

Tilghman, J., Schiapparelli, P., Lal, B., Ying, M., Quinones-Hinojosa, A., Xia, S. and Lathera, J. (2016) 'Regulation of Glioblastoma Tumor-Propagating Cells by the Integrin Partner Tetraspanin CD151', *Neoplasia*, 18(3), pp. 185-98.

Tome-Garcia, J., Erfani, P., Nudelman, G., Tsankov, A. M., Katsyv, I., Tejero, R., Bin Zhang, Walsh, M., Friedel, R. H., Zaslavsky, E. and Tsankova, N. M. (2018) 'Analysis of chromatin accessibility uncovers TEAD1 as a regulator of migration in human glioblastoma', *Nat Commun*, 9(1), pp. 4020.

Tompitak, M., Vaillant, C. and Schiessel, H. (2017) 'Genomes of Multicellular Organisms Have Evolved to Attract Nucleosomes to Promoter Regions', *Biophys J*, 112(3), pp. 505-511.

Tung, B., Ma, D., Wang, S., Oyinlade, O., Lathera, J., Ying, M., Lv, S. Q., Wei, S. and Xia, S. (2018) 'Krüppel-like factor 9 and histone deacetylase inhibitors synergistically induce cell death in glioblastoma stem-like cells', *BMC Cancer*, 18(1), pp. 1025.

Ulz, P., Perakis, S., Zhou, Q., Moser, T., Belic, J., Lazzeri, I., Wölfler, A., Zebisch, A., Gerger, A., Pristauz, G., Petru, E., White, B., Roberts, C. E. S., John, J. S., Schimek, M. G., Geigl, J. B., Bauernhofer, T., Sill, H., Bock, C., Heitzer, E. and Speicher, M. R. (2019) 'Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection', *Nat Commun*, 10(1), pp. 4666.

Underhill, H. R., Kitzman, J. O., Hellwig, S., Welker, N. C., Daza, R., Baker, D. N., Gligorich, K. M., Rostomily, R. C., Bronner, M. P. and Shendure, J. (2016) 'Fragment Length of Circulating Tumor DNA', *PLoS Genet*, 12(7), pp. e1006162.

Vainshtein, Y., Rippe, K. and Teif, V. B. (2017a) 'NucTools: analysis of chromatin feature occupancy profiles from high-throughput sequencing data', *BMC Genomics*, 18(1), pp. 158.

Vainshtein, Y., Rippe, K. and Teif, V. B. (2017b) 'NucTools: analysis of chromatin feature occupancy profiles from high-throughput sequencing data', *BMC Genomics*, 18(1), pp. 158.

Vasseur, P., Tonazzini, S., Ziane, R., Camasses, A., Rando, O. J. and Radman-Livaja, M. (2016) 'Dynamics of Nucleosome Positioning Maturation following Genomic Replication', *Cell Rep*, 16(10), pp. 2651-2665.

Waitkus, M. S., Diplas, B. H. and Yan, H. (2018) 'Biological Role and Therapeutic Potential of IDH Mutations in Cancer', *Cancer Cell*, 34(2), pp. 186-195.

Wan, J., Su, Y., Song, Q., Tung, B., Oyinlade, O., Liu, S., Ying, M., Ming, G. L., Song, H., Qian, J., Zhu, H. and Xia, S. (2017) 'Methylated', *Elife*, 6.

Wang, F., He, L., Huangyang, P., Liang, J., Si, W., Yan, R., Han, X., Liu, S., Gui, B., Li, W., Miao, D., Jing, C., Liu, Z., Pei, F., Sun, L. and Shang, Y. (2014a) 'JMJD6 promotes colon carcinogenesis through negative regulation of p53 by hydroxylation', *PLoS Biol*, 12(3), pp. e1001819.

Wang, H., Zang, C., Taing, L., Arnett, K. L., Wong, Y. J., Pear, W. S., Blacklow, S. C., Liu, X. S. and Aster, J. C. (2014b) 'NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers', *Proc Natl Acad Sci U S A*, 111(2), pp. 705-10.

Wang, J., Lawry, S. T., Cohen, A. L. and Jia, S. (2014c) 'Chromosome boundary elements and regulation of heterochromatin spreading', *Cell Mol Life Sci*, 71(24), pp. 4841-52.

Wang, Y., Bibi, M., Min, P., Deng, W., Zhang, Y. and Du, J. (2019) 'SOX2 promotes hypoxia-induced breast cancer cell migration by inducing NEDD9 expression and subsequent activation of Rac1/HIF-1 $\alpha$  signaling', *Cell Mol Biol Lett*, 24, pp. 55.

Wang, Z., Hao, Y., Zhang, C., Liu, X., Li, G., Sun, L., Liang, J., Luo, J., Zhou, D., Chen, R. and Jiang, T. (2017) 'The Landscape of Viral Expression Reveals Clinically Relevant Viruses with Potential Capability of Promoting Malignancy in Lower-Grade Glioma', *Clin Cancer Res*, 23(9), pp. 2177-2185.

Weth, O., Paprotka, C., Günther, K., Schulte, A., Baierl, M., Leers, J., Galjart, N. and Renkawitz, R. (2014) 'CTCF induces histone variant incorporation, erases the H3K27me3 histone mark and opens chromatin', *Nucleic Acids Res*, 42(19), pp. 11941-51.

Williamson, E. A., Wray, J. W., Bansal, P. and Hromas, R. (2012a) 'Overview for the histone codes for DNA repair', *Prog Mol Biol Transl Sci*, 110, pp. 207-27.

Williamson, E. A., Wray, J. W., Bansal, P. and Hromas, R. (2012b) 'Overview for the histone codes for DNA repair', *Prog Mol Biol Transl Sci*, 110, pp. 207-27.

Wong, M., Sun, Y., Xi, Z., Milazzo, G., Poulos, R. C., Bartenhagen, C., Bell, J. L., Mayoh, C., Ho, N., Tee, A. E., Chen, X., Li, Y., Ciaccio, R., Liu, P. Y., Jiang, C. C., Lan, Q., Jayatilleke, N., Cheung, B. B., Haber, M., Norris, M. D., Zhang, X. D., Marshall, G. M., Wang, J. Y., Hüttelmaier, S., Fischer, M., Wong, J. W. H., Xu, H., Perini, G., Dong, Q., George, R. E. and Liu, T. (2019) 'JMJD6 is a tumorigenic factor and therapeutic target in neuroblastoma', *Nat Commun*, 10(1), pp. 3319.

Xie, Q., Wu, Q., Kim, L., Miller, T. E., Liau, B. B., Mack, S. C., Yang, K., Factor, D. C., Fang, X., Huang, Z., Zhou, W., Alazem, K., Wang, X., Bernstein, B. E., Bao, S. and Rich, J. N. (2016a) 'RBPJ maintains brain tumor-initiating cells through CDK9-mediated transcriptional elongation', *J Clin Invest*, 126(7), pp. 2757-72.

Xie, Q., Wu, Q., Kim, L., Miller, T. E., Liau, B. B., Mack, S. C., Yang, K., Factor, D. C., Fang, X., Huang, Z., Zhou, W., Alazem, K., Wang, X., Bernstein, B. E., Bao, S. and Rich, J. N. (2016b) 'RBPJ maintains brain tumor-initiating cells through CDK9-mediated transcriptional elongation', *J Clin Invest*, 126(7), pp. 2757-72.

Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G. J., Friedman, H., Friedman, A., Reardon, D., Herndon, J., Kinzler, K. W., Velculescu, V. E., Vogelstein, B. and Bigner, D. D. (2009) 'IDH1 and IDH2 mutations in gliomas', *N Engl J Med*, 360(8), pp. 765-73.

Yang, X., Deng, S., Wei, X., Yang, J., Zhao, Q., Yin, C., Du, T., Guo, Z., Xia, J., Yang, Z., Xie, W., Wang, S., Wu, Q., Yang, F., Zhou, X., Nauen, R., Bass, C. and Zhang, Y. (2020) 'MAPK-directed activation of the whitefly transcription factor', *Proc Natl Acad Sci U S A*, 117(19), pp. 10246-10253.

Yazdi, P. G., Pedersen, B. A., Taylor, J. F., Khattab, O. S., Chen, Y. H., Chen, Y., Jacobsen, S. E. and Wang, P. H. (2015) 'Increasing Nucleosome Occupancy Is Correlated with an Increasing Mutation Rate so Long as DNA Repair Machinery Is Intact', *PLoS One*, 10(8), pp. e0136574.

Ying, M., Tilghman, J., Wei, Y., Guerrero-Cazares, H., Quinones-Hinojosa, A., Ji, H. and Lattera, J. (2014) 'Kruppel-like factor-9 (KLF9) inhibits glioblastoma stemness through global transcription repression and integrin  $\alpha 6$  inhibition', *J Biol Chem*, 289(47), pp. 32742-56.

Zhang, G., Tanaka, S., Jiapaer, S., Sabit, H., Tamai, S., Kinoshita, M. and Nakada, M. (2020a) 'RBPJ contributes to the malignancy of glioblastoma and induction of proneural-mesenchymal transition via IL-6-STAT3 pathway', *Cancer Sci*.

Zhang, L. H., Yin, Y. H., Chen, H. Z., Feng, S. Y., Liu, J. L., Chen, L., Fu, W. L., Sun, G. C., Yu, X. G. and Xu, D. G. (2020b) 'TRIM24 Promotes Stemness and Invasiveness of Glioblastoma Cells via Activating SOX2 Expression', *Neuro Oncol*.

Zhang, S. and Cui, W. (2014) 'Sox2, a key factor in the regulation of pluripotency and neural differentiation', *World J Stem Cells*, 6(3), pp. 305-11.

Zhang, Y., Dube, C., Gibert, M., Cruickshanks, N., Wang, B., Coughlan, M., Yang, Y., Setiady, I., Deveau, C., Saoud, K., Grello, C., Oxford, M., Yuan, F. and Abounader, R. (2018a) 'The p53 Pathway in Glioblastoma', *Cancers (Basel)*, 10(9).

Zhang, Y., Yang, L., Kucherlapati, M., Chen, F., Hadjipanayis, A., Pantazi, A., Bristow, C. A., Lee, E. A., Mahadeshwar, H. S., Tang, J., Zhang, J., Seth, S., Lee, S., Ren, X., Song, X., Sun, H., Seidman, J., Luquette, L. J., Xi, R., Chin, L., Protopopov, A., Li, W., Park, P. J., Kucherlapati, R. and Creighton, C. J. (2018b) 'A Pan-Cancer Compendium of Genes

Deregulated by Somatic Genomic Rearrangement across More Than 1,400 Cases', *Cell Rep*, 24(2), pp. 515-527.

Zhong, Z., Zhou, F., Wang, D., Wu, M., Zhou, W., Zou, Y., Li, J., Wu, L. and Yin, X. (2018) 'Expression of KLF9 in pancreatic cancer and its effects on the invasion, migration, apoptosis, cell cycle distribution, and proliferation of pancreatic cancer cell lines', *Oncol Rep*, 40(6), pp. 3852-3860.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S. O., Nitta, K. R., Morgunova, E., Taipale, M., Cramer, P. and Taipale, J. (2018) 'The interaction landscape between transcription factors and the nucleosome', *Nature*, 562(7725), pp. 76-81.



## Appendix: Supplementary Figures and Computer Scripts

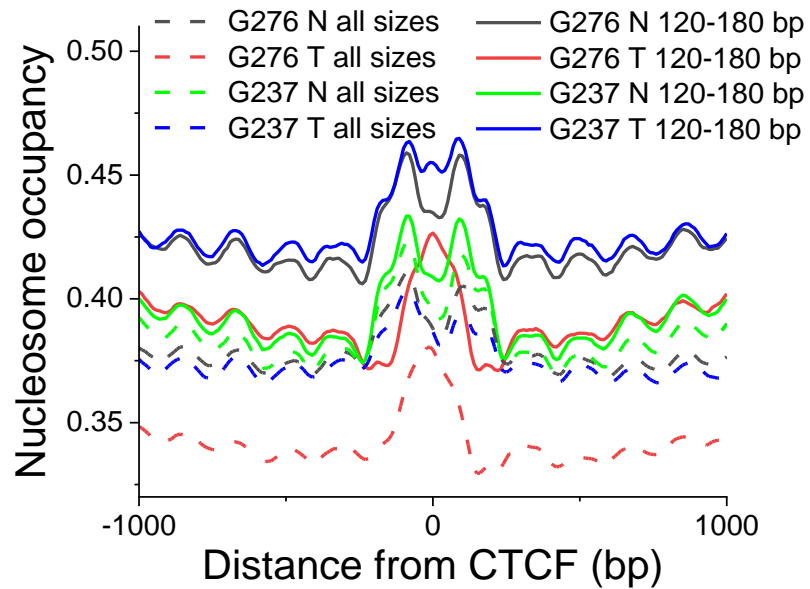


Figure S1. Nucleosome occupancy profiles changes around CTCF for all nucleosomal DNA fragment sizes without filtering versus nucleosome occupancy profiles only for fragment sizes 120-180.

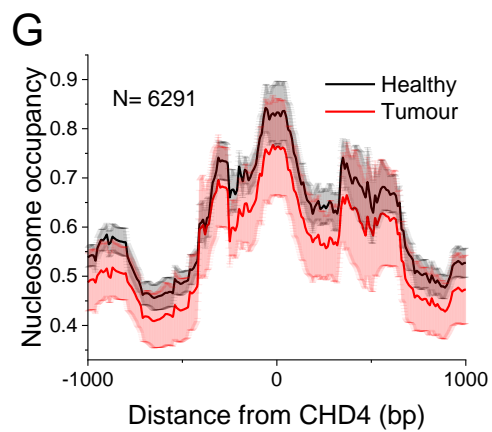
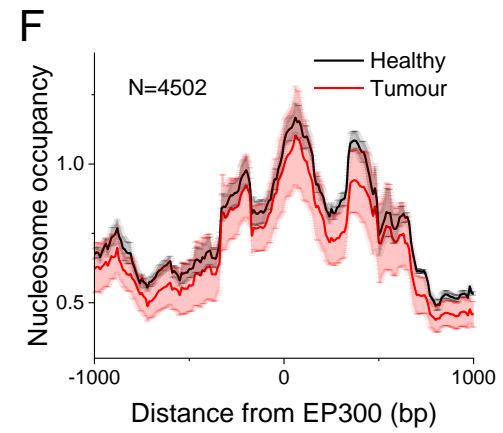
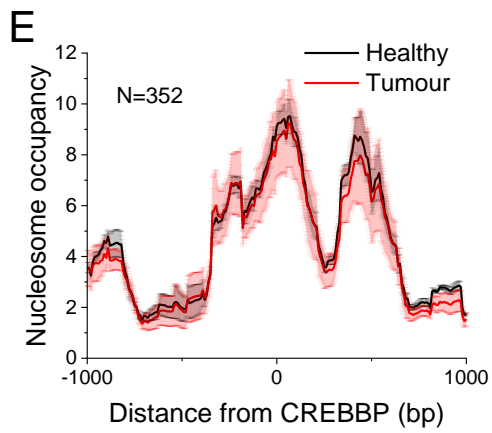
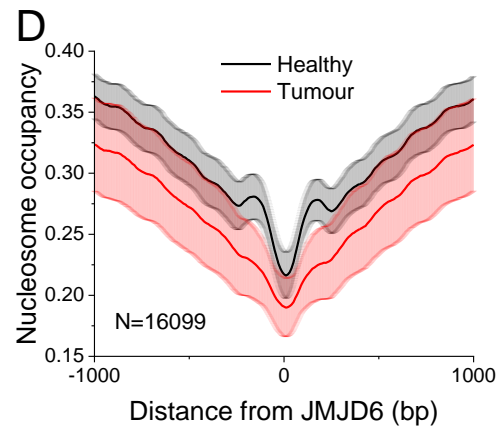
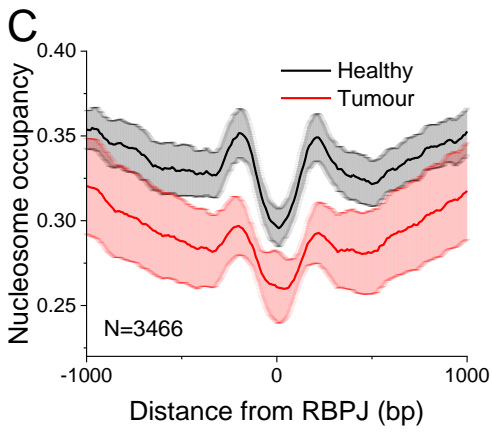
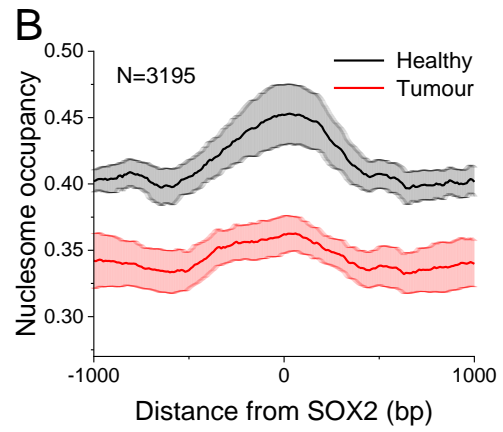
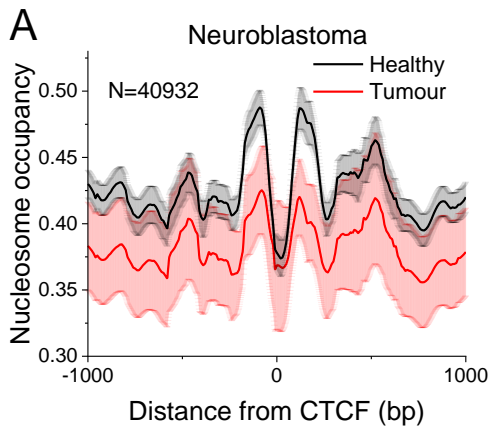


Figure S2. Average nucleosome occupancy profile around 5 transcription factor motifs for 120-180 base pair nucleosomal DNA fragment sizes averaged over 4 GBM patients from the experiments performed in Essex. (A) Neuroblastoma cell line CTCF (GSM803333), (B) SOX2 (GSE58345), (C) RBPJ (GSM2101765), (D) CHD4, (E) CREBBP and (F) EP300. Averaged profiles for healthy (black) and GBM (red). Lighter areas correspond to the standard errors of averaging. The number of regions (N) for each TF file is indicated on the graph.

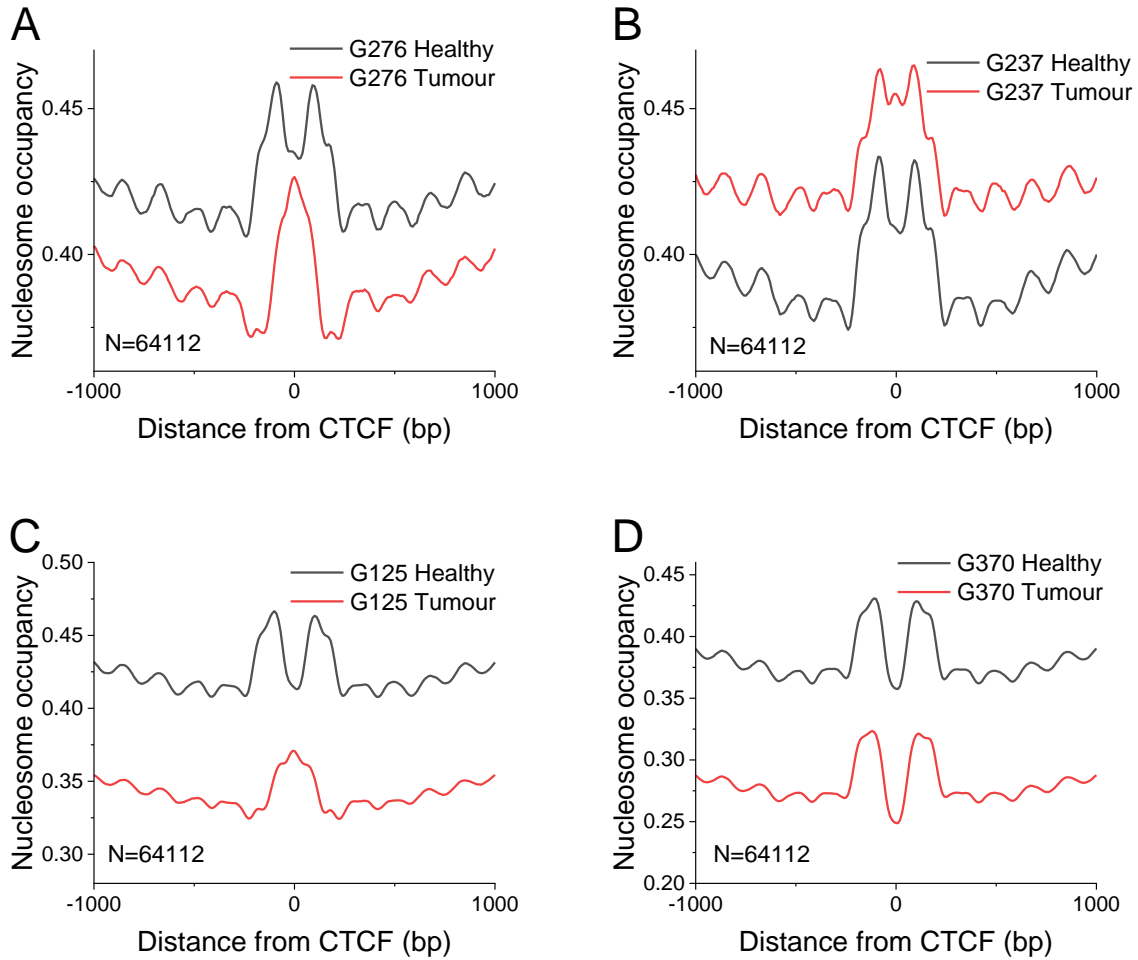


Figure S3. Average genome-wide nucleosome occupancy around CTCF (GSM822303) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

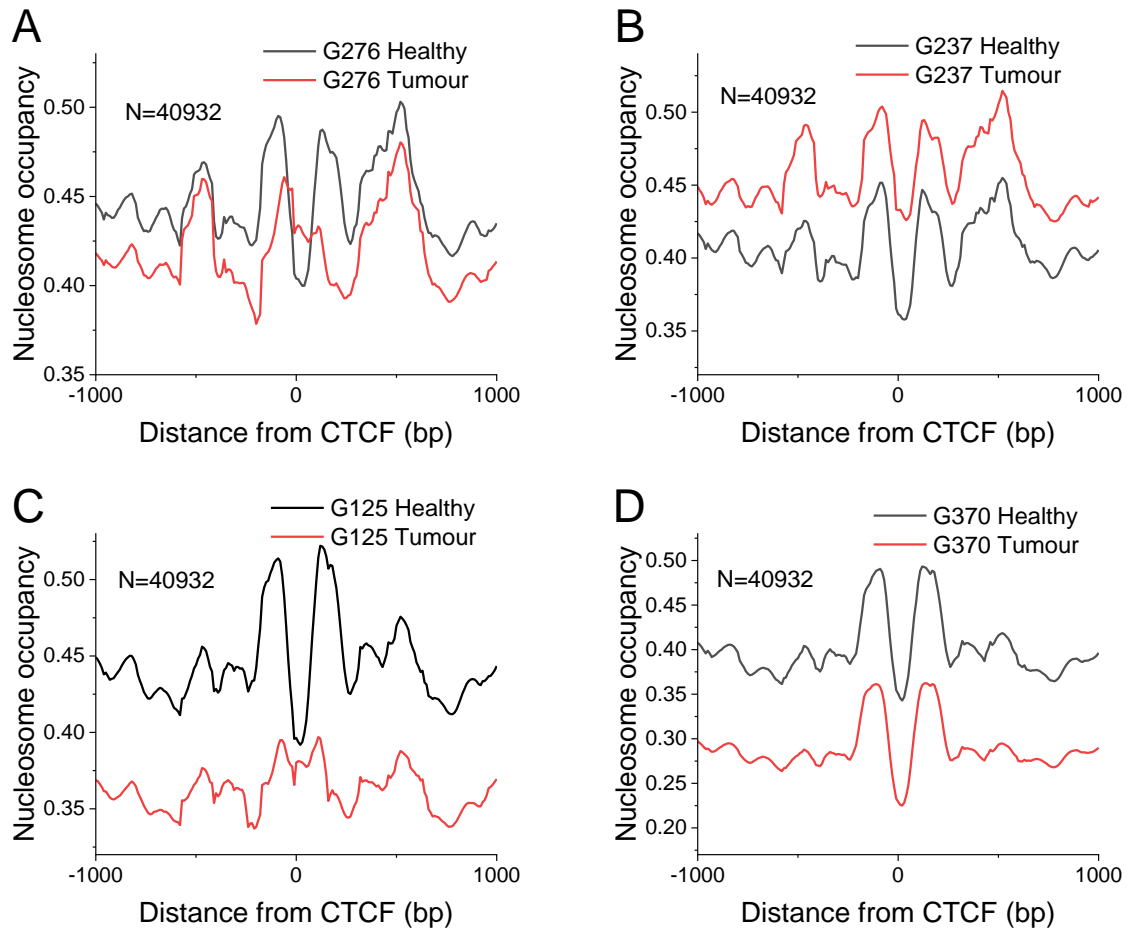


Figure S4. Average genome-wide nucleosome occupancy around CTCF (GSM803333) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

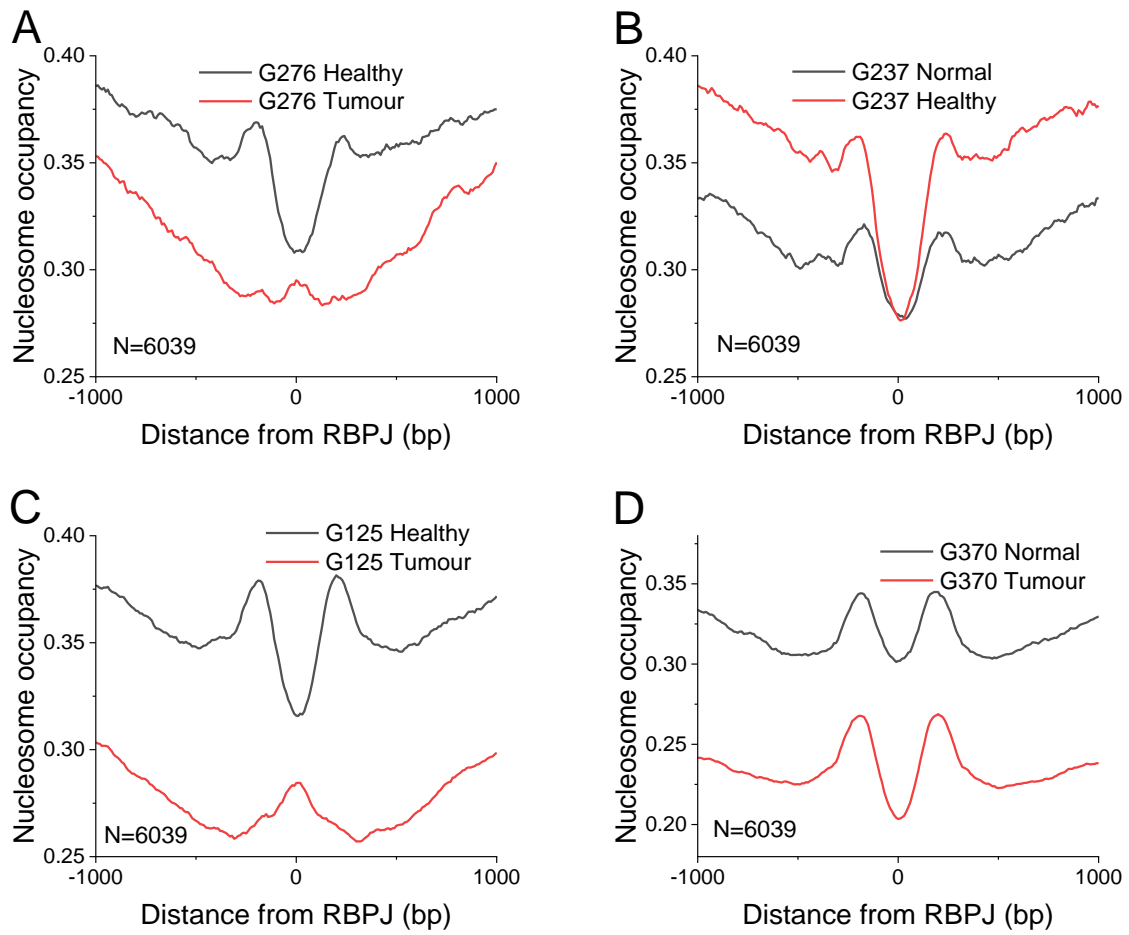


Figure S5. Average genome-wide nucleosome occupancy around RBPJ binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

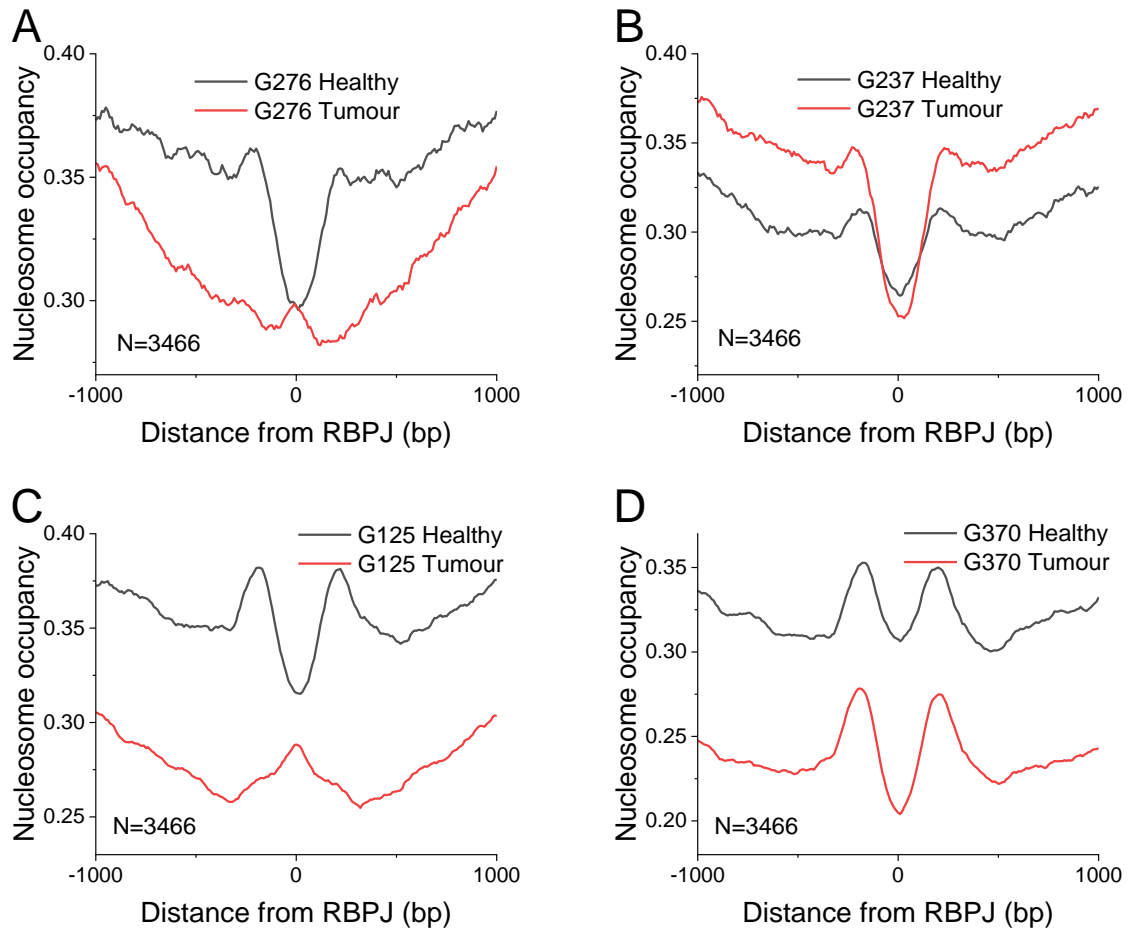


Figure S6. Average genome-wide nucleosome occupancy around RBPJ (GSM2101765) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

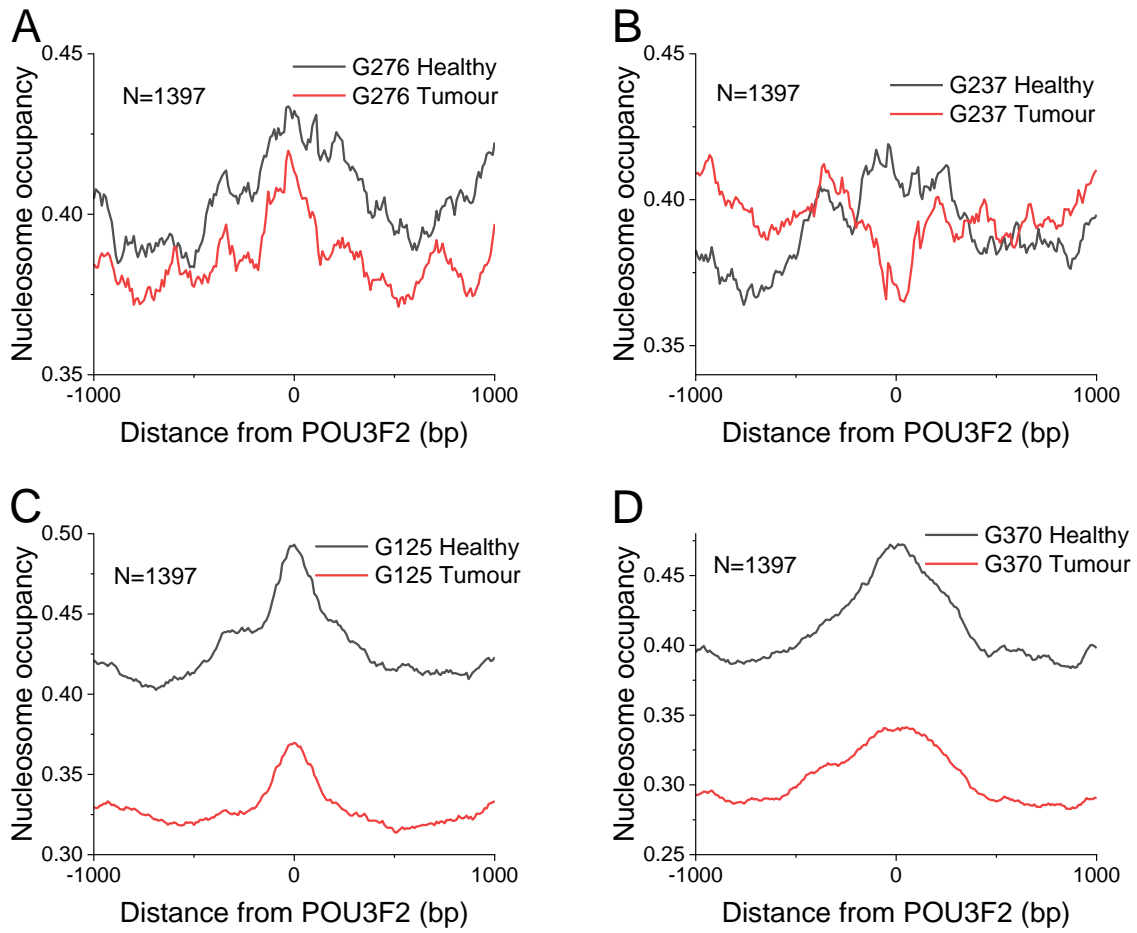


Figure S7. Average genome-wide nucleosome occupancy around POU3F2 binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient GG; Panel B: Patient JJ; Panel C: Patient G; Panel D: Patient G370.



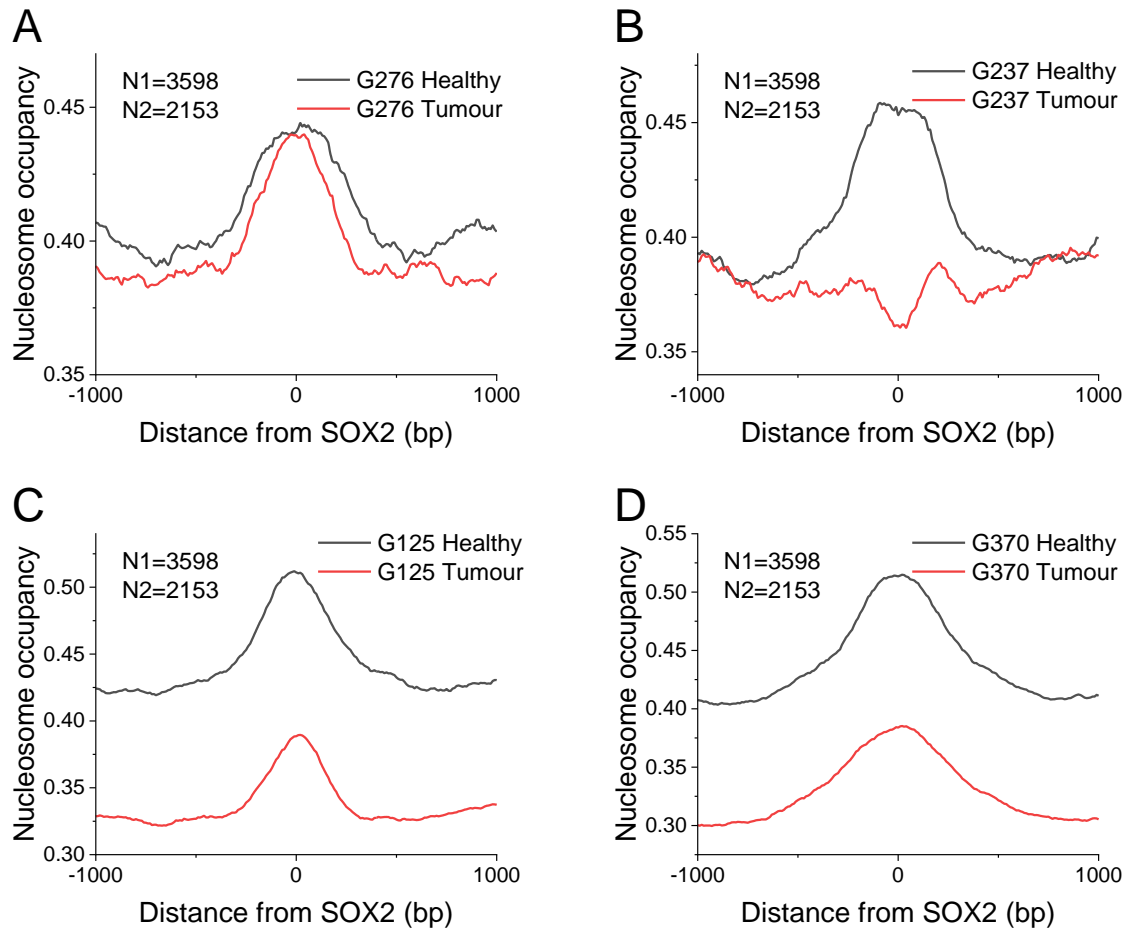


Figure S8. Average genome-wide nucleosome occupancy around SOX2 replicate 1 (GSM1306360) and replicate 2 (GSM1306362) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

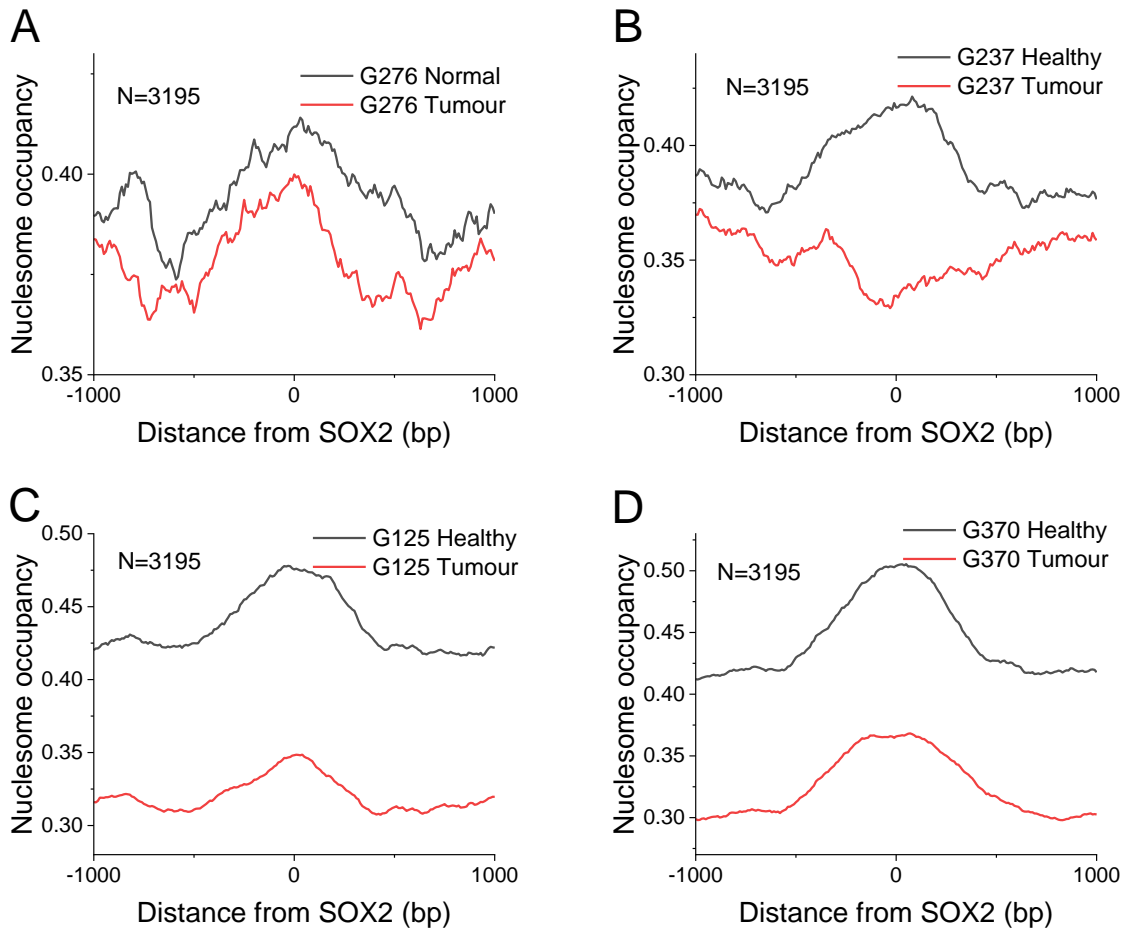


Figure S9. Average genome-wide nucleosome occupancy around SOX2 (GSE58345) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

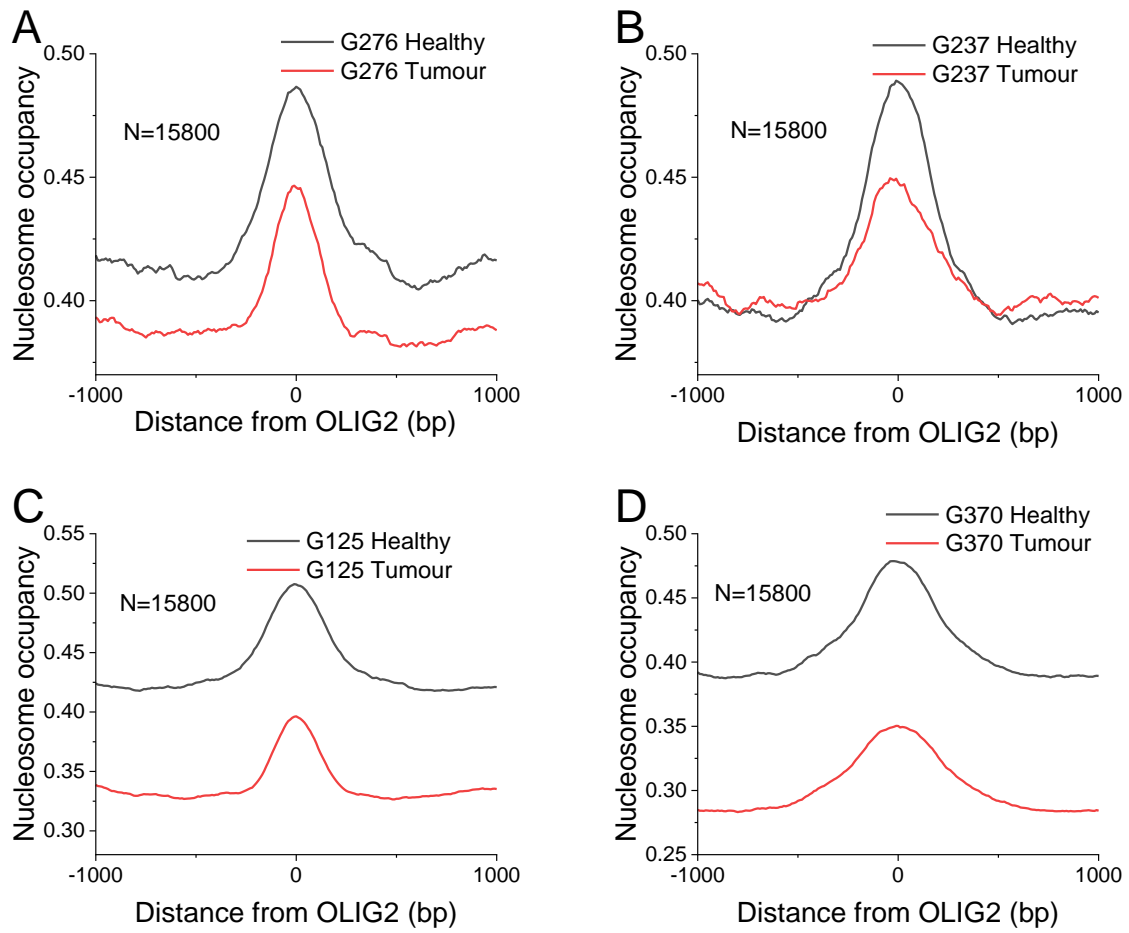


Figure S10. Average genome-wide nucleosome occupancy around OLIG2 (GSM1306367) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

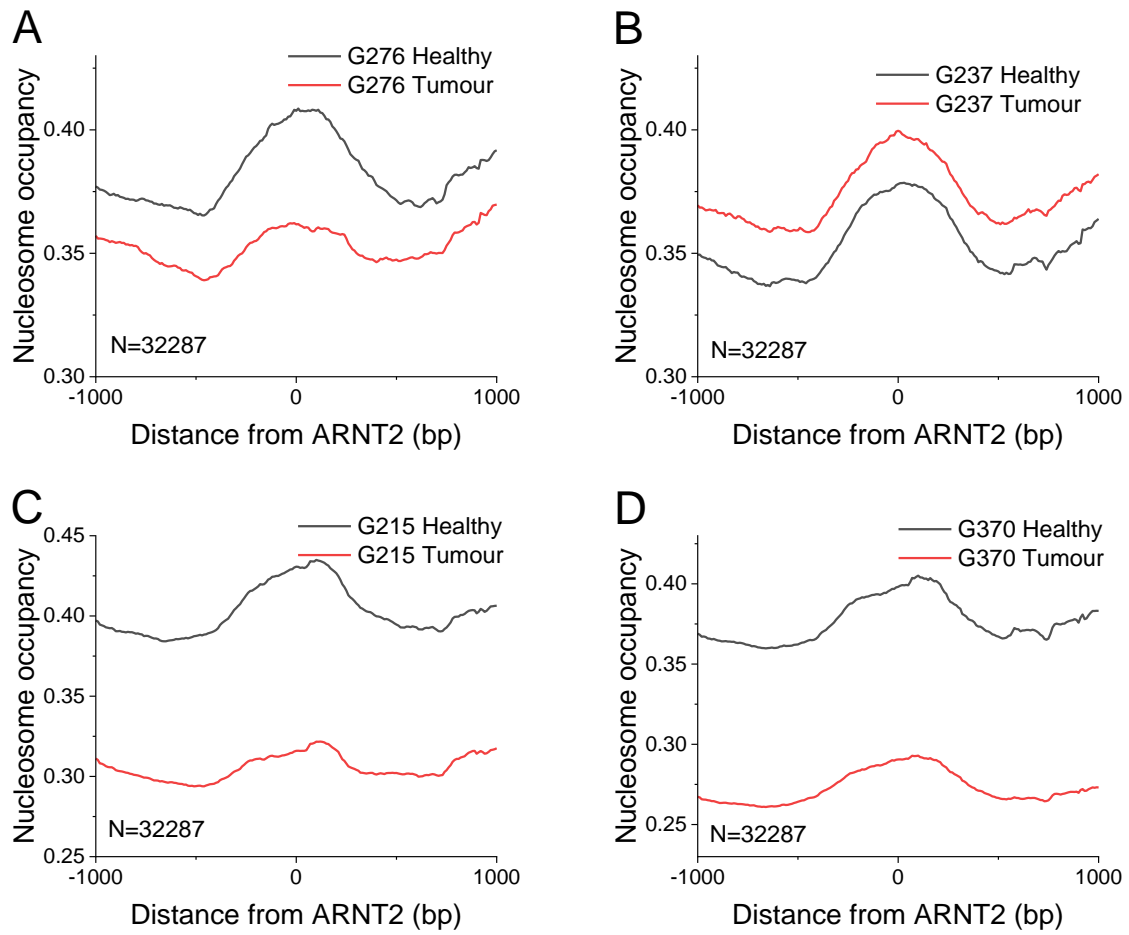


Figure S11. Average genome-wide nucleosome occupancy around ARNT2 binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

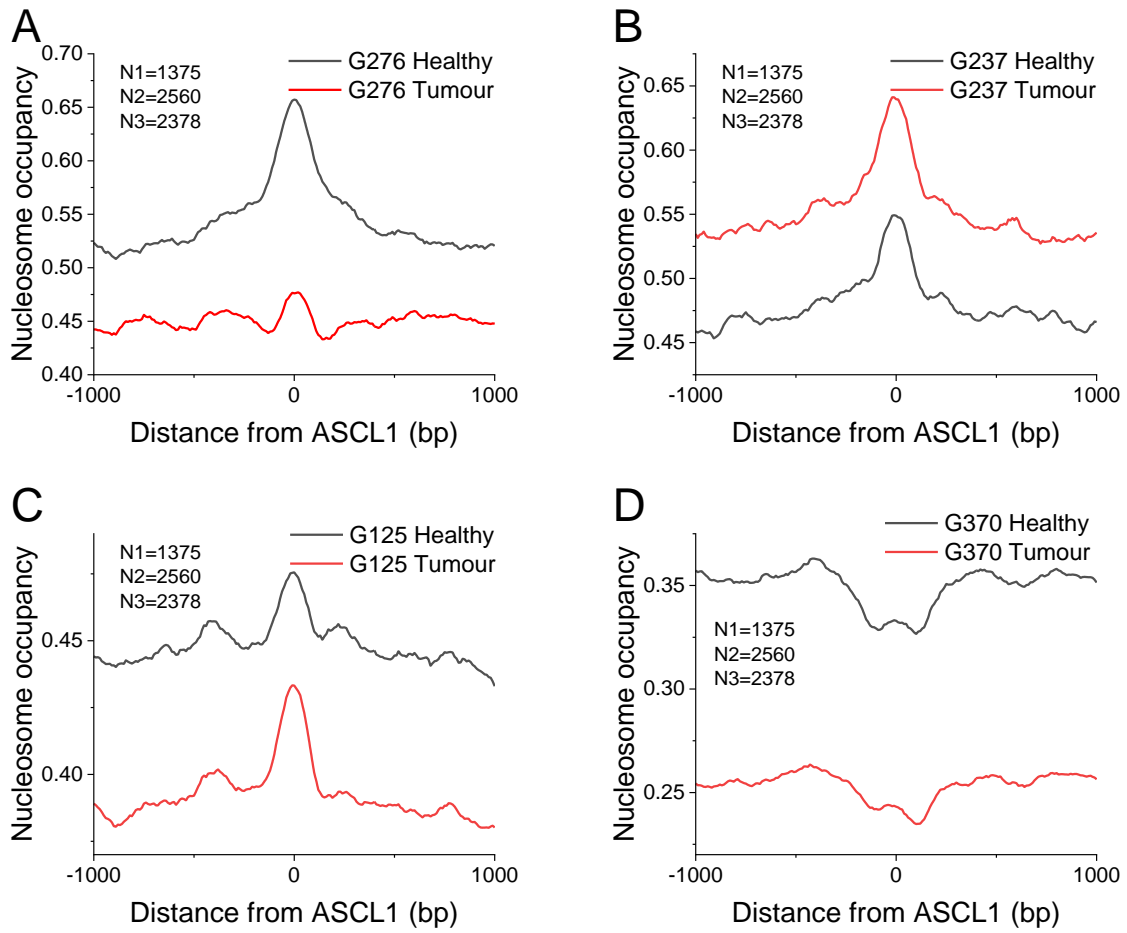


Figure S12. Average genome-wide nucleosome occupancy around ASCL1 replicate 1 (r1: GSM2335531), ASCL1 replicate 2 (r2: GSM2335532) and ASCL1 replicate 3 (r3: GSM2335533) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

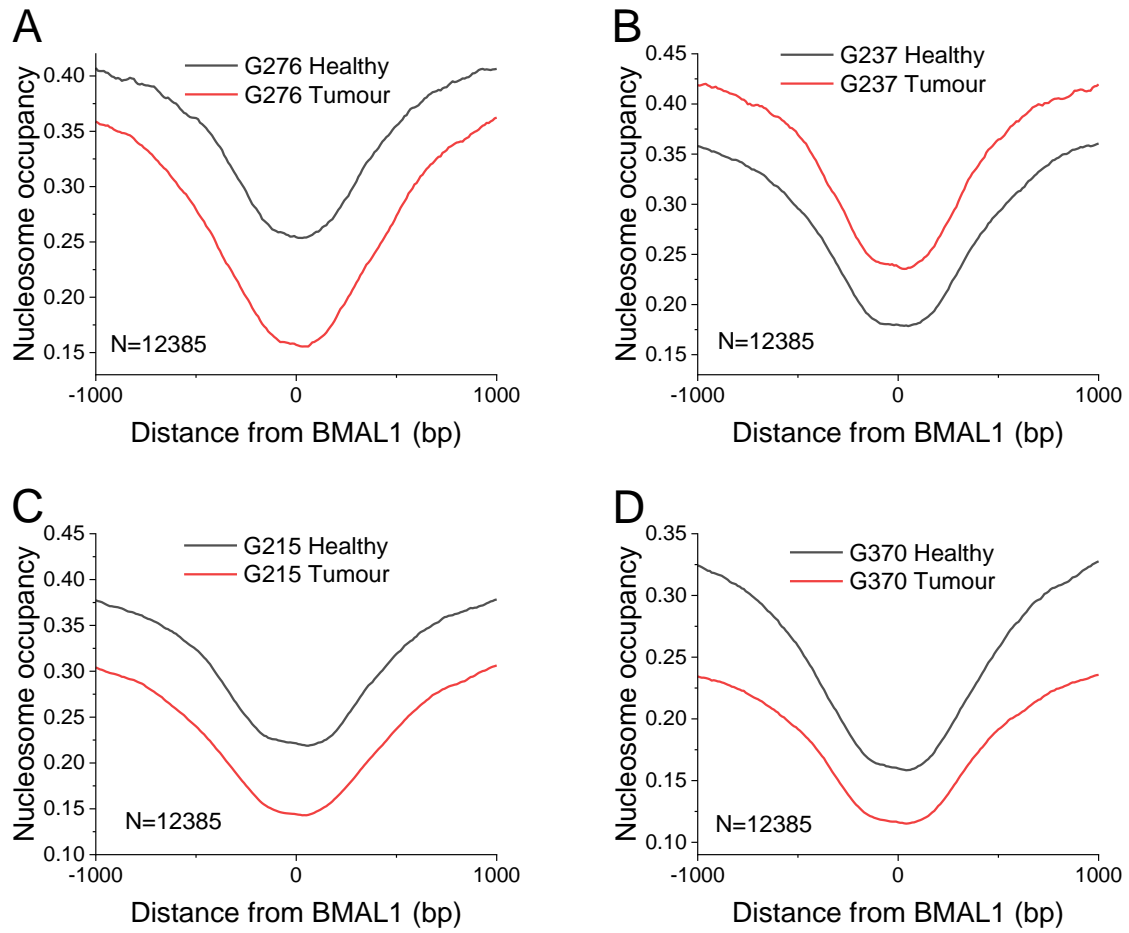


Figure S13. Average genome-wide nucleosome occupancy around BMAL1 binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

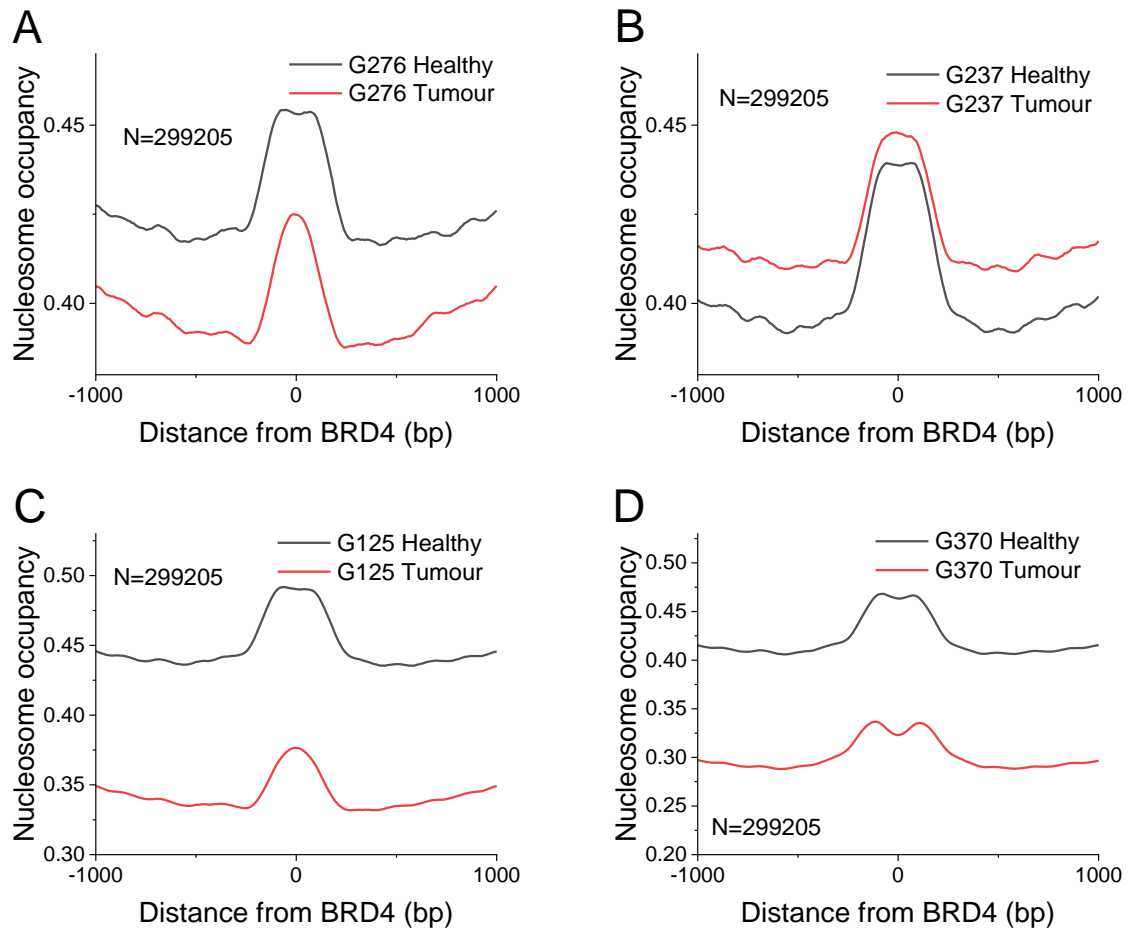


Figure S14. Average genome-wide nucleosome occupancy around BRD4 binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

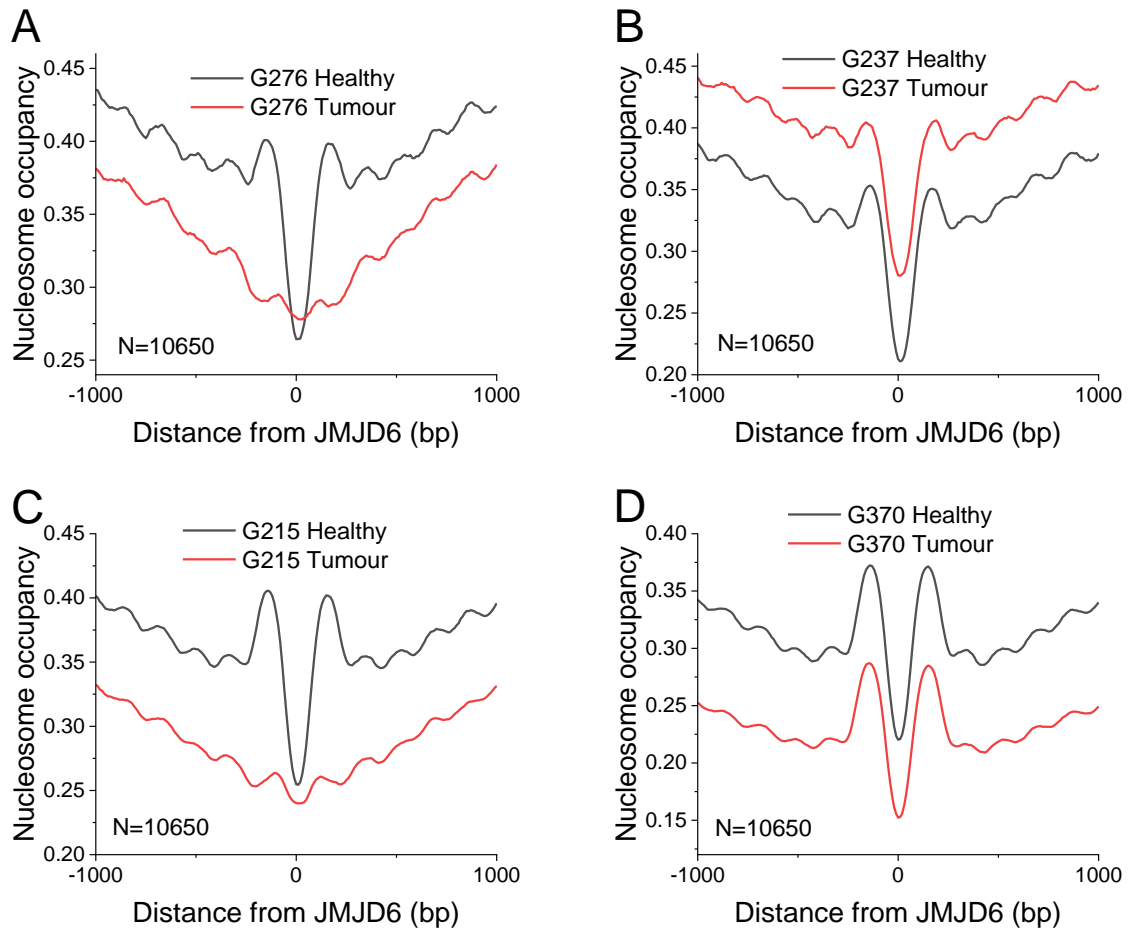


Figure S15. Average genome-wide nucleosome occupancy around JMJD6-528 (GSM1922076) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.



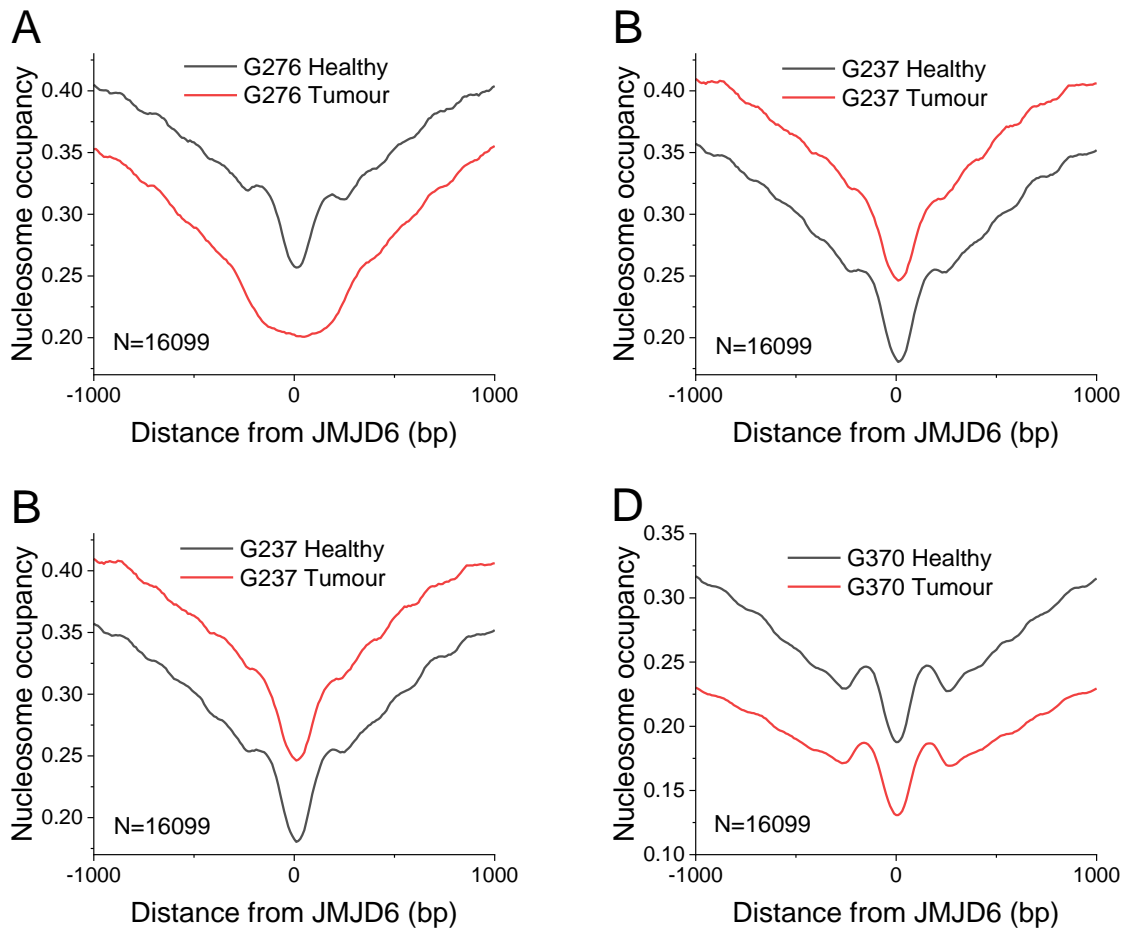


Figure S16. Average genome-wide nucleosome occupancy around JMJD6-3565 (GSM2360990) binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

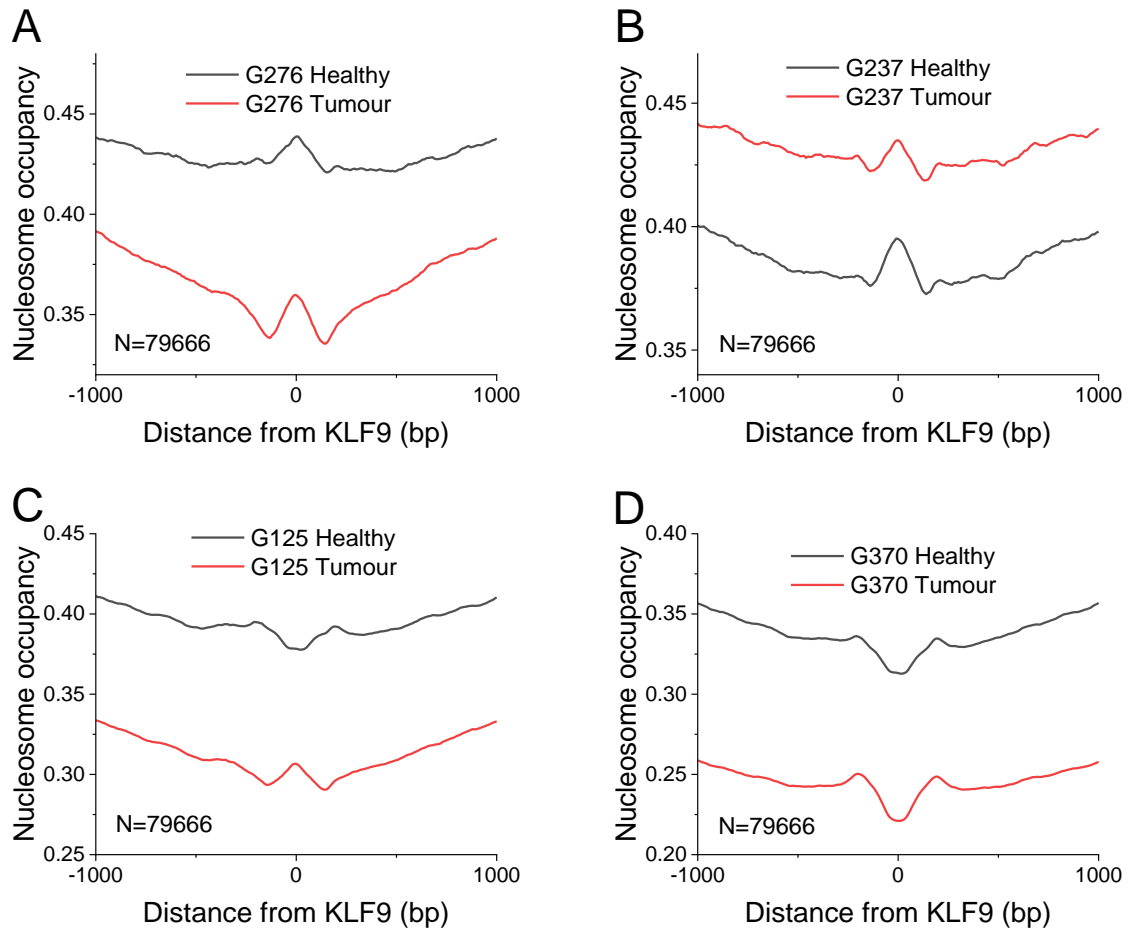


Figure S17. Average genome-wide nucleosome occupancy around KLF9 binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

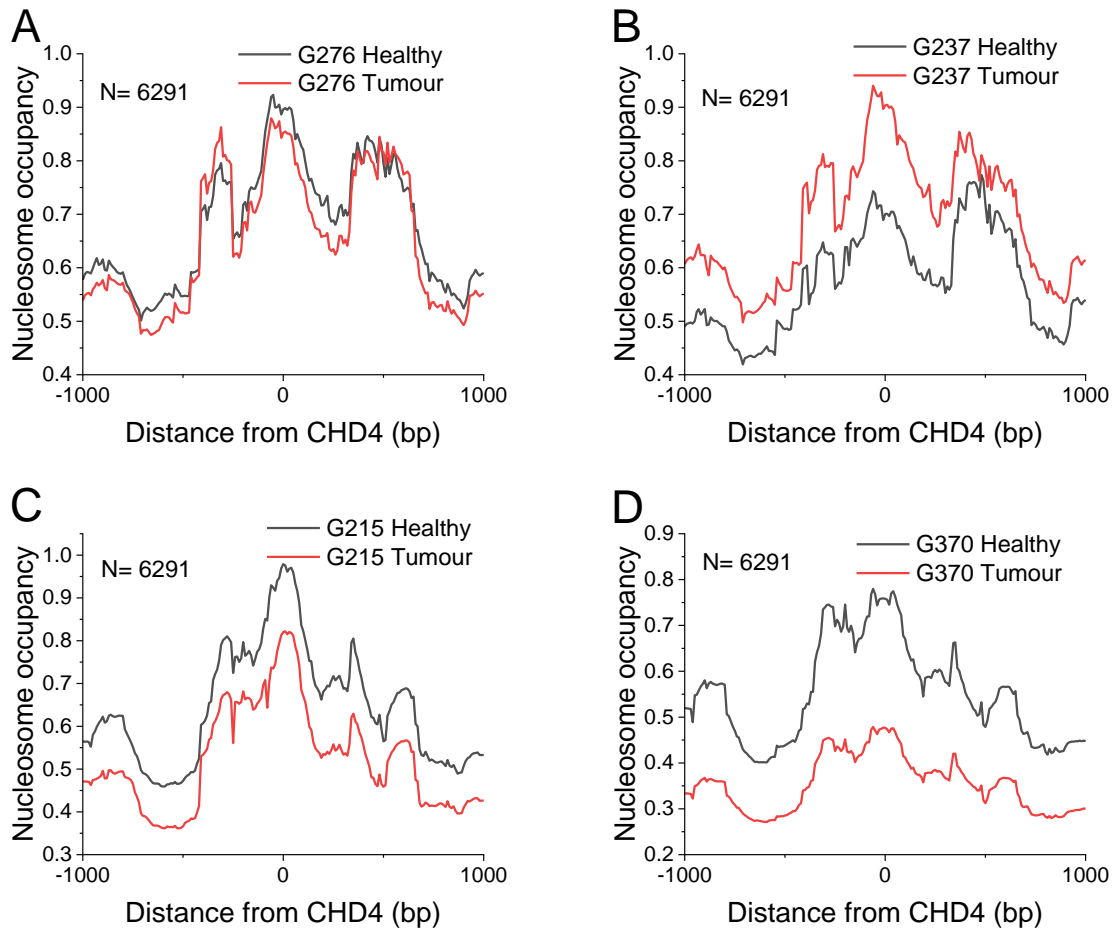


Figure S18. Average genome-wide nucleosome occupancy around CHD4 binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

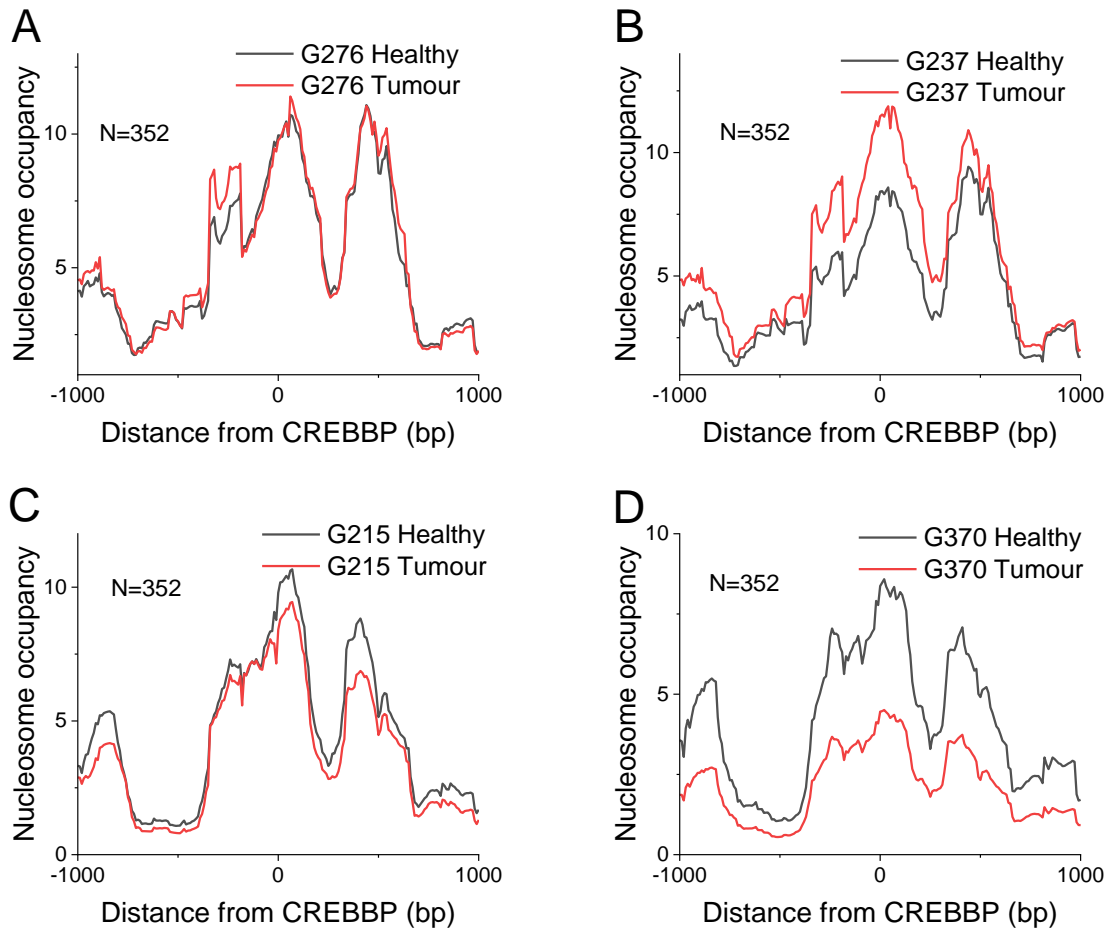


Figure S19. Average genome-wide nucleosome occupancy around CREBBP binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

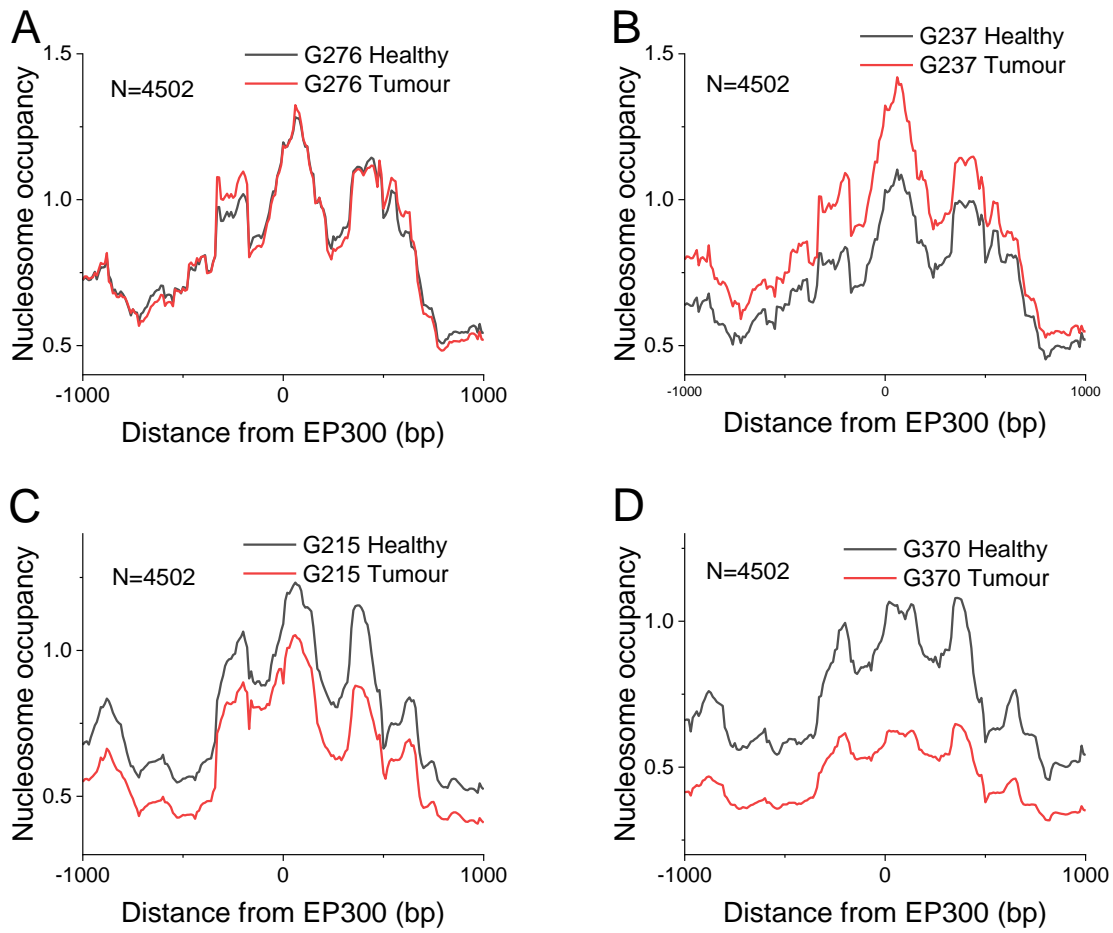


Figure S20. Average genome-wide nucleosome occupancy around EP300 binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

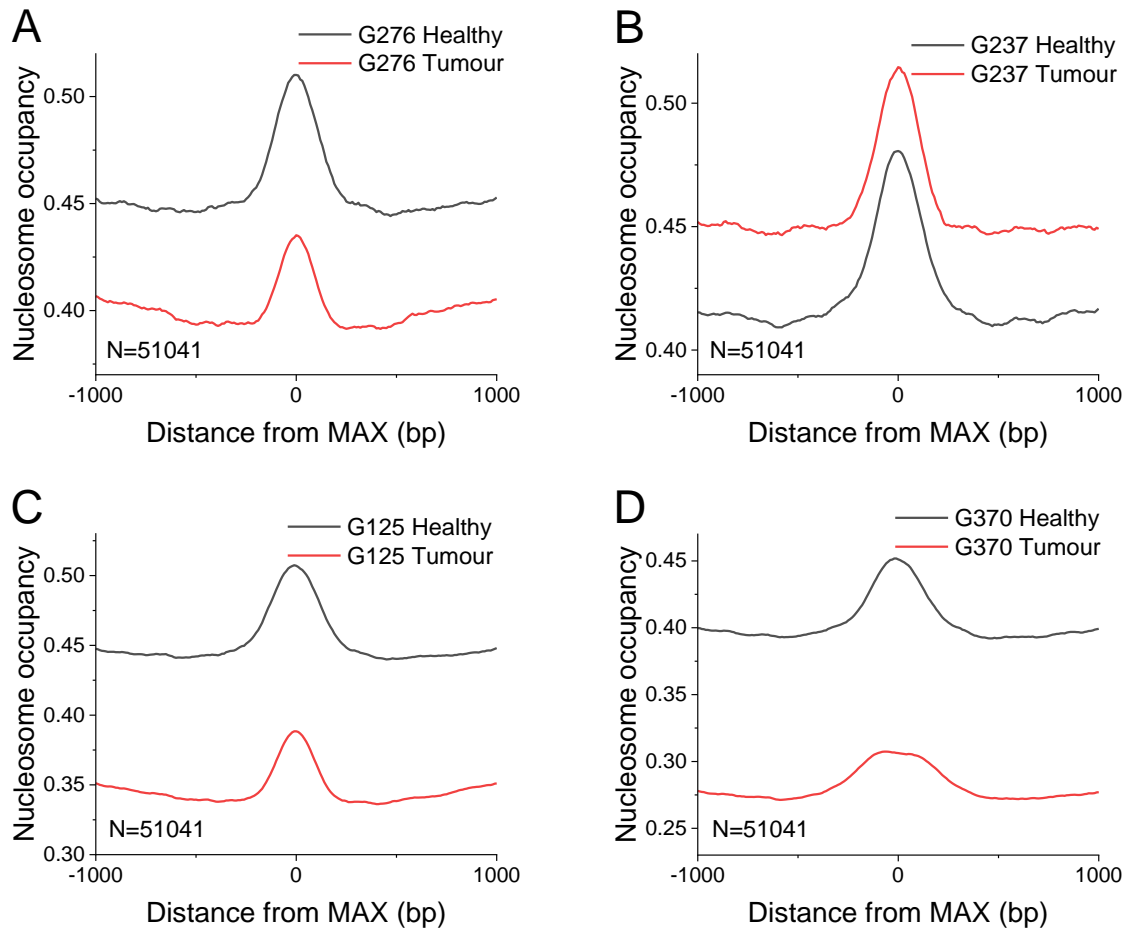


Figure S21. Average genome-wide nucleosome occupancy around MAX binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

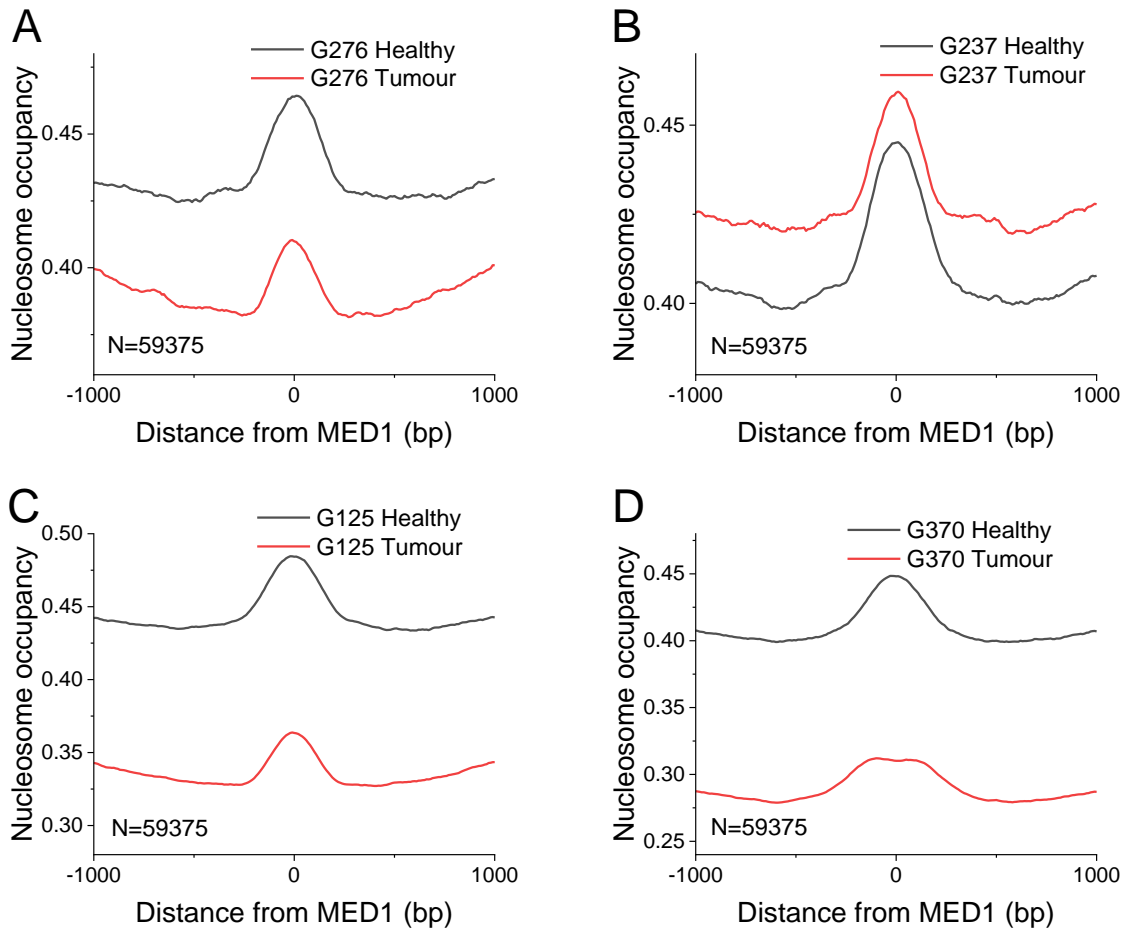


Figure S22. Average genome-wide nucleosome occupancy around MED1 binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.

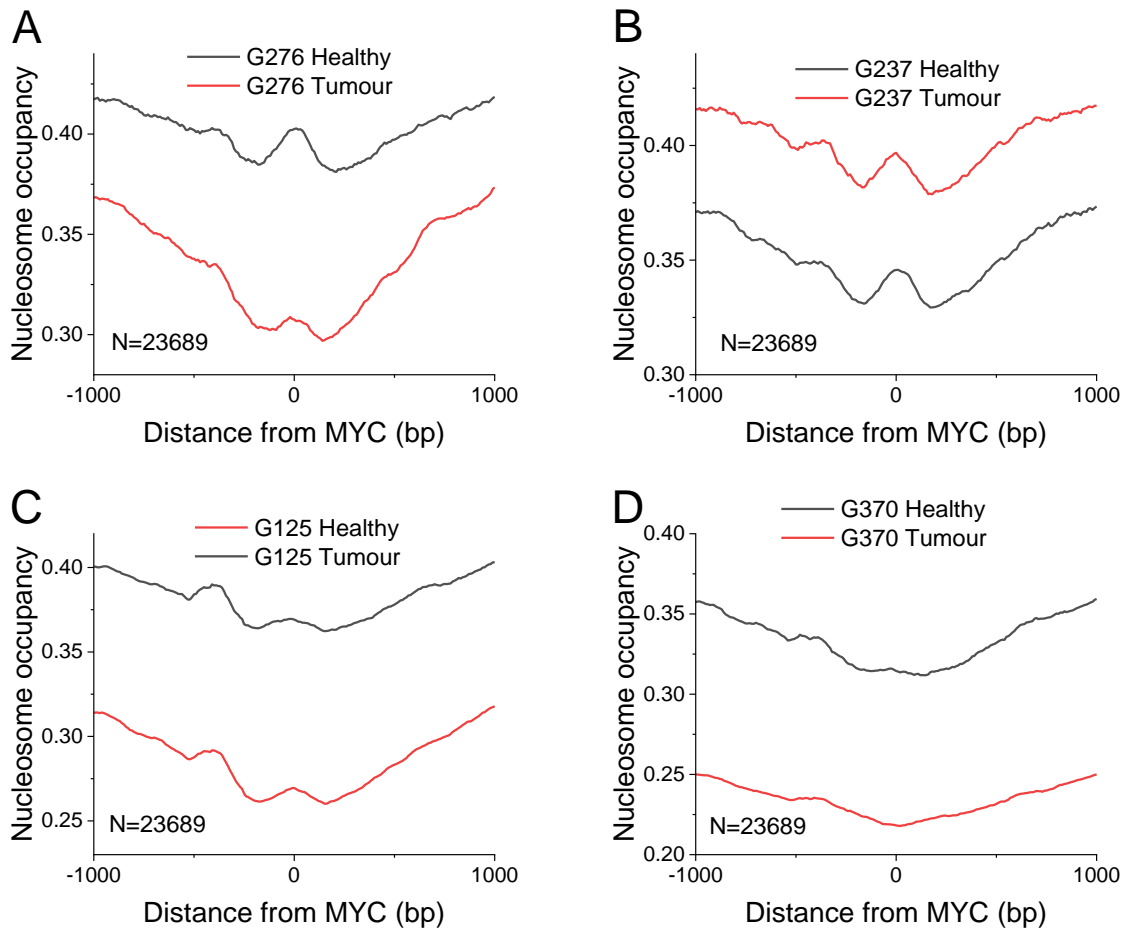


Figure S23. Average genome-wide nucleosome occupancy around MYC binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient 237; Panel C: Patient G125; Panel D: Patient G370.



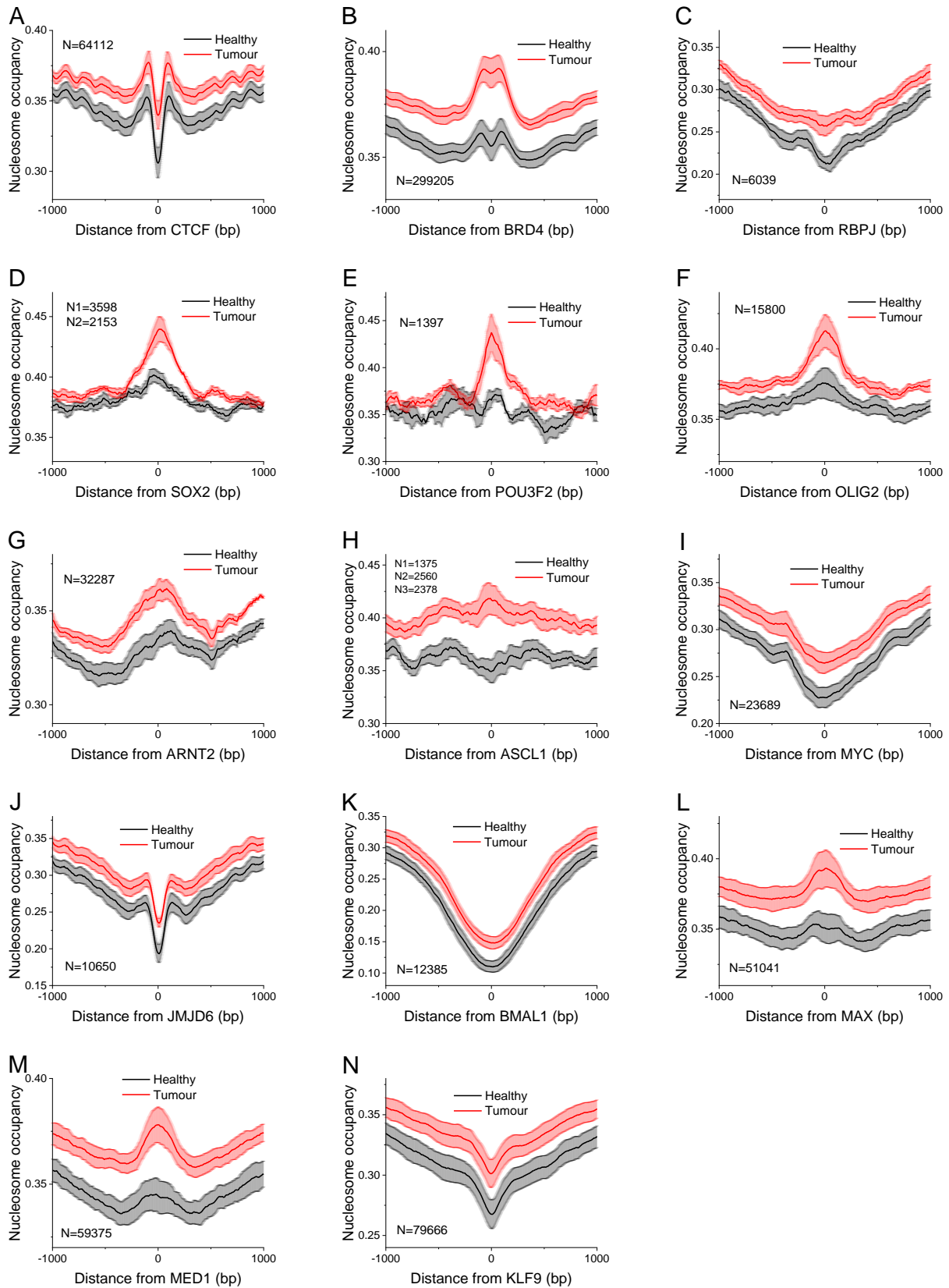


Figure S24. Average cfDNA nucleosome occupancy profile around 16 transcription factor motifs for 120-180 base pair nucleosomal DNA fragment sizes averaged over four healthy people and four GBM patients from Song et al. (<https://www.nature.com/articles/cr2017106>).

Averaged profiles for healthy (black) and GBM (red). Lighter areas correspond to the standard errors of averaging. The number of regions (N) for each TF file is indicated on the graph.

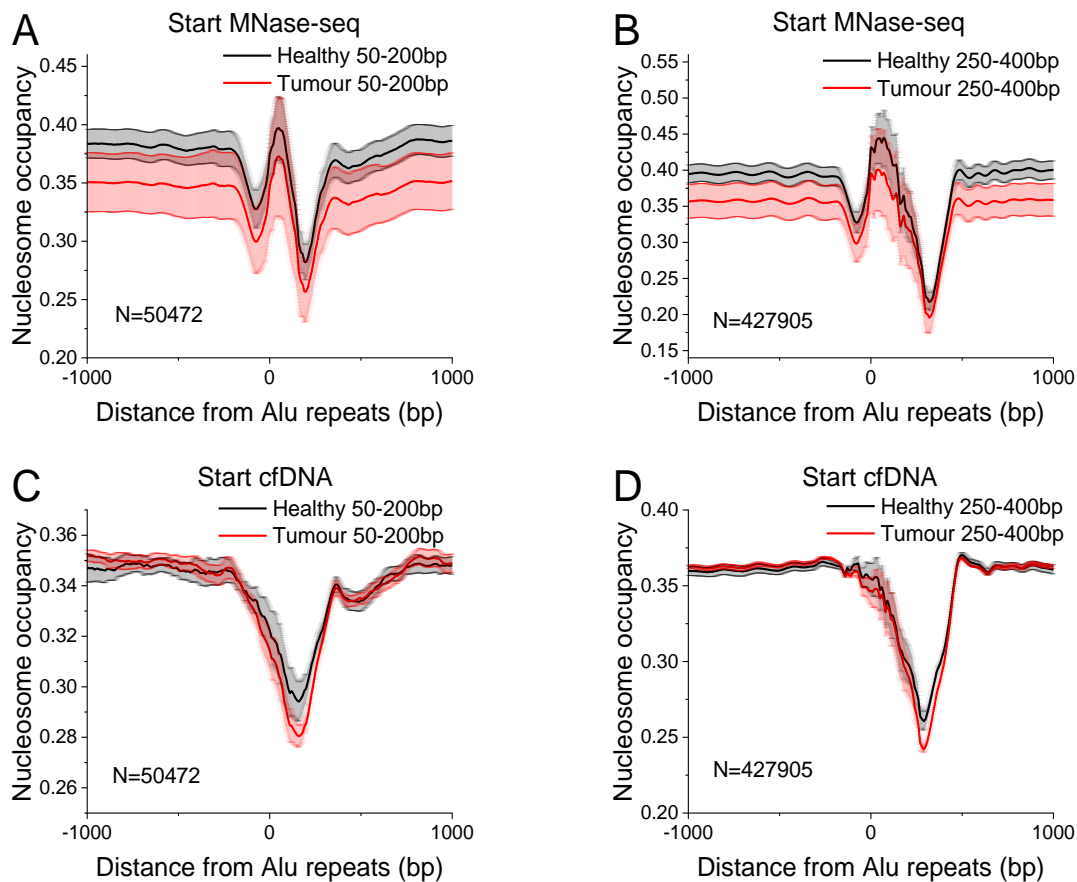


Figure S25. Average genome-wide nucleosome occupancy around ALU repeat elements. Nucleosome occupancy around ALU repeats averaged over the 4 GBM patients healthy tissue MNase (black) and GBM tissue MNase (red) from the experiments performed in Essex separated for ALU sizes (A) 50-200bp and (B) 250-400bp. Nucleosome occupancy around ALU repeats averaged over the 4 samples of cfDNA from healthy people (black) and 4 samples of cfDNA from GBM patients (red) from Song et al. (<https://www.nature.com/articles/cr2017106>) for ALU sizes (C) 50-200bp and (D) 250-400bp. Calculations for Panels A-D were made by aligning the signal around the left end (start) of ALU repeats. Grey areas correspond to the standard errors of averaging.

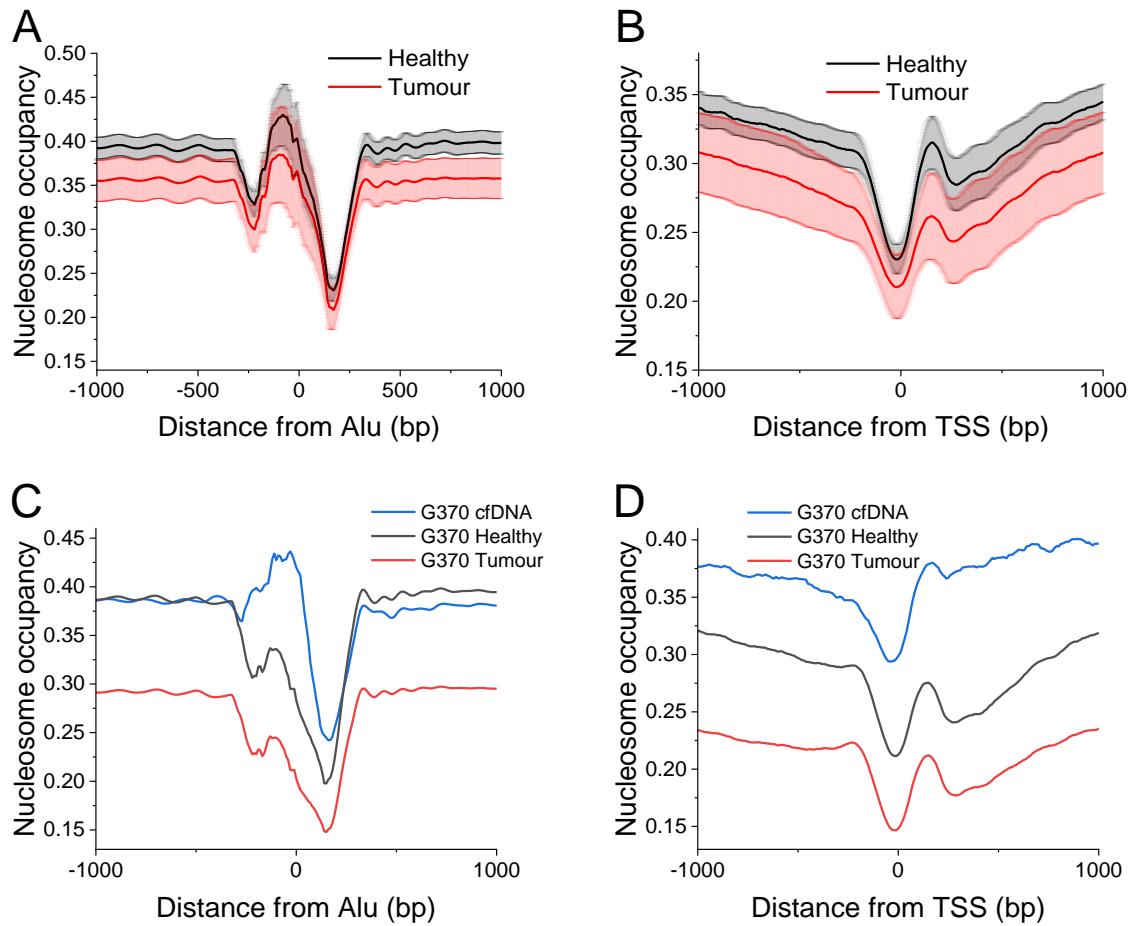


Figure S26. Average genome-wide nucleosome occupancy around Alu elements and Transcription Start Sites. (A) Nucleosome occupancy around Alu elements and (B) Transcription Starting Sites (TSS) averaged over the 4 GBM patients healthy tissue MNase (black) and GBM tissue MNase (red) from the experiments performed in Essex. Grey areas correspond to the standard errors of averaging. (C) Nucleosome occupancy around Alu elements and (B) Transcription Starting (TSS) for patient G370 cfDNA (blue), healthy tissue MNase (black) and GBM tissue MNase (red) from the experiments performed in Essex. Left panels: ALU repeats, right panels: TSS.

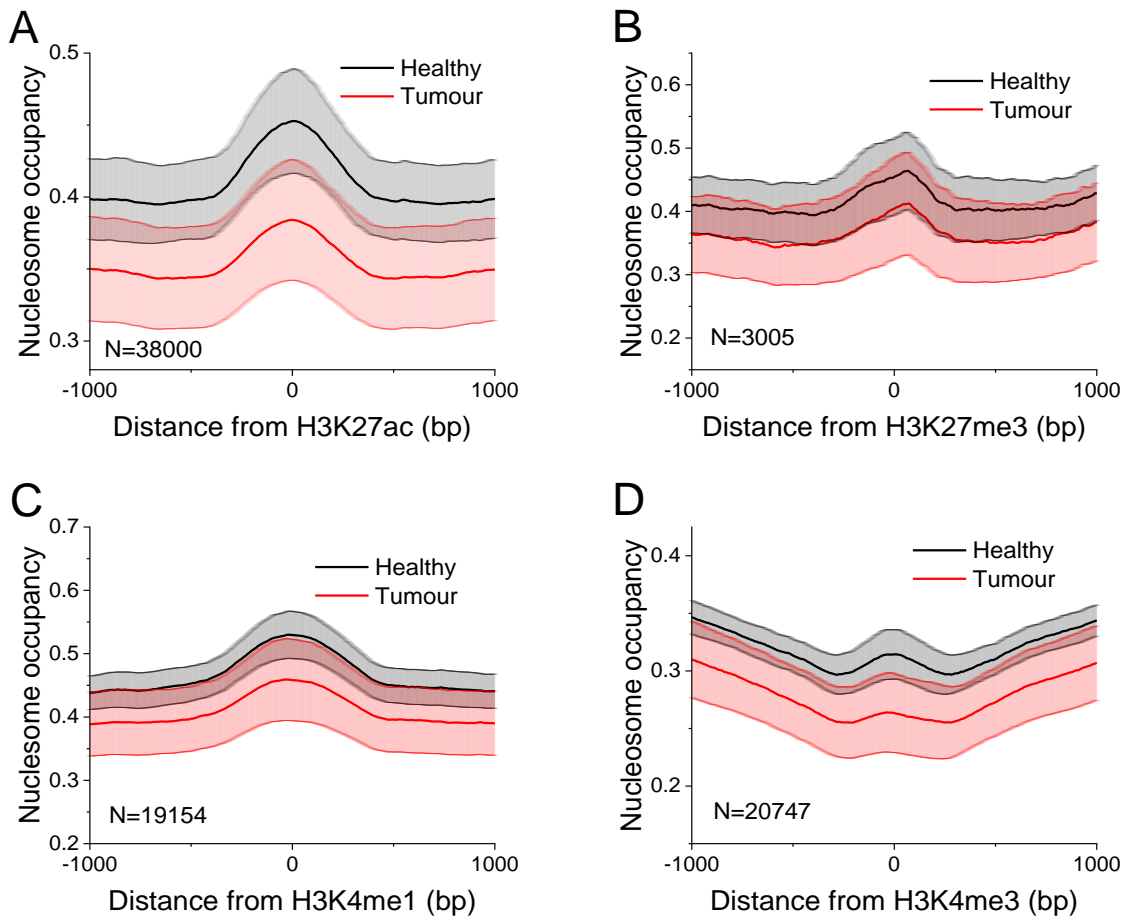


Figure S27. Average genome-wide nucleosome occupancy around (A) H3K26ac, (B), H3K27me3, (C) H3K4me1 and (D) H3K4me3 binding sites found in IDH GBM subtype averaged over the 4 GBM patients from the experiments performed in Essex. Grey areas correspond to the standard errors of averaging.

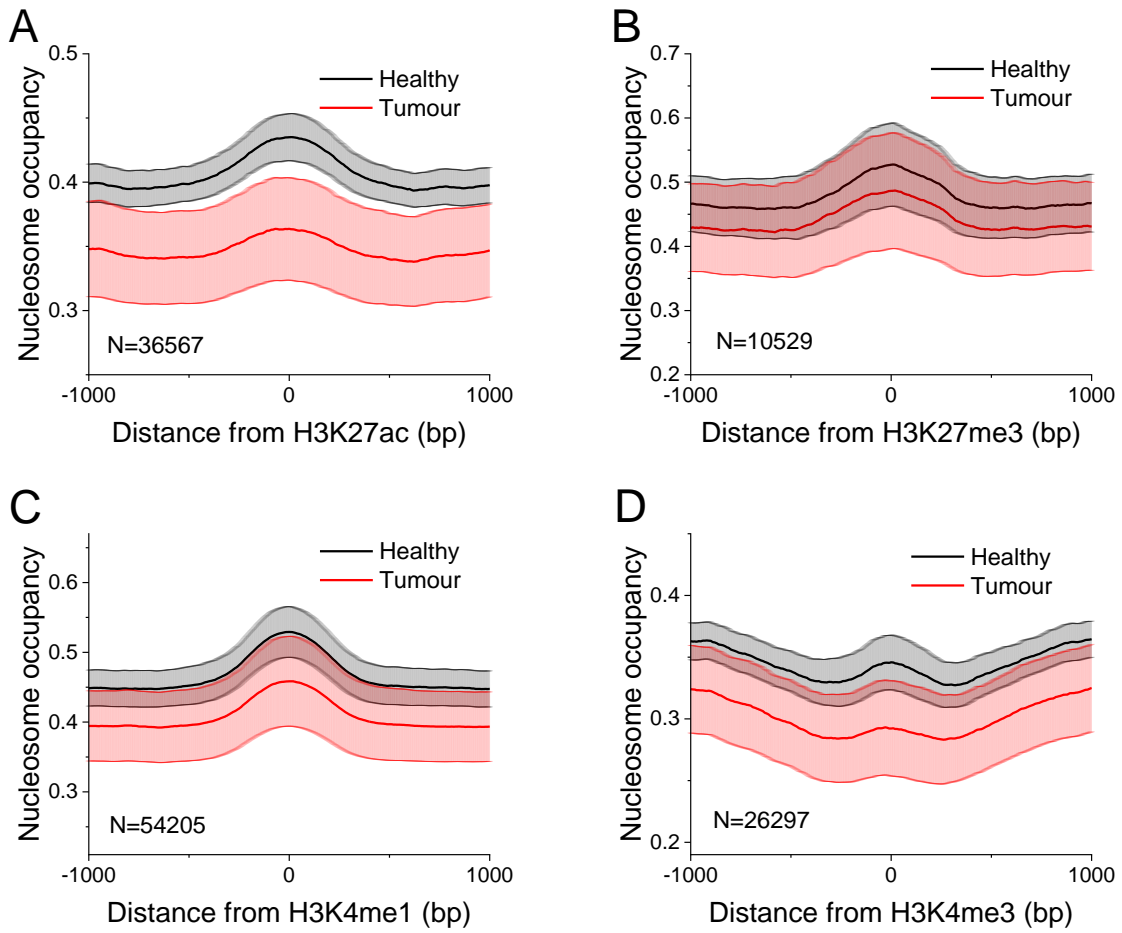


Figure S28. Average genome-wide nucleosome occupancy around (A) H3K26ac, (B), H3K27me3, (C) H3K4me1 and (D) H3K4me3 binding sites found in RTKI GBM subtype averaged over the 4 GBM patients from the experiments performed in Essex. Grey areas correspond to the standard errors of averaging.

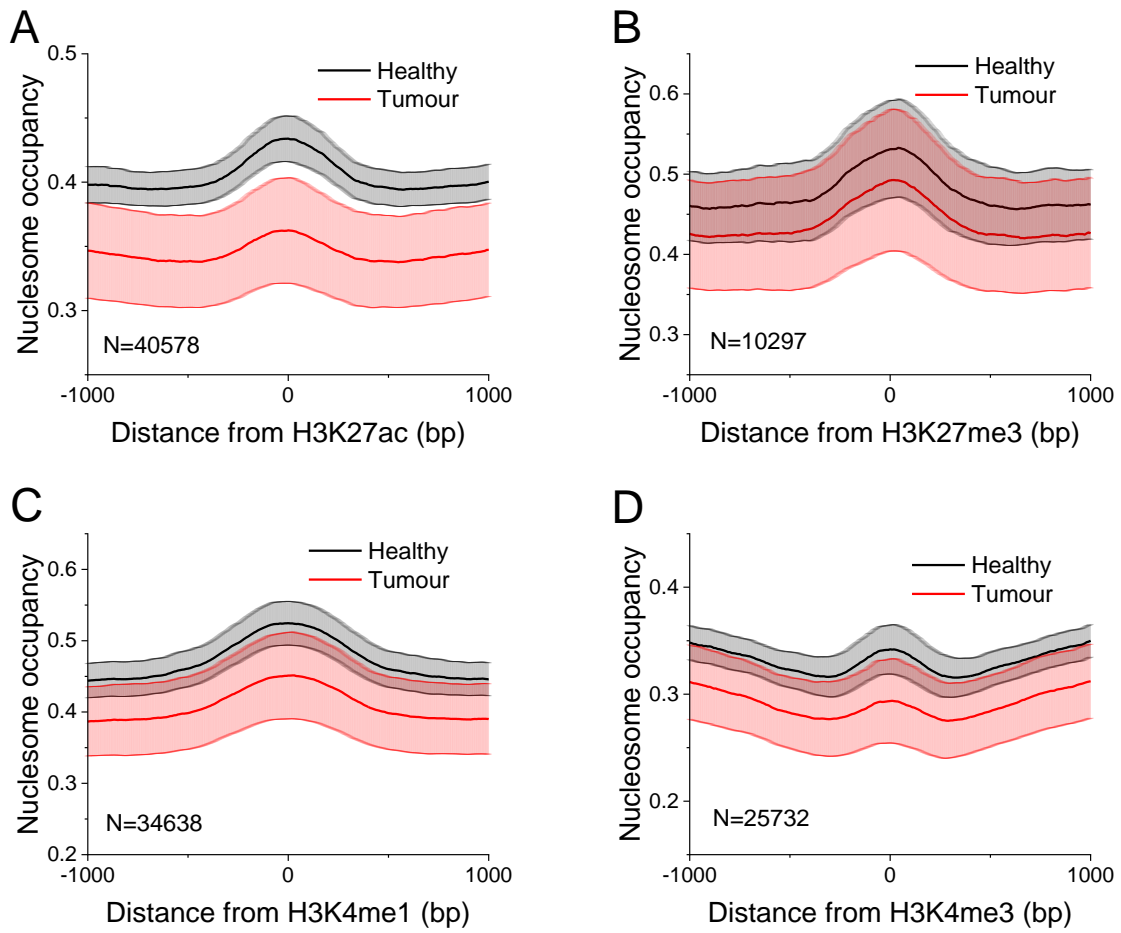


Figure S29. Average genome-wide nucleosome occupancy around (A) H3K26ac, (B), H3K27me3, (C) H3K4me1 and (D) H3K4me3 binding sites found in RTKII GBM subtype averaged over the 4 GBM patients from the experiments performed in Essex. Grey areas correspond to the standard errors of averaging.

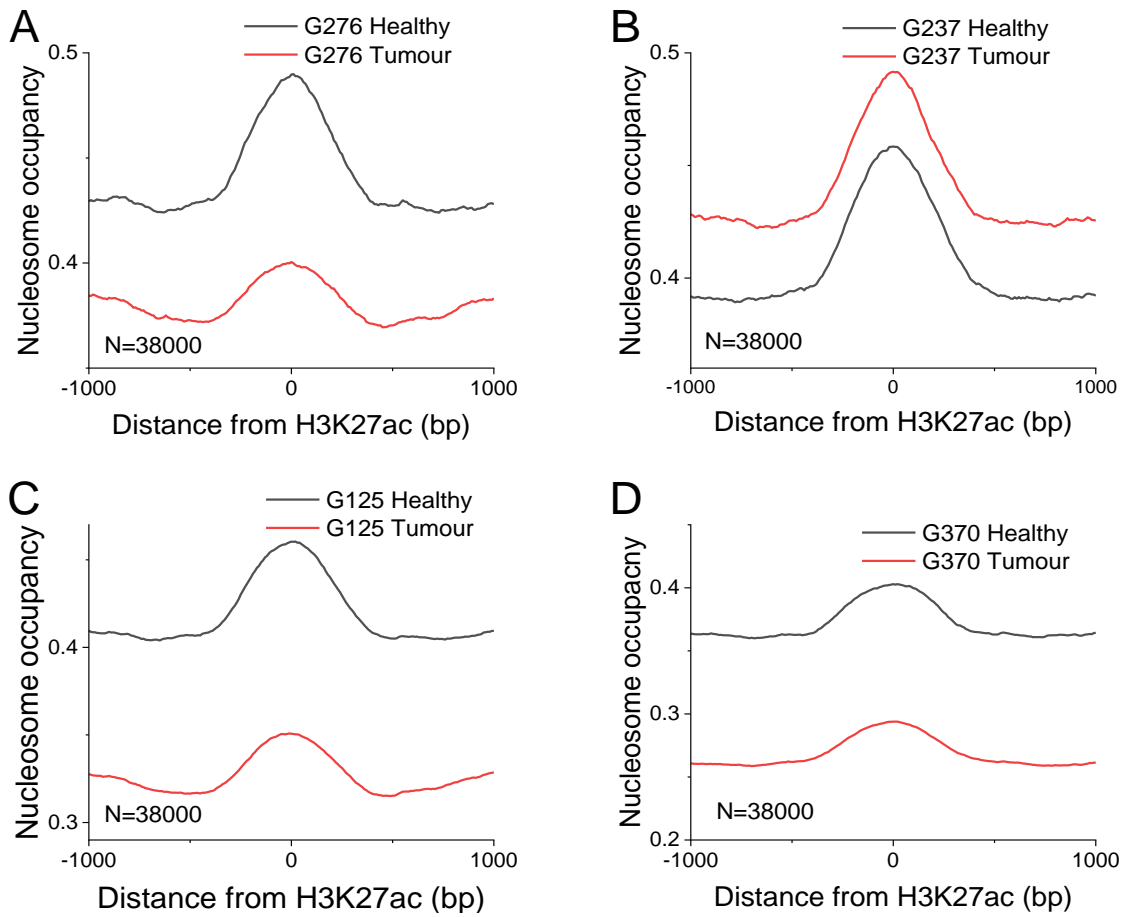


Figure S30. Average genome-wide nucleosome occupancy around H3K27ac found in IDH GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

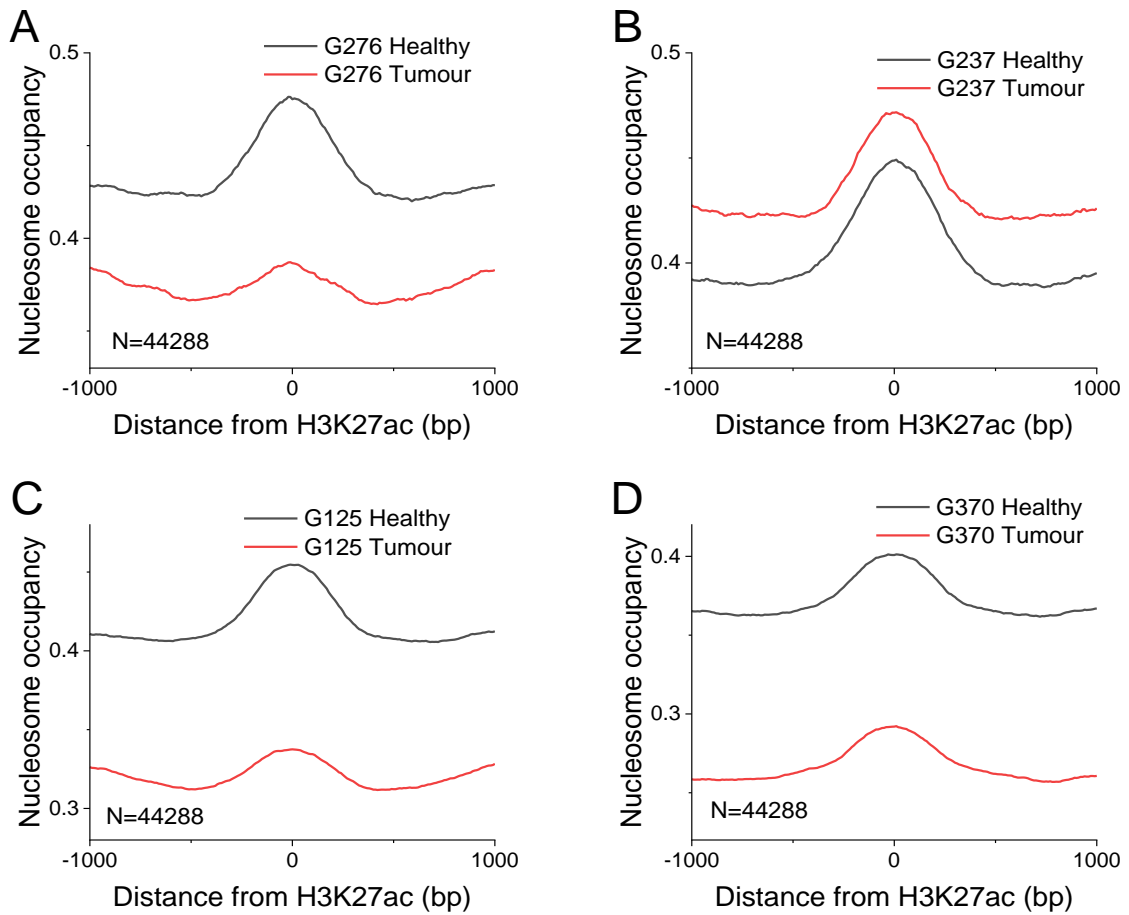


Figure S31. Average genome-wide nucleosome occupancy around H3K27ac found in MES GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.



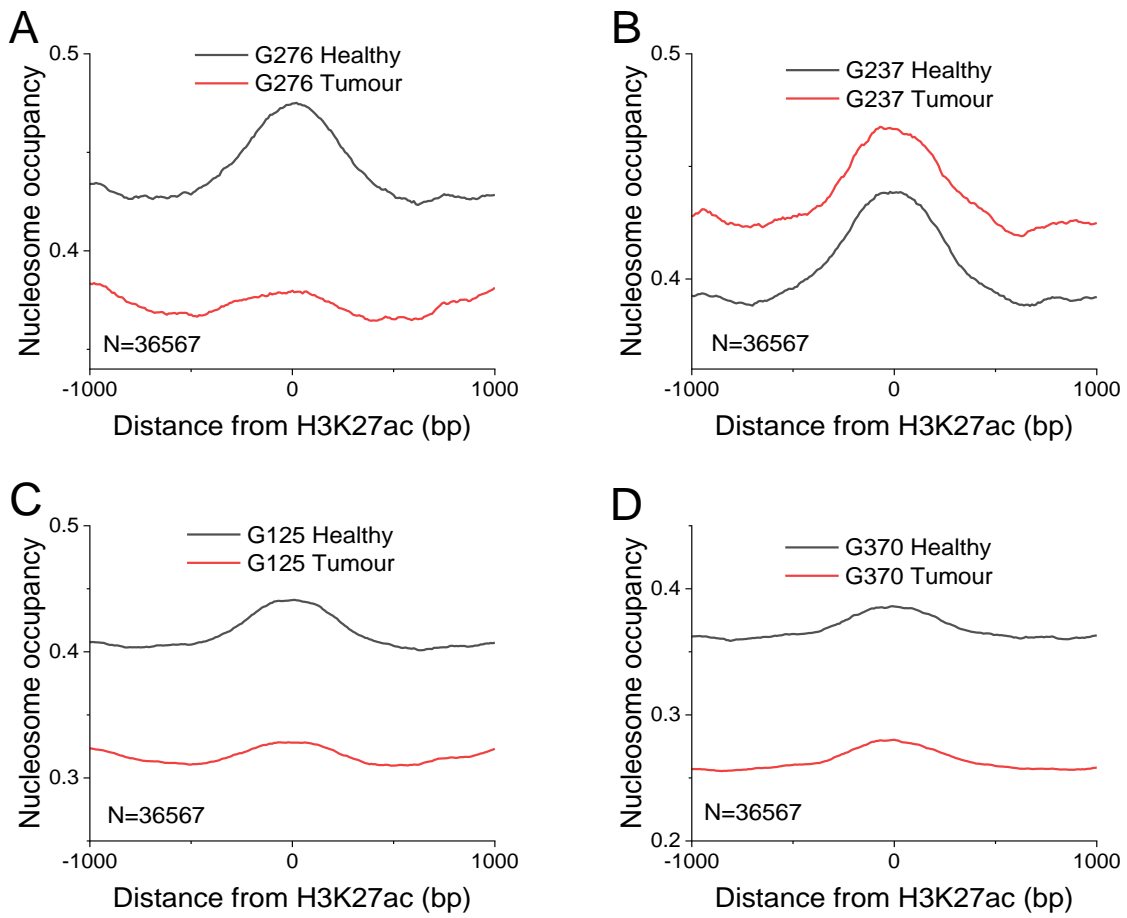


Figure S32. Average genome-wide nucleosome occupancy around H3K27ac found in RTKI GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

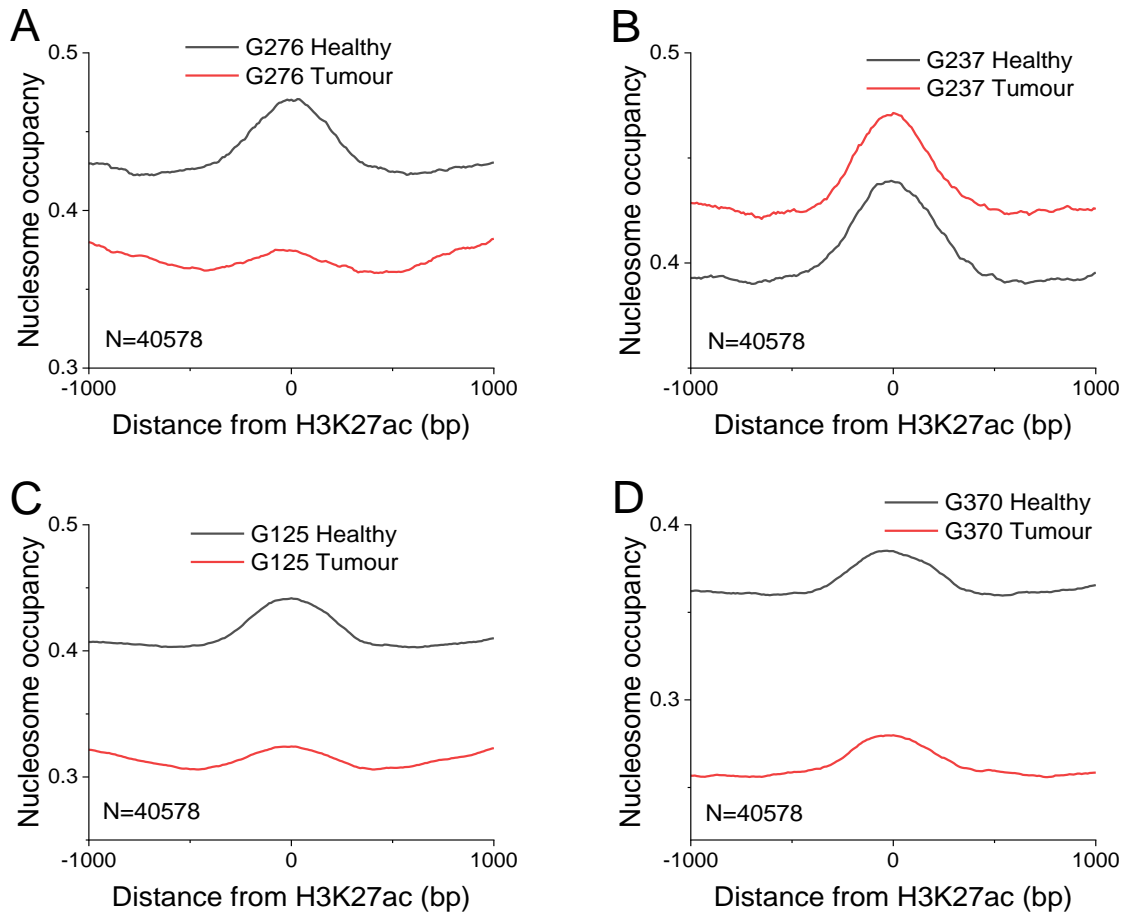


Figure S33. Average genome-wide nucleosome occupancy around H3K27ac found in RTKII GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

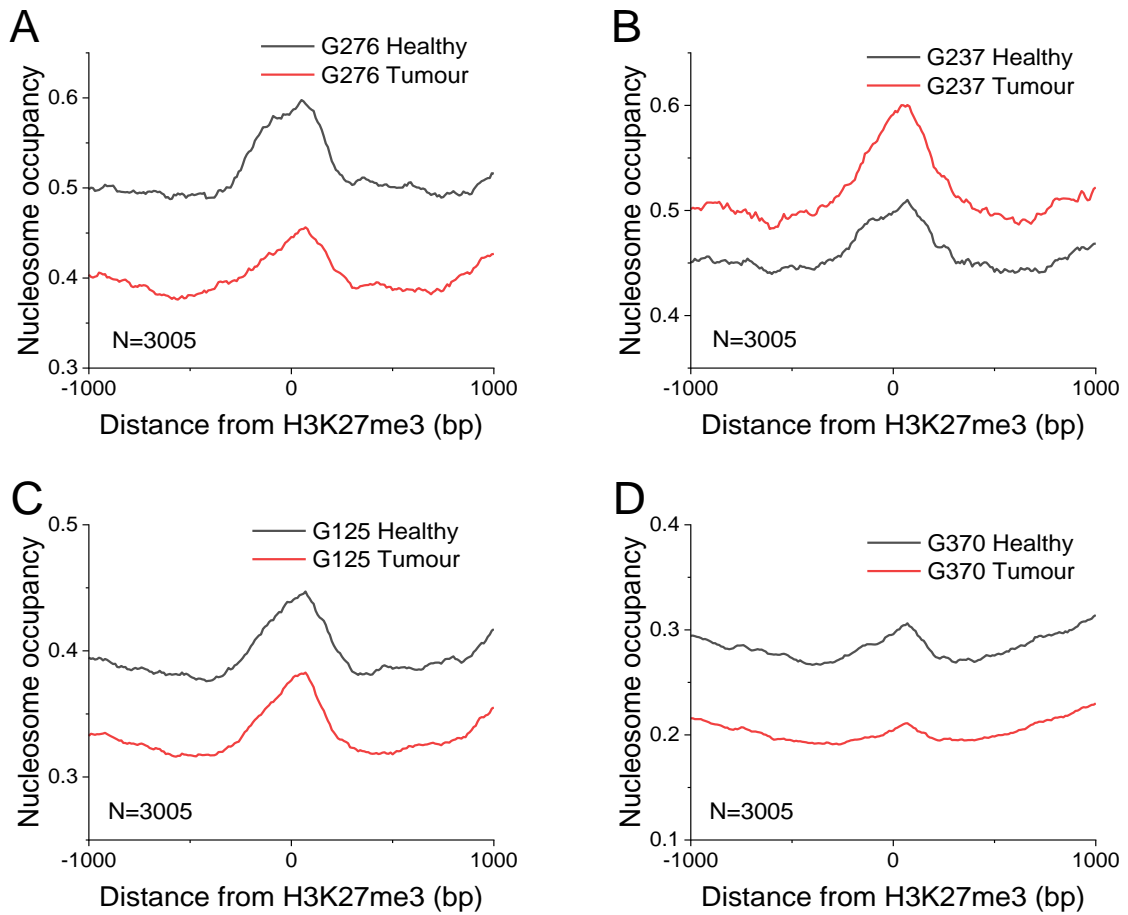


Figure S34. Average genome-wide nucleosome occupancy around H3K27me3 found in IDH GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

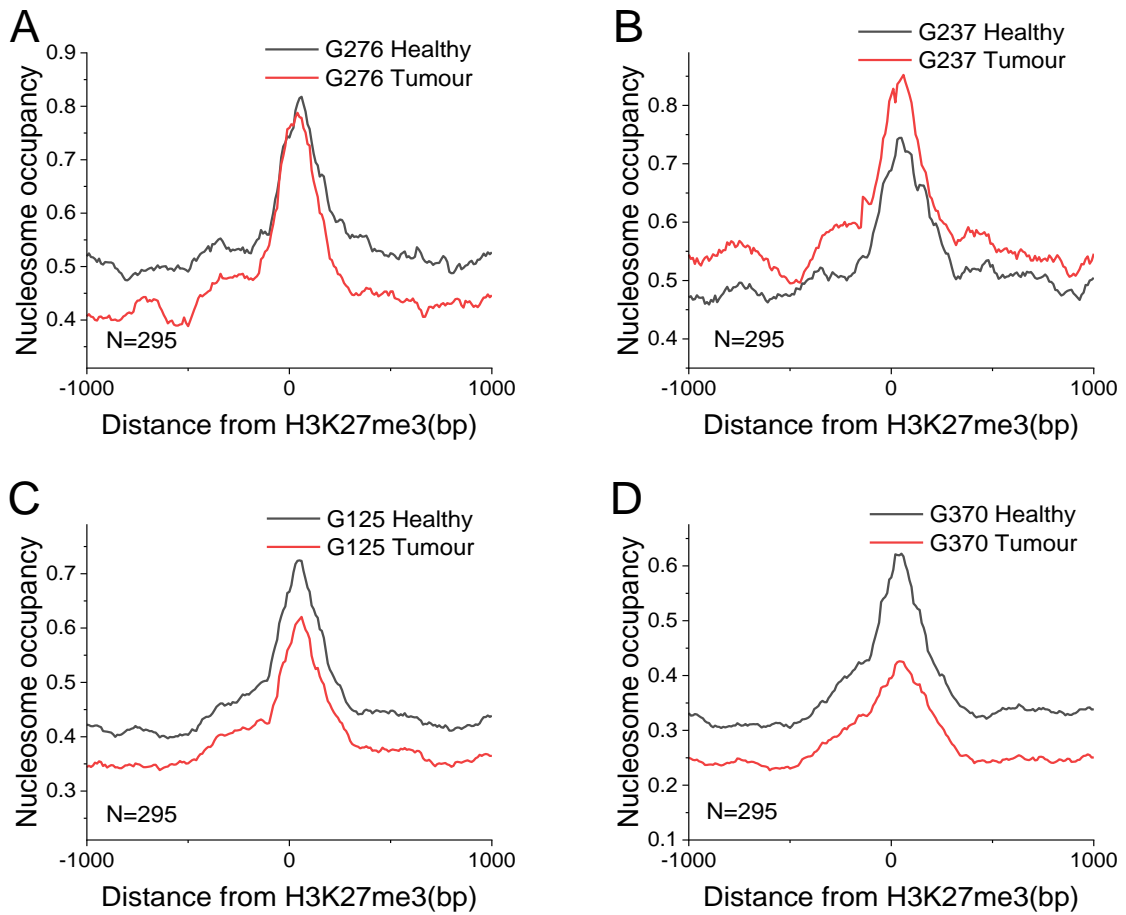


Figure S35. Average genome-wide nucleosome occupancy around H3K27me3 found in MES GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

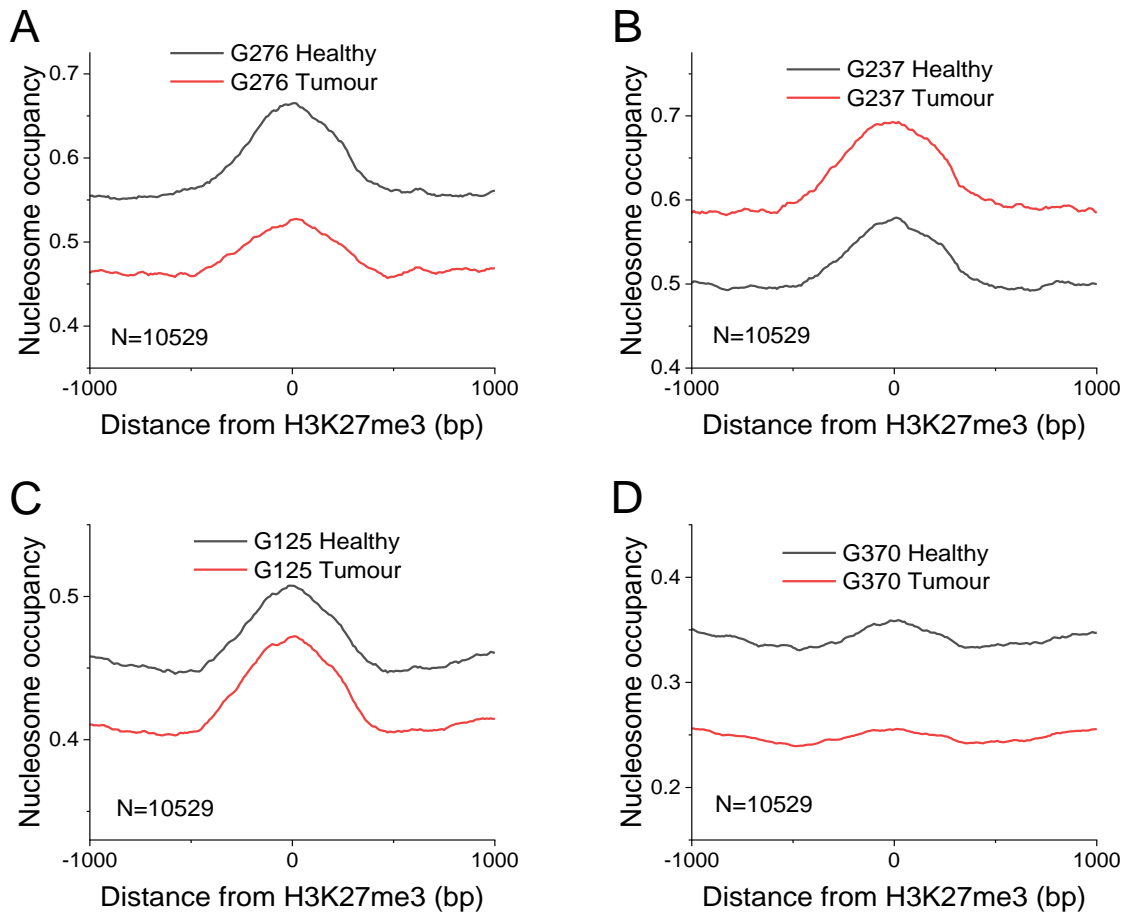


Figure S36. Average genome-wide nucleosome occupancy around H3K27me3 found in RTK1 GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

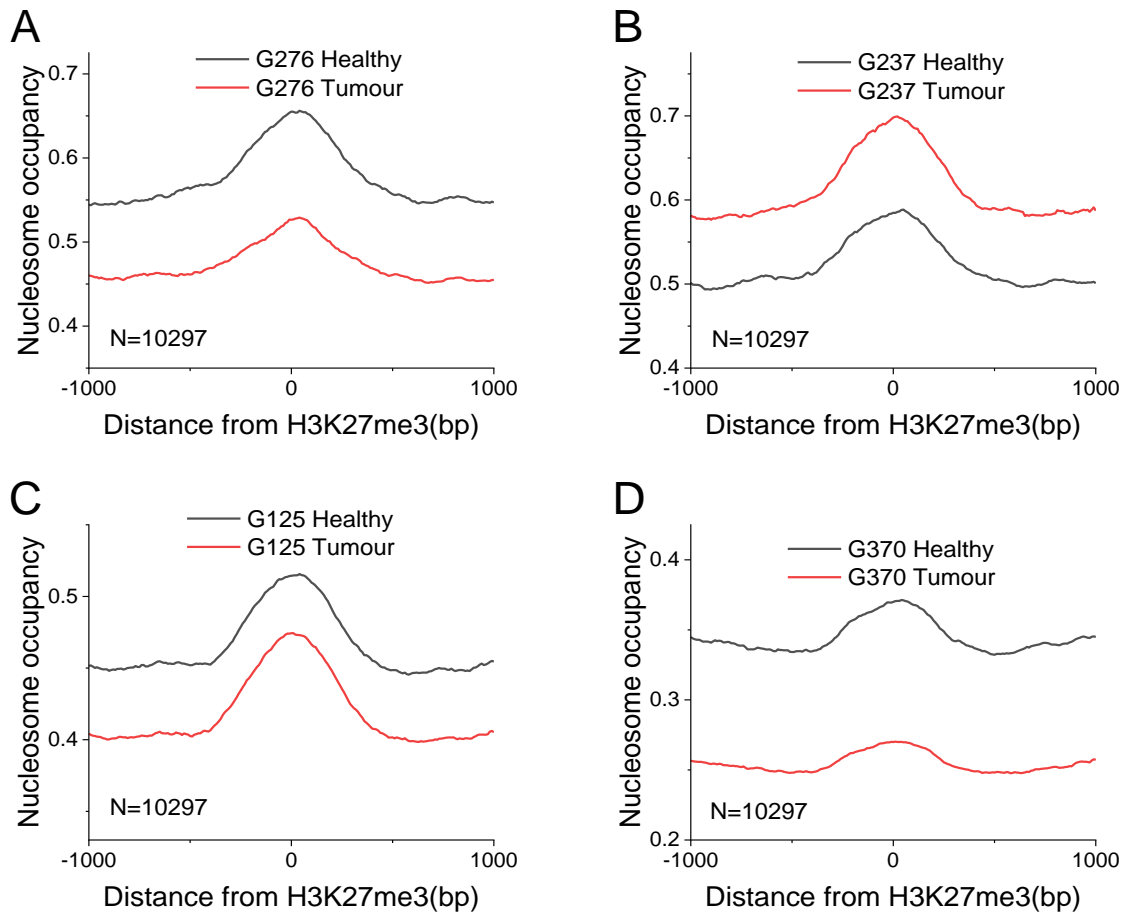


Figure S37. Average genome-wide nucleosome occupancy around H3K27me3 found in RTKII GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

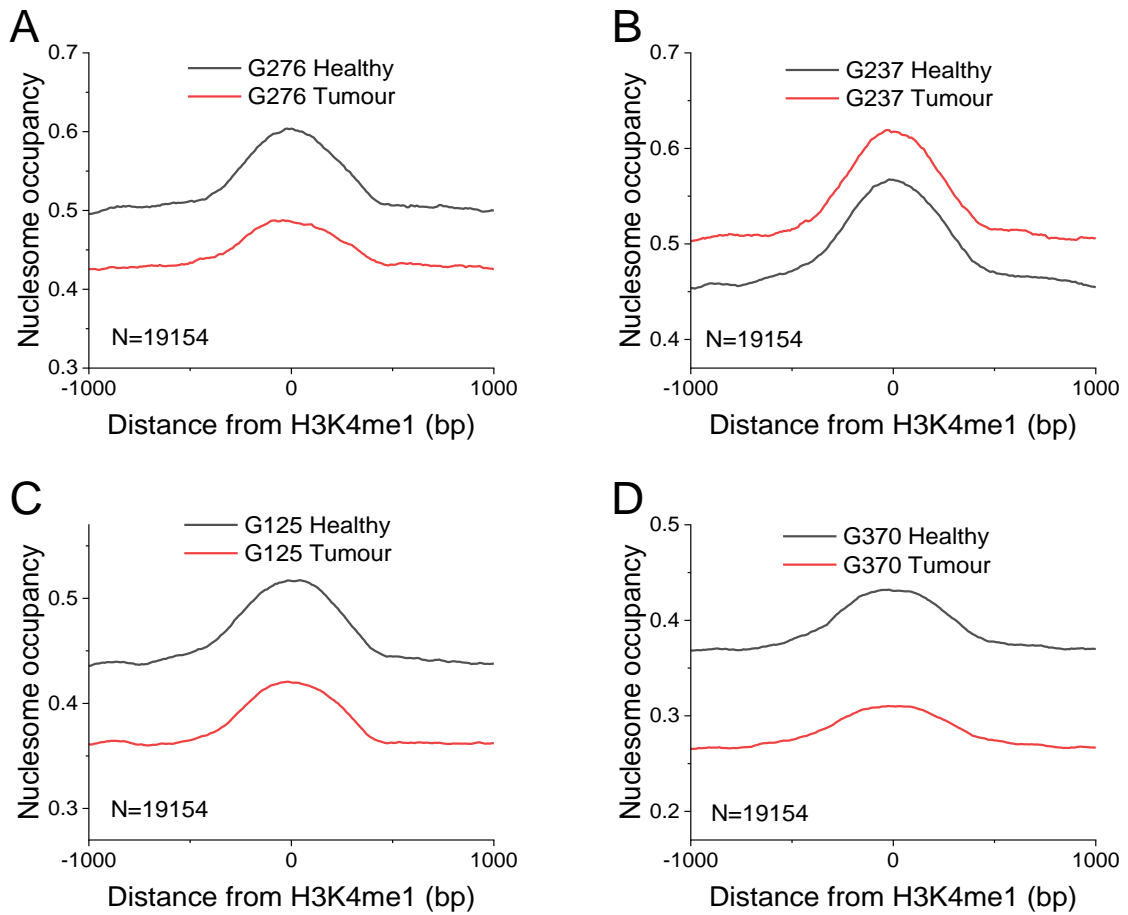


Figure S38. Average genome-wide nucleosome occupancy around H3K4me1 found in IDH GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

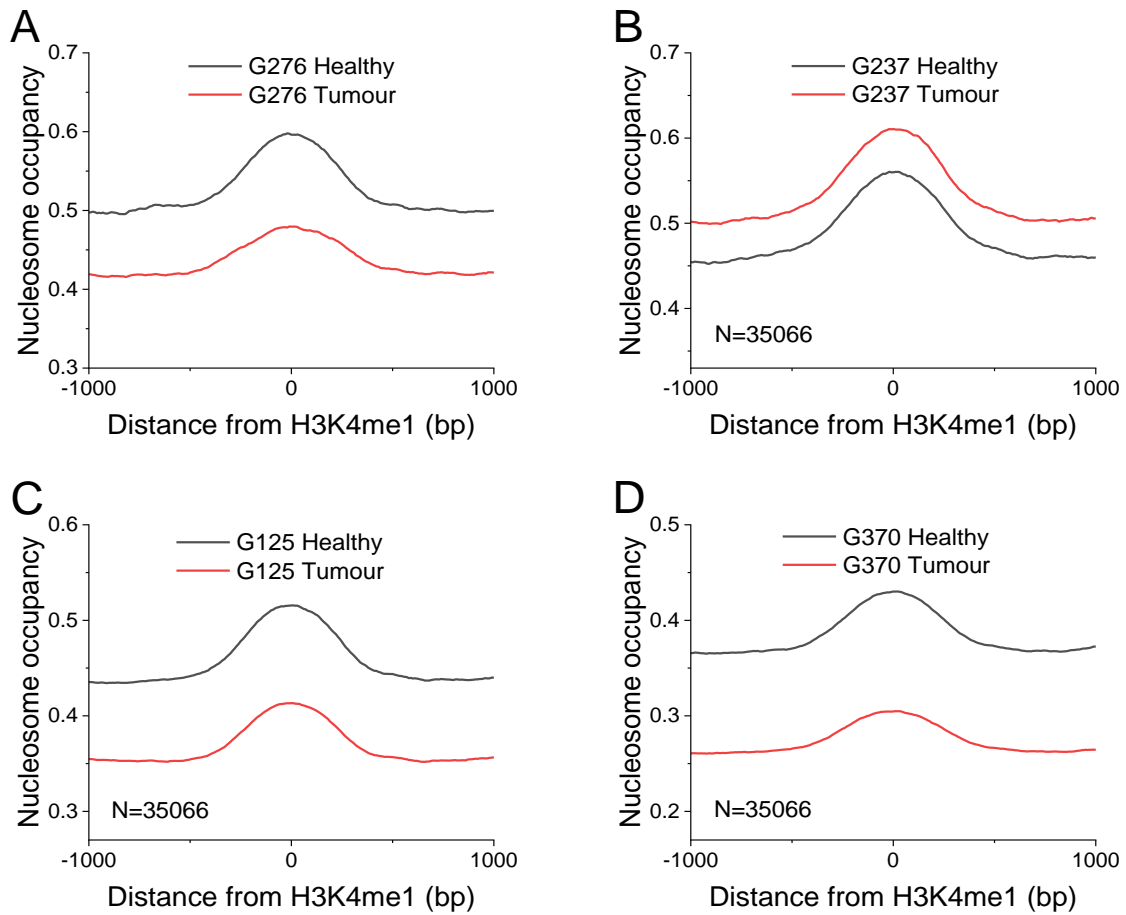


Figure S39. Average genome-wide nucleosome occupancy around H3K4me1 found in MES GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.



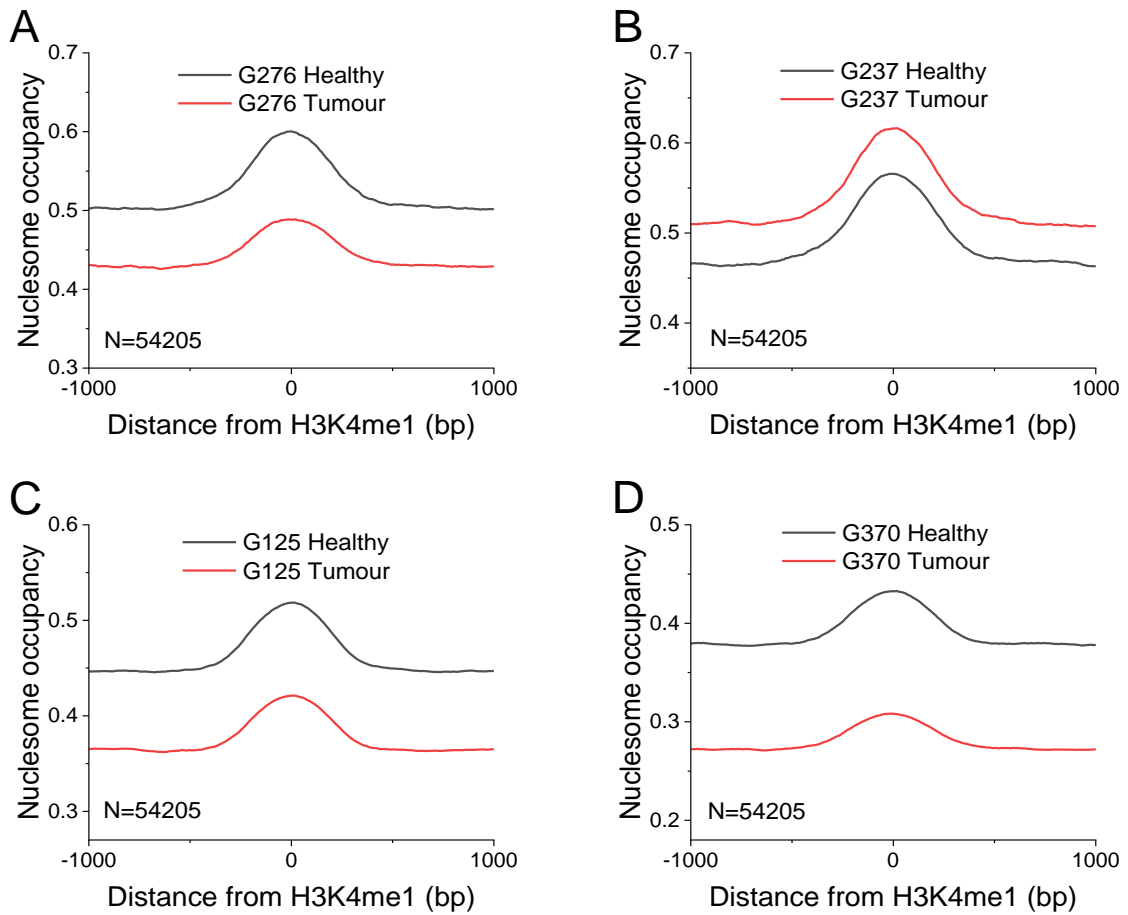


Figure S40. Average genome-wide nucleosome occupancy around H3K4me1 found in RTK1 GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

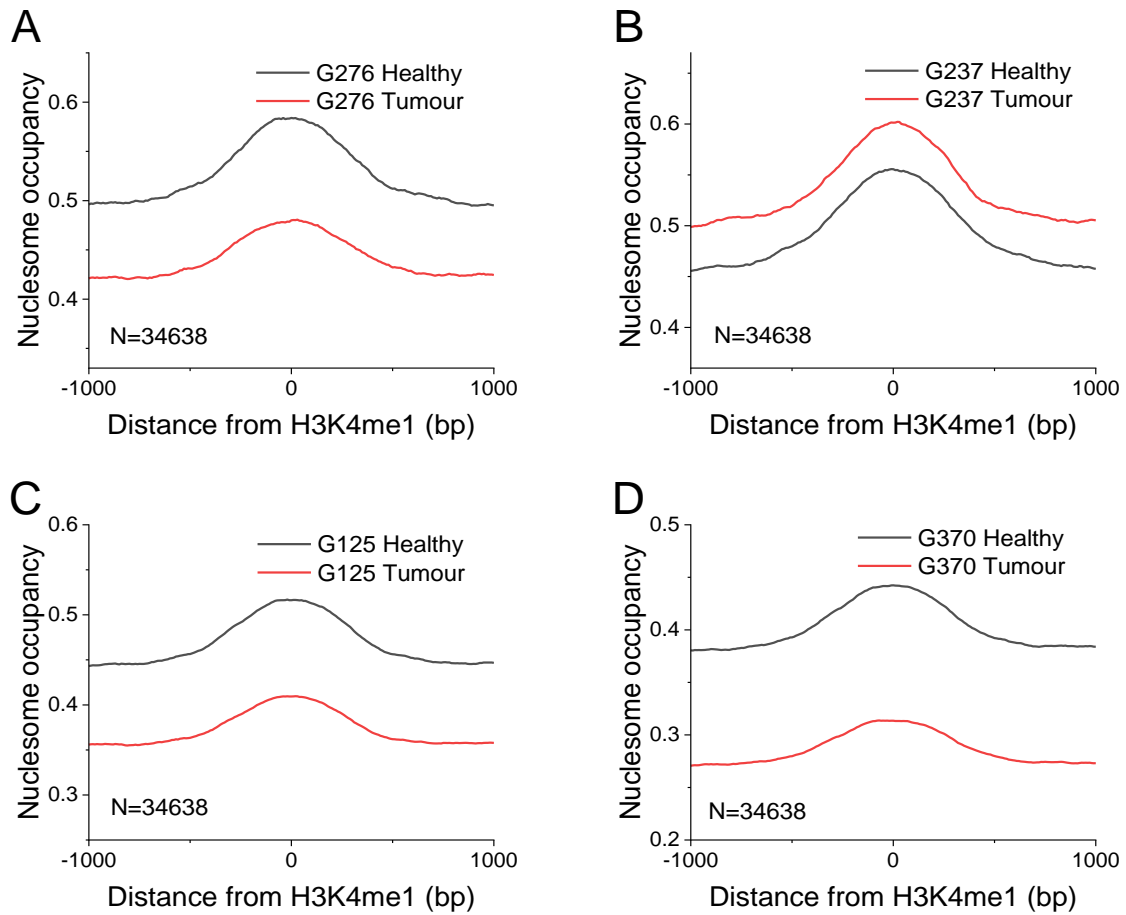


Figure S41. Average genome-wide nucleosome occupancy around H3K4me1 found in RTKII GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

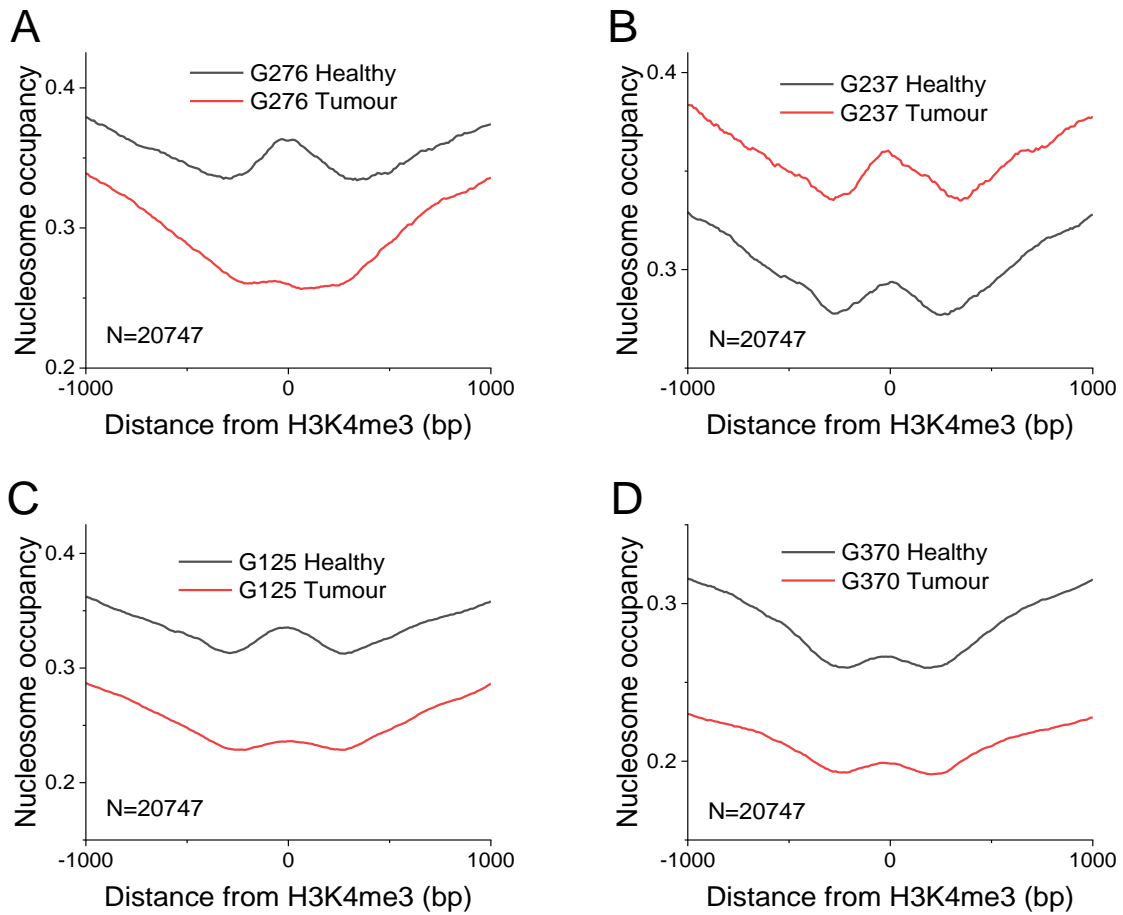


Figure S42. Average genome-wide nucleosome occupancy around H3K4me3 found in IDH GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

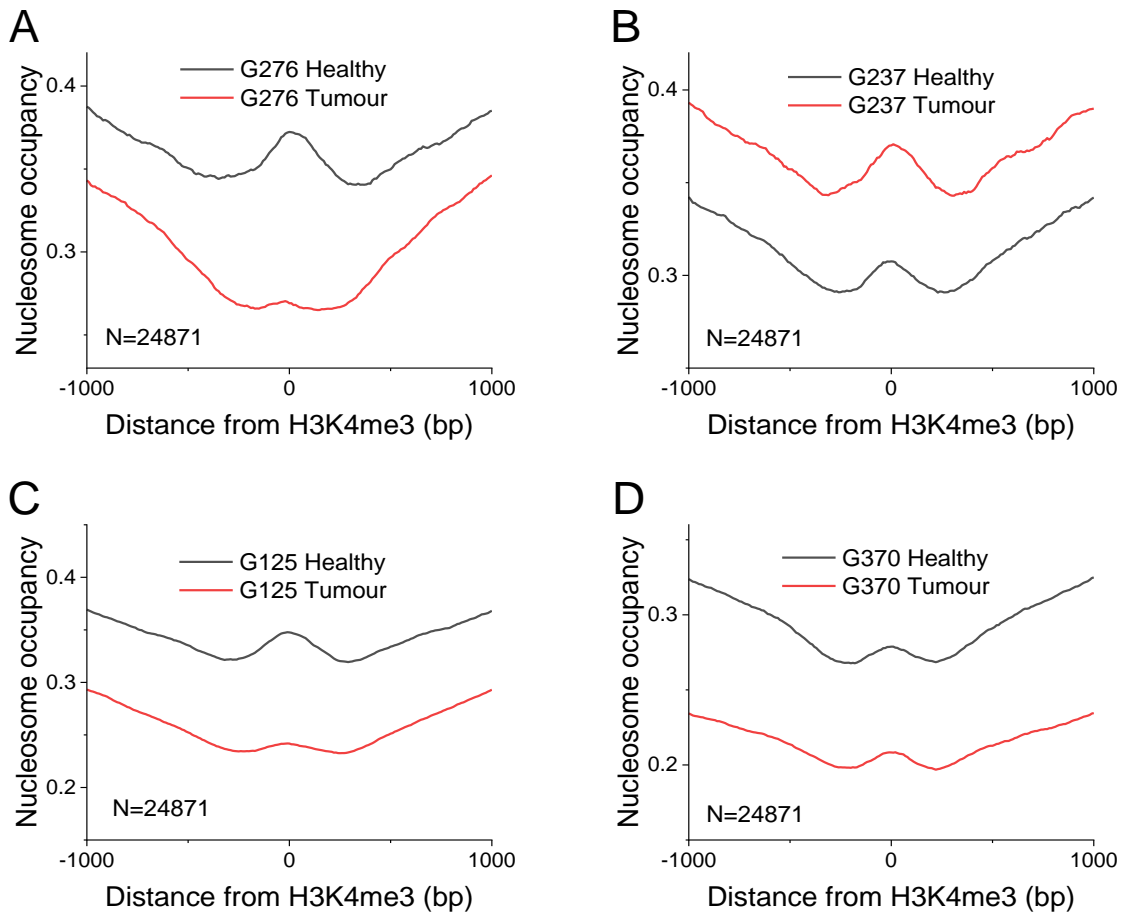


Figure S43. Average genome-wide nucleosome occupancy around H3K4me3 found in MES GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

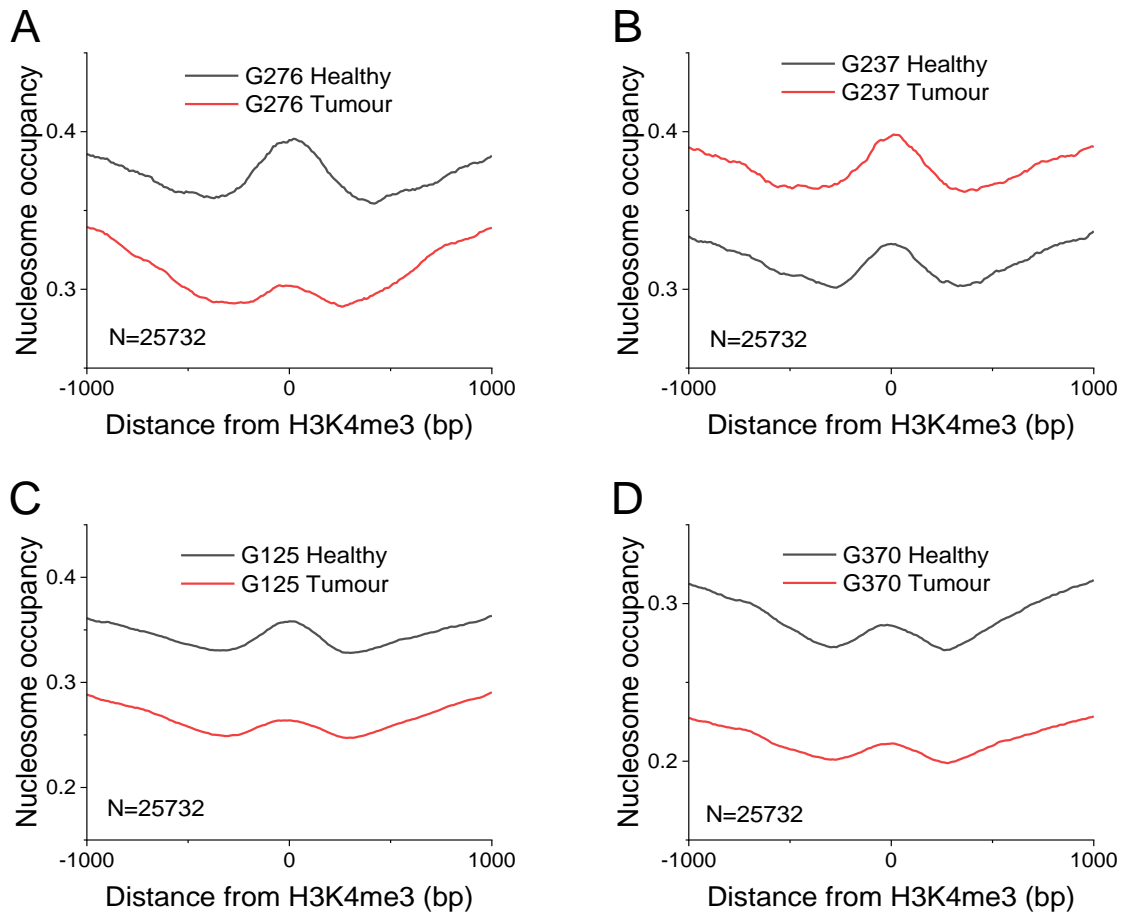


Figure S44. Average genome-wide nucleosome occupancy around H3K4me3 found in RTK1 GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

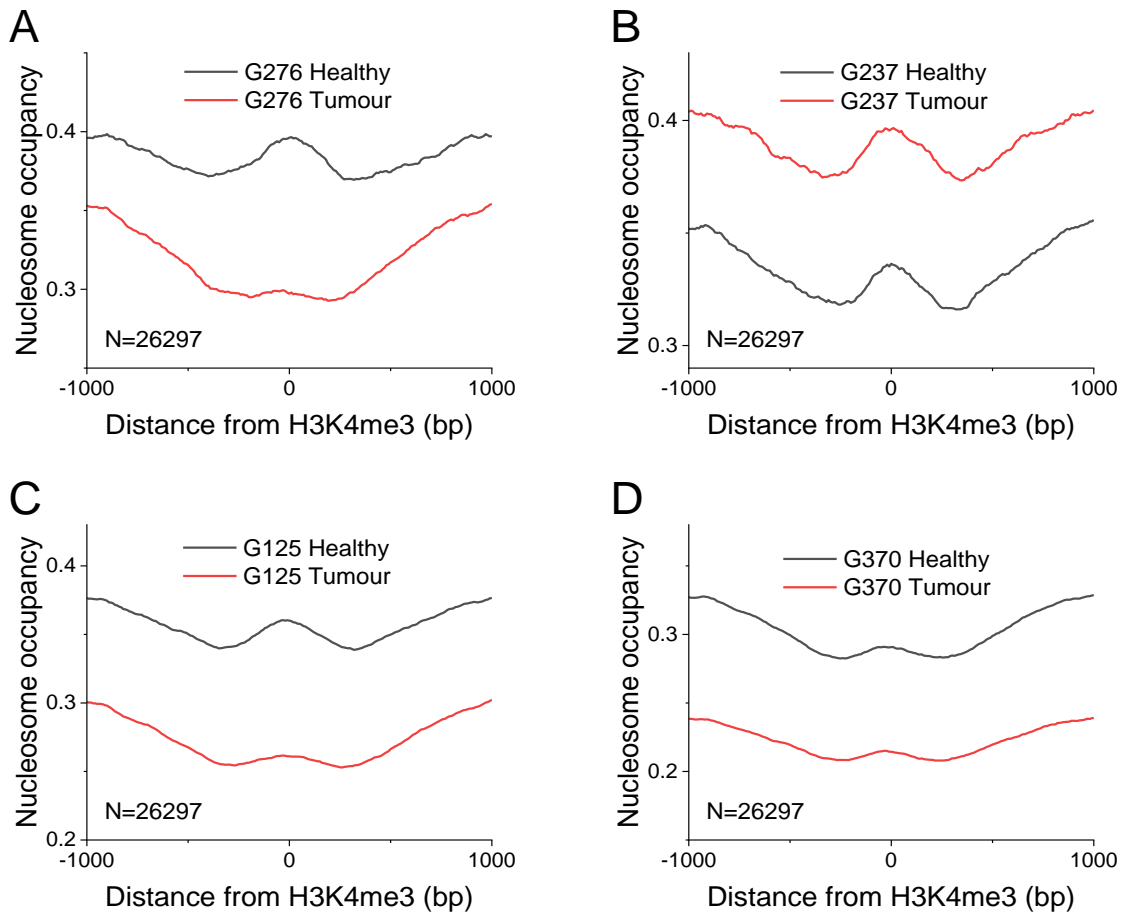


Figure S45. Average genome-wide nucleosome occupancy around H3K4me3 found in RTKII GBM subtype binding sites for each of GBM patients from the experiments performed in Essex. Panel A: Patient G276; Panel B: Patient G237; Panel C: Patient G125; Panel D: Patient G370.

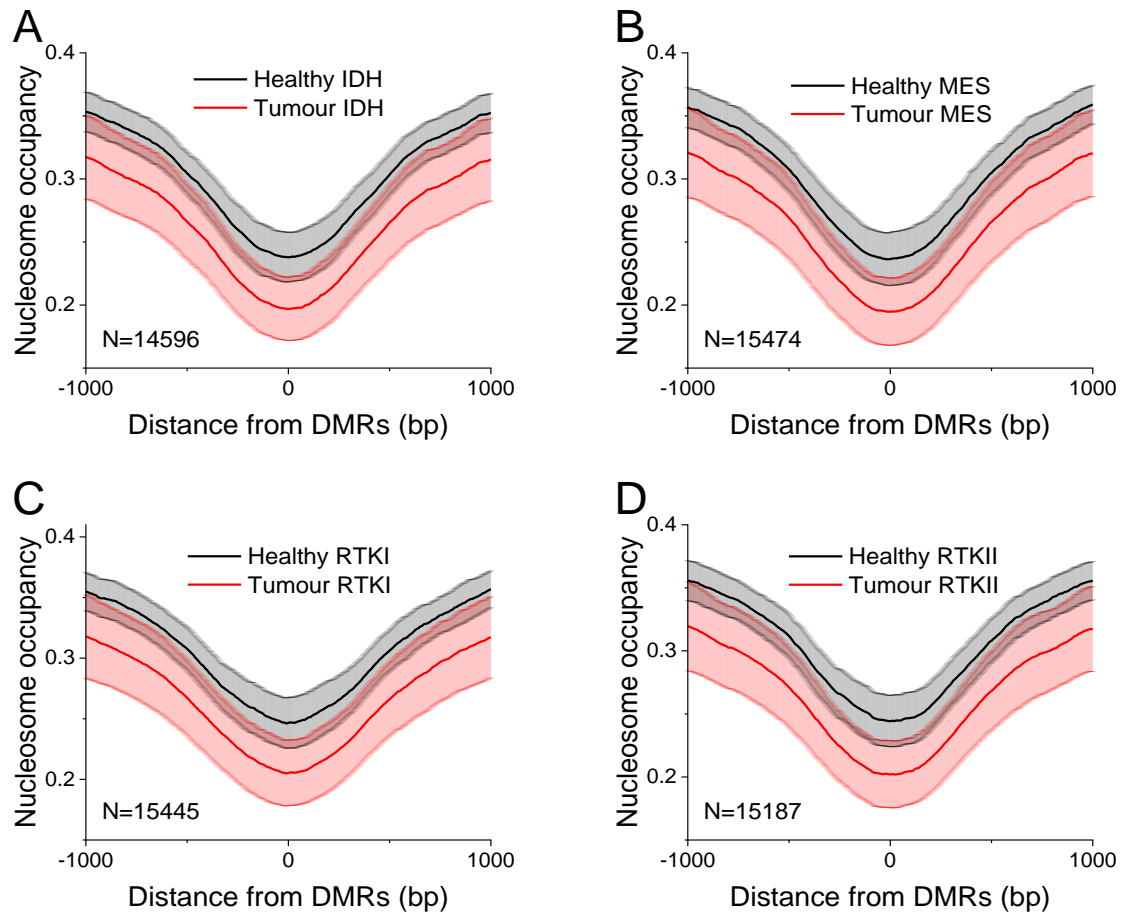


Figure S46. Average genome-wide nucleosome occupancy around DNA methylation regions (DMRs) found in IDH, MES, RTKI and RKTII GBM subtypes averaged over the 4 GBM patients from the experiments performed in Essex. Grey areas correspond to the standard errors of averaging.

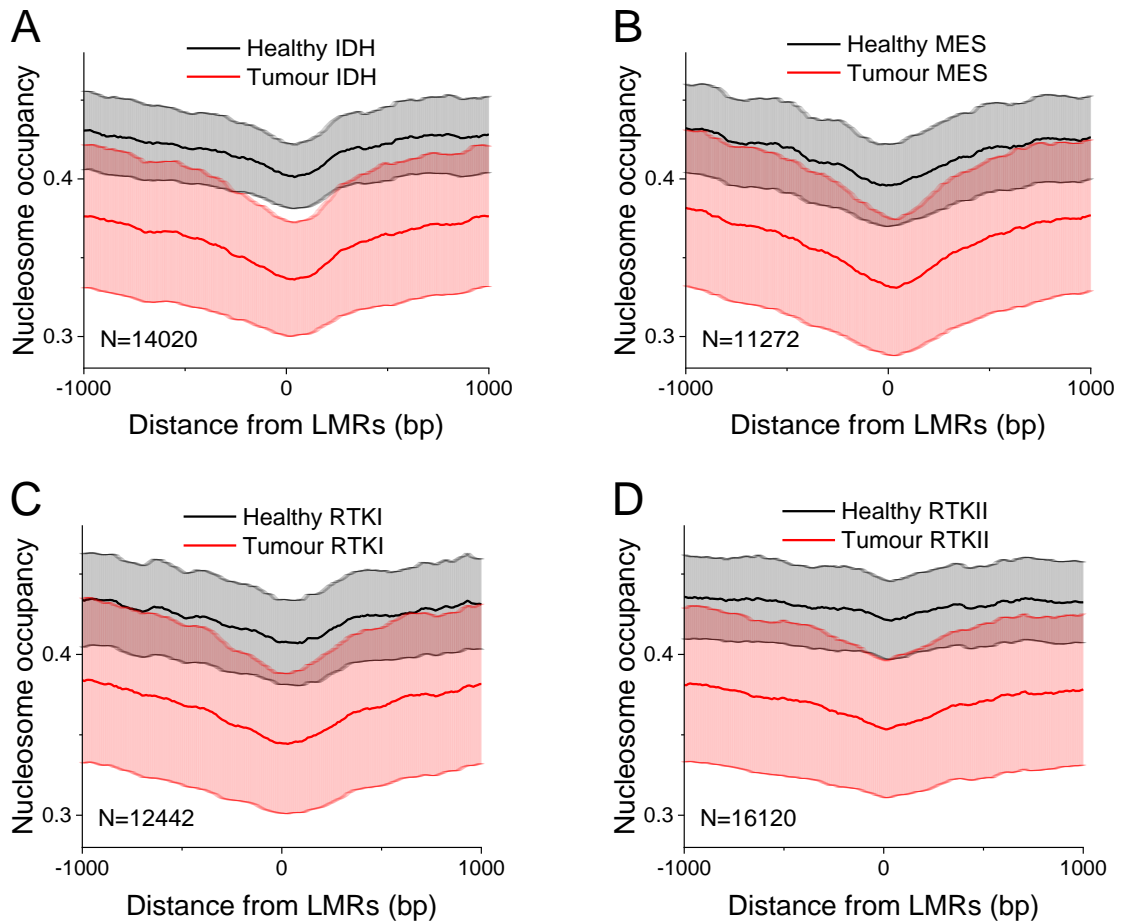


Figure S47. Average genome-wide nucleosome occupancy around low-methylated regions (LMRs) found in IDH, MES, RTKI and RKTII GBM subtypes averaged over the 4 GBM patients from the experiments performed in Essex. Grey areas correspond to the standard errors of averaging.



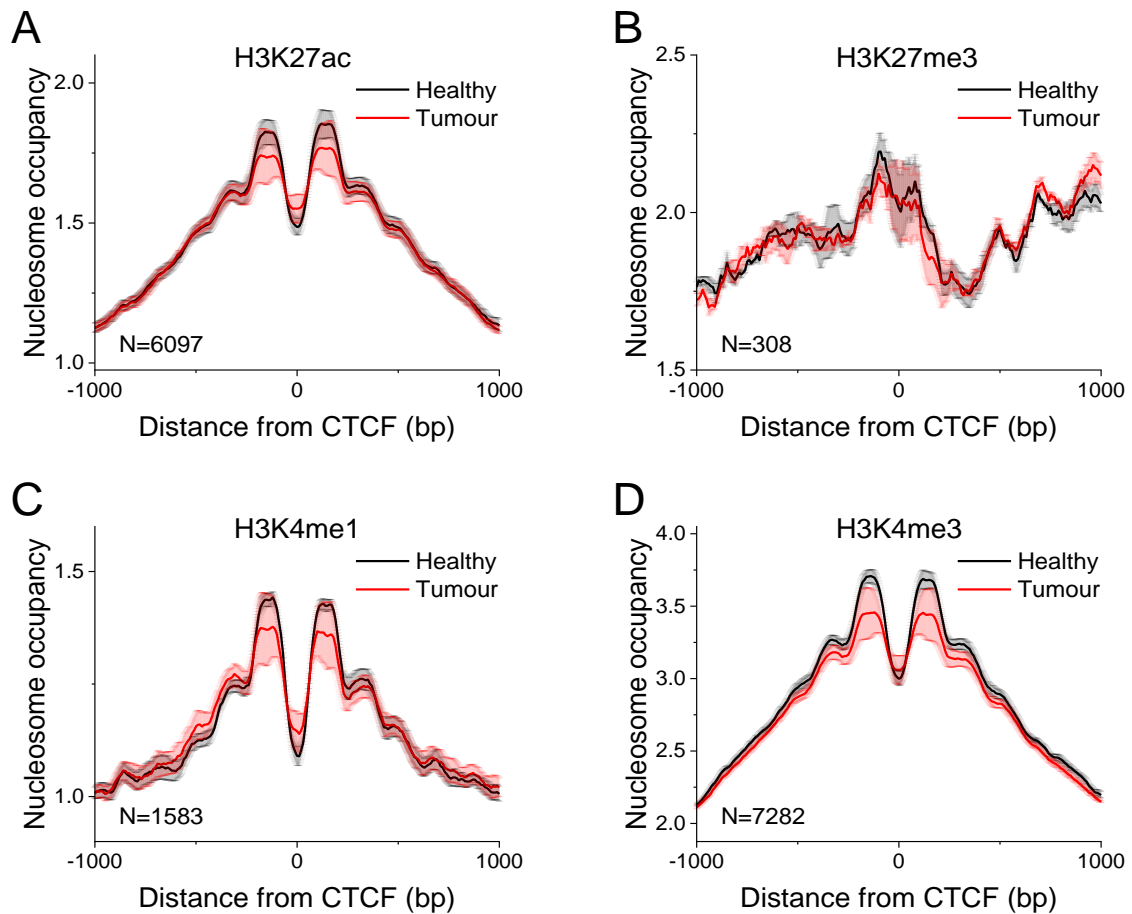


Figure S48. Histone mark enrichment profiles around CTCF in tumour for healthy and GBM tissue from the 4 GBM patients. (A) Histone mark H3K4me1, (B) Histone mark H3K4me3, (C) Histone mark H3K27ac, (D) Histone mark H3K27me3. Averaged profiles for healthy (black) and GBM (red). Lighter areas correspond to the standard errors of averaging. The number of regions (N) for each TF file is indicated on the graph.

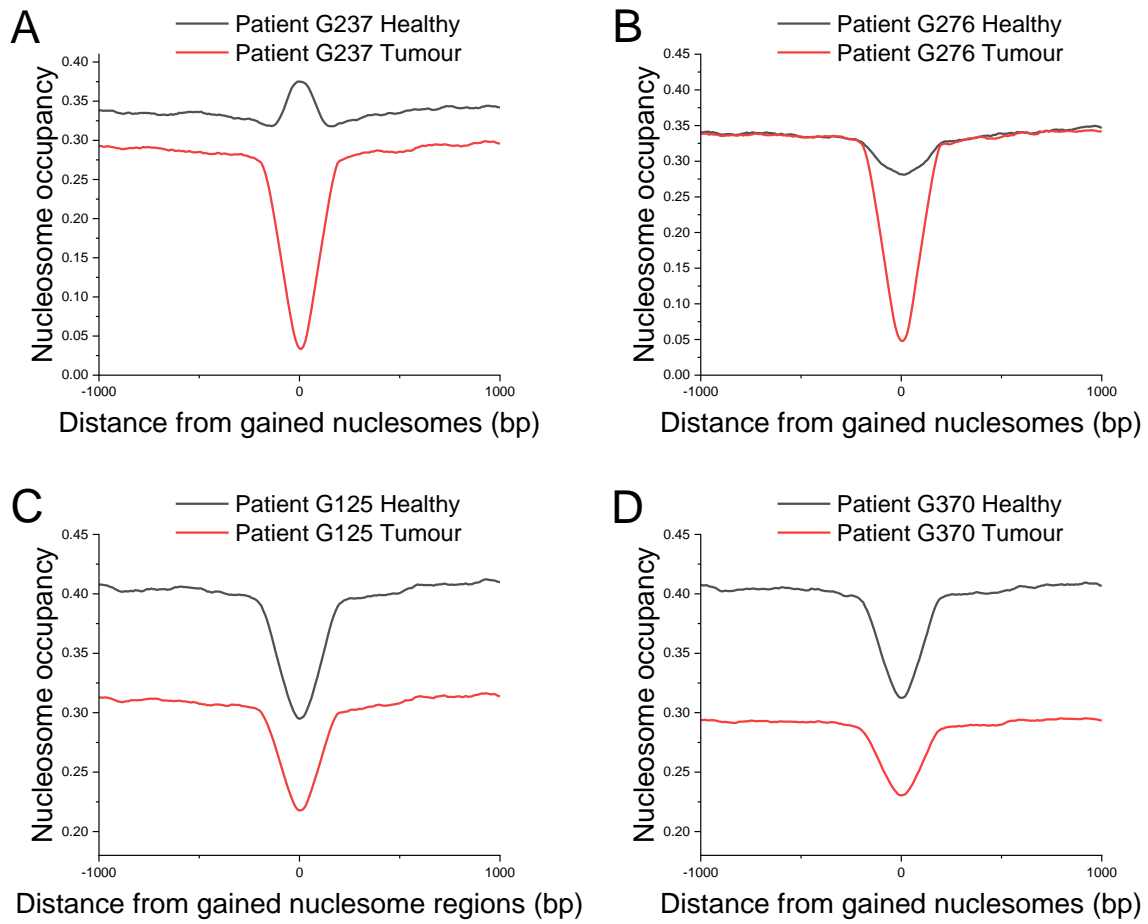


Figure S49. Average nucleosome occupancy profiles around gained nucleosomes in tumour for healthy and GBM tissue from the 4 GBM patients. (A) Patient G276 (B) Patient G237 (C) Patient G215 (D) Patient G370. Healthy profiles can be seen in black and GBM profiles in red.

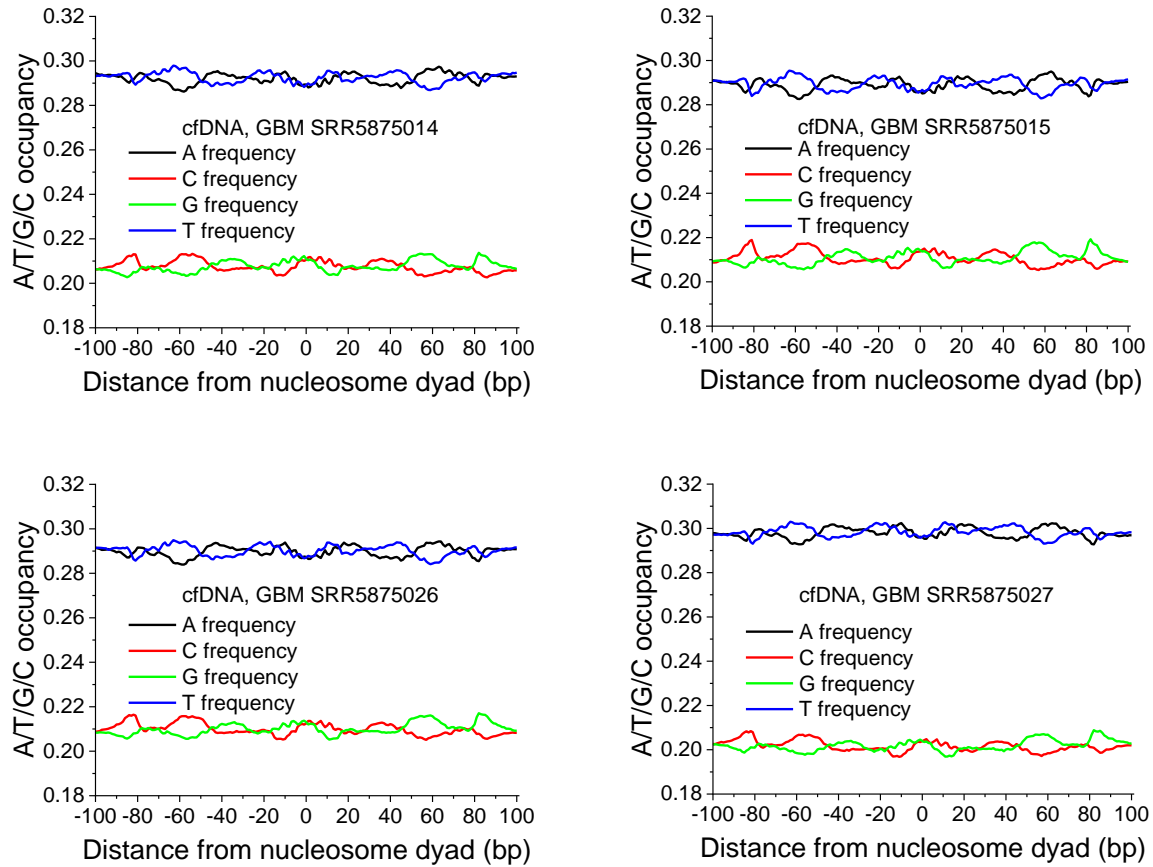


Figure S50. Nucleotide patterns of cfDNA fragments from Song et al, 2017. Each panel corresponds to one GBM patient. These patients are not related to the GBM patients studied using MNase-seq.

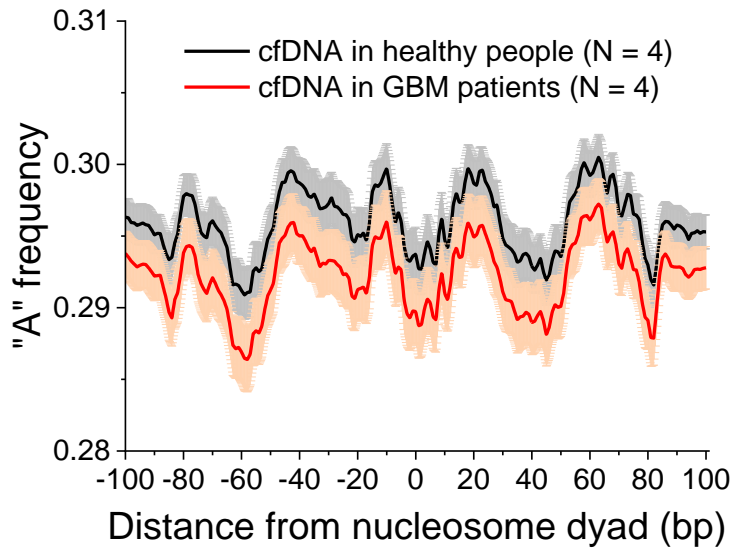


Figure S51. Average cfDNA nucleotide patterns over four GBM patients from Figure 1, and similarly over four healthy people from Song et al, 2017. Grey areas correspond to the standard errors of averaging.

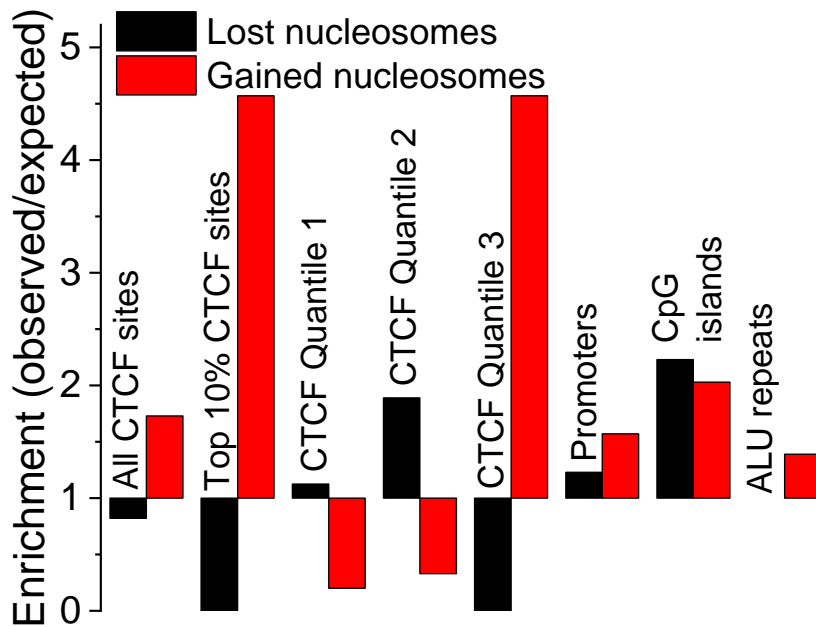


Figure S52. Enrichment of 100bp regions of lost (black) and gained (red) nucleosomes in GBM around genomic regions and CTCF binding sites including different binding strength quantiles, 1 being the weakest and 3 the strongest.

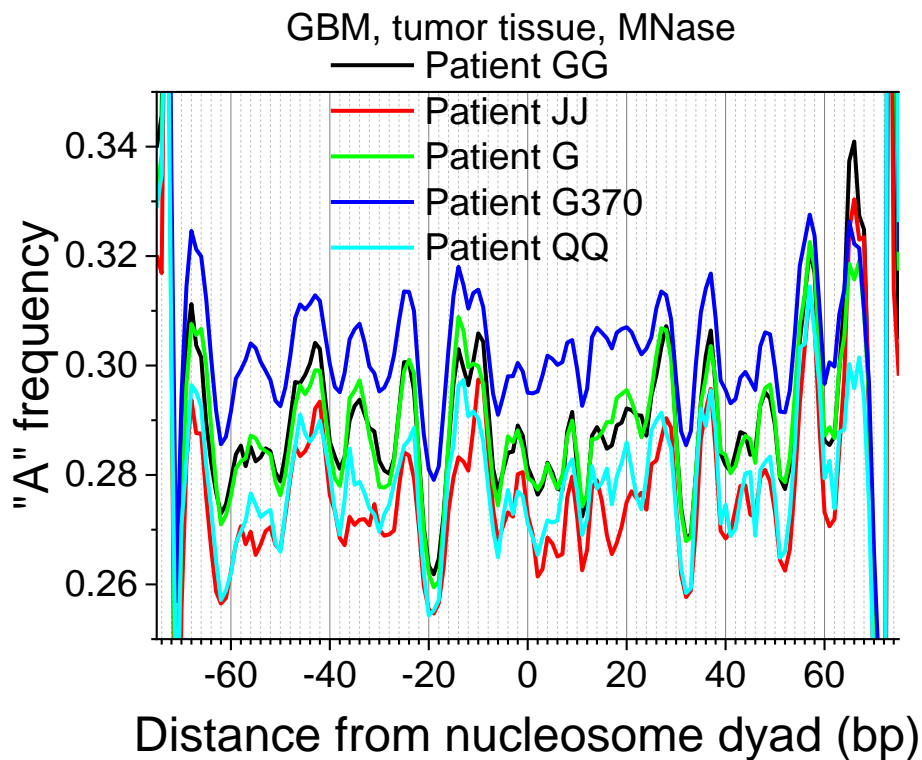
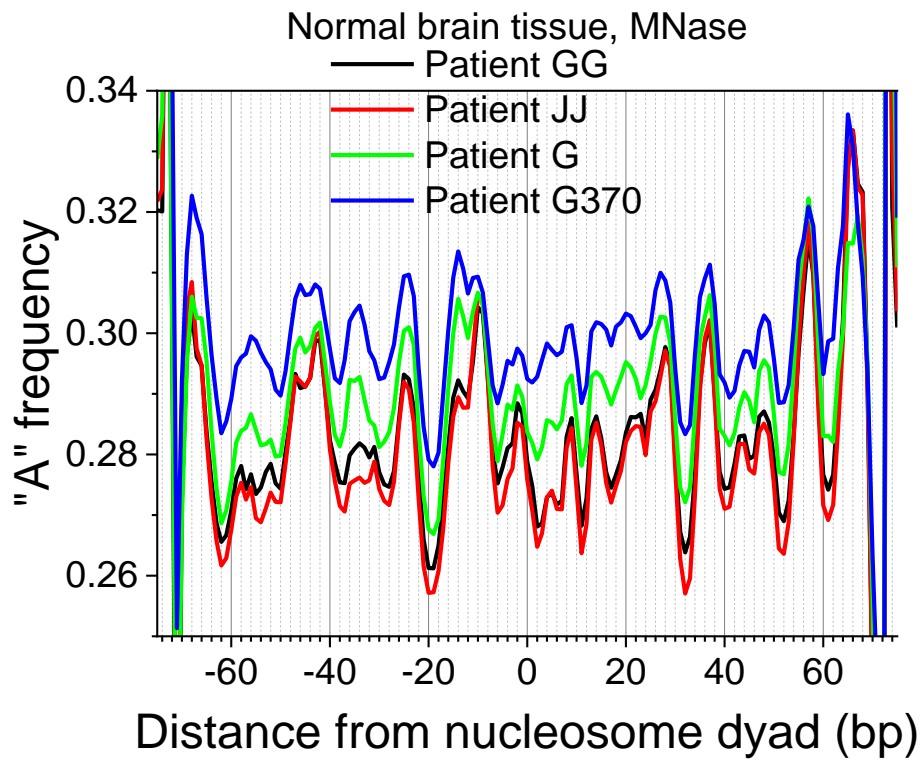


Figure S53. Frequency of "A" nucleotide for each of GBM patients from the experiments performed in Essex. Top panel: normal brain tissues; bottom panel: tumour tissues.

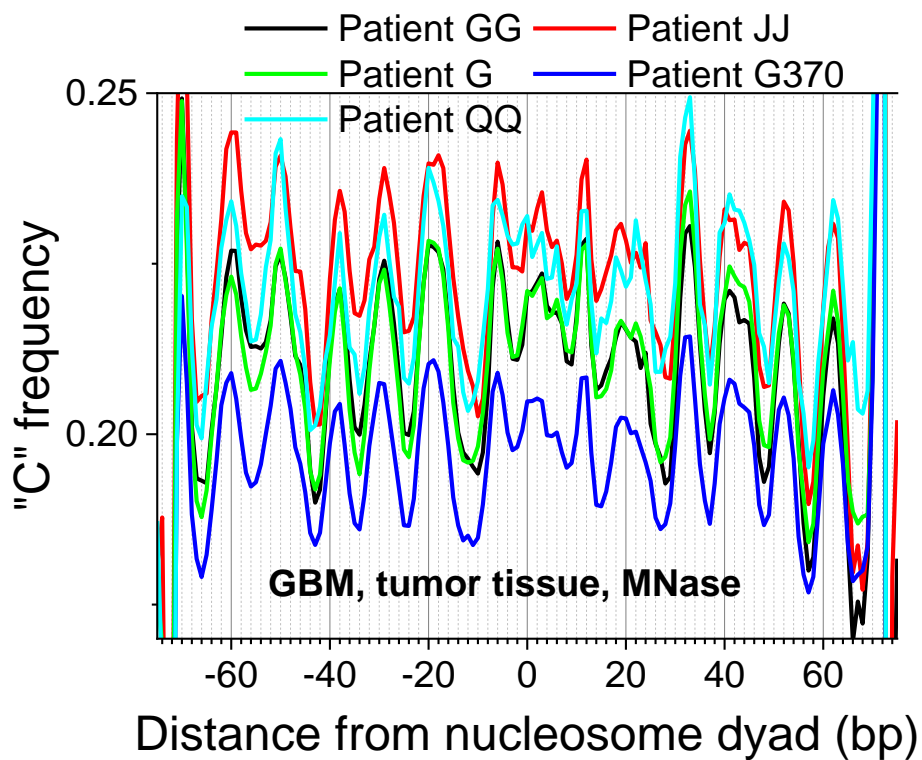
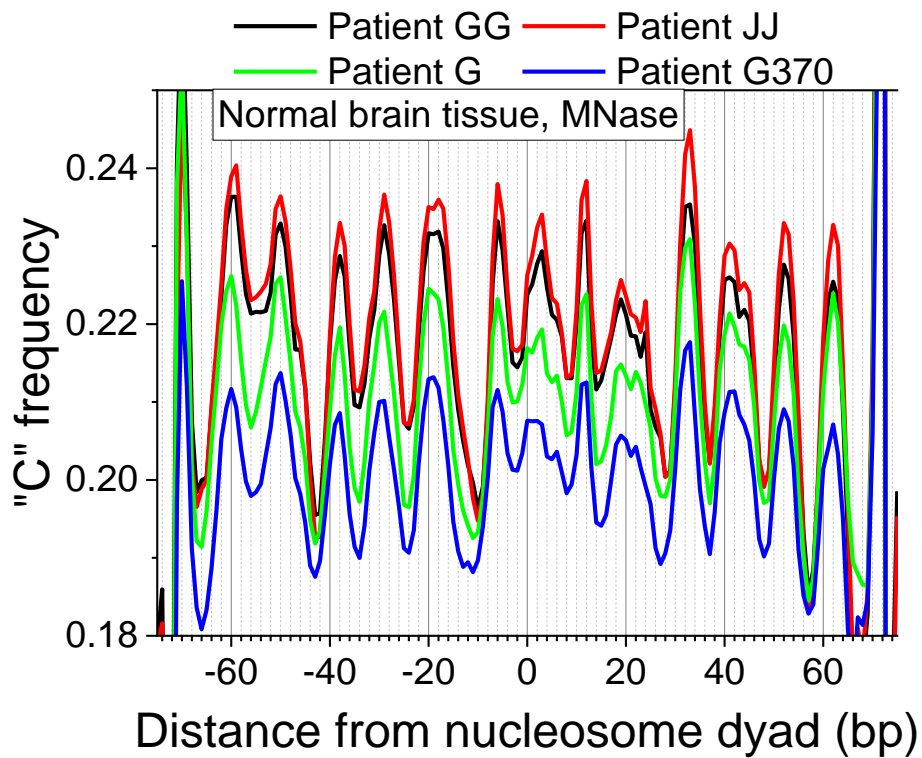


Figure S54. Frequency of "C" nucleotide for each of GBM patients from the experiments performed in Essex. Top panel: normal brain tissues; bottom panel: tumour tissue

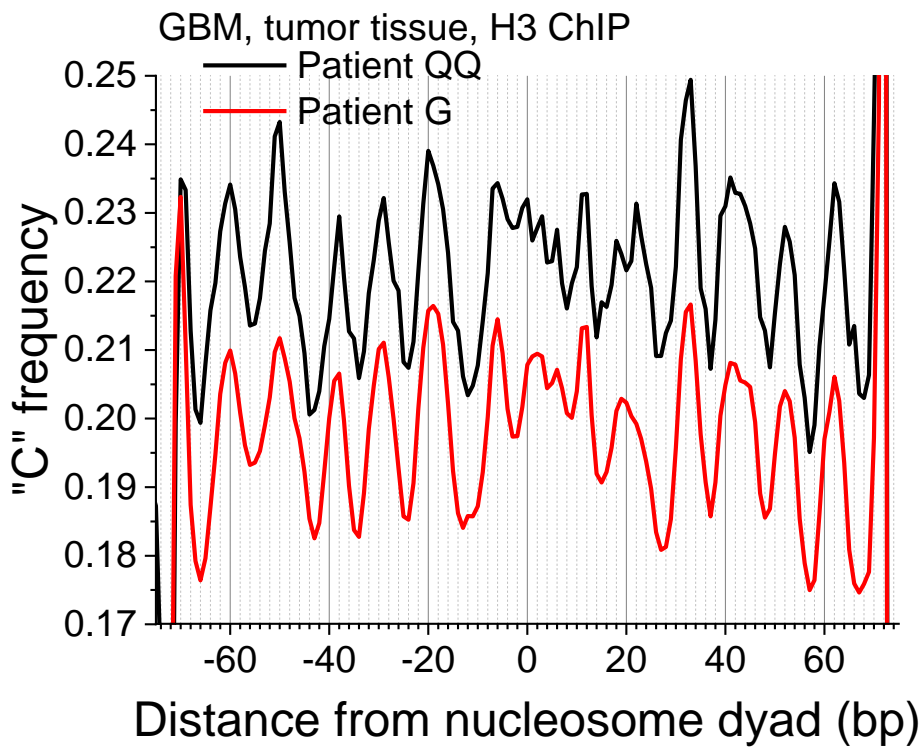
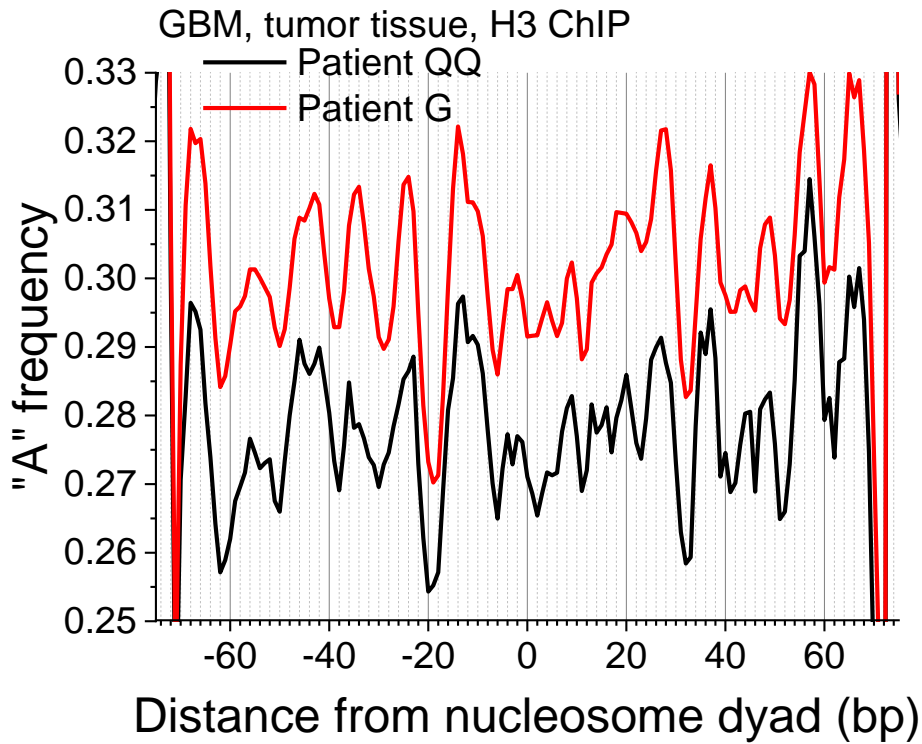


Figure S55. Average nucleotide profiles for MNase-assisted histone H3 ChIP-seq DNA fragments in patients QQ and G. Top panel – A frequency; bottom panel – C frequency.

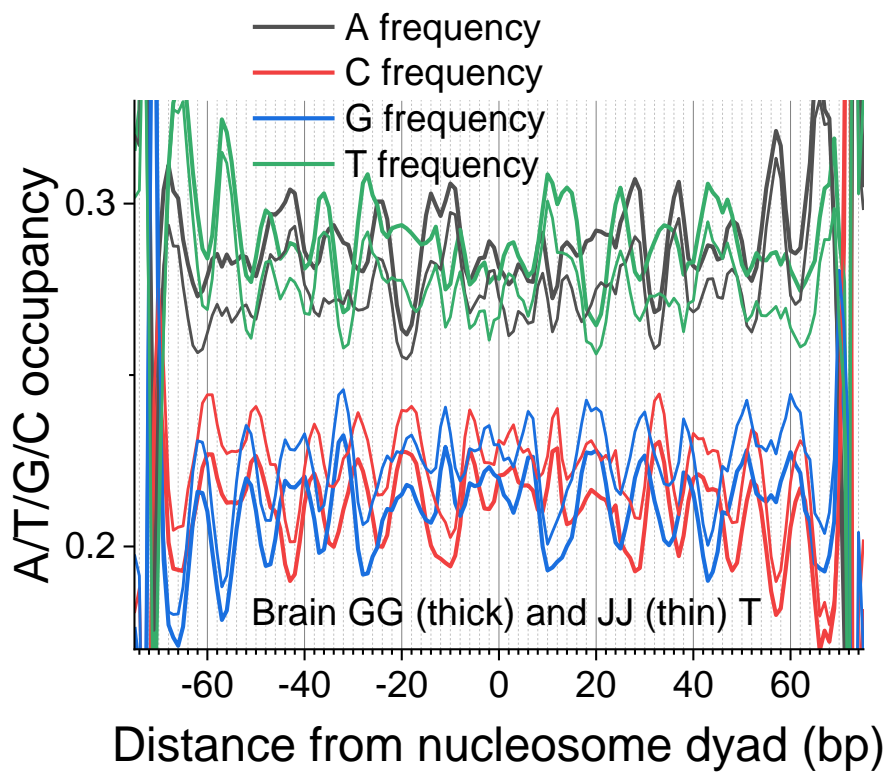
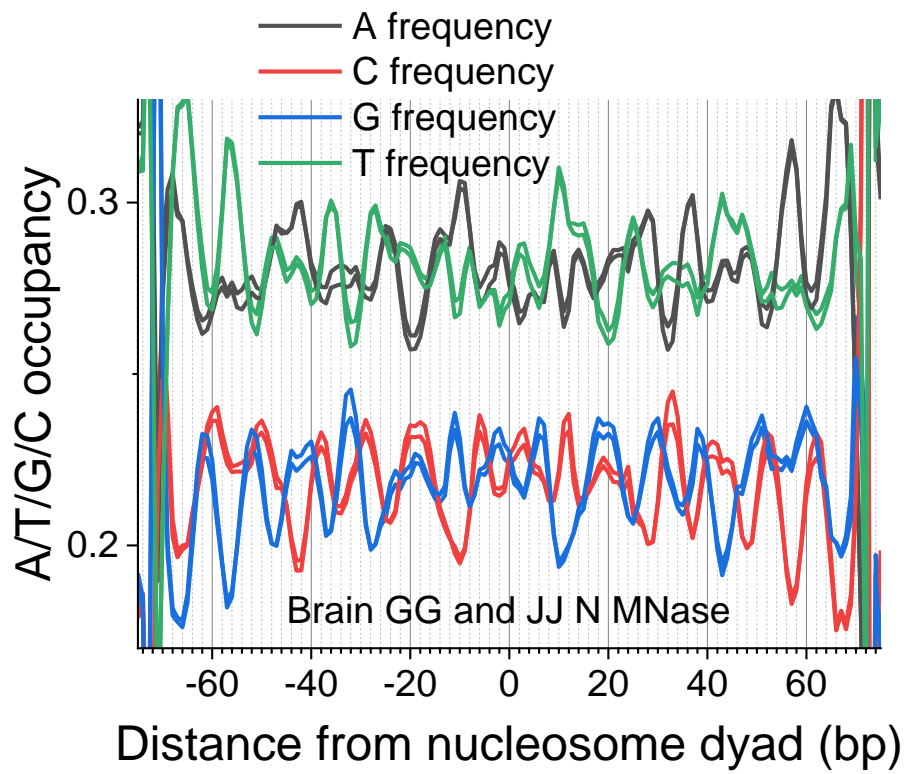


Figure S56. Individual samples GG and JJ with their A/C/G/T profiles



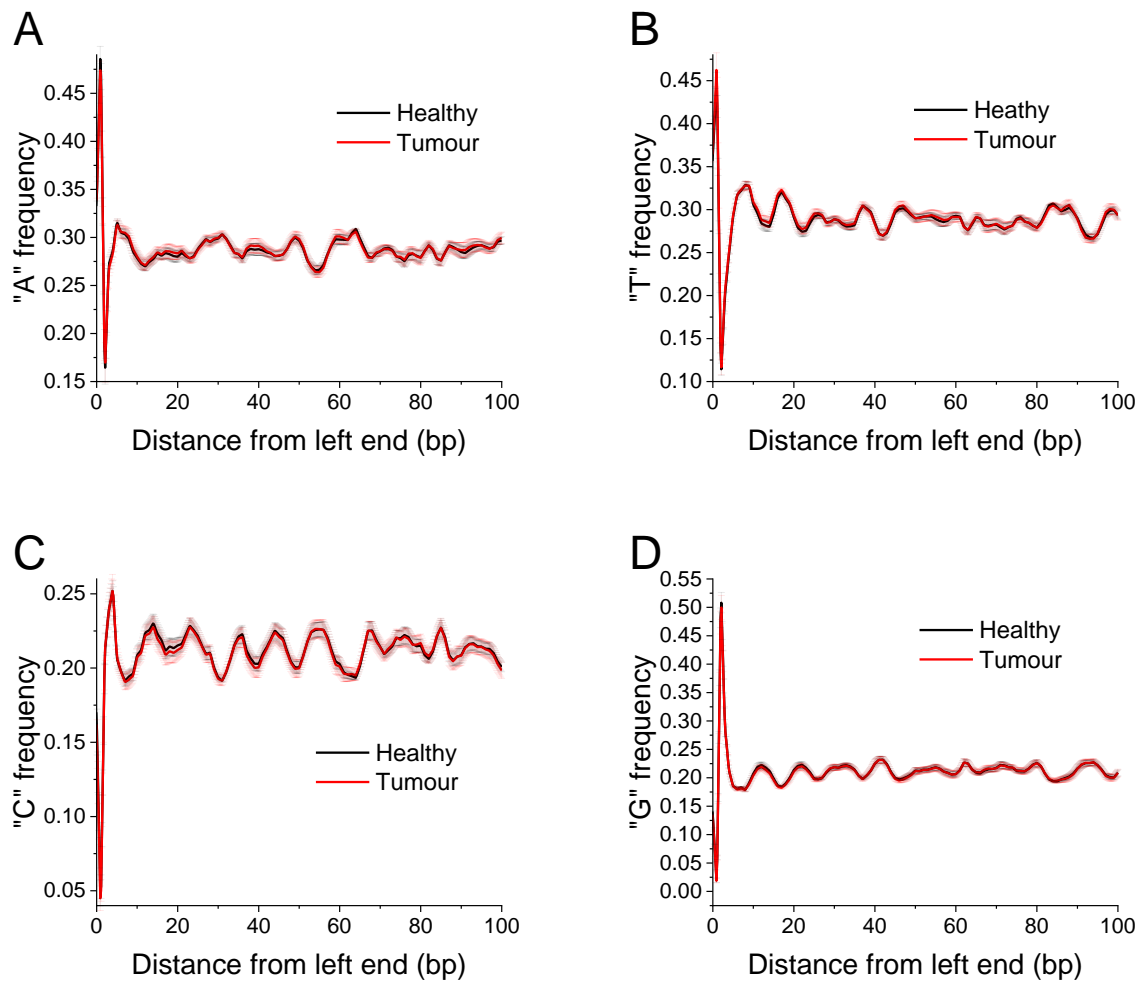


Figure S57. Nucleotide frequencies around nucleosomal DNA fragments (146-148bp) based on MNase-seq in paired brain samples from GBM patients (blue – normal; red – tumour tissues). A – adenine, B – thymine, C – cytosine, D – guanine. All calculations were made by aligning the signal around the start binding sites of nucleosome dyad.

# Computer scripts used in the project

**Script 1.** R script that calculates frequency of nucleosome fragments in bed files using the NucTools package to measure distribution of nucleosomal DNA fragments.

```
args = commandArgs(trailingOnly=TRUE); # allows for one to specify
input and output file
file_in=args[1] #input name
file_out=args[2] #output name
library(readr) # install this with "install.packages('readr')"
nucs=read_delim(file_in, delim="\t", col_names=F) #reads the file
print(head(nucs)) #shows first 6 lines of the file
colnames(nucs)=c("chr", "start", "end", "frag_length")
h=hist(nucs$frag_length, breaks=200, plot=F) #change the number of
bins with the 'breaks' parameter
dataoi=cbind(h$breaks, c(h$counts, NA), c(h$density, NA))
colnames(dataoi)=c("Breaks", "Counts", "Density")
write.table(dataoi, file_out, sep="\t", row.names=F) #writes the
histogram data to a text file
png("histogram.png")
plot(dataoi[,1],dataoi[,2],type='l',xlab='frag_lengths',ylab='Freque
ncy')
dev.off()
```

**Script 2.** Perl script that extracts specific base pair lengths from a bed file according to the input inserted at -min= x and -max= y.

```
#!/local/bin/perl
###
### extract_rows_occup.pl
### version: 2.01
###
#####

use strict "vars";
my $usage = "$0 -input=<in.bed> -output=<out.bed> -
min=<min_fragment> -max=<max_fragment> --help\n";
my $infile;
my $outfile;
my $min_fragment;
my $max_fragment;

if (@ARGV != 0) {
    foreach my $comand_line_flag (@ARGV) {
        if ($comand_line_flag =~ /-input=(.*)/i) { $infile = $1; }
        if ($comand_line_flag =~ /-output=(.*)/i) { $outfile = $1;}
        if ($comand_line_flag =~ /-min=(\d*)/i) { $min_fragment =
$1;}
        if ($comand_line_flag =~ /-max=(\d*)/i) { $max_fragment =
$1;}
    }
}
```

```

        if ($comand_line_flag =~ /--help/i) {
            print STDOUT <DATA>;
            print STDOUT "\nPress <ENTER> button to exit... ";
            <STDIN>;
            exit;
        }
    }
}
else { warn "$usage"; exit;}

open(OUT, ">$outfile") || die "error: can't open $outfile for
writing!\n";
open(IN, $infile) or die "error: $infile cannot be opened: $!\n";

print "min_fragment=", $min_fragment, "\n";
print "max_fragment=", $max_fragment, "\n";

my $buffer = "";
my $sz_buffer = 0;
my $timer2 = time();
# counter for the markers we see
my $marker_count = 0;

my $regex_split_newline='\n';

my $filesize_in_bytes = -s $infile; #determine file size in bytes
my $size_counter_step=int($filesize_in_bytes/100);
my $filesize = int($filesize_in_bytes/1048576); # filesize in
megabytes

print STDERR "Reading $infile file of $filesize MBs. Please
wait...\n";
my $processed_memory_size = 0;
my $offset=0;
my $BUFFER_SIZE = 1024;
my $old_coordinate=1;
my $counter=0;

while ((my $n = read(IN, $buffer, $BUFFER_SIZE)) !=0) {
    if (($n >= $BUFFER_SIZE) or (($n == $filesize_in_bytes))) {
        $buffer .= <IN>;
    }
    my @lines = split(/$regex_split_newline/o, $buffer);
    # process each line in zone file
    foreach my $line (@lines) {
        chomp($line);
        my @newline=split(/\t/, $line);

        #print "newline[3]=", $newline[3], "newline[3]-
min_fragment=", $newline[3]-$min_fragment, "newline[3]-
max_fragment=", $newline[3]-$max_fragment, "\n";

        if (($newline[3] >= $min_fragment) and ($newline[3] <=
$max_fragment))

```

```

        #if (($newline[3] >= 0) and ($newline[3] < 1e10))
        {
            print OUT join("\t",@newline), "\n";
            $counter++;
        }
    }
    $processed_memory_size += $n;
    $offset += $n;
    if(int($processed_memory_size/1048576)>= $filesize/100) {
        print STDERR int($offset/1048576), " Mbs processed in ",
time()-$timer2, " seconds.\n"; $processed_memory_size=0;
        #last;
    }
    undef @lines;
    $buffer = "";
}
close(IN);
close(OUT);

print "counter=", $counter, "\n";

print STDERR "done!\nJob finished in ", time()-$timer2, "
seconds.\n";
exit;

```

#### DATA

```

=====
##### Calculates nucleosomes occupancy frequency distribution around a
TSS
##### (c) 2012 Algorithmus, BioQuant Heidelberg. All Rights Reserved.
##### Authors: Yevhen Vainshtein
#####
##### extract_rows_occup.pl
##### version: 1.0.0.0
=====
=====

```

```

perl -w extract_rows_occup.pl -input=<in.bed> -output=<out.bed> -
min=<min_fragment> -stop=<max_fragment> --help

```

```

-input=<input *.bed file>
-output=<output *.bed file containing only selected region>
-strat=<selected min_fragment>
-max=<selected max_fragment>

```

### Script 3. R script used for Principle Component Analysis (PCA).

```
library(openxlsx)##downloading libraries needed for the analysis
library(reshape2)
library(plyr)
library(scales)
library(ggplot2)
library(devtools)
library(grid)
install_github("vqv/ggbiplot")
library(ggbiplot)
getwd()
data.gained<-read.table("chr1_1000bp_nucs_at_gained.bed",fill=TRUE)
data.gained<-data.gained[,c(4,8,12,16,20,24,28,32)]
colnames(data.gained)<-
c("G237_N", "G276_N", "G370_N", "G125_N", "G237_T", "G276_T", "G370_T", "G1
25_T")
data.gained2<-t(data.gained)
data.gained2<-data.gained2[,-c(749:758)]
n<-ncol(data.gained2)
colnames(data.gained2)<-c(1:n)
data.gained.pca<-prcomp(data.gained2,center=TRUE,scale=TRUE)
summary(data.gained.pca)

##plot PCA
data.group<-c(rep("Healthy",4),rep("Tumour",4))

pdf("PCA_nucs_at_gained_final.pdf",height=10,width=10) #opens a pdf
file
ggbiplot(data.gained.pca,var.axes=F,groups=data.group)+
stat_ellipse(geom="polygon",level=0.9,alpha=0.4,aes(fill=data.group)
)+
scale_fill_manual(values=c("gray62","palevioletred1","lightskyblue")
)+
guides(colour=guide_legend("Group"),fill=F)+
scale_colour_manual(name="Group",values=c("black","firebrick1","roya
lblue1"))+
scale_shape_manual(name="Group",values=c(16,16,16))+
geom_point(aes(colour=data.group,shape=data.group),size=5)+
geom_text(aes(colour=data.group,
label=rownames(data.gained2)),hjust=-0.2, vjust=0.1)+
ggtitle("PCA of nucleosome occupancy at regions gaining
nucleosomes")+
theme_classic()+
theme(plot.title=element_text(hjust=0.7,size=25),
legend.position="top",legend.title=element_blank(),legend.text=eleme
nt_text(size=25),axis.text.x=element_text(size=20),axis.text.y=eleme
nt_text(size=20),axis.title.x=element_text(size=30),axis.title.y=ele
ment_text(size=30))
dev.off()

##lost nucleosomes
data.lost<-read.table("chr1_1000bp_nucs_at_lost.bed",fill=TRUE)
data.lost<-data.lost[,c(4,8,12,16,20,24,28,32)]
```

```

colnames(data.lost)<-
c("G237_N", "G276_N", "G370_N", "G125_N", "G237_T", "G276_T", "G370_T", "G1
25_T")
data.lost2<-t(data.lost)
data.lost2<-data.lost2[,-c(749:758)]
n<-ncol(data.lost2)
colnames(data.lost2)<-c(1:n)
data.lost.pca<-prcomp(data.lost2,center=TRUE,scale=TRUE)
summary(data.lost.pca)

##plot PCA
data.group<-c(rep("Healthy",4),rep("Tumour",4))

pdf("PCA_nucs_at_lost_final.pdf",height=10,width=10) #opens a pdf
file
ggbiplot(data.lost.pca,var.axes=F,groups=data.group)+
stat_ellipse(geom="polygon",level=0.9,alpha=0.4,aes(fill=data.group)
)+
scale_fill_manual(values=c("gray62","palevioletred1","lightskyblue")
)+
guides(colour=guide_legend("Group"),fill=F)+
scale_colour_manual(name="Group",values=c("black","firebrick1","roya
lblue1"))+
scale_shape_manual(name="Group",values=c(16,16,16))+
geom_point(aes(colour=data.group,shape=data.group),size=5)+
geom_text(aes(colour=data.group,label=rownames(data.lost2)),hjust=-
0.2,vjust=0.1)+
ggtitle("PCA of nucleosome occupancy at regions losing
nucleosomes")+
theme_classic()+
theme(plot.title=element_text(hjust=0.9,size=25),
legend.position="top",legend.title=element_blank(),legend.text=eleme
nt_text(size=25),axis.text.x=element_text(size=20),axis.text.y=eleme
nt_text(size=20),axis.title.x=element_text(size=30),axis.title.y=ele
ment_text(size=30))
dev.off()

```